TRANSCRIPTOMIC ANALYSES OF *CLOSTRIDIUM BEIJERINCKII* NCIMB 8052 DURING TRANSITION FROM ACIDOGENESIS TO SOLVENTOGENESIS AND UNDER BUTYRATE SUPPLEMENTED CONDITIONS

BY

YI WANG

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Agricultural and Biological Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral committee:

Professor Hans P. Blaschek, Chair
Associate Professor Hao Feng
Professor Yuanhui Zhang
Associate Professor Nathan D. Price, Institute for Systems Biology
Dr. Zonglin (Lewis) Liu, USDA-Agricultural Research Service

ABSTRACT

Today, the exhaustion of fossil fuel resources and the deterioration of the natural environment drive people to seek alternative bio-based fuels and chemicals from renewable sources. Biobutanol produced through microbial fermentation of biomass has been of great interest because of its various advantages as a biofuel and considerable value as an industrial chemical feedstock. *Clostridium beijerinckii* is among the prominent species for biobutanol production as it demonstrates a broad substrate range for growth and solvent production. Although the transcriptome structure and transcriptional profiling are essential for understanding the functional and regulatory network of the genome and specific gene functions and regulations associated with the cell physiology, the physical structure of the transcriptome and the transcriptional profiles were not well understood for *C. beijerinckii*.

In this study, a single-nucleotide resolution analysis of the *C. beijerinckii* 8052 transcriptome was conducted using high-throughput RNA-Seq technology. The transcription start sites and operon structure throughout the genome were identified. The structure of important gene operons involved in metabolic pathways for acid and solvent production in *C. beijerinckii* 8052 were confirmed, including *pta-ack*, *ptb-buk*, *hbd-etfA-etfB-crt* (*bcs*) and *ald-ctfA-ctfB-adc* (*sol*) operons; important operons related to chemotaxis/motility, transcriptional regulation, stress response and fatty acids biosynthesis along with others were also defined. In addition, 20 previously non-annotated regions were discovered to be with significant transcriptional activities and 15 genes whose translation start codons were found to be likely mis-annotated. As a consequence, the accuracy of existing *C. beijerinckii* genome annotation was significantly enhanced.

The genome-wide transcriptional dynamics of *C. beijerinckii* 8052 over a batch fermentation process was revealed in detail based on the RNA-Seq data. The gene expression profiles indicated that the glycolysis genes were highly expressed throughout the fermentation, with comparatively more active expression during acidogenesis phase. The expression of acid formation genes was down-regulated at the onset of solvent formation, in accordance with the metabolic pathway shift from acidogenesis to solventogenesis. The *sol* operon genes, were highly-coordinately expressed and up-regulated at the onset of solventogenesis. Out of the 20 genes encoding alcohol dehydrogenase in *C. beijerinckii*, Cbei_1722 and Cbei_2181 were highly up-regulated at the initiation of solvent production, corresponding to their key roles in primary alcohol production. Most sporulation genes in *C.*

beijerinckii 8052 demonstrated similar temporal expression patterns to those observed in *B. subtilis* and *C. acetobutylicum*, while sporulation sigma factor genes *sigE* and *sigG* exhibited accelerated and stronger expression in *C. beijerinckii* 8052, which is consistent with the more rapid forespore and endspore development in this strain. Global expression patterns for specific gene functional classes demonstrated general expression profiles corresponding to the cell physiological variation and metabolic pathway switch during the fermentation.

Butyrate has long been suggested as a potential triggering factor for the solventogenesis switch during the ABE fermentation. In this study, sodium butyrate was added to the chemically defined MP2 medium for ABE fermentation with *C. beijerinckii* 8052 in order to investigate the effects of butyrate on solvent production and metabolic pathway switch. The results indicated that with butyrate supplementation, the solvent production was triggered early in the mid-exponential phase and finished quickly in < 50 h from the time of innoculation. While in the control, solventogenesis initiated during late exponential phase and took > 90 h to reach the maximum butanol level. Butyrate supplementation led to a 31% improvement in the final butanol titer, 58% improvement in butanol yield and 133% improvement in butanol productivity, comparing to the control without butyrate addition. It also led to a higher butanol/acetone ratio compared to the control, indicating a metabolism shift towards butanol production. The addition of butyrate also triggered much earlier and stronger sporulation during the fermentation. Genome-wide transcriptional analysis was performed with RNA-Seq over the course of fermentation with and without added butyrate. With butyrate addition, the gene expression in central metabolic pathway related to solventogenesis was induced about 10 hours earlier and accelerated, and the maximum expression levels of butyryl-CoA and solvent formation genes were elevated comparing to the control, correlating with the faster and higher solvent production. In both conditions, the sporulation genes were induced at the onset of solventogenesis, although this occurred much earlier in the reactor with butyrate addition, and went through a similar temporal expression patterns with much faster down-regulation under the condition with butyrate addition. The motility related genes were generally down-regulated to lower levels just prior to the stationary phase under both conditions. However, in the control it took much longer and the gene expression remained at comparatively higher levels after the microbe entered stationary phase. Supplemented butyrate may cause feedback inhibition to butyrate formation and be re-assimilated through the reversed butyrate formation pathway, and thus result in an elevated

level of intracellular butyryl phosphate, which may act as a phosphate donor to Spo0A and consequently triggered solventogenesis and related events.

The results from this work provided insights for further *C. beijerinckii* strain improvement employing system biology-based strategies and metabolic engineering approaches. Furthermore, this work is also an essential methodology reference for conducting transcriptional analysis employing next-generation sequencing technology.

To my family

ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor Dr. Hans Blaschek, for all his support on my work. He provided me the best research resources, abundant financial support, and helpful instruction and guidance. Without all of these, it is impossible for me to finish my research at UIUC. Meanwhile, I learned a lot from Dr. Blaschek concerning how to collaborate and get along with others, which will be useful for my whole career.

Secondly, I would like to give my special thanks to Dr. Xiangzhen Li, who is now a professor in Chengdu Institute of Biology in China. It was enjoyable to work together with him on several projects when he was working with Dr. Roderick Mackie in ASL. He not only taught me on how to do experiments in the lab, but also taught me a lot on how to do research as an academic researcher. Meanwhile, I also thank Dr. Yuejian Mao and other group members from Dr. Roderick Mackie and Dr. Isaac Cann's group for their altruistic help and spending the happy time together on the four floor of ASL.

Thirdly, I would like to thank Dr. Hao Feng, Dr. Yong-Su Jin, Dr. Nathan Price and many of their group members for the collaboration and assistance with my experiments. I thank my other committee members Dr. Yuanhui Zhang and Dr. Lewis Liu for their valuable inputs and guidance on my projects. I also thank various undergraduate and PSM students who worked with me and assisted with my projects.

In addition, I would like to thank all the other friends and colleagues who helped. They are from ABE, IGB, FSHN, CEE, Keck Center, Crop Sciences, Animal Sciences departments and even outside of the campus. I am sorry that I cannot include all their names here, but without their tremendous help, I cannot finish my study in a timely fashion.

Last but not least, I would like to thank my family for their great support on my study. My wife, my parents, relatives from my family and my wife's family all made huge sacrifices. I also thank my little daughter, Sage Wang. Her birth brought me marvelous happiness, best luck and a continuous motivating force.

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION AND RESEARCH OBJECTIVES

1.1 Background

Today, the exhaustion of fossil fuels and increasing severity of environmental problems are becoming big concerns to the sustainable survival of the human beings. Harvesting bioenergy from the renewable resources is considered to be a feasible resolution. Biobutanol (specifically n-butanol, will be generally referred as "butanol" in this dissertation) produced from biomass resources through fermentation has been of great interest because of its various advantages as a biofuel and high value as a raw chemical feedstock. As a fuel source, the energy content of butanol is considerably higher than ethanol, and close to that of gasoline; butanol is six times less evaporative than ethanol which makes it safer to handle; butanol can be distributed through the existing fuel pipelines; butanol can be used as a replacement for gasoline gallon for gallon (or at any percentage blend with gasoline) (Dürre, 2007). As chemical feedstocks for industry, butanol can be used in latex surface coatings, enamels and lacquers industries, and can also be used as the diluent for brake fluid formulations and the solvent for manufacturing of antibiotics, vitamins and hormones (Lee et al., 2008). Biobutanol production through the acetone-butanol-ethanol (ABE) fermentation with solventogenic clostridia was at a large industrial scale during the early 20th century. However, it was later almost completely replaced by industrial synthesis of solvents because of the increasing demand of butanol and the rapid development of petrochemical industry (Jones and Woods, 1986). Recently, the skyrocketing price of oil, the increasing concern about environmental problems and the advancement of biotechnology have led to a renewed interest in biobutanol production through biological fermentation routes.

Clostridium acetobutylicum and *C. beijerinckii* are among the prominent solvent-producing clostridial species. Although *C. beijerinckii* is phenotypically similar to *C. acetobutylicum*, the saccharolytic strains were demonstrated to be phylogenetically distant from the amylolytic *C. acetobutylicum* ATCC 824 type strain (Keis et al., 1995). The genome of *C. beijerinckii* NCIMB 8052 was sequenced in 2007. The genome size is around 6.0 Mb, which is 50% larger than that of *C. acetobutylicum* ATCC 824. *C. beijerinckii* 8052 has all its solvent producing genes located on the chromosome and it exhibits a broader substrate range and optimum pH for growth and solvent production; thus it may have greater potential for biosolvent production than *C. acetobutylicum* ATCC 824 (Ezeji and Blaschek, 2008).

The genome of *C. beijerinckii* has been annotated using bioinformatics methods based on gene prediction algorithms. Although transcriptome profiling is important to reveal the functional and regulatory architecture of the genome, little information is yet available about the physical structure of the *C. beijerinckii* 8052 transcriptome, such as operon linkages and transcript boundaries. On the other hand, global gene transcriptional analyses are important to understand the fermentation pathway and solvent production mechanism, and thus improve the strain for better solvent production. While global gene expression patterns for *C. acetobutylicum* have been studied extensively (Alsaker and Papoutsakis, 2005; Alsaker et al., 2010; Alsaker et al., 2005; Alsaker et al., 2004; Janssen et al., 2010; Jones et al., 2008; Tomas et al., 2004), to our best knowledge, little information is available for the transcriptional analysis of *C. beijerinckii* other than the recently published study by Shi and Blaschek (Shi and Blaschek, 2008), utilizing a ca. 500-gene set DNA microarray where only the expression of genes orthologous to those previously found to be important to the physiology of *C. acetobutylicum* 824 was examined.

During a typical ABE batch fermentation, acetic acid and butyric acid are produced during the active exponential growth phase. As growth slows down, acids are re-assimilated by the culture and acetone, butanol and a small amount of ethanol are produced (Jones and Woods, 1986). The switch from acidogenesis to solventogenesis is usually associated with the induction of solvent formation gene transcription and decreased cell growth and motility. However, the mechanism that triggers the metabolic switch is still elusive. The level of intracellular undissociated butyric acid has long been reported as a possible controlling factor for the shift from acid production to solvent production in the metabolic pathway. Terracciano and Kashket determined that solvent production began in *C. acetobutylicum* when the undissociated butyric acid level reached 13-18 mM (Terracciano and Kashket, 1986). Husemann and Papoutsakis determined that 6–13 mM of undissociated butyric acid was required for the initiation of solventogenesis in *C. acetobutylicum* when the external pH was between 3.7 and 5.0 (Husemann and Papoutsakis, 1988). Recently, Lee et al. reported that the initial addition of butyrate into the culture media for ABE fermentation with *C. beijerinckii* 8052 enhanced solvent production (Lee et al., 2008). But in their study, detailed fermentation dynamics were not included, and no transcriptional analysis or enzyme activity assays were preformed, thus making it hard to draw conclusions concerning the effects of butyrate on the metabolic pathway switch at a gene transcription level.

RNA-Seq technology was developed based on next-generation high-throughput sequencing technology (Mortazavi et al., 2008; Nagalakshmi et al., 2008). With RNA-Seq, it would be easy to obtain millions of cDNA reads simultaneously, and these reads can be assembled against genome sequence, and expression values calculated based on the reads mapped to genes. Deep sequencing of the cDNA pool allows us to study the bacterial transcriptome structure and gene transcriptional profiles at an unprecedented resolution and depth (Passalacqua et al., 2009; Wurtzel et al., 2010).

1.2 Research objectives

Based on the background information discussed above, the first objective is, using RNA-Seq technology, to investigate the *C. beijerinckii* 8052 transcriptome structure at a single-nucleotide resolution. The transcriptome structural information such as operon linkages and transcript boundaries will be defined, and thus the accuracy of existing genome annotation will be significantly enhanced. Such information will provide an important supplement to the current genome annotation, and help to reveal the gene functions and regulation in *C. beijerinckii.*

The second objective of this study is to conduct a genome-wide transcriptional analysis of *C. beijerinckii* 8052 with respect to the solvent production during the course of a batch fermentation from an unbiased perspective using RNA-Seq technology. In comparison to the extensive studies on *C. acetobutylicum*, the transcriptional analysis for *C. beijerinckii* has been lagging due to limited availability of genomic information and genetic study tools. With the availability of the genome sequence for *C. beijerinckii* 8052 and the newly developed RNA-Seq technology, a genome-wide gene expression profiling can provide essential information for further strain improvement through metabolic engineering and system biology approaches.

The third objective of this study is to examine the effect of supplementary butyrate in the growth medium on solvent production and the metabolic pathway switch in *C. beijerinckii* 8052. Throughout the fermentation, high-resolution genome-wide transcriptional analysis will be performed using RNA-Seq. In this way, further insights are expected to be revealed about the biochemical effects of butyrate on solventogenesis-related events from the gene regulation level.

1.3 References

Alsaker, K. V., and E. T. Papoutsakis. 2005. Transcriptional program of early sporulation and stationary-phase events in *Clostridium acetobutylicum*. *Journal of Bacteriology* 187(20):7103-7118.

Alsaker, K. V., C. Paredes, and E. T. Papoutsakis. 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*. *Biotechnology and Bioengineering* 105(6):1131-1147.

Alsaker, K. V., C. J. Paredes, and E. T. Papoutsakis. 2005. Design, optimization and validation of genomic DNA microarrays for examining the *Clostridium acetobutylicum* transcriptome. *Biotechnology and Bioprocess Engineering* 10(5):432-443.

Alsaker, K. V., T. R. Spitzer, and E. T. Papoutsakis. 2004. Transcriptional analysis of *spo0A* overexpression in *Clostridium acetobutylicum* and its effect on the cell's response to butanol stress. *Journal of Bacteriology* 186(7):1959-1971.

Dürre, P. 2007. Biobutanol: An attractive biofuel. *Biotechnology Journal* 2(12):1525-1534.

Ezeji, T., and H. P. Blaschek. 2008. Fermentation of dried distillers' grains and solubles (DDGS) hydrolysates to solvents and value-added products by solventogenic clostridia. *Bioresource Technology* 99(12):5232-5242.

Husemann, M. H. W., and E. T. Papoutsakis. 1988. Solventogenesis in *Clostridium acetobutylicum* fermentation related to carboxylic acid and proton concentration. *Biotechnology and Bioengineering* 32(7):843-852.

Janssen, H., C. Doring, A. Ehrenreich, B. Voigt, M. Hecker, H. Bahl, and R. J. Fischer. 2010. A proteomic and transcriptional view of acidogenic and solventogenic steady-state cells of *Clostridium acetobutylicum* in a chemostat culture. *Applied Microbiology and Biotechnology* 87(6):2209-2226.

Jones, D. T., and D. R. Woods. 1986. Acetone-butanol fermentation revisited. *Microbiological Reviews* 50(4):484-524.

Jones, S. W., C. J. Paredes, B. Tracy, N. Cheng, R. Sillers, R. S. Senger, and E. T. Papoutsakis. 2008. The transcriptional program underlying the physiology of clostridial sporulation. *Genome Biology* 9(7):R114.

Keis, S., C. F. Bennett, V. K. Ward, and D. T. Jones. 1995. Taxonomy and phylogeny of industrial solvent-producing clostridia. *International Journal of Systematic Bacteriology* 45(4):693-705.

Lee, S. M., M. O. Cho, C. H. Park, Y. C. Chung, J. H. Kim, B. I. Sang, and Y. Um. 2008a. Continuous butanol production using suspended and immobilized *Clostridium beijerinckii* NCIMB 8052 with supplementary butyrate. *Energy & Fuels* 22(5):3459-3464.

Lee, S. Y., J. H. Park, S. H. Jang, L. K. Nielsen, J. Kim, and K. S. Jung. 2008b. Fermentative butanol production by clostridia. *Biotechnology and Bioengineering* 101(2):209-228.

Mortazavi, A., B. A. Williams, K. McCue, L. Schaeffer, and B. Wold. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* 5(7):621-628.

Nagalakshmi, U., Z. Wang, K. Waern, C. Shou, D. Raha, M. Gerstein, and M. Snyder. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320:1344-1349.

Passalacqua, K. D., A. Varadarajan, B. D. Ondov, D. T. Okou, M. E. Zwick, and N. H. Bergman. 2009. Structure and complexity of a bacterial transcriptome. *Journal of Bacteriology* 191(10):3203-3211.

Perkins, T. T., R. A. Kingsley, M. C. Fookes, P. P. Gardner, K. D. James, L. Yu, S. A. Assefa, M. He, N. J. Croucher, D. J. Pickard, D. J. Maskell, J. Parkhill, J. Choudhary, N. R. Thomson, and G. Dougan. 2009. A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus *Salmonella typhi*. *Plos Genetics* 5(7):e1000569.

Shi, Z., and H. P. Blaschek. 2008. Transcriptional analysis of *Clostridium beijerinckii* NCIMB 8052 and the hyper-butanol-producing mutant BA101 during the shift from acidogenesis to solventogenesis. *Applied and Environmental Microbiology* 74(24):7709-7714.

Terracciano, J. S., and E. R. Kashket. 1986. Intracellular conditions required for initiation of solvent production by *Clostridium acetobutylicum*. *Applied and Environmental Microbiology* 52(1):86-91.

Tomas, C. A., J. Beamish, and E. T. Papoutsakis. 2004. Transcriptional analysis of butanol stress and tolerance in *Clostridium acetobutylicum*. *Journal of Bacteriology* 186(7):2006-2018.

Wurtzel, O., R. Sapra, F. Chen, Y. W. Zhu, B. A. Simmons, and R. Sorek. 2010. A single-base resolution map of an archaeal transcriptome. *Genome Research* 20(1):133-141.

CHAPTER 2. LITERATURE REVIEW

In this chapter, the advantages of bio-butanol as a liquid fuel and a chemical are first summarized to demonstrate that butanol produced from renewable resources is a preferable biofuel and highly valuable industrial feedstock. Then, the history of ABE fermentation with solventogenic clostridia is revisited, and the current development and future prospect are also outlooked. The basic physiological aspects of solvent production with clostridia, the previous conclusions on the regulation factors affecting the switch from acidogenesis to solventogenesis and the effects of butyrate on solvent production are generalized. The previous transcriptional analyses on solventogenic clostridia (mostly for *C. acetobutylicum*) are also reviewed. RNA-Seq technology is a recently developed transcriptomic study tool based on next-generation DNA sequencing technology. The basic principles, advantages over hybridization-based approaches and the recent application of RNA-Seq are reviewed.

2.1 Bio-butanol as a fuel and chemical feedstock

As a liquid biofuel, butanol has many superior properties over other biofuels, including ethanol. Butanol has higher energy content, 30% more than that of ethanol and is closer to that of gasoline (29.2 MJ L^{-1} for butanol versus 21.2 MJ L^{-1} for ethanol; and 32.0 MJ L^{-1} for gasoline) (Table 2.1); butanol has low vapor pressure, six times less "evaporative" than ethanol and 13.5 times less than gasoline, which makes it safer to handle and specially when used in hot area and hot weather; butanol is less corrosive, and thus it can be delivered through existing gasoline supply infrastructure while ethanol must be transported via rail, barge or truck; butanol can be used as a replacement for gasoline gallon for gallon (that is, 100% or any other blend level) and no modification is needed for the internal combustion engine, while ethanol can only be an additive to gasoline up to about 85% following significant engine modifications (Davis and Morton, 2008; Dürre, 2007). In addition, butanol can also be used safely and advantageously up to high blending ratios with the diesel fuel in the diesel engine (Mehta et al., 2010; Rakopoulos et al., 2010); butanol is not hygroscopic (therefore does not pick up water), which allows being blended with gasoline at the refinery, ahead of storage and distribution, while blending ethanol with gasoline must occur shortly before use. This feature of butanol also prevents contamination of groundwater in case of spills (Dürre, 2007). In this sense, butanol is considered to be the next generation biofuel after ethanol, and a fuel source much superior over ethanol.

Table 2.1 Physical properties of butanol, comparing to ethanol and gasoline

	butanol	ethanol	gasoline
Density (g mL^{-1})	0.81	0.79	0.7-0.8
Boiling point (^oC)	118	78	27-221
Energy content (MJ L^{-1})	29.2	21.2	32.0
Water solubility at 25^oC, %	9.1	100.0	< 0.01

Besides its use as a promising biofuel, more importantly, butanol can also be used as a valuable feedstock chemical for various industries. The industrial applications of butanol include but are not limited to: (1) solvents: for paints, varnishes, resins, gums, dyes, camphor, vegetable oils, fats, waxes, resins, shellac, rubbers and alkaloids; (2) plasticizers: to improve processing of plastic materials; (3) coatings: for a variety of applications, such as curable lacquers and cross-linked baking finishes. Now, approximately half of the butanol is used in the latex surface coating, enamels and lacquers industries in the form of butyl acrylate and methacrylate esters (Kirschner, 2006); (4) chemical intermediate or raw material: for producing many other chemicals and plastics, including safety glass, hydraulic fluids and detergent formulations; (5) textiles: as a swelling agent and manufacturing garments from coated fabric; (6) cosmetics: including eye makeup, foundations, lipsticks, nail care products, personal hygiene products and shaving products; (7) many others, such as flotation agents, cleaners, floor polishes, brake fluid, drugs and antibiotics, hormones, and vitamins (The Dow Chemical Company, 2002). Currently, the US annual consumption of butanol is around 740,000 metric tons, representing a 1.47 billion dollar market at the current price. The projected annual growth rate for butanol in US is 2.2% (SRI Consulting, 2009). Global annual consumption of butanol is currently around 2.9 million metric tons, which accounts for a 5.7 billion dollar market, with a predicted annual growth of 4.7% (SRI Consulting, 2009). The butanol market demand is expected to increase dramatically if butanol can be produced economically from low-cost biomass resources.

2.2 History and development of the ABE fermentation

The ABE fermentation is an anaerobic fermentation process that uses bacteria (mainly solvent-producing clostridia) to produce acetone, n-butanol and ethanol from amylaceous or saccharous feedstocks. Since acetone and butanol account for the most (usually > 90%) in the fermentation products, it is sometimes also called AB fermentation.

2.2.1 History

The production of butanol through fermentation was first reported by Pasteur in 1862 (Pasteur, 1862), while the microbial fermentation production of acetone was first reported by Schardinger in 1905 (Schardinger, 1905). At the beginning of 20th century, the butanol fermentation was actively studied in England for rubber synthesis purpose (Gabriel, 1928; Killeffer, 1927). In 1911, Fernbach isolated a culture that could ferment potato starch to produce butanol (Gabriel, 1928; Gabriel and Crawford, 1930). Between 1912 and 1914, Weizmann isolated several cultures; one of them was called BY, which was later named *Clostridium acetobutylicum* (Gabriel, 1928). The focus of ABE fermentation was changed to acetone production after the outbreak of World War I in 1914, because the British troops wanted more cordite for munitions manufacture, during which acetone was required as the solvent (Killeffer, 1927). During this period, Weizmann continued his research at the University of Manchester and made significant progress. He was issued a patent in March 1915 (Gabriel, 1928; Jones and Woods, 1986). However, one year later, Weizmann was summoned to the British Admiralty and was asked to start a project to manufacture acetone for the war purpose (Jones and Woods, 1986; Weizmann, 1963). In 1916, the British government requested a change to the Weizmann process to use maize as a substrate after it took possession of the Strange and Graham Ltd. Plant at King's Lynn in east of England (Jones and Woods, 1986; Ross, 1961). As a result, the acetone production was greatly increased. Later on, the fermentation was disrupted because of a lack of feedstock due to the German blockade, and ultimately the fermentation was transferred to Canada which is a part of the British Empire at that time and where suitable substrates were readily available for the fermentation (Gabriel, 1928; Weizmann, 1963). During that time, acetone was also produced in France and another plant was built in India, which was completed after the war and subsequently converted to a distillery by the Bombay government (Gabriel, 1928; Jones and Woods, 1986). In March 1917, United States joined the war and a project to produce acetone was initiated in the Midwest corn belt of the United States; some plants in Terre Haute, IN were operated from May to November in 1918 to produce acetone for cordite production and airplane dope, and butanol was produced meanwhile and mostly stored as a useless by-product (Gabriel, 1928; Jones and Woods, 1986; Ross, 1961). After the war, all plants were closed because acetone was no longer needed.

After World War I, the automobile industry was expanding rapidly and this resulted in a big demand for butanol as a suitable solvent during the production of quick-drying lacquers (Gabriel and Crawford, 1930; Killeffer, 1927). Commercial Solvents Corporation of Maryland was formed and they restarted the solvent production at the Terre Haute plants in 1920. The butanol production escalated during this period until 1923 when a major problem occurred to the fermentation, which was later diagnosed as a bacteriophage infection (Gabriel, 1928; Ross, 1961). While the phage infection was being investigated by an extensive research department, the company decided to erect a new plant in Peoria, IL in 1923 in order to fulfill the increasing demand for butanol (Hastings, 1971; Ross, 1961). The plant in Peoria expanded rapidly between 1924 and 1927 and there were 96 fermentors in operation by the end of 1927. Research was also carried out on the use of fermentation by-products, for example, the use of produced gases for the manufacture of ammonia and methanol (Gabriel, 1928; Jones and Woods, 1986). New strains were isolated with the purpose to ferment higher concentrations of starch to compete with the newly developed industry for acetone synthesis from petroleum, but little success was achieved (Hastings, 1971). At the beginning of 1930s, with the isolation of new microorganisms for fermenting sugars and the availability of cheap molasses, the grain mash fermentation substrate was partly substituted with molasses along with the addition of soybean meal and other protein sources. Unfortunately, this attempt resulted in much longer fermentation time and little success was achieved (Jones and Woods, 1986; McCutchan and Hickey, 1954). Later, several strains were obtained for fermenting molasses sugars at high concentrations through screening of the existing culture collections. Among them, the famous CSC No. 8 strain was able to ferment media containing 6% sugar and produce nearly 2% solvents, which significantly decreased the distillation cost for solvent recovery (Hastings, 1971; Jones and Woods, 1986). A plant went into operation at the end of 1935 at Bromborough in England, using a phage-immunized strain of *C. saccharoacetobutylicum* developed in the research lab at Terre Haute (Hastings, 1971). After 1936, the production of acetone and butanol from the fermentation of molasses progressed rapidly, and a number of new strains were isolated (Beesch, 1952; McCutchan and Hickey, 1954). When World War II started, the AB fermentation was again given top priority because of the high level demand for acetone for the manufacture of munitions (Hastings, 1971). As a result, the acetone production capacity was increased dramatically at Bromborough and the capacity of the existing batch-distillation plant was not enough. Therefore, multiple-column

continuous distillation units were developed and employed for the alcohol production industry (Hastings, 1978). The fermentation of acetone in Britain was not able to continue because of the problem for importing molasses during the war; acetone was imported from the United States, where the fermentation plants switched back to maize mash as the substrate due to a shortage of molasses (Johnson, 1947; Silcox, 1948).

After 1936, solvent plants were also built in many other countries, including Japan, India, Australia, South Africa, Egypt, Brazil, USSR and China, etc (Jones and Woods, 1986; Ross, 1961). In Japan, the fermentation plants continued to operate during World War II, and fermentation production of solvents did not cease until the 1960s, and the research on the process was continued for many more years afterwards (Jones and Woods, 1986). In 1937, an AB fermentation plant using maize mash as the substrate was built at Germiston in South Africa. The plant switched to molasses as the substrate in the late part of World War II and continued the operation until 1983 (Jones and Woods, 1986). During this period, acetone and butanol were also produced on a large scale by fermentation in the USSR, China and also some other Eastern countries. Continuous fermentation of acetone and butanol were reported during 1950s and 1960s, and a full factory-scale acetone production process came into operation in USSR in 1960. Maize, wheat, and rye were all used for AB fermentation in Russia, and later other substrates were also used for substrates, such as mixtures of molasses and grain and corn cob hydrolysates (Jones and Woods, 1986; Nakhmanovich, 1959).

After World War II, the acetone and butanol production through fermentation declined rapidly and finally ceased by 1960 in the United States and Britain because of the competition from the petro-chemically produced solvents and also the shortage of cheap molasses (Hastings, 1971, 1978). The solvent production through AB fermentation in South Africa continued operating for a longer time thanks to the sufficient availability of cheap molasses and coal and the lack of cheap petroleum. However, the plants were forced to close because of a shortage of molasses in early 1980s (Jones and Woods, 1986).

2.2.2 Current status and future prospects

Over past few decades, almost all butanol has been produced via petrochemical routes. In recent years, increasing crude oil price and concerns about environmental problems have renewed the interests in butanol production through biological fermentation processes. Great efforts have been focused on the academic research on the ABE fermentation employing

solventogenic clostridia. A hyper-amylolytic mutant *C. beijerinckii* BA101 was generated using N-methyl-N'-nitro-N-nitrosoguanidine together with selective enrichment on the glucose analog 2-deoxyglucose, which can produce a butanol concentration 100% higher than its parent strain (Annous and Blaschek, 1991). Transcriptional analyses and metabolic engineering have been extensively performed on C. *acetobutylicum* (Alsaker and Papoutsakis, 2005; Alsaker et al., 2010; Alsaker et al., 2004; Green et al., 1996; Heap et al., 2007; Papoutsakis, 2008; Shao et al., 2007). Conversion of low-value feedstocks like agricultural residues and optimization of fermentation process and downstream manipulations have also been studied (Ezeji et al., 2007a, b; Ezeji et al., 2005; Mariano et al., 2011; Qureshi et al., 2008). In industry, in recent years, a number of companies have started to market butanol and develop the bio-butanol processes. Many venture companies have also been established focusing on the commercialization of bio-butanol. Several major players include Cathay Industrial Biotech (Shanghai, China), Cobalt Technologies (Mountain view, CA, USA), Green Biologics (Abingdon, England), METabolic EXplorer (Clermont-Ferrand, France), etc. (Table 2.2) (Mascal, 2012). With the blending wall issue for bioethanol industry, and the expiration of the blenders tax credit (NYTimes, 2012), many ethanol producers are considering retrofitting and transforming their existing ethanol production infrastructures for butanol production (Ethanolproducer.com, 2012; WSJ, 2009). In sum, the prospects for biological butanol production are very bright.

However, there are still many challenges for the large-scale industrial production of butanol through clostridia fermentation processes. For example, the high cost of substrates, the low yield and final titer for the traditional ABE fermentation process, diversion of substrates to other unwanted bioproducts (like acetone and acids), and the high energy requirement for butanol recovery from dilute streams etc. Great efforts on overcoming these challenges should be made in order to significantly improve the butanol production process. First, substrate normally accounts for most of the total cost associated with the ABE fermentation (Qureshi and Blaschek, 2001). So, the search and utilization of low-value and abundant substrates that do not compete with food and feed supplies would offer tremendous economic benefits. These include the cellulosic residues or byproducts from agricultural and industrial processes, municipal wastes, and other energy biomass. Effective pretreatment and biomass conversion techniques must be developed to satisfy the efficient fermentation requirement. Second, search or development of new strains and improvement of the existing

strains should be undertaken in order to acquire degradation capabilities on vast ranges of substrates, higher butanol yield, titer and productivity. In this process, transcriptional analysis, metabolic engineering, and system biology tools will play key roles. On one hand, butanol producers such as the solventogenic clostridia can be mutated or engineered for desired properties. On the other, butanol production pathways can be introduced into a desired industrial microorganism for butanol production purpose. For example, *Escherichia coli* has been used as a host and engineered for this objective (Ranganathan and Maranas, 2010; Shen et al., 2011). Adapted *Pseudomonas putida* strains that can grow in the presence of up to 6% (vol/vol) butanol have been recently reported, which could be potential hosts for biobutanol production (Ruhl et al., 2009). Third, development of novel fermentation processes and improvement of the downstream processing would be significant for improving the butanol production economics.

Due to high energy cost, the traditional distillation method for butanol separation from the dilute fermentation broth (approximate 1%) is not desirable. It was estimated that if the final butanol concentration in the fermentation broth is increased to 2%, the energy requirement for recovering butanol by distillation would be lowered by 62% (Phillips and Humphrey, 1983). Several fermentation processes and *in situ* recovery approaches have been examined in order to lower the recovery energy cost (Dürre, 1998; Ezeji et al., 2004). Other alternative possibilities include adsorption, gas stripping, liquid-liquid extraction, perstraction, pervaporation, and reverse osmosis, etc (Lee et al., 2008; Qureshi and Ezeji, 2008). Recently, a novel fermentation process with simultaneous ABE fermentation and *in situ* product recovery with vacuum was reported (Mariano et al., 2011, 2012). With the advancement of separation technology and fermentation engineering, more efficient and cost-effective downstream processes are expected in the future.

Table 2.2 Major companies currently marketing n-butanol

Company	Time founded	Location	Technology	Development status
Butylfuel, LLC	1991	Gahanna, Ohio, USA	Based on a two-stage fermentation that decouples butyric acid production from butanol formation.	A pilot plant incorporating this technology in a fibrous bed bioreactor is under development. Green Biologics merged with Butylfuel in January 2012 (Green Biologics press release, 2012).
Cathay Industrial Biotech	1997	Shanghai, China	Continuous clostridia ABE fermentation, with corn starch as the feedstock; optimized process reduced water and energy usage significantly.	Currently the largest biobutanol producer in the world, with a production capacity of 21 million gallons of biobutanol per year.
METabolic EXplorer	1999	Clermont-Ferrand, France	Using its combined expertise in molecular biology, metabolic engineering and bioinformatics to design high-performance microorganisms that can transform plant-derived raw materials into an existing bulk chemical, including butanol.	With a broad product portfolio, they are developing "an extremely flexible and competitive fermentation process for butanol production," which can utilize sugar, starch, and hemicellulosic feedstocks.
Green Biologics	2003	Abington, UK	The technology platform includes a large collection of native and genetically modified clostridia, fermentation process design, production strain to feedstock matching, and separations.	Merged with Butylfuel in January 2012; is aiming at retrofit existing ethanol plants and international market expansion including China, Brazil, etc (Ethanolproducer.com, 2012).
Cobalt Technologies	2005	Mountain View, CA, USA	An immobilized cell technology using a proprietary, non-genetically modified clostridia strain with high productivity and ability to convert both C5 and C6 sugars into butanol for operation in a continuous fermentation reactor.	Currently operating a 5000 GPY pilot facility since June 2009. Cobalt is now in the process to commission a 470,000 GPY cellulosic biobutanol facility.
Tetravitae Bioscience (Eastman Chemical)	2006	Chicago, IL, USA	The platform is based on a stable, non-genetically modified *Clostridium beijerinckii* strain which offers high butanol yields, reduced product inhibition, and the ability to utilize both C5 and C6 sugars.	Trying to adapt the technology to process cellulosic feedstocks; acquired by Eastman Chemical Company in November 2011 (Eastman press release, 2011)
Butalco	2007	Canton of Zug, Switzerland	Based on genetically optimized yeasts for production of advanced biofuels like ethanol and butanol by using pentose sugars	Working together with partners to develop integrated production processes (Butalco.com, 2012).

2.3 Formation of solvents in clostridia

Based on 16S rRNA gene sequencing and DNA-DNA reassociation results, the industrial solvent-producing clostridia have been recently definitively assigned into four species: *C. acetobutylicum*, *C. beijerinckii*, *C. saccharobutylicum* and *C. saccharoperbutylacetonicum* (Keis et al., 2001). *C. acetobutylicum* and *C. beijerinckii* are undoubtedly the prominent ones due to their hypersolvent-producing capability (Dürre, 2008). Extensive transcriptional analyses and genetic studies have been performed, especially for *C. acetobutylicum*.

2.3.1 Physiology of ABE fermentation by solventogenic clostridia

The metabolic pathways for butanol production in solventogenic clostridia are very similar (Zheng et al., 2009). A typical feature of an ABE batch fermentation by solventogenic clostridia is its biphasic nature. The biochemical pathways of the fermentation with the model microorganism *C. acetobutylicum* are shown in Figs. 2.1 & 2.2 (Jones and Woods, 1986). In the initial growth phase, the solvent-producing clostridia first mainly produce acetate, butyrate, hydrogen and carbon dioxide; this phase is also termed as acidogenic phase. After the culture enters into the stationary phase (solventogenic phase), the formation of acids decreases, and acetone and butanol become the dominant products (with a small fraction of ethanol). A portion of the previously produced acids is re-assimilated and converted into solvents. During these processes, the pH first decreases to lower levels due to the accumulation of acids and then increases after solventogenesis is initiated.

As shown in Figure 2.1, the hexose sugars are metabolized via glycolysis process with a conversion of 1 mole of hexose to 2 moles of pyruvate, with a net production of 2 moles of adenosine triphosphate (ATP) and 2 moles of reduced nicotinamide adenine dinucleotide (NADH). The pyruvate is then cleaved by pyruvate ferredoxin oxidoreductase to produce carbon dioxide, acetyl-CoA and reduced ferredoxin (Jones and Woods, 1986). During the acidogenic phase, acetyl-CoA is first phosphorylated to acetyl phosphate by phosphotransacetylase (pta) and then converted to acetate by acetate kinase (ack). Butyryl-CoA is formed from acetyl-CoA following a metabolic pathway with the assistance of four enzymes: thiolase (thl), 3-hydroxybutyryl-CoA dehydrogenase (hbd), crotonase (crt) and butyryl-CoA dehydrogenase (bcd). Butyryl-CoA is thereafter converted to butyryl phosphate by phosphotransbutyrylase (phosphate butyryltransferase, ptb). Butyrate is produced via

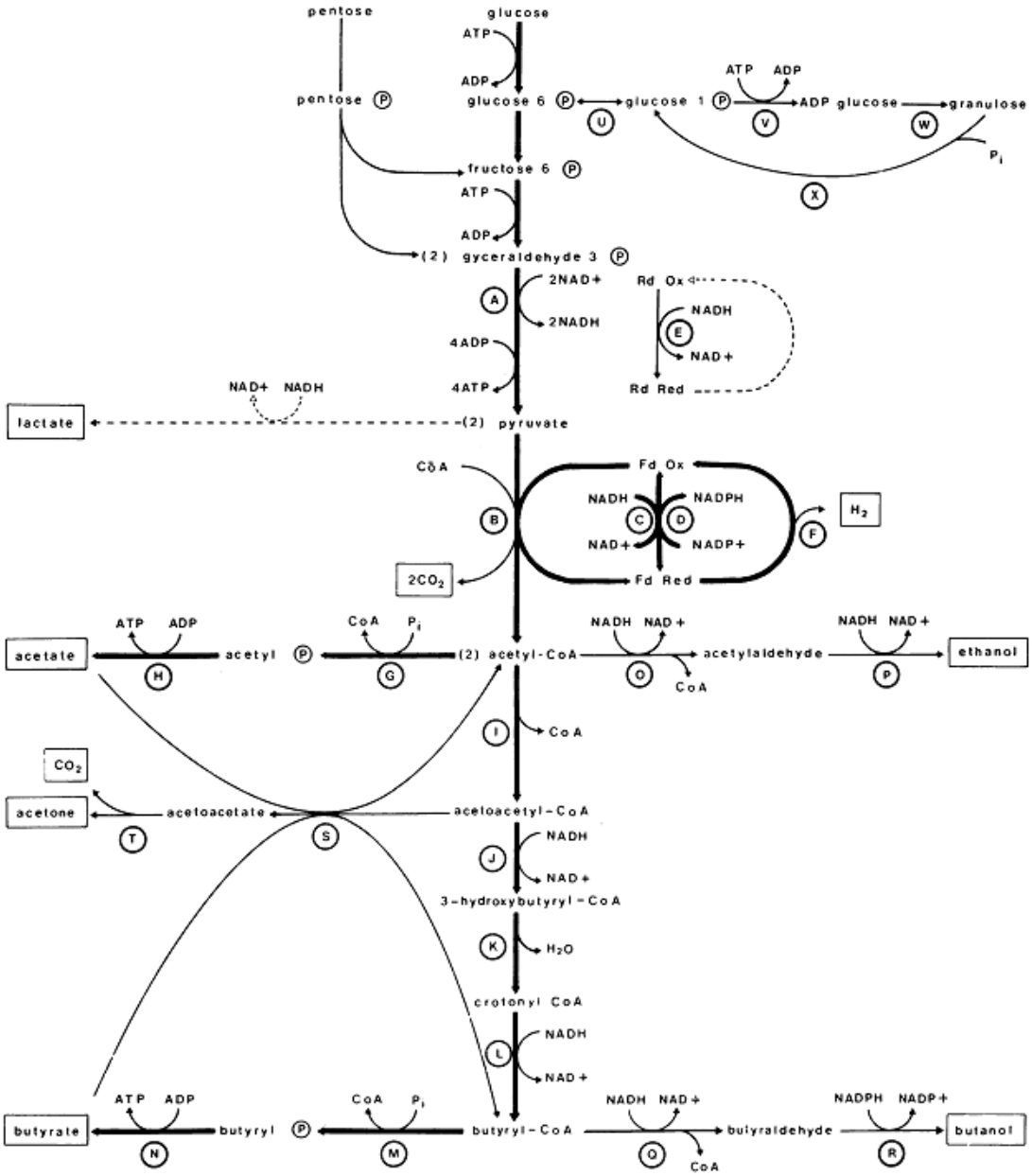


Figure 2.1 Biochemical pathways in solventogenic clostridia. Acids formation reactions predominate during the acidogenic phase (highlighted with thick arrows) (Jones and Woods, 1986). Enzymes: (A) glyceraldehyde 3-phosphate dehydrogenase; (B) pyruvate-ferredoxin oxidoreductase; (C) NADH-ferredoxin oxidoreductase; (D) NADPH-ferredoxin oxidoreductase; (E) NADH rubredoxin oxidoreductase; (F) hydrogenase; (G) phosphate acetyltransferase (phosphotransacetylase, pta); (H) acetate kinase (ack); (i) thiolase (acetyl-CoA acetyltransferase, thl); (J) 3-hydroxybutyryl-CoA dehydrogenase (hbd); (K) crotonase (crt); (L) butyryl-CoA dehydrogenase (bcd); (M) phosphate butyltransferase (phosphotransbutyrylase, ptb); (N) butyrate kinase (buk); (O) acetaldehyde dehydrogenase (adhE); (P) ethanol dehydrogenase (bdhA, bdhB); (Q) butyraldehyde dehydrogenase; (R) butanol dehydrogenase (bdhA, bdhB); (S) acetoacetyl-CoA:acetate/butyrate:CoA transferase (ctfA, ctfB); (T) acetoacetate decarboxylase (adc); (U) phosphoglucomutase; (V) ADP-glucose pyrophosphorylase; (W) granulose synthase; (X) granulose phosphorylase.

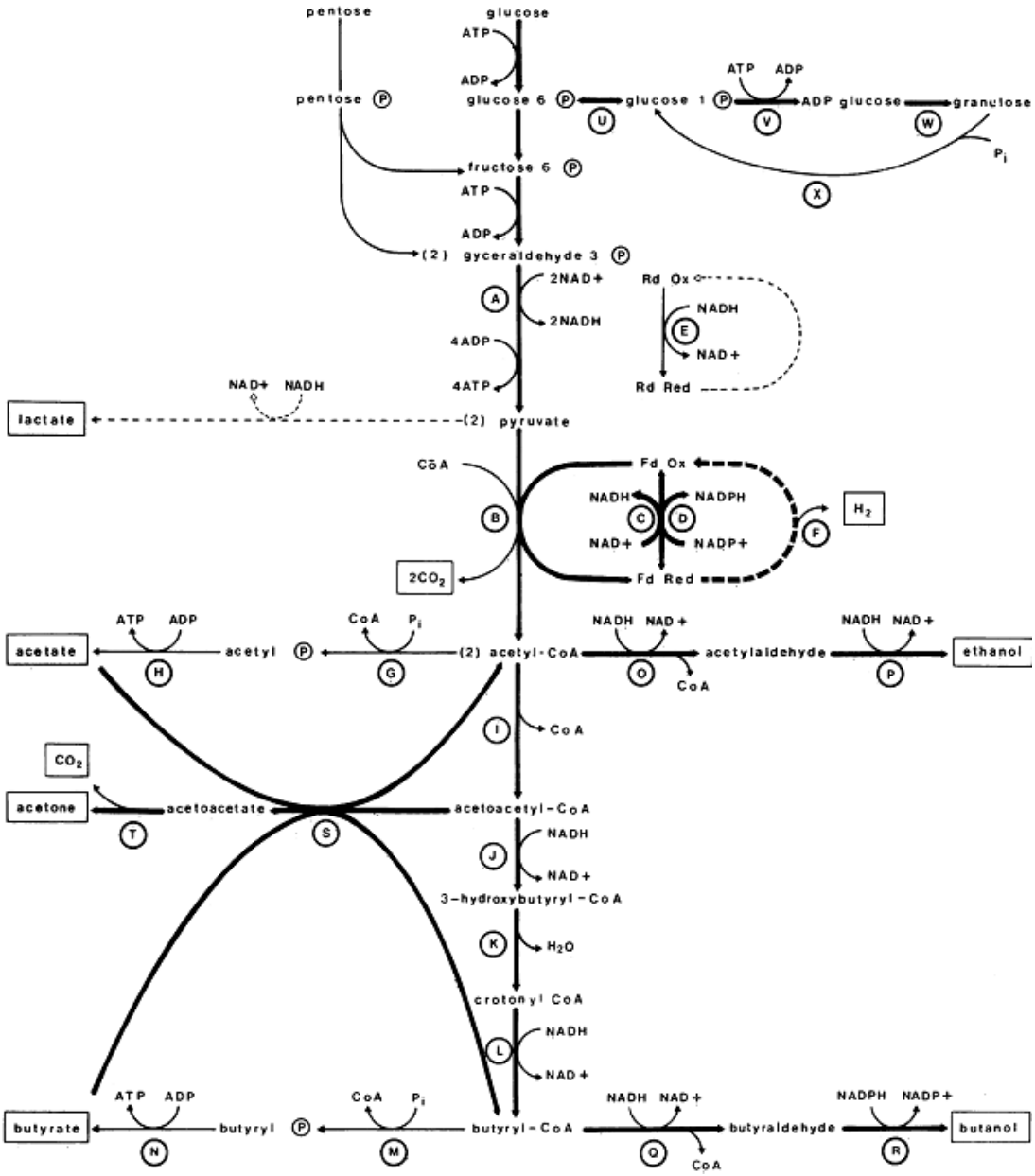


Figure 2.2 Biochemical pathways in solventogenic clostridia. Solvents formation reactions predominate during the solventogenic phase (highlighted with thick arrows) (Jones and Woods, 1986). Enzymes: (A) glyceraldehyde 3-phosphate dehydrogenase; (B) pyruvate-ferredoxin oxidoreductase; (C) NADH-ferredoxin oxidoreductase; (D) NADPH-ferredoxin oxidoreductase; (E) NADH rubredoxin oxidoreductase; (F) hydrogenase; (G) phosphate acetyltransferase (phosphotransacetylase, pta); (H) acetate kinase (ack); (i) thiolase (acetyl-CoA acetyltransferase, thl); (J) 3-hydroxybutyryl-CoA dehydrogenase (hbd); (K) crotonase (crt); (L) butyryl-CoA dehydrogenase (bcd); (M) phosphate butyltransferase (phosphotransbutyrylase, ptb); (N) butyrate kinase (buk); (O) acetaldehyde dehydrogenase (adhE); (P) ethanol dehydrogenase (bdhA, bdhB); (Q) butyraldehyde dehydrogenase; (R) butanol dehydrogenase (bdhA, bdhB); (S) acetoacetyl-CoA:acetate/butyrate:CoA transferase (ctfA, ctfB); (T) acetoacetate decarboxylase (adc); (U) phosphoglucomutase; (V) ADP-glucose pyrophosphorylase; (W) granulose synthase; (X) granulose phosphorylase.

dephosphorylation of butyryl phosphate by butyrate kinase (buk). The onset of solvent production involves a switch from the acid-producing pathways to the solvent-producing pathways (Figure 2.2). In solventogenic phase, acetate is metabolized by acetoactyl-CoA:acetate/butyrate-CoA transferase (ctf) into acetyl-CoA and acetoacetate. The latter is split into acetone and carbon dioxide by acetoacetate decarboxylase (adc). The acetyl-CoA then goes through two pathways. In one pathway, it is converted to ethanol by acetaldehyde dehydrogenase (adhE) and ethanol dehydrogenase (bdhA, bdhB); in another, it is converted to butyryl-CoA through the butyryl-CoA formation pathway. On the other hand, butyrate produced during acidogenic phase is converted to butyryl-CoA by the same ctf enzyme as described above. Butyryl-CoA is thereafter converted to butyraldehyde by butyraldehyde dehydrogenase (adhE) and finally to butanol by butanol dehydrogenase (bdhA, bdhB).

During the acidogenic phase, the accumulation of acids in the fermentation broth imposes a severe threat for the cells. At a low pH, the butyrate and acetate produced will present as the undissociated form and be able to pass through the cytoplasmic membrane via diffusion. Inside the cells, the acids will dissociate to form salts and protons due to the internal higher pH. Thus, the proton gradient across the membrane will be destroyed and the cells will die. Be decreasing the concentration of acids, solventogenesis lessens this effect. However, the produced solvents, especially high concentration of butanol is also toxic to the microorganism. Therefore, solventogenesis can only be a kind of emergency reaction to buy time for the cells to go through endospore formation for long-term survival. So, in solvent-producing clostridia, solventogenesis and sporulation are usually concurrently (Dürre, 2005).

2.3.2 Effects of butyric acid on the metabolic pathway switch

The level of undissociated butyric acid has long been suggested as the controlling factor for the shift from acidogenesis to solventogenesis in the metabolic pathway (Jones and Woods, 1986). Monot et al. (1983) reported that, when the concentration of undissociated butyric acid reached a level of 0.5 to 1.5 g L^{-1}, solvent production was initiated (Monot et al., 1983). Terracciano and Kashket determined that solvent production began when the undissociated butyric acid level reached 13-18 mM (Terracciano and Kashket, 1986). Husemann and Papoutsakis reported that 6–13 mM of undissociated butyric acid was required for the initiation of solventogenesis when the external pH was between 3.7 and 5.0 (Husemann and Papoutsakis, 1988). However, Harris et al. later suggested that butyrate might

not be involved in the induction of solventogenesis during a study in which they obtained high solvent production from the fermentation with a butyrate kinase inactivated mutant of *C. acetobutylicum* (Harris et al., 2000). Zhao et al. (2005) further indicated that butyryl phosphate (but not acetyl phosphate) correlated with (and possibly regulated) the initiation of solvent formation through a sensitive measurement of butyryl phosphate and acetyl phosphate levels in the wild type of *C. acetobutylicum* ATCC 824 and several mutant strains (Zhao et al., 2005).

2.4 Transcriptomic analysis for solventogenic clostridia

With the development of transcriptional analysis tools, transcriptomic analysis has been intensively performed on the solventogenic clostridia, especially for *C. acetobutylicum*. As early as 1999, a gene expression reporter system was developed using the *lacZ* gene of *Thermoanaerobacterium sulfurogenes* EM1 to examine the expression pattern of key metabolic pathway genes in *C. acetobutylicum* 824 (Tummala et al., 1999). In 2003, a reporter system *gusA* (designed with a beta-glucuronidase from *E. coli*) was employed to examine gene expression in *C. acetobutylicum* (Girbal et al., 2003). In 2002, Northern blotting was first used to study the expression of key genes involved in solvent production with *spo0A* inactivation and overexpression in *C. acetobutylicum* ATCC 824, although this technique can only analyze small numbers of genes and samples in a single experiment (Harris et al., 2002). With the development of microarray technology, its ability to simultaneously determine the expression levels of thousands of genes in one experiment was soon adopted for the transcriptional analysis of solventogenic clostridia, and it is meanwhile often a tool for confirming the results of metabolic engineering. Glass DNA arrays containing a selected set of 1,019 genes (including all 178 mega-plasmid pSOL1 genes) covering more than 25% of the whole genome were used for the transcriptional analysis of two *C. acetobutylicum* strains (SKO1: with an inactivated *spo0A* gene; M5: a degeneration strain due to the loss of pSOL1 plasmid) relative to that of the wide type parent strain. Results indicated that the SKO1 and M5 strains displayed down-regulation of most solvent formation genes, but high-regulation of several motility and chemotaxis genes (Tomas et al., 2003a). Transcriptional analysis of a *spo0A* overexpression strain 824(pMSPOA) against its plasmid control 824(pIMP1) with the same DNA microarray set confirmed that *spo0A* overexpression led to earlier induction of solventogenic gene expression. In addition, several fatty acid

metabolism genes, heat shock proteins, and the Fts family of cell division proteins were differentially expressed, suggesting their roles in sporulation and the solvent stress response, while many glycolytic genes during stationary phase were down-regulated (Alsaker et al., 2004). The same DNA microarray set containing selected genes covering more than 25% of the whole genome of *C. acetobutylicum* was also used for transcriptional analyses of the effects of *ctfA/B* downregulation by antisense RNA (asRNA) manipulation approach (Tummala et al., 2003), *groESL* overexpression and host-plasmid interactions (Tomas et al., 2003b), and the butanol stress and tolerance (Tomas et al., 2004). Later, a full-genome microarray was designed and used for gene expression changes during the shift from exponential phase and acidogenesis to stationary phase and solventogenesis of *C. acetobutylicum* (Alsaker and Papoutsakis, 2005; Alsaker et al., 2005). Recently, genome-wide microarray analyses were also performed to study the sporulation events and metabolite (butanol, butyrate and acetate) stress and tolerance (Alsaker et al., 2010; Jones et al., 2008).

In comparison, the transcriptional analysis for *C. beijerinckii* is far behind. In 2000, a gene expression reporter system was developed using a beta-1, 4-endoglucanase gene (*eglA*) for the transcriptional analysis of *C. beijerinckii* 8052 (Quixley and Reid, 2000). In 2008, Shi and Blaschek (Shi and Blaschek, 2008) conducted a transcriptional analysis for *C. beijerinckii* 8052 and its hyper-butanol-producing mutant BA101 during the shift from acidogenesis to solventogenesis using a ca. 500-gene set DNA microarray. To our best knowledge, this is the only reported transcriptional analysis for *C. beijerinckii* using microarray technology. Therefore, high-throughput genome-wide transcriptional analyses for *C. beijerinckii* under various conditions are desirable in order to better understand the physiology and metabolism of this strain and further develop this strain for improved solvent production and other benefits. The lag in transcriptional analysis for *C. beijerinckii* is to a large extent due to the previously limited availability of genomic and transcriptomic information for this species.

2.5 RNA-Seq technology

For the past > 10 years, genome sequencing and microarray technology provided us the power for genomics and transcriptional analyses. Now, high-throughput next generation DNA sequencing technology is making and will make substantial contribution to our

understanding of genomics, transcriptome organization, gene expression and regulation networks.

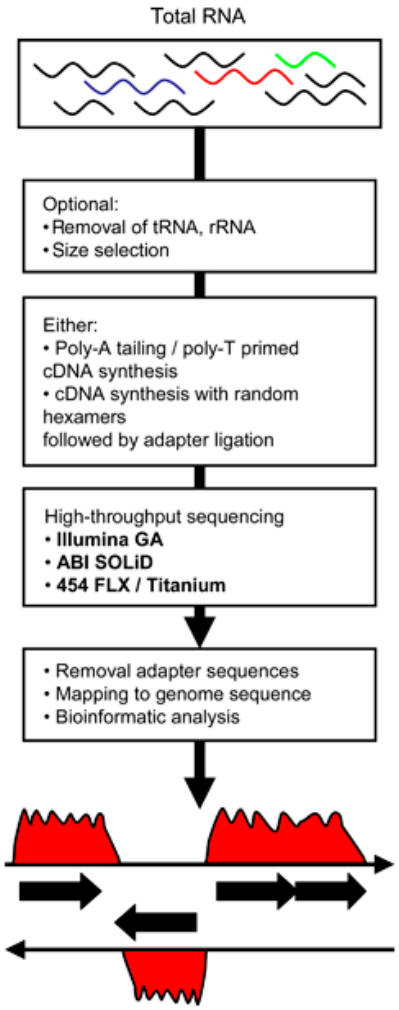


Figure 2.3 Flow diagram of the steps involved in RNA-Seq for microbial transcriptome sequencing (van Vliet, 2010).

2.5.1 Origination and principles

In the past few years, microarray has made it possible to examine the expression levels of the genomic annotated genes in the cell population. With the recent realization that RNAs from non-coding regions also play fundamental roles, new generations of microarrays, called tilling-arrays were designed to study the genome-wide transcriptome and transcription regardless of the gene annotation. Next-generation sequencing technology found its market at this time (Marguerat and Bahler, 2010). Next-generation sequencing technology enabled us to obtain massive amounts of sequence data rapidly and cost-effectively. RNA-Seq is thus developed for transcriptomic analysis based on the novel high-throughput DNA sequencing

technologies (Lister et al., 2009; Marguerat et al., 2008; Wang et al., 2009; Wilhelm and Landry, 2009).

Generally, the total RNA is first extracted from cells and tRNA and rRNA are removed in order to retrieve more valuable data from the remaining portion (mRNA) of the transcriptome. Then, mRNA is reverse transcribed and a library of cDNA fragments is constructed with barcoded adaptors attached to the ends. The templates are thereafter sequenced on a high-throughput sequencing platform (Roche 454 (Vera et al., 2008), Illumina (Filiatrault et al., 2010; Passalacqua et al., 2009; Wurtzel et al., 2010; Zenoni et al., 2010) or ABI SOLiD (Passalacqua et al., 2009; Yang et al., 2010)) to obtain a large amount of short sequence reads, which are usually 35-400 bp long depending on the employed DNA sequencing technology. The short reads can then be aligned back to the reference genome or reference transcripts to produce a genome-wide map consisting of both transcriptome structural and expression level information for each gene (Figure 2.3).

2.5.2 Advantages and challenges

RNA-Seq has several advantages over traditional hybridization-based techniques, such as the DNA microarray. First, microarray and other hybridization techniques have a relatively low dynamic range for the detection of transcription levels due to background, saturation and poor sensitivity for gene expression at very low or high levels (van Vliet, 2010), while RNA-Seq does not have a upper limit for quantification and has a large dynamic range of expression levels. In addition, compared to microarray technology, RNA-Seq has very low, if any, background signal because the sequence reads can be unambiguously mapped to unique regions along the genome (Wang et al., 2009). Furthermore, with proper normalization, gene expression can be quantified accurately with RNA-Seq technology, and this makes the comparison of transcription levels among experiments possible (Mortazavi et al., 2008); while this is challenging for microarray method and usually requires complex normalization (Hinton et al., 2004). Besides, RNA-Seq can map the precise location of transcript boundaries at a single base resolution, which is extremely useful for studying the complex transcriptome organization of microorganisms (Passalacqua et al., 2009). Last but not least, RNA-Seq can also detect and define transcripts from non-coding regions, including non-coding small RNAs or antisense RNAs which may play essential regulatory functions (Perkins et al., 2009).

Along with the rapid development of RNA-Seq, there are still several issues that may require enough attention. First, one of the problems for bacterial RNA is that rRNA and tRNA often accounts for 50-80% of the total RNA. The published sequencing-based microbial transcriptomics to date have partially removed these rRNA and tRNA sequences. It is unknown what effects this could have on the composition of the RNA fraction and following sequencing. So, in the future, it may be rational to keep this fraction and construct an unbiased cDNA library with the development of this technology (van Vliet, 2010). Second, the construction of cDNA library involves several manipulation stages, which can introduce biases before sequencing and complicate the use of RNA-Seq for profiling all types of transcripts (Wang et al., 2009). In addition, the large datasets produced by RNA-Seq face informatics challenges, including the requirement for large storage space, accurate sequence determination and mapping reads onto the genome, data normalization, visualization, analysis and interpretation. Finally, proper standardization and guidelines need to be developed. Like for microarray studies, the MIAME guidelines established the minimal requirements for publication of microarray datasets in journals and online databases (Brazma et al., 2001), similar guidelines are needed for RNA-Seq.

Despite the potential challenges, RNA-Seq technology already demonstrated its power for studying the high-definition transcription events for different microorganisms, and meanwhile improving and correcting the annotation of existing genome sequences (Denoeud et al., 2008; Passalacqua et al., 2009; Perkins et al., 2009). Using RNA-Seq, the transcription events of microorganisms can be studied in a high-throughput and quantitative manner, with both single-base resolution annotation and 'digital' gene expression levels over a genome scale.

2.6 References

Alsaker, K. V., and E. T. Papoutsakis. 2005. Transcriptional program of early sporulation and stationary-phase events in *Clostridium acetobutylicum*. *Journal of Bacteriology* 187(20):7103-7118.

Alsaker, K. V., C. Paredes, and E. T. Papoutsakis. 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*. *Biotechnology and Bioengineering* 105(6):1131-1147.

Alsaker, K. V., C. J. Paredes, and E. T. Papoutsakis. 2005. Design, optimization and validation of genomic DNA microarrays for examining the *Clostridium acetobutylicum* transcriptome. *Biotechnology and Bioprocess Engineering* 10(5):432-443.

Alsaker, K. V., T. R. Spitzer, and E. T. Papoutsakis. 2004. Transcriptional analysis of *spo0A* overexpression in *Clostridium acetobutylicum* and its effect on the cell's response to butanol stress. *Journal of Bacteriology* 186(7):1959-1971.

Annous, B. A., and H. P. Blaschek. 1991. Isolation and characterization of *Clostridium acetobutylicum* mutants with enhanced amylolytic activity. *Applied and Environmental Microbiology* 57(9):2544-2548.

Beesch, S. C. 1952. Acetone-butanol fermentation of sugars. *Engineering and Process Development* 44:1677-1682.

Brazma, A., P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, W. Ansorge, C. A. Ball, H. C. Causton, T. Gaasterland, P. Glenisson, F. C. P. Holstege, I. F. Kim, V. Markowitz, J. C. Matese, H. Parkinson, A. Robinson, U. Sarkans, S. Schulze-Kremer, J. Stewart, R. Taylor, J. Vilo, and M. Vingron. 2001. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nature Genetics* 29(4):365-371.

Butalco.com. 2012. http://www.butalco.com.

Davis, S. E., and S. A. Morton. 2008. Investigation of ionic liquids for the separation of butanol and water. *Separation Science and Technology* 43(9-10):2460-2472.

Denoeud, F., J. M. Aury, C. Da Silva, B. Noel, O. Rogier, M. Delledonne, M. Morgante, G. Valle, P. Wincker, C. Scarpelli, O. Jaillon, and F. Artiguenave. 2008. Annotating genomes with massive-scale RNA sequencing. *Genome Biology* 9(12):R175.

Dürre, P. 1998. New insights and novel developments in clostridial acetone/butanol/isopropanol fermentation. *Applied Microbiology and Biotechnology* 49(6):639-648.

Dürre, P. 2005. Formation of solvents in clostridia. In *Handbook on clostridia*, 671-693. P. Dürre, ed. London: CRC press.

Dürre, P. 2007. Biobutanol: An attractive biofuel. *Biotechnology Journal* 2(12):1525-1534.

Dürre, P. 2008. Fermentative butanol production-Bulk chemical and biofuel. In *Incredible Anaerobes: From Physiology to Genomics to Fuels*, 353-362. J. Wiegel, R. J. Maier, and M. W. W. Adams, eds. Oxford: Blackwell Publishing.

Eastman press release. 2011. http://www.eastman.com/Company/News_Center/2011/Pages/Eastman_Acquires_TetraVitae_Bioscience_Assets.aspx.

Ethanolproducer.com. 2012. http://www.ethanolproducer.com/articles/8578/target-for-transformation.

Ezeji, T., N. Qureshi, and H. P. Blaschek. 2007a. Butanol production from agricultural residues: impact of degradation products on *Clostridium beijerinckii* growth and butanol fermentation. *Biotechnology and Bioengineering* 97(6):1460-1469.

Ezeji, T., N. Qureshi, and H. P. Blaschek. 2007b. Production of acetone-butanol-ethanol (ABE) in a continuous flow bioreactor using degermed corn and *Clostridium beijerinckii*. *Process Biochemistry* 42(1):34-39.

Ezeji, T. C., P. M. Karcher, N. Qureshi, and H. P. Blaschek. 2005. Improving performance of a gas stripping-based recovery system to remove butanol from *Clostridium beijerinckii* fermentation. *Bioprocess and Biosystems Engineering* 27(3):207-214.

Ezeji, T. C., N. Qureshi, and H. P. Blaschek. 2004. Butanol fermentation research: upstream and downstream manipulations. *Chemical Record* 4(5):305-314.

Filiatrault, M. J., P. V. Stodghill, P. A. Bronstein, S. Moll, M. Lindeberg, G. Grills, P. Schweitzer, W. Wang, G. P. Schroth, S. J. Luo, I. Khrebtukova, Y. Yang, T. Thannhauser, B. G. Butcher, S. Cartinhour, and D. J. Schneider. 2010. Transcriptome analysis of *Pseudomonas syringae* identifies new genes, noncoding RNAs, and antisense activity. *Journal of Bacteriology* 192(9):2359-2372.

Gabriel, C. L. 1928. Butanol fermentation process. *Industrial & Engineering Chemistry* 20:1063-1067.

Gabriel, C. L., and F. M. Crawford. 1930. Development of the butyl-acetonic fermentation industry. *Industrial & Engineering Chemistry* 22:1163-1165.

Girbal, L., I. Mortier-Barriere, F. Raynaud, C. Rouanet, C. Croux, and P. Soucaille. 2003. Development of a sensitive gene expression reporter system and an inducible promoter-repressor system for *Clostridium acetobutylicum*. *Applied and Environmental Microbiology* 69(8):4985-4988.

Green Biologics press release. 2012. http://www.greenbiologics.com/pr-butylfuel.php.

Green, E. M., Z. L. Boynton, L. M. Harris, F. B. Rudolph, E. T. Papoutsakis, and G. N. Bennett. 1996. Genetic manipulation of acid formation pathways by gene inactivation in *Clostridium acetobutylicum* ATCC 824. *Microbiology-Sgm* 142:2079-2086.

Harris, L. M., R. P. Desai, N. E. Welker, and E. T. Papoutsakis. 2000. Characterization of recombinant strains of the *Clostridium acetobutylicum* butyrate kinase inactivation mutant: Need for new phenomenological models for solventogenesis and butanol inhibition? *Biotechnology and Bioengineering* 67(1):1-11.

Harris, L. M., N. E. Welker, and E. T. Papoutsakis. 2002. Northern, morphological, and fermentation analysis of *spo0A* inactivation and overexpression in *Clostridium acetobutylicum* ATCC 824. *Journal of Bacteriology* 184(13):3586-3597.

Hastings, J. H. J. 1971. Development of the fermentation industries in Great Britain. In *Advances in Applied Microbiology*. D. Perlman, ed. New York: Academic Press, Inc.

Hastings, J. H. J. 1978. Acetone-butyl alcohol fermentation. In *Economic Microbiology*, 31-45. A. H. Rose, ed. New York: Academic Press, Inc.

Heap, J. T., O. J. Pennington, S. T. Cartman, G. P. Carter, and N. P. Minton. 2007. The ClosTron: A universal gene knock-out system for the genus Clostridium. *Journal of Microbiological Methods* 70(3):452-464.

Hinton, J. C. D., I. Hautefort, S. Eriksson, A. Thompson, and M. Rhen. 2004. Benefits and pitfalls of using microarrays to monitor bacterial gene expression during infection. *Current Opinion in Microbiology* 7(3):277-282.

Husemann, M. H. W., and E. T. Papoutsakis. 1988. Solventogenesis in *Clostridium acetobutylicum* fermentation related to carboxylic acid and proton concentration. *Biotechnology and Bioengineering* 32(7):843-852.

Johnson, M. J. 1947. Industrial fermentations. *Annual Review of Microbiology* 1:159-172.

Jones, D. T., and D. R. Woods. 1986. Acetone-butanol fermentation revisited. *Microbiological Reviews* 50(4):484-524.

Jones, S. W., C. J. Paredes, B. Tracy, N. Cheng, R. Sillers, R. S. Senger, and E. T. Papoutsakis. 2008. The transcriptional program underlying the physiology of clostridial sporulation. *Genome Biology* 9(7):R114.

Keis, S., R. Shaheen, and D. T. Jones. 2001. Emended descriptions of *Clostridium acetobutylicum* and *Clostridium beijerinckii*, and descriptions of *Clostridium saccharoperbutylacetonicum* sp. nov. and *Clostridium saccharobutylicum* sp. nov. *International Journal of Systematic and Evolutionary Microbiology* 51:2095-2103.

Killeffer, D. H. 1927. Butanol and acetone from corn-A description of the fermentation process. *Industrial & Engineering Chemistry* 19:46-50.

Kirschner, M. 2006. n-Butanol. In *Chemical Market Reporter*, 42. ABI/INFORM Global.

Lee, S. Y., J. H. Park, S. H. Jang, L. K. Nielsen, J. Kim, and K. S. Jung. 2008. Fermentative butanol production by clostridia. *Biotechnology and Bioengineering* 101(2):209-228.

Lister, R., B. D. Gregory, and J. R. Ecker. 2009. Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. *Current Opinion in Plant Biology* 12(2):107-118.

Marguerat, S., and J. Bahler. 2010. RNA-seq: from technology to biology. *Cellular and Molecular Life Sciences* 67(4):569-579.

Marguerat, S., B. T. Wilhelm, and J. Bahler. 2008. Next-generation sequencing: applications beyond genomes. *Biochemical Society Transactions* 36:1091-1096.

Mariano, A. P., N. Qureshi, R. Maciel, and T. C. Ezeji. 2011. Bioproduction of butanol in bioreactors: new insights from simultaneous in situ butanol recovery to eliminate product toxicity. *Biotechnology and Bioengineering* 108(8):1757-1765.

Mariano, A. P., N. Qureshi, R. Maciel, and T. C. Ezeji. 2012. Assessment of in situ butanol recovery by vacuum during acetone butanol ethanol (ABE) fermentation. *Journal of Chemical Technology and Biotechnology* 87(3):334-340.

Mascal, M. 2012. Chemicals from biobutanol: technologies and markets. *Biofuels Bioproducts & Biorefining-Biofpr* 6:in press.

McCutchan, W. N., and R. J. Hickey. 1954. The butanol acetone fermentations. *Industrial Fermentation* 1:347-388.

Mehta, R. N., M. Chakraborty, P. Mahanta, and P. A. Parikh. 2010. Evaluation of fuel properties of butanol-biodiesel-diesel blends and their impact on engine performance and emissions. *Industrial & Engineering Chemistry Research* 49(16):7660-7665.

Monot, F., J. M. Engasser, and H. Petitdemange. 1983. Regulation of acetone butanol production in batch and continuous cultures of *Clostridium acetobutylicum*. *Biotechnology and Bioengineering* 25(S13):207-216.

Mortazavi, A., B. A. Williams, K. McCue, L. Schaeffer, and B. Wold. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* 5(7):621-628.

Nakhmanovich, B. M., and N. A. Shcheblykina. 1959. Fermentation of pentoses of corn cob hydrolyzates by *Clostridium acetobutylicum*. *Mikrobiologiya* 28:99-104.

NYTimes. 2012. http://www.nytimes.com/2012/01/02/business/energy-environment/after-three-decades-federal-tax-credit-for-ethanol-expires.html.

Papoutsakis, E. T. 2008. Engineering solventogenic clostridia. *Current Opinion in Biotechnology* 19(5):420-429.

Passalacqua, K. D., A. Varadarajan, B. D. Ondov, D. T. Okou, M. E. Zwick, and N. H. Bergman. 2009. Structure and complexity of a bacterial transcriptome. *Journal of Bacteriology* 191(10):3203-3211.

Pasteur, L. 1862. Quelques resultats nouveaux relatifs aux fermentations acetique et butyrique. *Bull Soc Chim Paris* May:52-53.

Perkins, T. T., R. A. Kingsley, M. C. Fookes, P. P. Gardner, K. D. James, L. Yu, S. A. Assefa, M. He, N. J. Croucher, D. J. Pickard, D. J. Maskell, J. Parkhill, J. Choudhary, N. R. Thomson, and G. Dougan. 2009. A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus *Salmonella typhi*. *Plos Genetics* 5(7):e1000569.

Phillips, J. A., and A. E. Humphrey. 1983. An overview of process technology for the production of liquid fuels and chemical feedstocks via fermentation. In *Organic chemicals from biomass*, 249-304. W. DL, ed. Menlo Park, California: Benjamins/Cummings Publishing.

Quixley, K. W. M., and S. J. Reid. 2000. Construction of a reporter gene vector for *Clostridium beijerinckii* using a *Clostridium endoglucanase* gene. *Journal of Molecular Microbiology and Biotechnology* 2(1):53-57.

Qureshi, N., and H. P. Blaschek. 2001. ABE production from corn: a recent economic evaluation. *Journal of Industrial Microbiology & Biotechnology* 27(5):292-297.

Qureshi, N., and T. C. Ezeji. 2008. Butanol, 'a superior biofuel' production from agricultural residues (renewable biomass): recent progress in technology. *Biofuels Bioproducts & Biorefining-Biofpr* 2(4):319-330.

Qureshi, N., B. C. Sahaa, R. E. Hector, S. R. Hughes, and M. A. Cotta. 2008. Butanol production from wheat straw by simultaneous saccharification and fermentation using *Clostridium beijerinckii*: Part I-Batch fermentation. *Biomass & Bioenergy* 32(2):168-175.

Rakopoulos, D. C., C. D. Rakopoulos, E. G. Giakoumis, A. M. Dimaratos, and D. C. Kyritsis. 2010. Effects of butanol-diesel fuel blends on the performance and emissions of a high-speed DI diesel engine. *Energy Conversion and Management* 51(10):1989-1997.

Ranganathan, S., and C. D. Maranas. 2010. Microbial 1-butanol production: Identification of non-native production routes and in silico engineering interventions. *Biotechnology Journal* 5(7):716-725.

Ross, D. 1961. The acetone-butanol fermentation. *Progress in Industrial Microbiology* 3:73-85.

Ruhl, J., A. Schmid, and L. M. Blank. 2009. Selected *Pseudomonas putida* strains able to grow in the presence of high butanol concentrations. *Applied and Environmental Microbiology* 75(13):4653-4656.

Schardinger, F. 1905. Bacillus macerans, ein Aceton bildender Rottebacillus. *Zentralbl Bakteriol Parasitenkd Infektionskr* II(Abt 4):772-781.

Shao, L., S. Hu, Y. Yang, Y. Gu, J. Chen, W. Jiang, and S. Yang. 2007. Targeted gene disruption by use of a group II intron (targetron) vector in *Clostridium acetobutylicum*. *Cell Research* 17(11):963-965.

Shen, C. R., E. I. Lan, Y. Dekishima, A. Baez, K. M. Cho, and J. C. Liao. 2011. Driving forces enable high-titer anaerobic 1-butanol synthesis in *Escherichia coli*. *Applied and Environmental Microbiology* 77(9):2905-2915.

Shi, Z., and H. P. Blaschek. 2008. Transcriptional analysis of *Clostridium beijerinckii* NCIMB 8052 and the hyper-butanol-producing mutant BA101 during the shift from acidogenesis to solventogenesis. *Applied and Environmental Microbiology* 74(24):7709-7714.

Silcox, H. E., and S. B. Lee. 1948. Fermentation. *Industrial & Engineering Chemistry* 1948:1602-1603.

SRI Consulting. 2009. World petrochemicals report, oxo alcohols. SRI Consulting, Englewood, CA.

Terracciano, J. S., and E. R. Kashket. 1986. Intracellular conditions required for initiation of solvent production by *Clostridium acetobutylicum*. *Applied and Environmental Microbiology* 52(1):86-91.

The Dow Chemical Company. 2002. n-Butanol product information. The Dow Chemical Company, Form No. 327-00014-1001.

Tomas, C. A., K. V. Alsaker, H. P. J. Bonarius, W. T. Hendriksen, H. Yang, J. A. Beamish, C. J. Paredes, and E. T. Papoutsakis. 2003a. DNA array-based transcriptional analysis of asporogenous, nonsolventogenic *Clostridium acetobutylicum* strains SKO1 and M5. *Journal of Bacteriology* 185(15):4539-4547.

Tomas, C. A., J. Beamish, and E. T. Papoutsakis. 2004. Transcriptional analysis of butanol stress and tolerance in *Clostridium acetobutylicum*. *Journal of Bacteriology* 186(7):2006-2018.

Tomas, C. A., N. E. Welker, and E. T. Papoutsakis. 2003b. Overexpression of *groESL* in *Clostridium acetobutylicum* results in increased solvent production and tolerance, prolonged metabolism, and changes in the cell's transcriptional program. *Applied and Environmental Microbiology* 69(8):4951-4965.

Tummala, S. B., S. G. Junne, and E. T. Papoutsakis. 2003. Antisense RNA downregulation of coenzyme A transferase combined with alcohol-aldehyde dehydrogenase overexpression leads to predominantly alcohologenic *Clostridium acetobutylicum* fermentations. *Journal of Bacteriology* 185(12):3644-3653.

Tummala, S. B., N. E. Welker, and E. T. Papoutsakis. 1999. Development and characterization of a gene expression reporter system for *Clostridium acetobutylicum* ATCC 824. *Applied and Environmental Microbiology* 65(9):3793-3799.

van Vliet, A. H. M. 2010. Next generation sequencing of microbial transcriptomes: challenges and opportunities. *Fems Microbiology Letters* 302(1):1-7.

Vera, J. C., C. W. Wheat, H. W. Fescemyer, M. J. Frilander, D. L. Crawford, I. Hanski, and J. H. Marden. 2008. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Molecular Ecology* 17(7):1636-1647.

Wang, Z., M. Gerstein, and M. Snyder. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10(1):57-63.

Weizmann, C. 1963. *Trial and error*. Hamish Hamilton, London.

Wilhelm, B. T., and J. R. Landry. 2009. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods* 48(3):249-257.

WSJ. 2009. http://online.wsj.com/article/SB125427984449051595.html.

Wurtzel, O., R. Sapra, F. Chen, Y. W. Zhu, B. A. Simmons, and R. Sorek. 2010. A single-base resolution map of an archaeal transcriptome. *Genome Research* 20(1):133-141.

Yang, Z. L., D. P. Bruno, C. A. Martens, S. F. Porcella, and B. Moss. 2010. Simultaneous high-resolution analysis of vaccinia virus and host cell transcriptomes by deep RNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America* 107(25):11513-11518.

Zenoni, S., A. Ferrarini, E. Giacomelli, L. Xumerle, M. Fasoli, G. Malerba, D. Bellin, M. Pezzotti, and M. Delledonne. 2010. Characterization of transcriptional complexity during berry development in *Vitis vinifera* using RNA-Seq. *Plant Physiology* 152(4):1787-1795.

Zhao, Y. S., C. A. Tomas, F. B. Rudolph, E. T. Papoutsakis, and G. N. Bennett. 2005. Intracellular butyryl phosphate and acetyl phosphate concentrations in *Clostridium acetobutylicum* and their implications for solvent formation. *Applied and Environmental Microbiology* 71(1):530-537.

Zheng, Y., L. Li, M. Xian, Y. Ma, J. Yang, X. Xu, and D. He. 2009. Problems with the microbial production of butanol. *Journal of Industrial Microbiology & Biotechnology* 36(9):1127-1138.

CHAPTER 3. SINGLE-NUCLEOTIDE RESOLUTION ANALYSIS OF THE TRANSCRIPTOME STRUCTURE OF *CLOSTRIDIUM BEIJERINCKII* 8052 USING RNA-SEQ[1]

3.1 Abstract

Clostridium beijerinckii is an important solvent producing microorganism. The genome of *C. beijerinckii* NCIMB 8052 has recently been sequenced. Although transcriptome structure is important in order to reveal the functional and regulatory architecture of the genome, the physical structure of transcriptome for this strain, such as the operon linkages and transcript boundaries are not well understood. In this study, a single-nucleotide resolution analysis of the *C. beijerinckii* NCIMB 8052 transcriptome was conducted using high-throughput RNA-Seq technology. The transcription start sites and operon structure throughout the genome were identified. The structure of important gene operons involved in metabolic pathways for acid and solvent production in *C. beijerinckii* 8052 were confirmed, including *pta-ack*, *ptb-buk*, *hbd-etfA-etfB-crt* (*bcs*) and *ald-ctfA-ctfB-adc* (*sol*) operons; important operons related to chemotaxis/motility, transcriptional regulation, stress response and fatty acids biosynthesis along with others were also defined. In addition, 20 previously non-annotated regions were discovered to be with significant transcriptional activities and 15 genes whose translation start codons were found to be likely mis-annotated. As a consequence, the accuracy of existing genome annotation was significantly enhanced. Furthermore, 78 putative silent genes and 177 putative housekeeping genes were identified based on normalized transcription measurement with the sequence data. More than 30% of pseudogenes were observed to have significant transcriptional activities during the fermentation process. Strong correlations exist between the expression values derived from RNA-Seq analysis and microarray data or qRT-PCR results. Transcriptome structural profiling in this research provided important supplemental information on the accuracy of genome annotation, and revealed additional gene functions and regulation in *C. beijerinckii.*

3.2 Introduction

Solvents such as acetone, butanol and ethanol (ABE) produced through microbial fermentation represent important potential renewable fuels and chemicals (Wu et al., 2008).

[1] A modified version of Chapter 3 was published in BMC Genomics, 2011, 12:479. (Y. Wang lead author with co-authors X. Li, Y. Mao and H.P. Blaschek).

Clostridium acetobutylicum and *C. beijerinckii* are among the prominent solvent-producing species. Although *C. beijerinckii* is phenotypically similar to *C. acetobutylicum*, the saccharolytic strains are phylogenetically distant from the amylolytic *C. acetobutylicum* ATCC 824 type strain (Keis et al., 1995). *C. beijerinckii* exhibits a broader substrate range and optimum pH for growth and solvent production (Ezeji and Blaschek, 2008); thus it may have greater potential for biosolvent production than *C. acetobutylicum*.

The genome of *C. beijerinckii* NCIMB 8052 was sequenced by the DOE Joint Genome Institute in 2007 (JGI project ID 3634512). The genome size is 6.0 Mb, which is 50% larger than that of *C. acetobutylicum* ATCC 824. The *C. beijerinckii* 8052 solvent-producing genes are all located on the chromosome, as opposed to the location of these genes on a mega-plasmid in *C. acetobutylicum* 824. Although transcriptome structural organization is important in order to reveal the functional and regulatory architecture of the genome, such annotation for the *C. beijerinckii* 8052 genome is far from complete. Current genome annotation was made by computational analysis based on gene prediction algorithms. Although this allows for the determination of the complete set of gene loci and intergenic regions of the genome, it does not provide sufficient information concerning the transcriptional organization on a genome-wide level. For example, transcriptome structures such as operon linkages and transcript boundaries, etc. are not well understood and lack confirmation with experimental approaches. Next-generation high-throughput sequencing technology enabled us to obtain millions of cDNA reads simultaneously. In RNA-Seq analysis, these reads can be assembled against the genome sequence, and expression values calculated based on the reads mapped to genes. Deep sequencing of cDNA pool allowed us to study the bacterial transcriptome structure and gene expression at an unprecedented resolution and depth (Mortazavi et al., 2008; Nagalakshmi et al., 2008; Passalacqua et al., 2009; Perkins et al., 2009; Wurtzel et al., 2010). In this study, with RNA-Seq technology, the *C. beijerinckii* 8052 transcriptome structure was investigated at a single-nucleotide resolution. The transcription start sites and operon structure were identified throughout the genome. The structure of important gene operons involved in metabolic pathways for acid and solvent production were confirmed in *C. beijerinckii* 8052. Important operons related to chemotaxis/motility, transcriptional regulation, stress response and fatty acids biosynthesis were defined. Twenty previously non-annotated regions with significant transcriptional activities and 15 genes whose translation start codons were likely mis-annotated have been

discovered. The results from this study significantly enhanced the accuracy of current genome annotation, and provided an essential reference point for other researchers working in related fields.

3.3 Materials and methods

3.3.1 Bacterial culture and fermentation experiment

Laboratory stocks of *C. beijerinckii* 8052 spores were stored in sterile H_2O at 4 ^oC (Annous and Blaschek, 1991). Spores were heat-shocked at 80 °C for 10 min, followed by cooling on ice for 5 min. The heat-shocked spores were inoculated into tryptone–glucose–yeast extract (TGY) medium containing 30 g L^{-1} tryptone, 20 g L^{-1} glucose, 10 g L^{-1} yeast extract and 1 g L^{-1} L-cysteine at a 1% inoculum level. The TGY culture was incubated at 35±1 ^oC for 12-14 h in an anaerobic chamber under N_2:CO_2:H_2 (volume ratio of 85:10:5) atmosphere. Subsequently, actively growing culture was inoculated into a model solution containing 60 g L^{-1} glucose, 1 g L^{-1} yeast extract, and filter-sterilized P2 medium (Jesse et al., 2002; Qureshi et al., 2001) in a Sixfors bioreactor system (Infors AG, Bottmingen, Switzerland). Oxygen-free nitrogen was flushed through the broth to initiate anaerobiosis until the culture initiated its own gas production (CO_2 and H_2). Temperature was controlled at 35±1 ^oC. A stirring at 50 rpm was employed for mixing. During the course of fermentation, samples were collected for cell density and product concentration measurements. For sequencing purpose, RNA samples were taken over the early exponential, late exponential and stationary phases (samples 1-6 at 2, 4.5, 10, 14, 17 and 26.5 h respectively as shown in Figure 3.1A).

3.3.2 Culture growth and fermentation products analysis

Culture growth was measured by following optical density (OD) in the fermentation broth at A_{600} using a BioMate 5 UV-Vis Spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA). ABE, acetic acid, and butyric acid concentrations were quantified using gas chromatography (GC) system as previously described (Ezeji et al., 2003).

3.3.3 RNA isolation, library construction and sequencing

In preparation for RNA isolation, 10 ml cultures were harvested at six time points, and centrifuged at 4,000 × g for 10 min at 4 °C. Total RNA was extracted from the cell pellet using Trizol reagent based on manufacture's protocol (Invitrogen, Carlsbad, CA) and further

purified using RNeasy minikit (Qiagen, Valencia, CA). DNA was removed using a DNA-*free*™ kit (Ambion Inc., Austin, TX). RNA quality was assessed using a nanochip on a model 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA). RNA concentration was determined with a nanodrop (Biotek Instruments, Winooski, VT). Bacterial 16S and 23S ribosomal RNAs were removed with a MICROBExpress™ kit (Ambion Inc., Austin, TX). The enriched mRNA was converted to a RNA-Seq library using the mRNA-Seq library construction kit (Illumina Inc., San Diego, CA) following manufacturer's protocols. For samples 1 to 6, two samples were pooled and sequenced on one single lane of an eight-lane flow cell with the Genome Analyzer IIx system (Illumina Inc., San Diego, CA). However, sample 6 yielded a poor read quality following the first sequencing. In order to obtain enough sequencing depth, sample 6 was sequenced again using one single lane under otherwise identical conditions. The derived sequence reads were 75 nt long. The overall error rate of the control DNA was < 0.6%. The total number of reads generated from each library is summarized in Table 3.1.

3.3.4 Sequence mapping and visualization

The generated 75-nt reads were mapped to the *C. beijerinckii* 8052 genome using MAQ, and those that did not align uniquely to the genome were discarded (Li et al., 2008; Perkins et al., 2009). The quality parameter (-q) used in MAQ pileup was set to 30. Each base was assigned a value based on the number of mapped sequence coverage. The coverage plot files were read into Artemis and visualized as sequence coverage profiles over the entire genome (Carver et al., 2008; Perkins et al., 2009). Transcription start sites (TSS) were manually identified as described in Passalacqua *et al.* (Passalacqua et al., 2009). The region between determined TSS and the annotated translation start site was defined as the 5'-untranslated region (5'-UTR).

3.3.5 Measurement of gene expression

The quantitative gene expression value, RPKM (reads/Kb/Million), was calculated using custom Perl scripts by normalizing the sequence coverage over the gene length and total unambiguously mapped reads in each library (Mortazavi et al., 2008; Severin et al., 2010).

3.3.6 End-point RT-PCR for operon structure assessment

End-point RT-PCR was performed as described in Passalacqua *et al.* (Passalacqua et al., 2009). For each selected co-operonic gene pair, four primers 1F, 1R, 2F and 2R were designed. Primers 1F and 1R amplify a region within gene 1, while 2F and 2R amplify a region within gene 2. When a continuous transcript exists, 1F and 2R amplify across the intergenic region between genes 1 and 2. Therefore, the size of the end-point RT-PCR product will be the same using either cDNA or genomic DNA as templates. The reaction products were visualized on 1.5% agarose gels stained with ethidium bromide.

3.3.7 Real time qRT-PCR

Quantitative reverse transcription PCR (qRT-PCR) was performed in order to validate the quantification of gene expression level by RNA-Seq. Twenty-five genes were chosen to represent a large range of RPKM values (from ~ 40 to > 18000). Triplicate reactions were performed using Power SYBR green PCR master mix (Applied Biosystems, Carlsbad, CA) on an ABI Prism 7900HT fast real-time PCR machine (Applied Biosystems). Detected genes and primer sequences are listed in Table A3.9 in Appendix A.

3.3.8 RNA-Seq data accession number

The RNA-Seq sequencing data have been deposited in the NCBI Sequence Read Archive (SRA) under the accession number SRA045799, and also included in the Additional file A3.1 in Appendix A.

3.4 Results and discussion

3.4.1 Growth kinetics and ABE fermentation

C. beijerinckii 8052 grew rapidly with a very short lag phase in the batch fermentation with P2 medium supplemented with yeast extract and glucose (Figure 3.1A). The fermentation experienced a shift from acidogenesis to solventogenesis at approximately 4.5-8h. Formation of solvents was detected at between 4.5-6.5 h after the start of fermentation, which corresponded to the late exponential growth phase. Butanol continued to increase throughout the stationary phase (Figure 3.1B). Samples for RNA isolation were collected at time points during acidogenesis (2 and 4.5 h) and solventogenesis (after 6.5 h) (Figure 3.1A). The combined information from these samples collected at different growth phases is representative for transcriptome structural analysis.

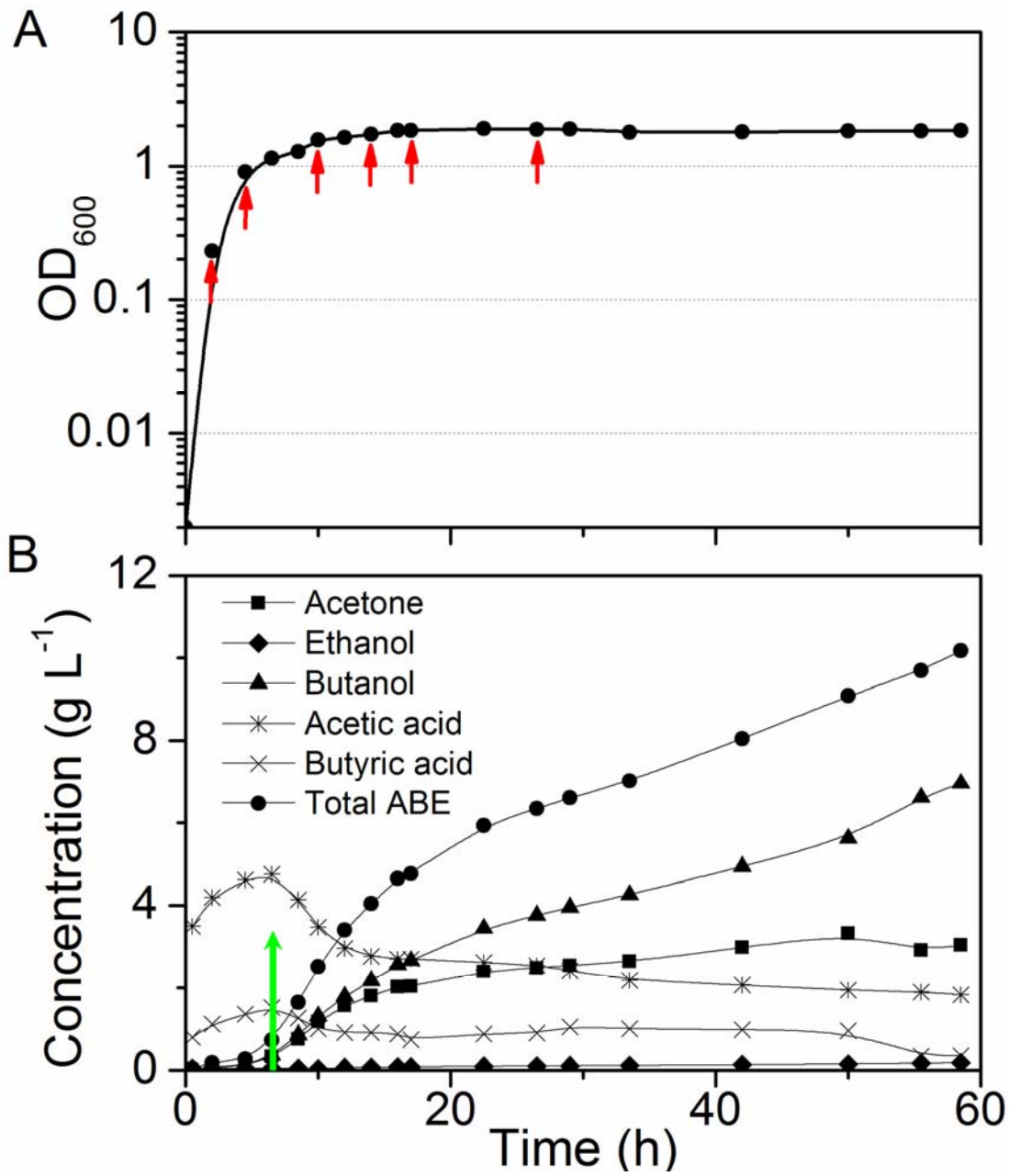


Figure 3.1 Fermentation kinetics of *C. beijerinckii* 8052 batch culture. (A) Cell growth curve with sampling points for RNA-Seq indicated by red arrows. (B) Solvents and acids production over time with the onset of solvent production indicated by a green arrow.

3.4.2 Transcriptome definition and structure

The 75-nt cDNA reads were mapped to the *C. beijerinckii* 8052 genome. Only those reads that mapped unambiguously to the genome were used for further analyses (Table 3.1). Collectively, around 75.5% of the *C. beijerinckii* 8052 genome was transcribed in at least one sample, even though the fraction in each single sample was much less.

Table 3.1 Summary of RNA-Seq sequencing and data analysis results

Sample	1	2	3	4	5	6[1]	Total
Time collected (h)	2	4.5	10	14	17	26.5	
Total number of reads	8988633	9457480	8011531	8448929	10363535	38574501	83844609
No. of reads mapped	8473125	9037616	7514804	7730815	9842491	37676913	80275764
No. of reads unambiguously mapped	6776544	7274568	6096405	6189652	8096169	35027722	69461060
No. of bases unambiguously mapped	508240800	545592600	457230375	464223900	607212675	2627079150	5209579500
Percentage of genome represented	38.71	36.59	43.84	40.60	41.34	50.67	75.46
No. of genes with detectable expression[2]	4219	4082	4496	4453	4487	4750	5024
Range in expression levels (RPKM)	3.2×10^{-1} ~ 2.5×10^{4}	5.8×10^{-2} ~ 6.0×10^{4}	4.5×10^{-2} ~ 2.5×10^{4}	7.0×10^{-2} ~ 9.0×10^{4}	1.0×10^{-1} ~ 8.6×10^{4}	3.6×10^{-2} ~ 9.8×10^{4}	3.6×10^{-2} ~ 9.8×10^{4}

[1]Sample 6 was sequenced with 1 sample/lane, while samples 1-5 with 2 samples/lane. See Section 3.3 for more details;

[2]Pseudogenes included.

The sequence coverage per base was plotted and visualized using the genome browser Artemis and DNAPlotter (Carver et al., 2008; Carver et al., 2009; Perkins et al., 2009) (Figure A3.1 in Appendix A). The sequence coverage of the genes (based on number of mapped reads) for different samples was observed to be different in the overall sequence coverage profiles, which might be a reflection of the physiological states as the cell transited from one growth phase to another. For example, sample 2 represented a time period that was at the beginning of the transition from acidogenesis to solventogenesis. In this sample, genes encoding butyrate kinase (*buk*, 233080-234147 nt; Gene ID is listed in Table A3.10, similarly hereinafter), acetyl-CoA acetyltransferase (*thl*,499121-500302 nt), glyceraldehyde-3-phosphate dehydrogenase (*gap*, 710763-711764 nt), fructose-bisphosphate aldolase (*fba*, 2199060-2199926 nt), aldehyde dehydrogenase (*ald*, 4399026-4400432 nt), acetoacetate decarboxylase (*adc*, 4401916-4402656 nt) were all actively expressed (high peaks in Figure A3.1). These genes are involved in acid and butyryl-CoA formation, solventogenesis and glycolysis. The time frame of sample 4 was consistent with the transition to non-active growth and clostridial spores formation. In this sample, genes encoding stage V sporulation protein T (*spoVT*, 115288-115836 nt), spore coat protein cotJC (*cotJC*, 2411617-2412183 nt), spore coat peptide assembly protein cotJB (*cotJB*, 2412190-2412453 nt), small acid-soluble spore protein, sspA (*sspA*, 3596975-3597184 nt), small acid-soluble spore protein, sspC2 (*sspC2*, 3814296-3814505 nt) were actively expressed among others (Figure A3.1). These genes are involved in the regulation of spore coat assembly and other sporulation-related processes (Steil et al., 2005; Wang et al., 2006). These results were in good agreement with our previous observation using a 500-gene set DNA microarray (Shi and Blaschek, 2008).

The distinct transcript boundaries and multigene operon structure in a genome are essential for elucidating a genome's regulatory mechanisms. Through transcriptome profiling analysis, the transcript boundaries and operon structures were identified across the entire *C. beijerinckii* 8052 genome (Table A3.1 in Appendix A). Of all the > 5000 genes, 2151 genes (42.2%) were determined as being parts of multi-gene operon structures. The largest number of genes in a single operon was 32 (ribosomal proteins operon, Cbei_0150-0181, 192128-207934 nt). A refined genome annotation file (in GenBank format) was generated (Additional file A3.2) based on the findings from this work and the current *C. beijerinckii* 8052 genome annotation in NCBI. The GenBank file can also be downloaded from https://netfiles.uiuc.edu/blaschek/www/Wang-BMC2011.

In addition, 5'-untranslated regions (5'-UTRs) were determined at the same time. Similar to other bacteria (McGrath et al., 2007; Passalacqua et al., 2009), most of the identified 5'-UTRs were very short. Among all the 1605 5'-UTRs that were estimated with high confidence (that is, valid 5'-UTRs identified in the 6 samples with a ≤ 50 nt window), 1262 were ≤ 50 nt, and only 65 were ≥ 100 nt (listed in Table A3.2 in Appendix A). RibEx (Abreu-Goodger and Merino, 2005) was used to test for the putative regulatory elements among the 5'-UTRs over 100bp. Three Predicted Riboswitch-like Elements (do not belong to any Rfam (Griffiths-Jones et al., 2005); yet present very significant associations with either a COG or a KEGG pathway) were found in the 5'-UTRs of Cbei_0465, Cbei_1518 and Cbei_3542, respectively. In addition, a Known Riboswitch-like Element (flavin mononucleotide (FMN) riboswitch) was found in the 5'-UTR of Cbei_1224. More details about the identified putative riboswitches were described in Table A3.3 in Appendix A.

Previously, an operon containing genes encoding glyceraldehyde-3-phosphate dehydrogenase (*gap*), phosphoglycerate kinase (*pgk*) and triosephosphate isomerase (*tpi*) was revealed by transcriptional analyses for *C. acetobutylicum*, while the gene encoding phosphoglycerate mutase (CA_C0712, *pgm*) was identified as not a member of this operon (Schreiber and Dürre, 2000). In this study, employing the single-nucleotide resolution sequence data, a similar *gap-pgk-tpi* operon was observed, upstream and apart from *pgm* for *C. beijerinckii* 8052. In addition, it is very interesting that the upstream gene Cbei_0596 was also found to be linked with this operon (Figure 3.2A), which was validated by end-point RT-PCR as discussed below (Figure A3.2 and Table A3.8). The upstream gene (CA_C0708) of *gap* in *C. acetobutylicum* 824 genome is conjectured to be the counterpart of Cbei_0596 in *C. beijerinckii* 8052, since they both encode putative transcriptional regulator proteins and share 64% sequence identity. However, the intergenic distance between CA_C0708 and CA_C0709 in *C. acetobutylicum* 824 genome is 146 nt, while that between Cbei_0596 and Cbei_0597 in *C. beijerinckii* 8052 genome is only 71 nt. Although Cbei_0596-0599 are organized in the same operon, the sequence coverage depth downstream of *gap* was unexpectedly lower than that of the anterior (Figure 3.2A). An intrinsic termination was predicted adjacent and downstream of *gap* (711818-711852 nt) using the bacterial genome transcription terminator prediction software TransTerm (Ermolaeva et al., 2000). The terminator may have played a key role in attenuating the transcription of the downstream genes in this operon. Further experiments need to be carried out to confirm the activity and function of this element.

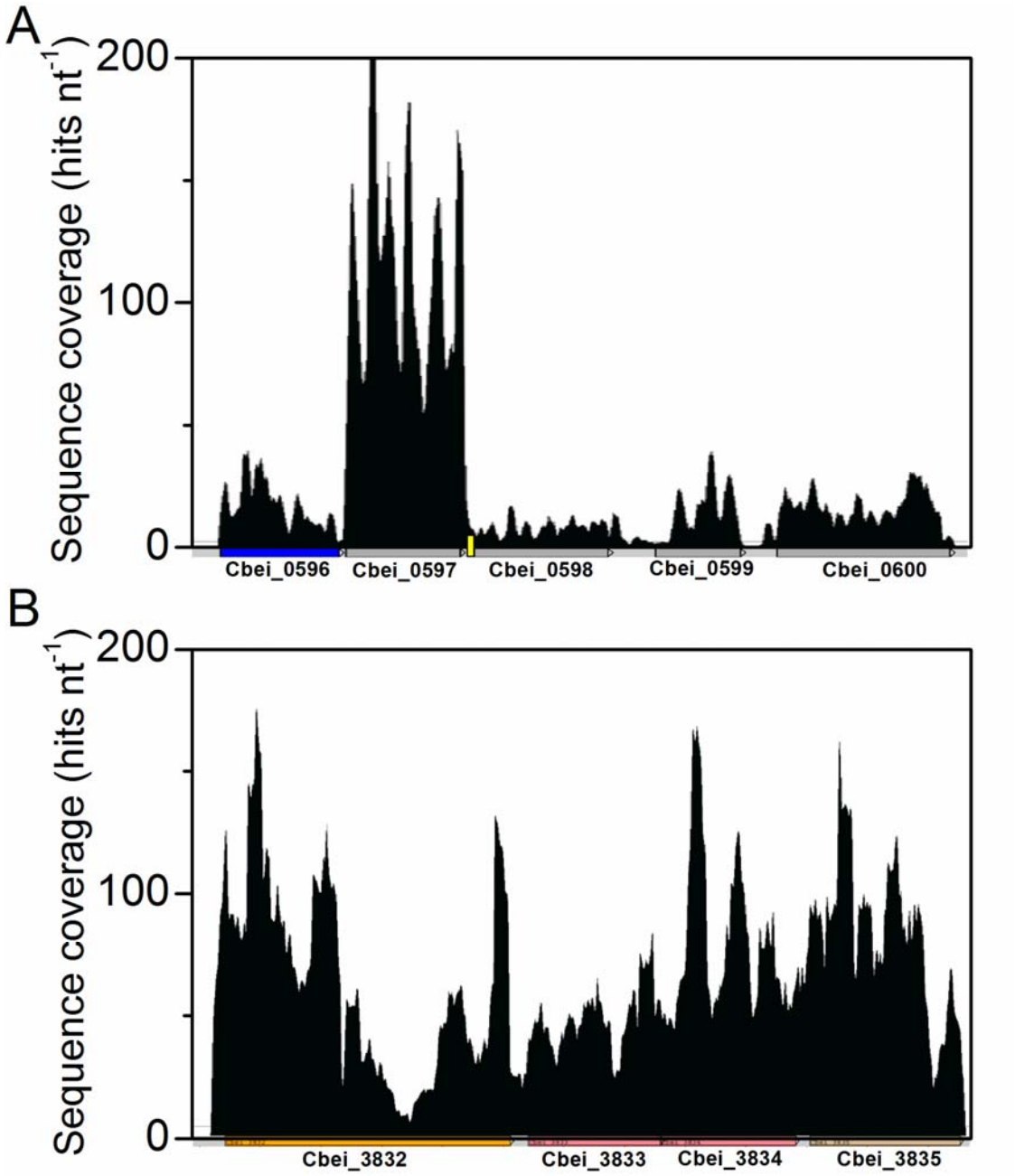


Figure 3.2 Single-nucleotide resolution-sequencing data revealed important operon structures in *C. beijerinckii* 8052. (A) The glycolytic genes (*gap-pgk-tpi*) operon structure (from sample 1). The upstream gene Cbei_0596 encoding a putative transcriptional regulator protein was also linked with this operon. The intrinsic terminator predicted with TransTerm (Ermolaeva et al., 2000) is represented with a small yellow bar. (B) The solventogenic genes are organized in the *sol* operon in the order of *ald-ctfA-ctfB-adc*. Gene positions and directions are shown beneath by arrows. The color of arrows indicates the COG functional class of genes as defined in Figure A3.1 in Appendix A.

The butyryl-CoA formation related genes in *C. acetobutylicum* 824 encoding 3-hydroxybutyryl-CoA dehydrogenase (*hbd*), crotonase (*crt*), and butyryl-CoA dehydrogenase (*bcd*) were identified as a *bcs* (butyryl-CoA synthesis) operon in the order of *crt-bcd-etfB* (encoding electron transfer flavoprotein subunit beta)-*etfA* (encoding electron transfer flavoprotein subunit alpha)-*hbd* (CA_C2708) (Bennett and Rudolph, 1995; Boynton et al., 1996a). A homologous operon structure was observed for *C. beijerinckii* 8052 (Cbei_0321-325) transcribed in the same order but the opposite direction (Table A3.1). In addition, based on the sequencing data, similar *pfk* (phosphofructokinase)-*pyk* (pyruvate kinase), *pta* (phosphotransacetylase)-*ack* (acetate kinase), *ptb* (phosphate butyryltransferase)-*buk* gene operons were identified in *C. beijerinckii* 8052 as those described in *C. acetobutylicum* 824 (Belouski et al., 1998; Boynton et al., 1996b; Walter et al., 1993) (Table A3.1). Among them, *ptb-buk* operon was also verified by end-point RT-PCR experiment as discussed below (Figure A3.2 and Table A3.8).

Nearly all the genes encoding solventogenic enzymes have been cloned and characterized in *C. acetobutylicum* (Düerre, 2005). In *C. acetobutylicum* 824, the genes *adhE* (acetaldehyde-CoA/alcohol dehydrogenase), *ctfA* (acetoacetyl-CoA: acetate/butyrate-CoA transferase subunit A) and *ctfB* (acetoacetyl-CoA: acetate/butyrate-CoA transferase subunit B) are located in the *sol* (solvent formation) operon on the mega-plasmid pSOL1 whose loss leads to the degeneration of the strain (Cornillot et al., 1997), while *adc* (acetoacetate decarboxylase) is organized in a monocistronic operon in the opposite direction (Gerischer and Dürre, 1990; Petersen et al., 1993). In this study, a *sol* operon organized in the order of *ald-ctfA-ctfB-adc* was observed (Figure 3.2B and Table A3.1). Previously, Chen and Blaschek (1999) speculated that the *ald*, *ctfA*, *ctfB* and *adc* genes were located in an operon following a Northern hybridization analysis of *C. beijerinckii* 8052 total RNA (Chen and Blaschek, 1999). With direct sequencing data, this study successfully confirmed the above hypothesis, and identified the transcriptional start sites (TSS) upstream of *ald* gene. This solventogenic gene arrangement in *C. beijerinckii* 8052 is consistent with that observed in *C. beijerinckii* NRRL B593 and *C. saccharoperbutylacetonicum* N1-4 (Kosaka et al., 2007; Rui, 1999).

A flagellar/chemotaxis gene operon (CA_C2225-C2215) was previously defined in *C. acetobutylicum* (Paredes et al., 2004); a counterpart organized in exactly the same order (Cbei_4312-4302) in *C. beijerinckii* 8052 was observed in this study (Table A3.1). Similarly

to *C. acetobutylicum* 824, the transcriptional regulator *sigF* operon was also confirmed in *C. beijerinckii* 8052 with the sequencing data that includes the forespore-specific sigma factor gene *sigF*, the anti-sigF factor gene *spoIIAB*, and the anti-anti-sigF factor gene *spoIIAA* (Table A3.1). In *Bacillus subtilis* and *C. acetobutylicum*, the class I heat shock genes are organized in *dnaKJ* (organized in the order of *hrcA-grpE-dnaK-dnaJ* in *C. acetobutylicum* 824) and *groESL* (*groES-groEL*) operons (Bahl et al., 1995; Tomas et al., 2003). The similar organization of *dnaKJ* and *groESL* operons for *C. beijerinckii* 8052 was also observed in this study (Table A3.1). The fatty acid biosynthesis genes are organized in a single *fab* operon in *C. acetobutylicum* (CA_C3568-C3580) (Tomas et al., 2004), while in this study, the *fab* genes (Cbei_1067-1077) in *C. beijerinckii* 8052 were observed to be organized in four operons, which is similar to those of *B. subtilis* (de Mendoza et al., 2002) (Table A3.1).

While the current *C. beijerinckii* 8052 genome was annotated based on bioinformatical predictions, the RNA-Seq sequencing approach provides additional experimental evidence for genome annotation. By comparing the sequence coverage data to the genome annotation, 20 non-annotated regions were found to have significant transcriptional activities (Figure 3.3A, see also the supplemental text in Appendix A). These regions may represent potential new genes or regulatory RNAs. Twelve potential new genes were predicted in these regions using GeneMark (Besemer and Borodovsky, 2005). Additional details about this test and the predicted genes were summarized in Table A3.4.

In addition, 15 transcripts were identified with TSS that are downstream of the current annotated translation start sites (Figure 3.3B, see also the supplemental text in Appendix A). For these regions, re-annotation is needed since the current start codons may have been mis-annotated.

3.4.3 Putative silent genes

Based on RNA-Seq sequence data, 78 protein-encoding genes demonstrated no transcripts over all six sampling time points, and these genes are likely silent (Supplemental Table A3.5). Thirty-one out of them were genes encoding hypothetical proteins. In addition, half of the genes encoding transposases (14 out of the total 29 in the genome) were silent. Although transposases may have played important roles during the evolution of *C. beijerinckii*, most of them are not as functional any more during the course of a batch fermentation. Besides, several genes encoding the subunits (such as sorbose-specific

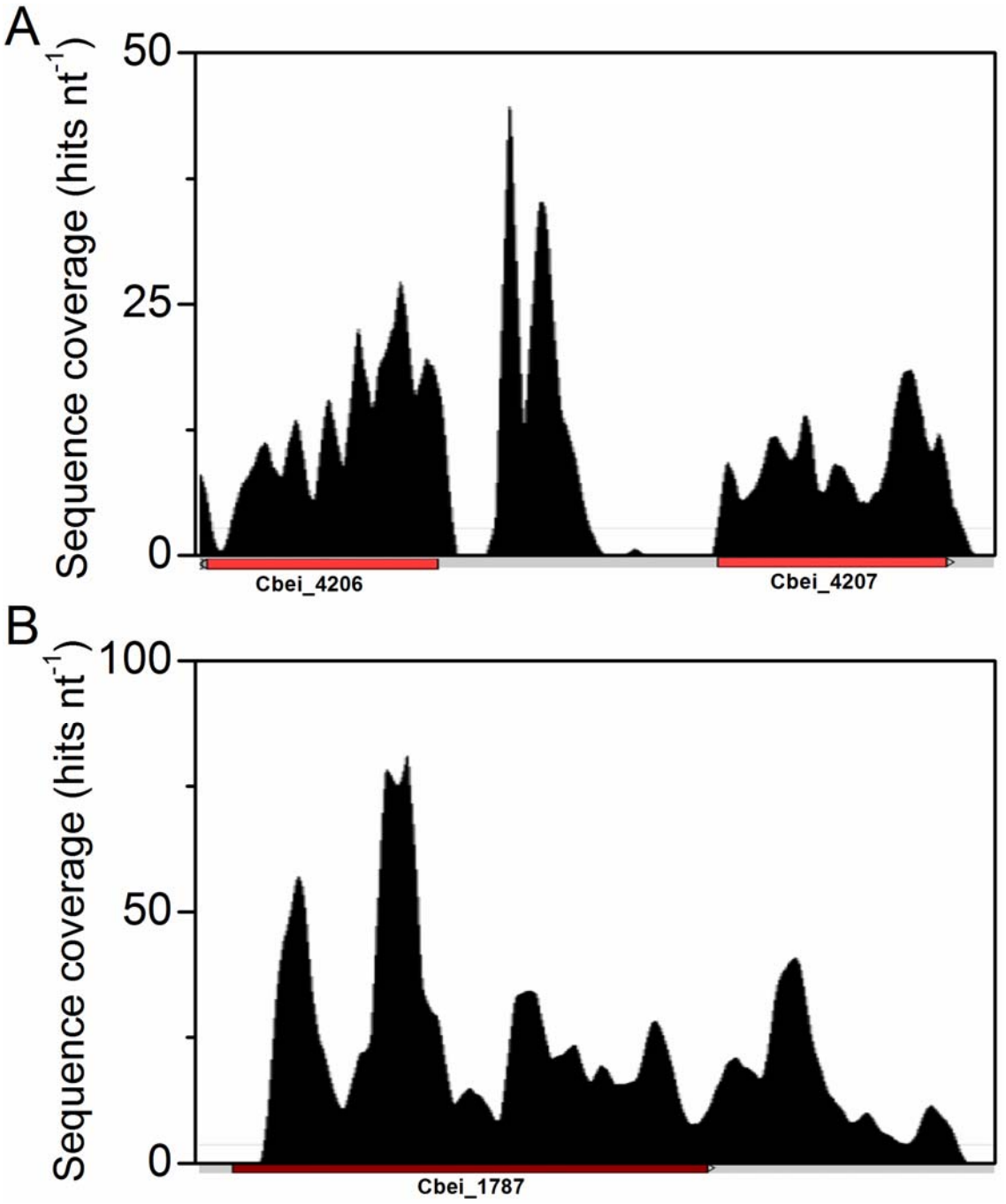


Figure 3.3 Mapping the high resolution RNA-Seq data back to the genome allows the test of current *C. beijerinckii* 8052 genome annotation. (A) Sequence coverage in the Cbei_4206-4207 region, showing transcription activity across a region not included in the current genome annotation (from sample 2). (B) Sequence coverage near Cbei_1787, with the sequence coverage starts downstream the annotated translation start codon (from sample 5). Gene positions and directions are shown beneath by arrows. The color of arrows indicates the COG functional class of genes as defined in Figure A3.1 in Appendix A.

subunits, lactose/cellobiose-specific subunits) of phosphotransferase system (PTS) related to the transport of sugars other than glucose were also among the list of silent genes. Since glucose was the only carbohydrate used in this study, these enzymes were not induced during the fermentation process.

3.4.4 Putative housekeeping genes (HKGs)

Some genes have little variation in expression level through the entire fermentation process, and they are regarded as putative housekeeping genes (HKGs). For accurate gene expression quantification, normalization of gene expression against HKGs (endogenous control or reference gene) is generally required. In this study, 177 protein-coding genes were identified as putative HKGs with the lowest coefficient of variation (CV) in RPKM values among all the sampling points (CV = standard deviation/mean; < 30% for listed genes in Supplemental Table A3.6) (Severin et al., 2010). A COG functional group analysis by Fisher's exact test found that COG functional category D (Cell cycle control, mitosis and meiosis, 10.3%), L (Replication, recombination and repair, 9.2%) and E (Amino acid transport and metabolism, 6.6%) were overrepresented in this list (Fisher, 1949; Severin et al., 2010). This list of putative HKGs was generated based on the transcription data obtained from the six samples under the certain batch fermentation conditions employed in this study. This list can be considered as a starting point for identifying HKGs for *C. beijerinckii*. However, whether these genes are stably expressed under different experimental conditions requires further study.

A prediction of the promoters for primary sigma factors for all the putative HKGs was carried out using BPROM (http://linux1.softberry.com/berry.phtml). In total, 88 promoters along with their corresponding -10 box and -35 box sequences were predicted (Table A3.6).

3.4.5 Pseudogenes

Pseudogenes in bacteria genomes are non-functional copies of gene fragments usually created by random mutations and chance events (Kuo and Ochman, 2010; Lerat and Ochman, 2005). Although pseudogenes are usually non-functional, it has been reported that quite a few of them can still go through the transcription process (Hirotsune et al., 2003; Svensson et al., 2006; Zheng et al., 2007). In this study, 26 out of the 82 pseudogenes in *C. beijerinckii* 8052 genome were found to have significant transcriptional activities over the course of fermentation (Table A3.7). However, although globally the 82 pseudogenes comprise about

0.6% of the predicted CDS of the genome, only < 0.1% of all the RNA-Seq reads mapped to these genes, indicating a significantly lower transcriptional activity when compared to the regular protein coding genes. In addition, six pseudogenes were found to be completely silent throughout the fermentation process (Table A3.7). This is overrepresented among the silent genes when compared to the protein-coding genes based on a Fisher's exact test ($p = 0.0015$).

When pseudogenes share high sequence identity with other functional genes in the genome, it is usually on one hand very difficult to design probes to detect the transcription of pseudogenes with traditional methods, and on the other pseudogenes can lead to amplification bias during genetic studies of the functional genes with high sequence identity to pseudogenes (Zheng et al., 2007). Apparently such problems can be avoided with RNA-Seq method, which is one of the unprecedented advantages of RNA-Seq technique.

For summarization and easy reference to other researchers, a table (Table A3.11) summarizing the various supplementary information provided by RNA-Seq study to the current genome annotation was listed in Appendix A.

3.4.6 End-point RT-PCR

End-point RT-PCR was used to validate the operon structure results obtained from RNA-Seq analysis (Passalacqua et al., 2009). One gene pair that is highly likely to be co-operonic with an intergenic distance of 7 bp was chosen as a positive control (Cbei_0341-0342). Ten other gene pairs with long intergenic distances (71-440 bp) where there were high transcription levels were chosen. All the operon structures determined by sequence data in the chosen gene pairs were confirmed by end-point RT-PCR results, indicating that RNA-Seq is a valid approach for operon structure identification (Figure A3.2 and Table A3.8). Among them, the *ptb-buk* operon and the linkage between Cbei_0596 and Cbei_0597 (*gap*) as discussed above were also confirmed.

3.4.7 Correlations of RNA-Seq data with microarray and qRT-PCR data

The RNA-Seq data obtained in this study were compared to the results obtained previously using a 500-gene set DNA microarray (Shi and Blaschek, 2008). The RNA-Seq data in this study were compared to those microarray data. A small number of genes that did not allow for unambiguous mapping in RNA-Seq analysis were not included in the comparison. A good correlation was observed between the normalized coverage depth for

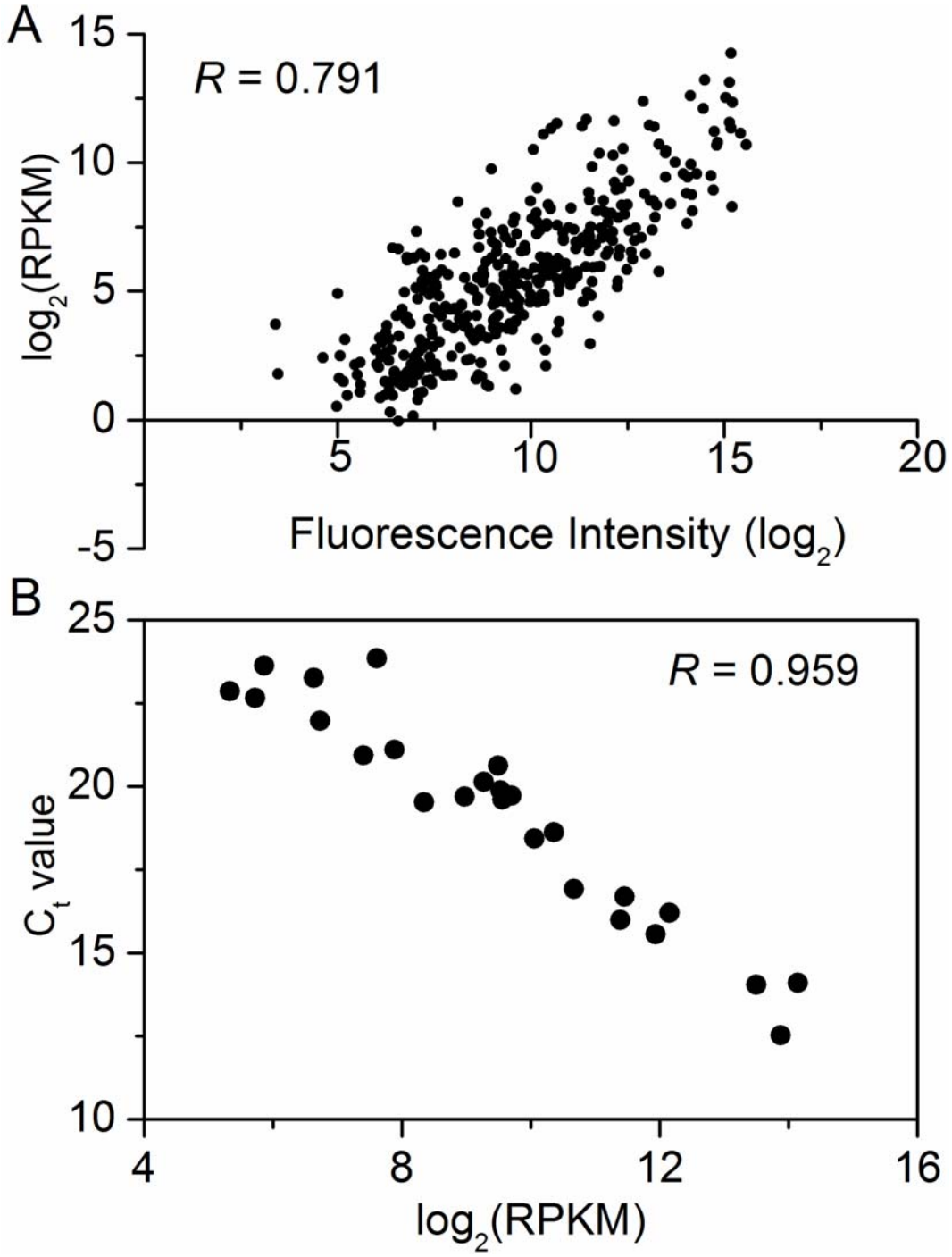


Figure 3.4 Correlations of RNA-Seq data with microarray and qRT-PCR data. (A) Comparison of normalized sequence coverage depth from RNA-Seq and absolute microarray fluorescence intensity. Shown is a plot of $\log_2$-transformation of RPKM for sample 2 (at 4.5 h) and raw microarray intensities for an equivalent sample (at 5 h) from Shi and Blaschek (Shi and Blaschek, 2008). (B) Real time qRT-PCR verification of RNA-Seq results for selected genes (from sample 4, see also Table A3.9).

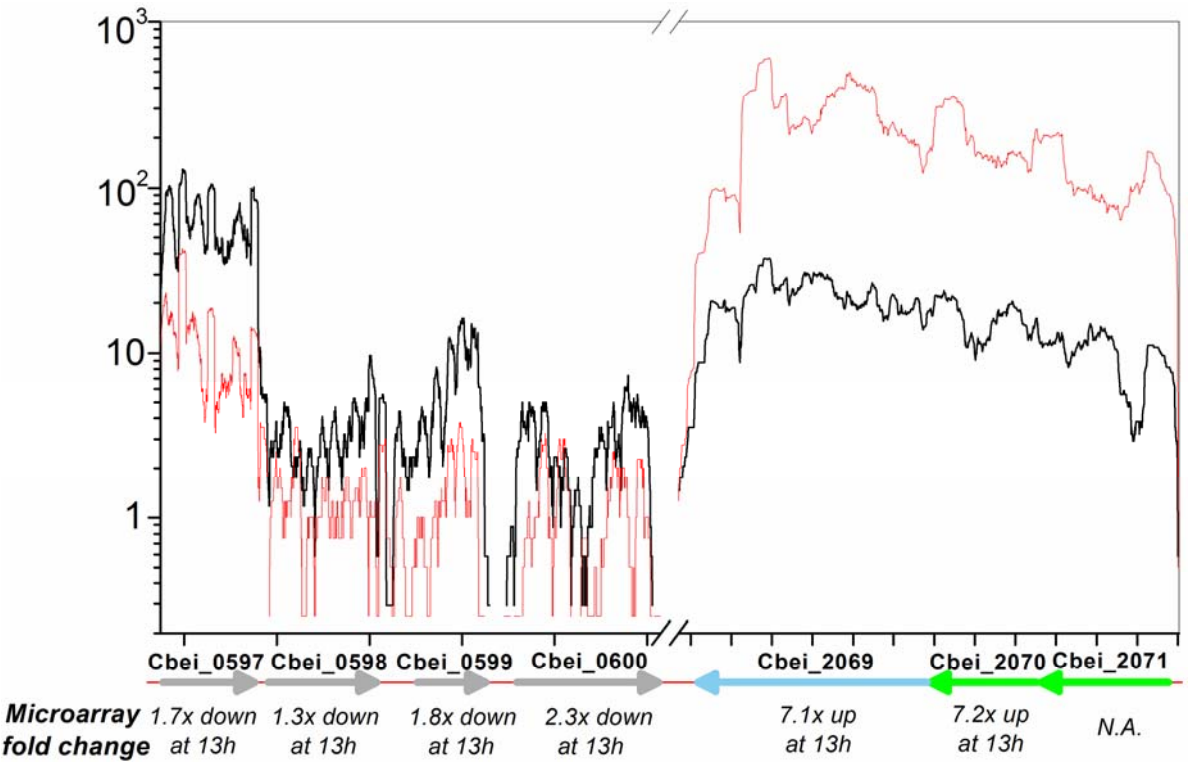


Figure 3.5 Comparison of transcriptional profiles for sample 2 (4.5h, black line) and sample 4 (14h, red line), normalized to the gene length and genome-wide total number of unambiguously mapped reads for that sample. Shown are an ~5.4 kb segment surrounding the *gap-pgk-tpi* glycolytic gene operon (Cbei_0597-0600) and an ~1.2 kb segment surrounding the *cotJC-cotJB* (Cbei_2069-2071) sporulation gene operon. Specific genes are indicated with arrows, whose colors indicate the COG functional class of genes as defined in Figure A3.1 in Appendix A. Below each gene is the expression level fold change that was measured by microarray data for equivalent samples (5h vs. 13h in Shi and Blaschek (Shi and Blaschek, 2008); microarray data for Cbei_2071 are not available).

sample 2 (collected at 4.5 h) and the raw microarray fluorescence intensities for an equivalent sample (collected at 5 h under the same growth condition) (Figure 3.4A). Specifically, the gene expression patterns of glycolytic genes Cbei_0597-0600 and sporulation genes Cbei_2069-2071 from samples 2 (4.5h) and 4 (14h) measured by sequencing data were compared. Similar differential expression patterns by microarray data from equivalent samples (5h vs. 13h) were observed (Figure 3.5). Although only a limited number of genes were compared, the good correlation between two methods further demonstrated the effectiveness of RNA-Seq approach. An additional advantage of RNA-Seq is that the

sequence data can measure expression levels for every gene without bias caused by sequence-specific differences in hybridization efficiency in microarray-based methods (Passalacqua et al., 2009; Wilhelm et al., 2008), and sequence data allow one to measure gene expression more accurately with a much higher dynamic range as indicated in this study (Table A3.1) and other references (Nagalakshmi et al., 2008; Passalacqua et al., 2009; t Hoen et al., 2008; Wilhelm et al., 2008).

The expression measurement with RNA-Seq data was further validated using real time quantitative reverse transcription PCR (qRT-PCR). Twenty-five genes were selected for the test (Table A3.9). A high degree of correlation (R = 0.959) between the threshold value (Ct) and the $\log_2$-tranformation of RPKM was observed (Figure 3.4B).

3.5 Conclusions

A single-nucleotide resolution analysis of the *C. beijerinckii* NCIMB 8052 transcriptome structure was conducted using high-throughput RNA-Seq technology. The transcription start sites and operon structures were identified throughout the genome. The structure of operons involved in metabolic pathways for acid and solvent production in *C. beijerinckii* 8052 were confirmed. Important operons related to chemotaxis/motility, transcriptional regulation, stress response and fatty acids biosynthesis along with others were defined. Twenty previously non-annotated regions were discovered with significant transcriptional activities and 15 genes were identified whose translation start codons were likely mis-annotated. As a consequence, the accuracy of existing genome annotation was significantly enhanced. Moreover, 78 silent genes and 177 putative housekeeping genes were identified based on normalized transcription measurement with the sequence data. More than 30% of the pseudogenes were observed to have significant transcriptional activities during the fermentation process. Strong correlations exist between the expression values derived from RNA-Seq analysis and microarray data or qRT-PCR results. Transcriptome structural profiling in this study provided important supplemental information on the accuracy of annotation of the *C. beijerinckii* genome.

3.6 References

Abreu-Goodger, C., and E. Merino. 2005. RibEx: a web server for locating riboswitches and other conserved bacterial regulatory elements. *Nucleic Acids Research* 33:W690-W692.

Annous, B. A., and H. P. Blaschek. 1991. Isolation and characterization of *Clostridium acetobutylicum* mutants with enhanced amylolytic activity. *Applied and Environmental Microbiology* 57(9):2544-2548.

Bahl, H., H. Muller, S. Behrens, H. Joseph, and F. Narberhaus. 1995. Expression of heat shock genes in *Clostridium acetobutylicum*. *Fems Microbiology Reviews* 17(3):341-348.

Belouski, E., D. E. Watson, and G. N. Bennett. 1998. Cloning, sequence, and expression of the phosphofructokinase gene of *Clostridium acetobutylicum* ATCC 824 in *Escherichia coli*. *Current Microbiology* 37(1):17-22.

Bennett, G. N., and F. B. Rudolph. 1995. The central metabolic pathway from acetyl-CoA to butyryl-CoA in *Clostridium acetobutylicum*. *Fems Microbiology Reviews* 17(3):241-249.

Besemer, J., and M. Borodovsky. 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Research* 33(Suppl 2):W451-W454.

Boynton, Z. L., G. N. Bennett, and F. B. Rudolph. 1996a. Cloning, sequencing, and expression of clustered genes encoding β-hydroxybutyryl-coenzyme A (CoA) dehydrogenase, crotonase, and butyryl-CoA dehydrogenase from *Clostridium acetobutylicum* ATCC 824. *Journal of Bacteriology* 178(11):3015-3024.

Boynton, Z. L., G. N. Bennett, and F. B. Rudolph. 1996b. Cloning, sequencing, and expression of genes encoding phosphotransacetylase and acetate kinase from *Clostridium acetobutylicum* ATCC 824. *Applied and Environmental Microbiology* 62(8):2758-2766.

Carver, T., M. Berriman, A. Tivey, C. Patel, U. Bohme, B. G. Barrell, J. Parkhill, and M.-A. Rajandream. 2008. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* 24(23):2672-2676.

Carver, T., N. Thomson, A. Bleasby, M. Berriman, and J. Parkhill. 2009. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 25(1):119-120.

Chen, C. K., and H. P. Blaschek. 1999. Effect of acetate on molecular and physiological aspects of *Clostridium beijerinckii* NCIMB 8052 solvent production and strain degeneration. *Applied and Environmental Microbiology* 65(2):499-505.

Cornillot, E., R. V. Nair, E. T. Papoutsakis, and P. Soucaille. 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. *Journal of Bacteriology* 179(17):5442-5447.

de Mendoza, D., G. E. Schujman, and P. S. Aguilar. 2002. Biosynthesis and function of membrane lipids. In *Bacillus subtilis and its closest relatives: from genes to cells*, 43-55. A. Sonenshein, J. Hoch, and R. Losick, eds. Washington, D.C.: ASM Press.

Düerre, P. 2005. Formation of solvents in clostridia. In *Handbook on clostridia*, 671-693. P. Düerre, ed. London: CRC press.

Ermolaeva, M. D., H. G. Khalak, O. White, H. O. Smith, and S. L. Salzberg. 2000. Prediction of transcription terminators in bacterial genomes. *Journal of Molecular Biology* 301(1):27-33.

Ezeji, T., and H. P. Blaschek. 2008. Fermentation of dried distillers' grains and solubles (DDGS) hydrolysates to solvents and value-added products by solventogenic clostridia. *Bioresource Technology* 99(12):5232-5242.

Ezeji, T. C., N. Qureshi, and H. P. Blaschek. 2003. Production of acetone, butanol and ethanol by *Clostridium beijerinckii* BA101 and *in situ* recovery by gas stripping. *World Journal of Microbiology & Biotechnology* 19(6):595-603.

Fisher, R. A. 1949. A preliminary linkage test with Agouti and Undulated mice; The fifth linkage-group. *Heredity London* 3:229-241.

Gerischer, U., and P. Dürre. 1990. Cloning, sequencing, and molecular analysis of the acetoacetate decarboxylase gene region from *Clostridium acetobutylicum*. *Journal of Bacteriology* 172(12):6907-6918.

Griffiths-Jones, S., S. Moxon, M. Marshall, A. Khanna, S. R. Eddy, and A. Bateman. 2005. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Research* 33:D121-D124.

Hirotsune, S., N. Yoshida, A. Chen, L. Garrett, F. Sugiyama, S. Takahashi, K.-i. Yagami, A. Wynshaw-Boris, and A. Yoshiki. 2003. An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature* 423(6935):91-96.

Jesse, T. W., T. C. Ezeji, N. Qureshi, and H. P. Blaschek. 2002. Production of butanol from starch-based waste packing peanuts and agricultural waste. *Journal of Industrial Microbiology & Biotechnology* 29(3):117-123.

Keis, S., C. F. Bennett, V. K. Ward, and D. T. Jones. 1995. Taxonomy and phylogeny of industrial solvent-producing clostridia. *International Journal of Systematic Bacteriology* 45(4):693-705.

Kosaka, T., S. Nakayama, K. Nakaya, S. Yoshino, and K. Furukawa. 2007. Characterization of the *sol* operon in butanol-hyperproducing *Clostridium saccharoperbutylacetonicum* strain

N1-4 and its degeneration mechanism. *Bioscience Biotechnology and Biochemistry* 71(1):58-68.

Kuo, C. H., and H. Ochman. 2010. The extinction dynamics of bacterial pseudogenes. *Plos Genetics* 6(8).

Lerat, E., and H. Ochman. 2005. Recognizing the pseudogenes in bacterial genomes. *Nucleic Acids Research* 33(10):3125-3132.

Li, H., J. Ruan, and R. Durbin. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Research* 18(11):1851-1858.

McGrath, P. T., H. Lee, L. Zhang, A. A. Iniesta, A. K. Hottes, M. H. Tan, N. J. Hillson, P. Hu, L. Shapiro, and H. H. McAdams. 2007. High-throughput identification of transcription start sites, conserved promoter motifs and predicted regulons. *Nature Biotechnology* 25(5):584-592.

Mortazavi, A., B. A. Williams, K. McCue, L. Schaeffer, and B. Wold. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* 5(7):621-628.

Nagalakshmi, U., Z. Wang, K. Waern, C. Shou, D. Raha, M. Gerstein, and M. Snyder. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320:1344-1349.

Paredes, C. J., I. Rigoutsos, and E. T. Papoutsakis. 2004. Transcriptional organization of the *Clostridium acetobutylicum* genome. *Nucleic Acids Research* 32(6):1973-1981.

Passalacqua, K. D., A. Varadarajan, B. D. Ondov, D. T. Okou, M. E. Zwick, and N. H. Bergman. 2009. Structure and complexity of a bacterial transcriptome. *Journal of Bacteriology* 191(10):3203-3211.

Perkins, T. T., R. A. Kingsley, M. C. Fookes, P. P. Gardner, K. D. James, L. Yu, S. A. Assefa, M. He, N. J. Croucher, D. J. Pickard, D. J. Maskell, J. Parkhill, J. Choudhary, N. R. Thomson, and G. Dougan. 2009. A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus *Salmonella typhi*. *Plos Genetics* 5(7):e1000569.

Petersen, D. J., J. W. Cary, J. Vanderleyden, and G. N. Bennett. 1993. Sequence and arrangement of genes encoding enzymes of the acetone-production pathway of *Clostridium acetobutylicum* ATCC 824. *Gene* 123(1):93-97.

Qureshi, N., A. Lolas, and H. P. Biaschek. 2001. Soy molasses as fermentation substrate for production of butanol using *Clostridium beijerinckii* BA101. *Journal of Industrial Microbiology & Biotechnology* 26(5):290-295.

Rui, H. 1999. The cloning of a putative regulatory gene and the *sol* region from *Clostridium beijerinckii*. Virginia Polytechnic Institute and State University, Virginia Polytechnic Institute, Department of Biochemistry

Schreiber, W., and P. Dürre. 2000. Differential expression of genes within the *gap* operon of *Clostridium acetobutylicum*. *Anaerobe* 6(5):291-297.

Severin, A., J. Woody, Y.-T. Bolon, B. Joseph, B. Diers, A. Farmer, G. Muehlbauer, R. Nelson, D. Grant, J. Specht, M. Graham, S. Cannon, G. May, C. Vance, and R. Shoemaker. 2010. RNA-Seq Atlas of *Glycine max*: A guide to the soybean transcriptome. *BMC Plant Biology* 10(1):160.

Shi, Z., and H. P. Blaschek. 2008. Transcriptional analysis of *Clostridium beijerinckii* NCIMB 8052 and the hyper-butanol-producing mutant BA101 during the shift from acidogenesis to solventogenesis. *Applied and Environmental Microbiology* 74(24):7709-7714.

Steil, L., M. Serrano, A. O. Henriques, and U. Volker. 2005. Genome-wide analysis of temporally regulated and compartment-specific gene expression in sporulating cells of *Bacillus subtilis*. *Microbiology-Sgm* 151:399-420.

Svensson, Ö., L. Arvestad, and J. Lagergren. 2006. Genome-wide survey for biologically functional pseudogenes. *PLoS Computational Biology* 2(5):e46.

t Hoen, P. A. C., Y. Ariyurek, H. H. Thygesen, E. Vreugdenhil, R. Vossen, R. X. de Menezes, J. M. Boer, G. J. B. van Ommen, and J. T. den Dunnen. 2008. Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. *Nucleic Acids Research* 36(21):e141.

Tomas, C. A., J. Beamish, and E. T. Papoutsakis. 2004. Transcriptional analysis of butanol stress and tolerance in *Clostridium acetobutylicum*. *Journal of Bacteriology* 186(7):2006-2018.

Tomas, C. A., N. E. Welker, and E. T. Papoutsakis. 2003. Overexpression of *groESL* in *Clostridium acetobutylicum* results in increased solvent production and tolerance, prolonged metabolism, and changes in the cell's transcriptional program. *Applied and Environmental Microbiology* 69(8):4951-4965.

Walter, K. A., R. V. Nair, J. W. Cary, G. N. Bennett, and E. T. Papoutsakis. 1993. Sequence and arrangement of two genes of the butyrate-synthesis pathway of *Clostridium acetobutylicum* ATCC 824. *Gene* 134(1):107-111.

Wang, S. T., B. Setlow, E. M. Conlon, J. L. Lyon, D. Imamura, T. Sato, P. Setlow, R. Losick, and P. Eichenberger. 2006. The forespore line of gene expression in *Bacillus subtilis*. *Journal of Molecular Biology* 358(1):16-37.

Wilhelm, B. T., S. Marguerat, S. Watt, F. Schubert, V. Wood, I. Goodhead, C. J. Penkett, J. Rogers, and J. Bahler. 2008. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* 453(7199):1239-1243.

Wu, M., M. Wang, J. H. Liu, and H. Huo. 2008. Assessment of potential life-cycle energy and greenhouse gas emission effects from using corn-based butanol as a transportation fuel. *Biotechnology Progress* 24(6):1204-1214.

Wurtzel, O., R. Sapra, F. Chen, Y. W. Zhu, B. A. Simmons, and R. Sorek. 2010. A single-base resolution map of an archaeal transcriptome. *Genome Research* 20(1):133-141.

Zheng, D. Y., A. Frankish, R. Baertsch, P. Kapranov, A. Reymond, S. W. Choo, Y. T. Lu, F. Denoeud, S. E. Antonarakis, M. Snyder, Y. J. Ruan, C. L. Wei, T. R. Gingeras, R. Guigo, J. Harrow, and M. B. Gerstein. 2007. Pseudogenes in the ENCODE regions: Consensus annotation, analysis of transcription, and evolution. *Genome Research* 17(6):839-851.

CHAPTER 4. GENOME-WIDE TRANSCRIPTIONAL DYNAMICS IN *CLOSTRIDIUM BEIJERINCKII* 8052 REVEALED BY SINGLE-NUCLEOTIDE RESOLUTION RNA-SEQ[1]

4.1 Abstract

Clostridium beijerinckii is a prominent solvent-producing microbe that has great potential for biofuel and chemical industries. Although transcriptional analysis is essential to understand gene functions and regulation and thus elucidate proper strategies for further strain improvement, limited information is available on the genome-wide transcriptional analysis for *C. beijerinckii*. The genome-wide transcriptional dynamics of *C. beijerinckii* NCIMB 8052 over a batch fermentation process was investigated using high-throughput RNA-Seq technology. The gene expression profiles indicated that the glycolysis genes were highly expressed throughout the fermentation, with comparatively more active expression during acidogenesis phase. The expression of acid formation genes was down-regulated at the onset of solvent formation, in accordance with the metabolic pathway shift from acidogenesis to solventogenesis. The acetone formation gene (*adc*), as a part of the *sol* operon, exhibited highly-coordinated expression with the other *sol* genes. Out of the > 20 genes encoding alcohol dehydrogenase in *C. beijerinckii*, Cbei_1722 and Cbei_2181 were highly up-regulated at the onset of solventogenesis, corresponding to their key roles in primary alcohol production. Most sporulation genes in *C. beijerinckii* 8052 demonstrated similar temporal expression patterns to those observed in *B. subtilis* and *C. acetobutylicum*, while sporulation sigma factor genes *sigE* and *sigG* exhibited accelerated and stronger expression in *C. beijerinckii* 8052, which is consistent with the more rapid forespore and endspore development in this strain. Global expression patterns for specific gene functional classes were examined using self-organizing map analysis. The genes associated with specific functional classes demonstrated global expression profiles corresponding to the cell physiological variation and metabolic pathway switch. The results from this work provided insights for further *C. beijerinckii* strain improvement employing system biology-based strategies and metabolic engineering approaches.

[1] A modified version of Chapter 4 was published in BMC Genomics, 2012, 13:102. (Y. Wang lead author with co-authors X. Li, Y. Mao and H.P. Blaschek).

4.2 Introduction

Solvents such as acetone, butanol and ethanol (ABE) produced through microbial fermentation represent important potential renewable fuels and chemicals (Dürre, 2007). *Clostridium acetobutylicum* and *C. beijerinckii* are among the prominent solvent-producing species. *C. beijerinckii* exhibits a broader substrate range and optimum pH for growth and solvent production (Ezeji and Blaschek, 2008); thus it may have greater potential for biosolvent production than *C. acetobutylicum*.

The genome of *C. beijerinckii* NCIMB 8052 was sequenced by the DOE Joint Genome Institute in 2007. The genome size is 6.0 Mb, which is 50% larger than that of *C. acetobutylicum* ATCC 824. The *C. beijerinckii* 8052 solvent-producing genes are all located on the chromosome, as opposed to the location of these genes on a mega-plasmid in *C. acetobutylicum* 824. Transcriptional analysis is essential to understand gene functions and regulation and thus elucidate proper strategies for further strain improvement. While global gene expression patterns for *C. acetobutylicum* have been studied extensively (Alsaker and Papoutsakis, 2005; Alsaker et al., 2010; Alsaker et al., 2005; Alsaker et al., 2004; Janssen et al., 2010; Jones et al., 2008; Tomas et al., 2004), limited information is available on the genome-wide transcriptional analysis for *C. beijerinckii* (Shi and Blaschek, 2008). RNA-Seq is based on the high-throughput DNA sequencing technology and provides an approach to profile and quantify gene expression to an unprecedented resolution and depth (Mortazavi et al., 2008; Nagalakshmi et al., 2008; Passalacqua et al., 2009; Perkins et al., 2009). RNA-Seq technology has several advantages over DNA microarray methods for transcriptional analysis. First, microarray and other hybridization techniques exhibit a relatively low dynamic range for the detection of transcriptional levels due to background, saturation and poor sensitivity for gene expression at very low or high levels (van Vliet, 2010), while RNA-Seq does not have an upper limit for quantification and has a larger dynamic range of expression levels. In addition, when compared to microarray-based approaches, RNA-Seq has very low background noise because DNA sequence reads can be unambiguously mapped to unique regions along the genome (Wang et al., 2009). Several recent studies have proven that RNA-Seq is a powerful tool for transcriptional analysis (Legendre et al., 2010; Perkins et al., 2009; Severin et al., 2010). In this study, a genome-wide transcriptional analysis of *C. beijerinckii* 8052 over the course of a batch fermentation was described from an unbiased perspective using RNA-Seq technology. The findings from this work provided insights for further *C.*

beijerinckii strain improvement employing system biology-based strategies and metabolic engineering. Furthermore, this work is also an essential methodology reference for conducting transcriptional analysis employing next-generation sequencing technology.

4.3 Materials and methods

4.3.1 Bacterial culture and fermentation experiment

Laboratory stocks of *C. beijerinckii* 8052 spores were stored in sterile H_2O at 4 ^oC (Annous and Blaschek, 1991). Spores were heat-shocked at 80 °C for 10 min, followed by cooling on ice for 5 min. The heat-shocked spores were inoculated at a 1% inoculum level into tryptone–glucose–yeast extract (TGY) medium containing 30 g L^{-1} tryptone, 20 g L^{-1} glucose, 10 g L^{-1} yeast extract and 1 g L^{-1} L-cysteine. The TGY culture was incubated at 35±1 ^oC for 12-14 h in an anaerobic chamber under N_2:CO_2:H_2 (volume ratio of 85:10:5) atmosphere. Subsequently, actively growing culture was inoculated into a model solution containing 60 g L^{-1} glucose, 1 g L^{-1} yeast extract, and filter-sterilized P2 medium (Qureshi et al., 2001) in a Sixfors bioreactor system (Infors AG, Bottmingen, Switzerland). Oxygen-free nitrogen was flushed through the broth to initiate anaerobiosis until the culture initiated its own gas production (CO_2 and H_2). Temperature was controlled at 35±1 ^oC. A stirring at 50 rpm was employed for mixing. Cell density and product concentration were monitored through the course of fermentation. For sequencing purpose, samples were taken over the early exponential, late exponential and stationary phases (samples 1-6 at 2, 4.5, 10, 14, 17 and 26.5 h respectively as shown in Figure 4.1A).

4.3.2 Culture growth and fermentation products analysis

Culture growth was measured by following optical density (OD) in the fermentation broth at A_{600} using a BioMate 5 UV-Vis Spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA). ABE, acetic acid, and butyric acid concentrations were quantified using gas chromatography (GC) system as previously described (Qureshi et al., 2001).

4.3.3 RNA isolation, library construction and sequencing

For RNA isolation, 10 ml cultures were harvested at each time point, and centrifuged at 4,000 × g for 10 min at 4 °C. Total RNA was extracted from the cell pellet using Trizol reagent based on manufacture's protocol (Invitrogen, Carlsbad, CA) and further purified

using RNeasy mini kit (Qiagen, Valencia, CA). DNA was removed using a DNA-*free*™ kit (Ambion Inc., Austin, TX). RNA quality was assessed using a 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA). RNA concentration was determined with a nanodrop (Biotek Instruments, Winooski, VT). Bacterial 16S and 23S ribosomal RNAs were removed with a MICROBExpress™ kit (Ambion Inc., Austin, TX). The enriched mRNA was converted to a RNA-Seq library using the mRNA-Seq library construction kit (Illumina Inc., San Diego, CA) following manufacturer's protocols. For samples 1 to 6, two samples were pooled and sequenced on one single lane of an eight-lane flow cell with the Genome Analyzer IIx system (Illumina Inc., San Diego, CA). However, sample 6 yielded a poor read quality following the first sequencing. In order to obtain enough sequencing depth, sample 6 was sequenced again using one single lane under otherwise identical conditions. The derived sequence reads were 75 nt long. The overall error rate of the control DNA was < 0.6%. The total number of reads generated from each library is summarized in Table 4.1.

4.3.4 RNA-Seq data analysis

The generated 75-nt reads were mapped to the *C. beijerinckii* 8052 genome using MAQ, and those that did not align uniquely to the genome were discarded (Li et al., 2008). The quality parameter (-q) used in MAQ pileup was set to 30. Each base was assigned a value based on the number of mapped sequence coverage. The quantitative gene expression value, RPKM (reads/Kb/Million), was calculated using custom Perl scripts by normalizing the sequence coverage over the gene length and total unambiguously mapped reads in each library (Mortazavi et al., 2008; Severin et al., 2010).

Z score was calculated as $Z = (X - X_{av})/X_{SD}$. X is the $\log_2$-transformed RPKM value for a given gene in a specific sample, X_{av} is the mean of $\log_2$-transformed RPKM values for that gene in all six samples, and X_{SD} is the standard deviation of $\log_2$-transformed RPKM for that gene across all samples (Benedito et al., 2008; Severin et al., 2010).

The gene expression results were visualized colorimetrically using heatmap plots. R language and the open source software Bioconductor based on R were used for the data analysis and plots (Anders and Huber, 2010; Gentleman et al., 2004; Robinson et al., 2010). Specifically, functions in Heatplus package in Bioconductor and the R package gplots were employed for constructing the heatmap plots. Time course RPKM values were first transformed to $\log_2$-scale. The $\log_2$-transfromed RPKM values were then properly centered

for better representation of the data by the heatmap plots. More details about this method were summarized in the Additional file B4.1 in Appendix B.

4.3.5 Self-organizing maps (SOM) analysis

Global gene expression patterns were analyzed with self-organizing maps (SOM) analysis (Alsaker and Papoutsakis, 2005). The RPKM value of a gene at each time point was first divided by the square root of the sum of the squares of the gene's RPKM values at all six time points, and therefore the sum of the squares of each gene's standardized ratios within the experiment is 1. The preprocessed gene expression values were then standardized and grouped into 16 SOM clusters based on the expression pattern similarity (Figure 4.5). In order to examine the SOM clusters for enriched Clusters of Orthologous Groups of proteins (COG), the frequency at which the genes in each cluster fell in every COG category was calculated and plotted colorimetrically (Figure 4.5). If one gene belongs to more than one COG categories, the gene was counted once for each category (Alsaker and Papoutsakis, 2005). Data were processed and visualized with Gene Cluster 3.0 (Eisen et al., 1998) and the SOMClustering and Java TreeView packages in GenePattern (Reich et al., 2006).

4.3.6 Real time qRT-PCR

Quantitative reverse transcription PCR (qRT-PCR) was performed for selected genes in order to validate the quantification of gene expression levels obtained by RNA-Seq. Nine genes that are important involved in the central pathways for glycolysis, acidogenesis and solventogenesis were chosen for the examination. Cbei_2428 was used as the endogenous control gene based on its relatively constant expression levels throughout the fermentation process as discussed earlier. Specific primer sequences for tested genes are listed in Table B4.5. Triplicate reactions were performed using Power SYBR green PCR master mix (Applied Biosystems, Carlsbad, CA) on an ABI Prism 7900HT fast real-time PCR machine (Applied Biosystems). The normalized threshold value (ΔCt) of each gene along time course was compared to that of the same gene at the first time point (time point 1 in Figure 4.1A).

4.3.7 RNA-Seq data accession number

The RNA-Seq sequencing data have been deposited in the NCBI Sequence Read Archive (SRA) under the accession number SRA045799, and also included in the Additional file A3.1 in Appendix A.

4.4 Results and discussion

4.4.1 Growth kinetics and ABE fermentation

The *C. beijerinckii* 8052 growth response in the batch fermentation is illustrated in Figure 4.1A. The fermentation shifted from acidogenesis to solventogenesis at approximately 4.5-8h. Production of solvents was detected at between 4.5-6.5 h after the start of fermentation, which corresponded to the late exponential growth phase. Butanol continued to increase throughout the stationary phase (Figure 4.1B). Samples for RNA isolation were collected at various time points during acidogenesis (2 and 4.5 h) and solventogenesis (after 6.5 h) (Figure 4.1A). These samples were sequenced for dynamic transcriptional profiling for *C. beijerinckii* 8052.

4.4.2 Overall gene transcription dynamics

The 75-nt sequence reads were mapped to the *C. beijerinckii* 8052 genome. Only those reads that mapped unambiguously to the genome were used for further analyses (Table 4.1). Based on RNA-Seq sequence data, 3386 out of 5020 (67.4%) protein-coding genes had detectable expression in all six sampling time points, while 78 demonstrated no transcripts over all six time points, and are likely silent (Wang et al., 2011).

On the other hand, there were some genes that demonstrated little variation in expression levels throughout the entire fermentation process, and they are regarded as putative housekeeping genes (HKGs). In this study, 177 protein-coding genes were identified as putative HKGs with the lowest coefficient of variation (CV) in RPKM (reads/Kb/Million) values among all the sampling points (Wang et al., 2011). For accurate gene expression analysis, normalization of gene expression against HKGs is generally required. In the qRT-PCR test in this work, a putative HKG (Cbei_2428, encoding peptidase T) with highly constant expression levels across all six sampling points was selected as the endogenous control gene.

Detrended Correspondence Analysis (DCA) was conducted on the RNA-Seq data to generate overall dynamic transcription profiles throughout the batch fermentation process. DCA is an ordination technique that uses detrending to remove the ‘arch effect’ usually existing in correspondence analysis (Hill and Gauch, 1980). As shown in Figure 4.2A, an obvious temporal variation trend was observed for the overall gene transcription. While the

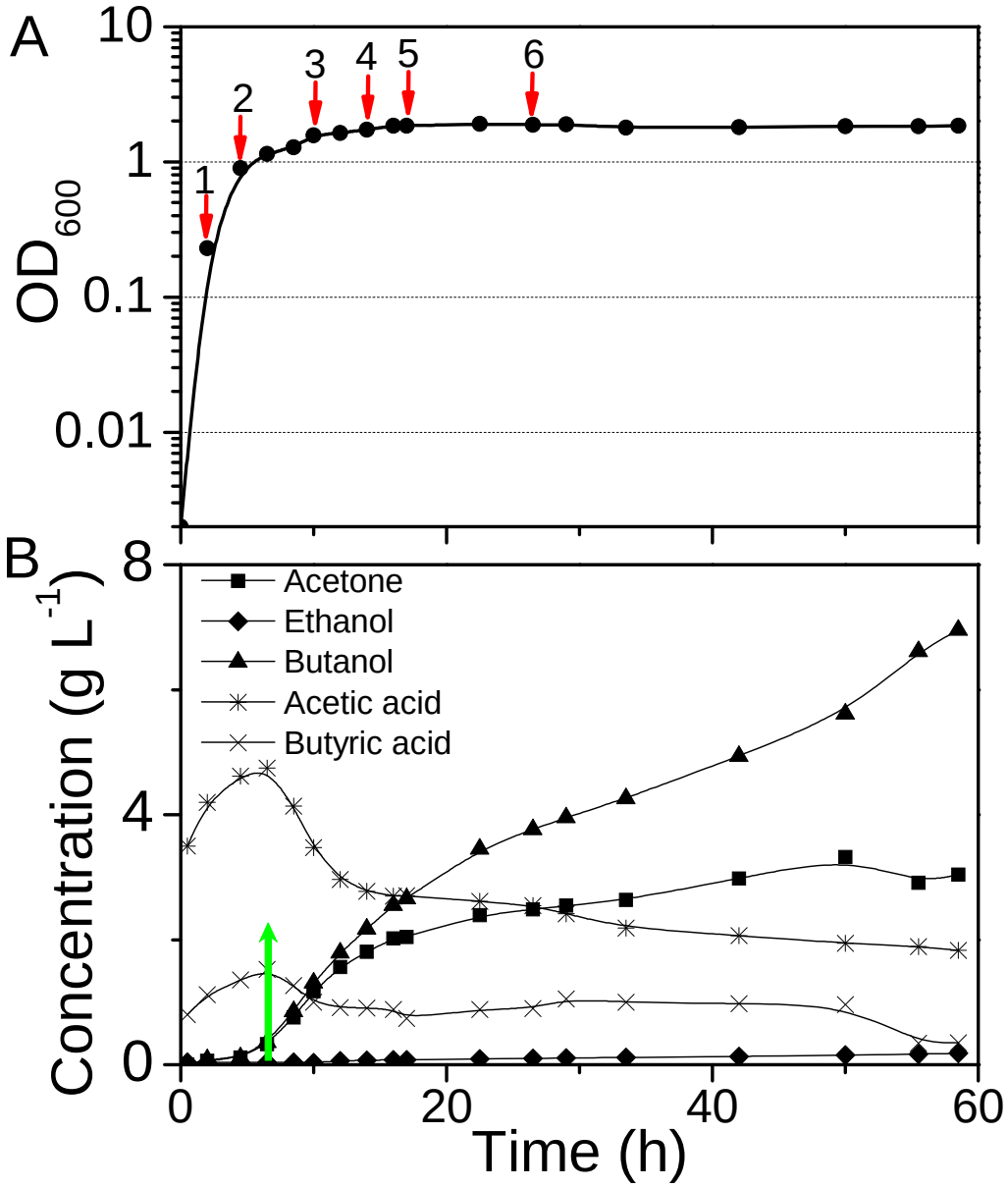


Figure 4.1 Fermentation kinetics of *C. beijerinckii* 8052 batch culture. (A) Cell growth curve with sampling points for RNA-Seq indicated by red arrows. (B) Solvents and acids production over time with the onset of solvent production indicated by a green arrow.

first three samples which were from the exponential and transition phases presented both significant first detrended correspondence (DC1, explained 40.6% of detrended correspondence) and second detrended correspondence (DC2, explained 15.2% of detrended correspondence), the last three samples which were from stationary phase demonstrated only significant DC2 but little DC1. The temporal variation trend of the overall gene expression is

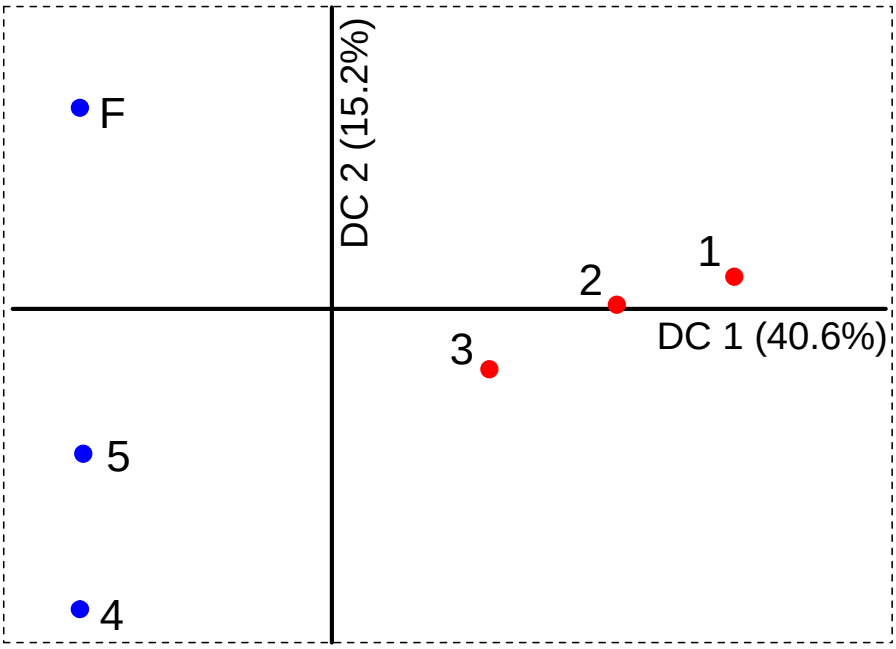
Table 4.1 Summary of RNA-Seq sequencing and data analysis results

Sample	1	2	3	4	5	6[1]	Total
Time collected (h)	2	4.5	10	14	17	26.5	
Total number of reads	8988633	9457480	8011531	8448929	10363535	38574501	83844609
No. of reads mapped	8473125	9037616	7514804	7730815	9842491	37676913	80275764
No. of reads unambiguously mapped	6776544	7274568	6096405	6189652	8096169	35027722	69461060
No. of genes with detectable expression[2]	4219	4082	4496	4453	4487	4750	5024
Range in expression levels (RPKM)	3.2×10^{-1} ~ 2.5×10^{4}	5.8×10^{-2} ~ 6.0×10^{4}	4.5×10^{-2} ~ 2.5×10^{4}	7.0×10^{-2} ~ 9.0×10^{4}	1.0×10^{-1} ~ 8.6×10^{4}	3.6×10^{-2} ~ 9.8×10^{4}	3.6×10^{-2} ~ 9.8×10^{4}

[1]Sample 6 was sequenced with 1 sample/lane, while samples 1-5 with 2 samples/lane. See Section 4.3 for more details;

[2]Pseudogenes included.

A



B

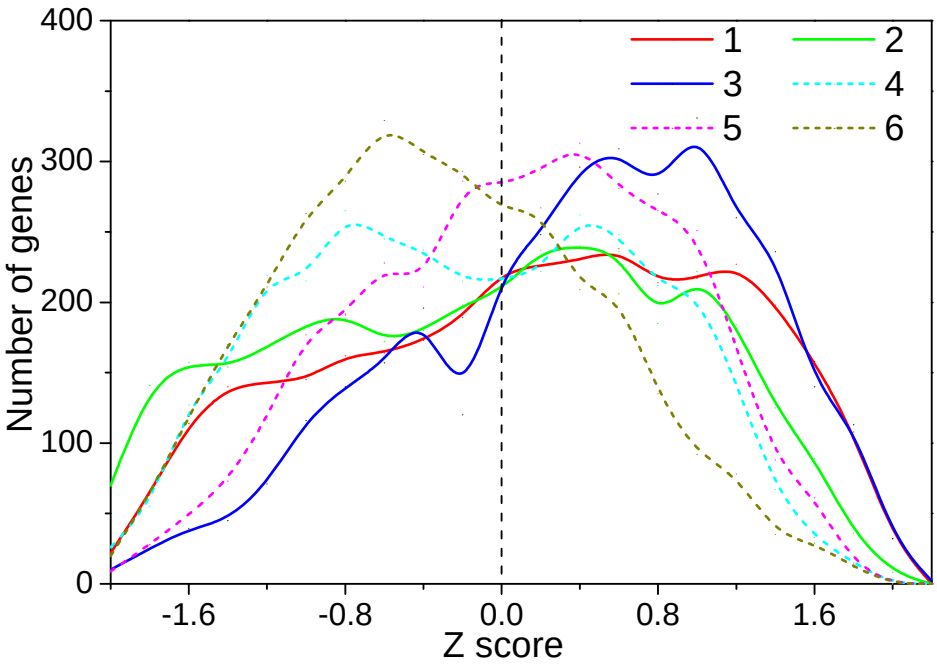


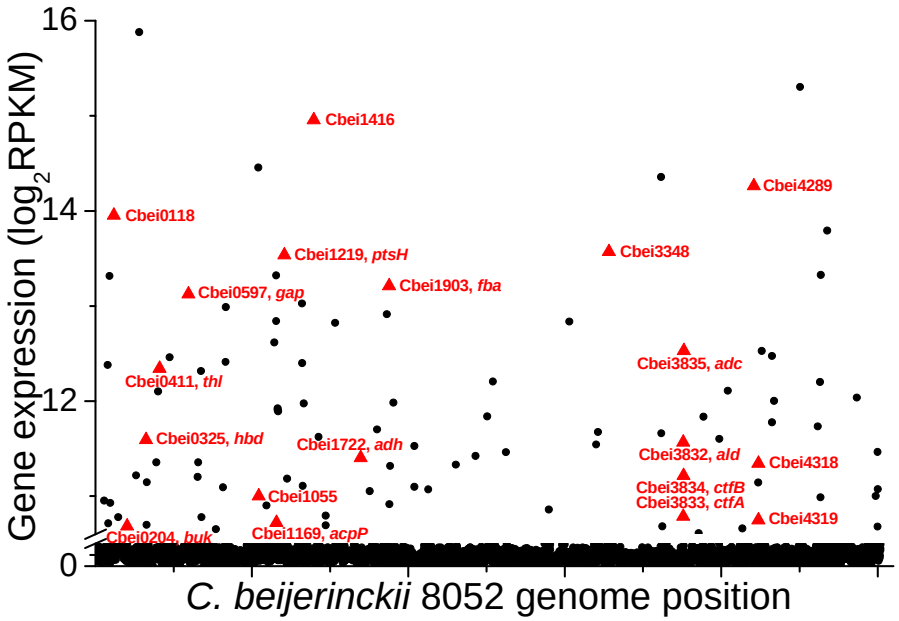
Figure 4.2 Overall gene transcription profiles during the fermentation based on RNA-Seq data. (A) Detrended Correspondence Analysis (DCA). Sampling points from exponential phase and transition phase (1-3) were indicated in red, while points from stationary phase (4-6) were indicated in blue. (B) Gene distribution based on Z-score analysis. Samples 1-6 were indicated by different line styles and colors.

consistent with different cell physiological states during transition from acidogenesis to solventogenesis.

A Z-score analysis was conducted to investigate the overall gene expression pattern at each time point. Z-score measures the number of standard deviations that a gene expression in a specific sample is from the mean expression of this gene over all samples. It is a dimensionless quantity derived by subtracting the mean expression of a specific gene over all samples from the individual expression level in a specific sample and then dividing the difference by the standard deviation (Benedito et al., 2008; Severin et al., 2010). The results demonstrated temporally dynamic patterns over the course of cell growth and batch fermentation (Figure 4.2B). Generally, during exponential and early stationary phases (2-10h), more genes shifted towards higher expression values (to the right on Figure 4.2B), while during late stationary phases, more genes shifted to lower expression levels (to the left on Figure 4.2B). A list of genes whose Z-score was > 1.7 in each sample is summarized in Table B4.1 in Appendix B. The highly expressed gene categories were different at different time points during the course of the fermentation. For example, translation related genes (COG J) were predominant (49 out of total 141 genes in this list) in sample 1 (at 2 h), indicating fast protein synthesis during this growth stage; while signal transduction mechanism involved genes (COG T) were overrepresented at 10 h, indicating the cellular reaction to the environmental changes.

The expression of different genes in a specific sample can be compared by plotting $\log_2$-transformed RPKM values of all genes against their loci along the genome (Figure 4.3). Sample 2 represented the beginning of the transition from acidogenesis to solventogenesis. In this sample, among the highest expressed genes, besides the genes encoding hypothetical proteins (31 out of the top 100 highest expressed), 12 are ribosomal protein genes, indicating active protein synthesis machinery in the cell at this time point. In addition, genes involved in acid and solvent formation were detected as highly expressed, including *adc* (Cbei_3835), *thl* (Cbei_0411), *hbd* (Cbei_0325), *ald* (Cbei_3832), *adh* (Cbei_1722), *ctfB* (Cbei_3834), *ctfA* (Cbei_3833) and *buk* (Cbei_0204). Highly expressed genes also included those encoding proteins that mediate electron transfer, such as Cbei_1416 (rubrerythrin), Cbei_0118, Cbei_3348 (desulfoferrodoxin), Cbei_4318, Cbei_4319 (flavodoxin), and those genes related to cell motility (Cbei_4289), phosphotransferase system (PTS) for sugar uptake (Cbei_1219) and glycolytic activity (Cbei_1903, *fba* and Cbei_0597, *gap*) (Figure 4.3 and Table B4.2 in

A



B

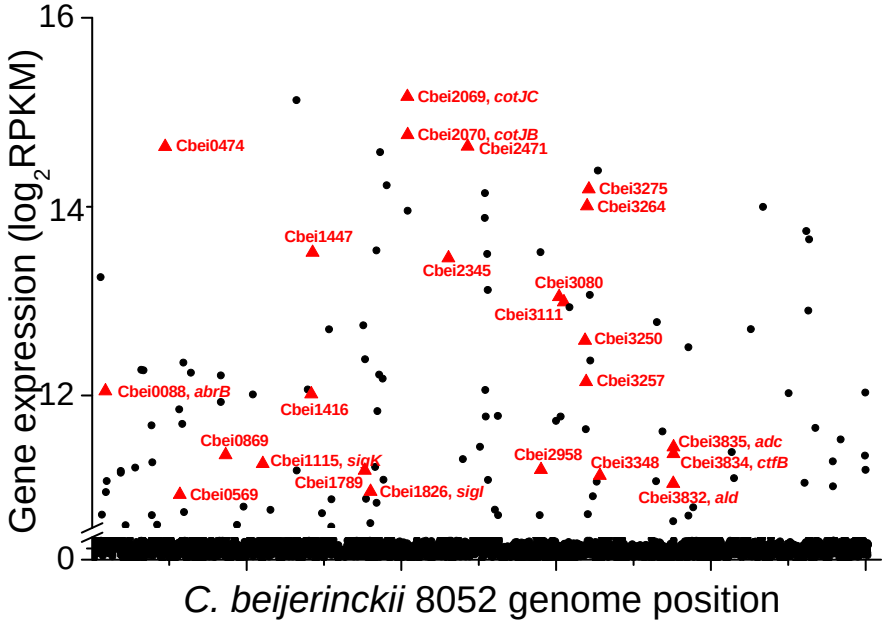


Figure 4.3 Identification of highly expressed genes in samples 2 and 4 by plotting the log_2-scaled gene expression values against the gene loci through the genome. (A) Plot of gene expression level (log_2(RPKM)) against the gene locus through the genome in sample 2 (at 4.5 h). (B) Plot of gene expression level (log_2(RPKM)) against the gene locus through the genome in sample 4 (at 14 h). The highly expressed genes involved in the main metabolic pathways during the fermentation were labeled with larger triangles.

Appendix B). The time frame of sample 4 was consistent with the transition to non-active growth and clostridial spore formation. While the solventogenesis genes (*adc*, *ald*, *ctfB, ctfA*) and several ferredoxin activity protein encoding genes (Cbei_3257, _3348 and _0569) were still actively expressed at this stage (Figure 4.3 and Table B4.2 in Appendix B), genes encoding sporulation-related proteins demonstrated the highest transcriptional activities. These genes included those encoding spore coat proteins (Cbei_2069, *cotJC* and Cbei_2070, *cotJB*), small acid-soluble spore proteins (Cbei_2471, _0474, _3275, _3264, _1447, _2345, _3080, _3111 and _3250), AbrB family transcriptional regulator (Cbei_0088, annotated as AbrB family stage V sporulation protein T in *C. acetobutylicum*) and sporulation sigma factor SigK (Cbei_1115).

It is interesting that the hypothetical genes account for a large fraction of the highly expressed genes (in both samples 2 and 4 presented above). As we mentioned previously (Wang et al., 2011), the genome annotation for *C. beijerinckii* 8052 is far from complete. Many of the hypothetical genes predicted by computational analysis algorithms may have important functions for the cell activities. The RNA-Seq results herein provided useful references for future studies in defining the specific functions of these genes.

4.4.3 Expression of glycolysis genes

Generally, the glycolysis genes encoding the pathway for conversion of glucose to pyruvate were expressed at high levels throughout the entire fermentation process (Figure 4.4A). Comparatively, the expression of most of these genes decreased slightly after entering the stationary phase, down-regulated by 2- to 4-fold. In *C. beijerinckii* 8052, three genes (Cbei_0584, Cbei_0998 and Cbei_4852) are annotated as encoding 6-phosphofructokinase (pfk). While Cbei_0584 and _4852 were down-regulated 2.3- and 2.7-fold respectively after entering stationary phase, Cbei_0998 was up-regulated 3.8-fold at time point 3 and expressed at higher levels throughout stationary phase. Cbei_0485, Cbei_1412 and Cbei_4851 are all annotated as encoding pyruvate kinase (pyk). While Cbei_0485 and _4851 decreased by 2.5-fold after entering stationary phase, Cbei_1412 was expressed at a higher level throughout stationary phase when compared to the exponential phase. These observations imply that different allelic genes may be induced at different specific stages and play different roles during the fermentation process. The expression patterns of glycolysis genes in this study are similar to those found for *C. beijerinckii* 8052 in Shi and Blaschek (2008)

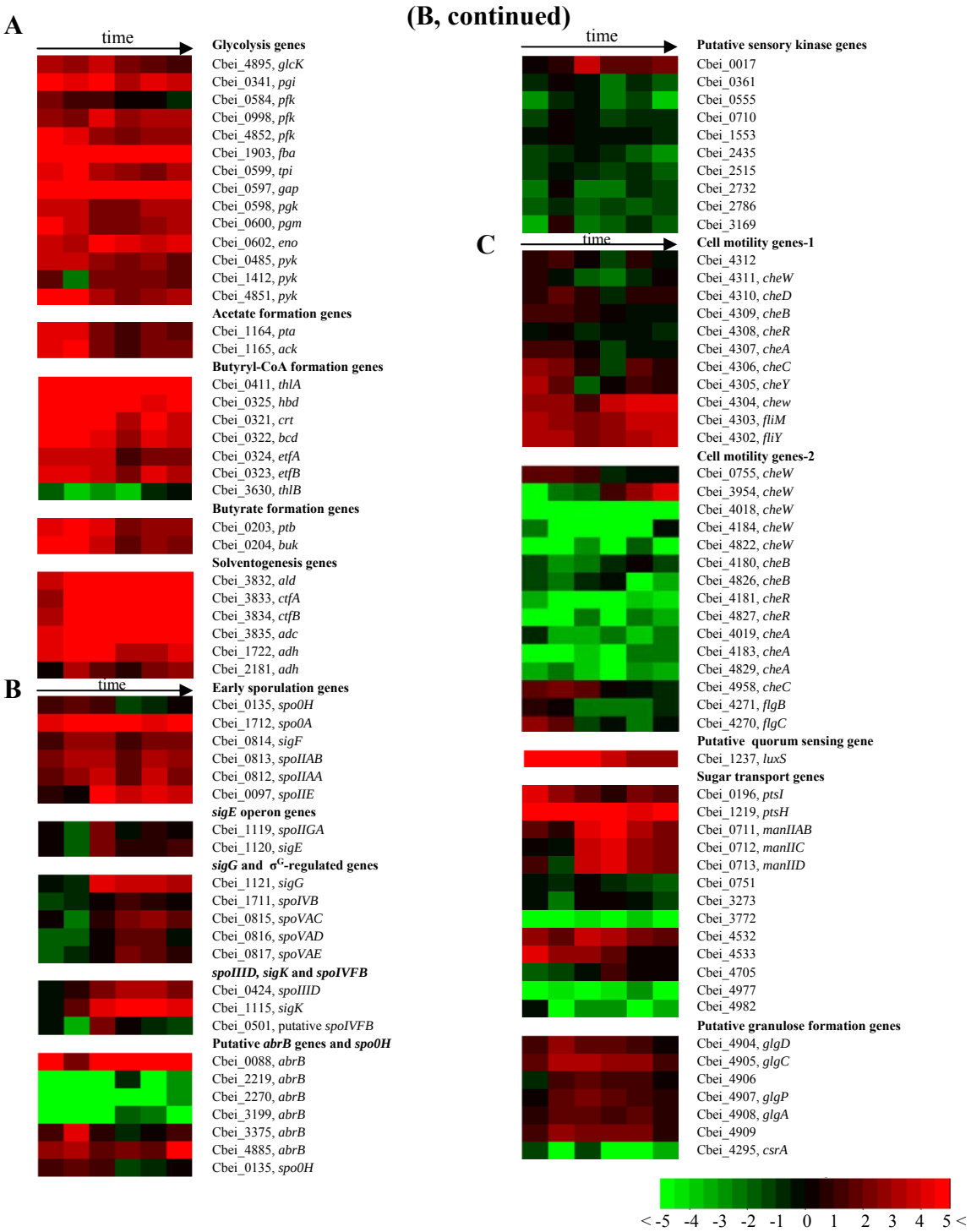


Figure 4.4 Time course expression profiles of important genes during the batch fermentation process. (A) Genes related to glycolysis, acidogenesis and solventogenesis. (B) Sporulation genes, putative *abrB* genes and putative sensory kinase genes. (C) Genes related to cell motility, putative quorum sensing, sugar transport and granulose formation.

(Shi and Blaschek, 2008). While in a time course transcriptional analysis of *C. acetobutylicum* 824 during a batch fermentation, most of the glycolysis genes showed higher expression during stationary-phase (Alsaker and Papoutsakis, 2005).

4.4.4 Expression of acidogenesis and solventogenesis genes

The acetate formation genes encoding phosphotransacetylase (*pta*, Cbei_1164) and acetate kinase (*ack*, Cbei_1165) and the butyrate formation genes encoding phosphate butyryltransferase (*ptb*, Cbei_0203) and butyrate kinase (*buk*, Cbei_0204) were up-regulated during the acidogenesis phase and peaked at time point 2, and declined 2- to 4-fold at the onset of solvent formation (Figure 4.4A). Based on the sequencing data, the *pta-ack* and *ptb-buk* gene operon structures were identified (Wang et al., 2011). The coordinated expression patterns of these two gene clusters further confirmed this transcriptional organization. Expression of genes encoding proteins involved in butyryl coenzyme A (butyryl-CoA) formation was induced early during the acidogenesis phase, and remained at high levels after cells entered stationary phase (time point 3), and afterwards decreased 3- to 7-fold at late stationary phase, although the absolute transcription level was still high (Figure 4.4A).

The product of the *thlA* (Cbei_0411) in *C. beijerinckii* has 77% amino acid sequence identity to the product of *thlA* (CAC2873) in *C. acetobutylicum*, and the product of a second thiolase gene (*thlB*, Cbei_3630) shows 90% amino acid sequence identity to that of *thlB* (or *thil*) in *C. acetobutylicum* (CAP0078). A transcriptional analysis with qRT-PCR for continuous cultures of *C. acetobutylicum* indicated that the level of *thlB* transcripts was much lower compared to the level of *thlA* transcript, and the expression of *thlB* was maximal at about 10 to 15 h after induction of solvent formation (Winzer et al., 2000). Whereas another transcriptional analysis for continuous cultures of *C. acetobutylicum* showed that *thlB* was highly induced only during the transition state from acidogenesis to solventogenesis (Grimmler et al., 2011). In this study, the expression level of *thlB*, although much lower than that of *thlA*, was found to be maximal in late stationary phase (time points 5 and 6), which is around 10 to 20 h after onset of solventogenesis (Figure 4.1). While thlA is involved in the primary metabolism of acid and solvent formation, the physiological function of thlB warrants further study. In the *C. beijerinckii* 8052 genome, there are different genes encoding for isoenzymes active in specific stages and playing different roles. Transcriptomic analysis

leads to insights for further biochemical characterization of the isoenzymes properties and functions.

In *C. acetobutylicum* 824, the genes *adhE* (bifunctional acetaldehyde-CoA/alcohol dehydrogenase, CA_P0162), *ctfA* (acetoacetyl-CoA: acetate/butyrate-CoA transferase subunit A, CA_P0163) and *ctfB* (acetoacetyl-CoA: acetate/butyrate-CoA transferase subunit B, CA_P0164) are located in the *sol* (solvent formation) operon on the mega-plasmid pSOL1, while *adc* (acetoacetate decarboxylase, CA_P0165) is organized in a monocistronic operon in the opposite direction (Cornillot et al., 1997; Gerischer and Dürre, 1990; Petersen et al., 1993). The regulation of *adc* was believed to differ from that of the *sol* operon, and the expression of *adc* was reported to be initiated earlier than the *sol* operon in both batch and continuous cultures of *C. acetobutylicum* (Alsaker and Papoutsakis, 2005; Feustel et al., 2004; Gerischer and Dürre, 1992; Grimmler et al., 2011; Sauer and Dürre, 1995). Based on our sequencing data, a *sol* operon organized in the order of *ald* (Cbei_3832, encoding aldehyde dehydrogenase)-*ctfA* (Cbei_3833)-*ctfB* (Cbei_3834)-*adc* (Cbei_3835) was revealed in *C. beijerinckii* 8052 (Wang et al., 2011). Consistent with this transcriptomic organization, coordinated expression was observed for the *sol* operon genes. Their expression was up-regulated at the onset of solventogenesis phase and maintained at high levels during the stationary phase (Figure 4.4A).

Two iron-containing alcohol dehydrogenase genes *adhA* (GenBank: AF497741.2) and *adhB* (GenBank: AF497742.1) were previously characterized in the solvent-producing strain *C. beijerinckii* NRRL B592 (Chen, 1995; Shi and Blaschek, 2008). The enzymes encoded by both genes were classified as primary alcohol dehydrogenases, which catalyze the reaction for primary alcohol (that is, butanol and ethanol, but not isopropanol) production (Chen, 1995). The product of Cbei_2181 in *C. beijerinckii* 8052 exists 99% and 97% amino acid sequence identity to that of *adhA* and *adhB*, and the product of Cbei_1722 has 97% and 94% amino acid sequence identity to that of *adhB* and *adhA*, respectively. Cbei_2181 and _1722 may play key roles during primary alcohol production in *C. beijerinckii* 8052. Both genes were induced to high levels (9 and 4-fold up respectively) when the fermentation transitioned from the acid production phase to the solvent production phase (time point 2), and then decreased. Similar transient up-regulation of *adh* genes and enzymes associated with active solvent production was previously reported for *C. beijerinckii* NRRL B592 and *C. acetobutylicum* 824 (Chen, 1995; Walter et al., 1992; Welch et al., 1989).

There are 20 annotated alcohol dehydrogenase genes in *C. beijerinckii* 8052, suggesting the great potential for solvent production of this microorganism. Most of these genes were induced early during the acidogenesis phase and expressed at high levels throughout the solvent production process, although some of them did not have detectable expression at some specific time points (Table B4.3 in Appendix B). The specific functions of different alcohol dehydrogenase genes associated with *C. beijerinckii* require further investigation.

4.4.5 Expression of sporulation genes

Production of solvents was first detected between 4.5-6.5 h after the start of fermentation, which corresponded to the late exponential growth phase. The initiation of sporulation in *C. beijerinckii* 8052 was concurrent with the onset of solventogenesis.

In *B. subtilis*, the gene *spo0H* encodes the earliest acting sigma factor associated with sporulation, σ^H, which regulates the early sporulation genes, including *spo0A* and *sigF* operon (Predich et al., 1992). The phosphorylated Spo0A (Spo0A ~ P) by phosphorelay system indirectly induces *spo0H* transcription by repressing *abrB* transcription (Dürre and Hollergschwandner, 2004). In clostridia, no similar phosphorelay components and regulation system have been observed. It have been reported that in *C. acetobutylicum*, *spo0H* and *spo0A* were both constitutively expressed at constant levels throughout the growth cycle, and the amount of *spo0H* transcript was much lower than that of *spo0A* (Dürre and Hollergschwandner, 2004). While a recent time-course transcriptional analysis of *C. acetobutylicum* indicated that *spo0A* was induced at the onset of sporulation and kept at high levels throughout the stationary phase (Alsaker and Papoutsakis, 2005). In this study, the expression of *spo0H* (Cbei_0135) was constant from time points 1-3, but was down-regulated 5.4-fold later during the fermentation (Figure 4.4B). The expression of *spo0A* (Cbei_1712) was induced early during the acidogenesis phase, and elevated by 2.2-fold during the transition phase and the onset of solventogenesis (time points 2 and 3), after which the expression was slightly down-regulated (1.7-fold). Comparatively, the transcription level of *spo0A* was much higher than that of *spo0H* (Figure 4.4B). Putative σ^H-dependent promoter and Spo0A-binding motif (0A box) have been identified upstream of the *spo0A* gene in *C. beijerinckii* 8052 (Wilkinson et al., 1995).

It is well established that Spo0A is the master regulator for sporulation events in both bacilli and clostridia (Alsaker and Papoutsakis, 2005; Dürre and Hollergschwandner, 2004).

Phosphorylated Spo0A has been reported to induce various targets, including the solventogenic operon and multiple sporulation sigma factor genes in *C. acetobutylicum* (Alsaker et al., 2004; Harris et al., 2002; Tomas et al., 2003). 0A boxes were also identified upstream of the operons related to chemotaxis/motility, suggesting the possible negative regulation of Spo0A on the expression of chemotaxis/motility genes (Molle et al., 2003; Paredes et al., 2005; Tomas et al., 2003). A previous study involving insertional inactivation of *spo0A* indicated that Spo0A does also control the formation of solvents, spores and granulose in *C. beijerinckii* (Ravagnani et al., 2000). A recently developed *spo0A* mutant of *C. beijerinckii* 8052 exhibited an asporogenous and non-septating phenotype (Heap et al., 2010)

The *sigF* operon (*spoIIAA-spoIIAB-sigF*) encodes the anti-anti-sigma factor (SpoIIAA), the anti-sigma factor (SpoIIAB) and the sigma factor σ^F. In the "crisscross regulation" system of *B. subtilis*, both σ^H and Spo0A ~ P are required for the expression of the genes in this operon (Losick and Stragier, 1992). In this study, the *sigF* operon genes were induced before the onset of sporulation (time point 2), and kept coordinated and up-regulated throughout the fermentation (Figure 4.4B).

SpoIIE is a phosphatase required for regulating the dephosphorylation of SpoIIAA, which helps to release the σ^F from SpoIIAB by forming an ADP-SpoIIAB-SpoIIAA complex (Arigoni et al., 1996). In *B. subtilis*, expression of *spoIIE* starts about 1 h after the onset of sporulation, and peaks approximately 1 h later (York et al., 1992). In the present study, high expression of *spoIIE* (Cbei_0097) in *C. beijerinckii* 8052 was induced at time point 3 (51.3-fold up-regulated) following the onset of solvent production, and then quickly decreased to a lower level (but is still 7-fold higher than that at time point 1) (Figure 4.4B).

σ^E is the first mother cell-specific sigma factor in *B. subtilis*, and the expression of *sigE* begins about 2 h after the onset of sporulation (Kenney and Moran, 1987). In *C. acetobutylicum*, the expression of *sigE* operon (*spoIIGA-sigE*, CAC1694 and CAC1695) does not exceed 1.3-fold higher than that at the initiation of fermentation (Alsaker and Papoutsakis, 2005). In the case of *C. beijerinckii*, the expression of *sigE* operon (*spoIIGA-sigE*, Cbei_1119 and Cbei_1120) was up-regulated around 20-fold at time point 3, and then decreased by 3- to 6-fold during late stationary phase (Figure 4.4B).

σ^G is a forespore-specific sigma factor and it is active after complete engulfment of the forespore by the mother cell (Cutting et al., 1989). In *B. subtilis*, significant expression of *sigG* begins 150 to 210 min after the onset of sporulation (Sun et al., 1991). In the experiment

presented here, the expression of *sigG* was up-regulated by 28.3-fold at time point 3, and kept at high levels throughout the stationary process. More than 80 genes in *B. subtilis* which are related to spore protection and maturation were identified to be regulated by σ^G (Steil et al., 2005; Wang et al., 2006). Among them, *spoIVB* is the gene encoding a signal protein for activating the later-acting sigma factor σ^K in the mother cell (Cutting et al., 1991); *spoVA* operon (*spoVAA-AE*) is a cluster of sporulation genes downstream of the *sigF* operon associated with dipicolinic acid transport into the developing spores (Kunst et al., 1997; Tovar-Rojo et al., 2002). In many clostridium species, *spoVAC*, *AD*, *AE* were found at the same locus, but *spoVAA* and *AB* are absent (Stragier, 2002). This is also the case for *C. beijerinckii* 8052, and *spoVAC*, *AD*, *AE* (Cbei_815-817) were found to be organized in a single operon based on RNA-Seq data (Wang et al., 2011). The expression of *spoIVB* and the *spoVA* operon was induced after the onset of high expression of *sigG*, and peaked at time points 4 and 5 (around 4- to 16-fold higher than that at time point 1). The expression profiles of *sigG* and the putative σ^G–regulated genes well confirmed their regulation manner in *C. beijerinckii* (Figure 4.4B).

Sporulation regulation factors genes *sigE* and *sigG* in *C. beijerinckii* in the present study were induced and up-regulated > 20-fold right after the onset of sporulation. While in *C. acetobutylicum*, no active expression of *sigE* and *sigG* was detected up to 7.5 h after the onset of sporulation (Alsaker and Papoutsakis, 2005). This dissimilarity coincides with the fact that forespores and endspores develop more rapidly in *C. beijerinckii* 8052 than in *C. acetobutylicum* 824 (Alsaker and Papoutsakis, 2005; Shi and Blaschek, 2008).

In *B. subtilis*, *sigK* gene is expressed from a σ^E-dependent promoter, and regulated by another transcription factor *spoIIID* (Driks and Losick, 1991; Halberg and Kroos, 1994; Kunkel et al., 1989). The *sigK* gene is first expressed as an inactive pro-σ^k factor, and then cleaved to activate by a membrane-localized protease, SpoIVFB, whose activity is induced by SpoIVB, the signaling protein produced in the forespore (Green and Cutting, 2000; Lu et al., 1990). The *sigK* gene in most clostridial species was also found to encode a pro-σ^k factor, and *spoIVFB* and *spoIVB* were identified in most clostridial species (Stragier, 2002). The product of Cbei_0501 (annotated as peptidase M50) in *C. beijerinckii* exhibits 36% amino acid identity (99 out of 273) and 62% positive (169 out of 273) to SpoIVFB in *C. acetobutylicum* by BLASTP. The expression of *spoIIID* (Cbei_0424) and *sigK* (Cbei_1115) was induced unexpectedly early at time point 2, and increased up to 5.6- and 24.1-fold respectively at time

point 4 (compared to time point 2). Afterwards, both genes maintained at high expression levels. Nevertheless, significant expression of Cbei_0501 (putative *spoIVFB*) was only observed at time point 3, and then the expression decreased to a lower level (Figure 4.4B).

In *B. subtilis*, *abrB* encodes a transition state regulator, whose transcription depression by Spo0A leads to a burst of σ^H synthesis at the initiation of sporulation (Weir et al., 1991). Maximal expression of *abrB* in *B. subtilis* was observed 2 h before the onset of sporulation (Jiang et al., 2000). In *C. acetobutylicum*, three *abrB* homologs (CAC0310, 1941 and 3647) were identified (Nolling et al., 2001), among which CAC0310 was suggested as the putative true transitional-state regulator (Scotcher et al., 2005). However, the decrease of expression of CAC0310 was not observed at initiation of sporulation but the highest expression was observed after the onset of sporulation during a time course transcriptional analysis for the wild type *C. acetobutylicum* 824 (Alsaker and Papoutsakis, 2005). In *C. beijerinckii* 8052, six loci were annotated as genes encoding AbrB family transcriptional regulator. Among them, expression of three (Cbei_2219, Cbei_2270 and Cbei_3199) are not detectable most of the time. Expression of Cbei_0088 decreased temporally by 10.8-fold at time point 2, but up-regulated back to even higher levels afterwards. The expression of Cbei_3375 was significantly depressed at time point 3 (10.5-fold from point 2), and kept at low levels during the stationary phase. Although Cbei_4885 was down-regulated by 3.2-fold at time point 3, following expression was kept at relatively high levels and increased to a peak level at time point 6 (Figure 4.4B). In summary, none of the *abrB* genes in *C. beijerinckii* 8052 displayed exact antagonistic expression pattern to that of *spo0H*. Whether there are *abrB* genes that play the similar role as *abrB* in *B. subtilis* warrants further study in respect that *spo0H* in this study had a very different transcriptional profile from that in *B. subtilis*.

As observed in other clostridium species, there was no orthologous phosphorelay system or sensory kinases in *C. beijerinckii* similar to those found in *B. subtilis* (KinA-KinE) for phosphorylating *spo0A* (Stephenson and Hoch, 2002; Stragier, 2002). In *C. acetobutylicum*, it was suggested that the gene CAC3319 might act in a similar fashion to that of the sensory kinase genes in *B. subtilis* (Tomas et al., 2003; Tomas et al., 2004). A recent study revealed evidence for two possible pathways for Spo0A activation in *C. acetobutylicum*, one dependent on CAC0323, and the other dependent on CAC0903 and CAC3319 (Steiner et al., 2011). In this study, with the time course transcriptional data, putative candidate proteins can be identified in *C. beijerinckii* that may play roles similar to sensory kinases for sporulation.

Out of all the genes annotated as encoding sensory histidine kinases, ten were observed to be momentarily upregulated during the onset of sporulation (Figure 4.4B). Further molecular biology work focusing on specific genes needs to be carried out to define the genes that play the key roles in fermentation regulation.

4.4.6 Expression of cell motility genes

The cell motility-related genes encode products responsible for chemotactic responses and flagellar assembly (Wadhams and Armitage, 2004). From the sequencing data, it was confirmed that a similar flagellar/chemotaxis multi-gene cluster exists in *C. beijerinckii* 8052 (Cbei_4312-4302) as that in *C. acetobutylicum* (CAC2225-2215) (Paredes et al., 2004; Wang et al., 2011). Motility-related genes in clostridia and bacilli are usually down-regulated in sporulation cells (Paredes et al., 2005). In this study, most genes in the flagellar/chemotaxis cluster were down-regulated by 2- to 4-fold at the onset of sporulation and were kept at lower expression levels thereafter. However, the last three genes at the end of the cluster (Cbei_4304-4302, encoding cheW, fliM and fliY, respectively) were initially down-regulated by 2- to 3-fold at the onset of sporulation, and then unexpectedly increased during late stationary phase (Figure 4.4C).

There are several other genes in *C. beijerinckii* 8052 annotated as motility-related. Among them, Cbei_0755 (*cheW*) and Cbei_4958 (*cheC*) were active during exponential phase, and were down-regulated by 4.1- and 3.1-fold respectively after cells entered the stationary phase (time point 4). Another *cheW* gene (Cbei_3954), on the other hand, exhibited an unexpected antagonistic pattern, which was strongly expressed throughout the entire stationary phase and peaked at time point 6 (Figure 4.4C). The other chemotaxis genes were not active over the course of the fermentation process. The flagellar motility related genes *flgB* (Cbei_4271) and *flgC* (Cbei_4270) were strongly expressed during exponential phase (time points 1 and 2), and down-regulated by 5.9- to 7.1-fold during the stationary phase. It has been reported that the expression of the flagellar/motility operon is negatively regulated by Spo0A in *B subtilis* (Molle et al., 2003).

4.4.7 Expression of quorum sensing gene

The gene *luxS* (Cbei_1237) encodes S-ribosylhomocysteinase, which produces autoinducers to function in the Luxs/AI-2 quorum sensing mechanism (Ohtani et al., 2002).

Maximal autoinducer activity was observed during mid-exponential phase in *C. perfringens* with autoluminesence assays (Ohtani et al., 2002), and the expression of *luxS* in *C. acetobutylicum* was found to be roughly proportional to cell density in a time-course transcriptional analysis of *C. acetobutylicum* 824 (Alsaker and Papoutsakis, 2005). Sequence alignment of the protein encoded by the putative *luxS* gene in *C. beijerinckii* (Cbei_1237, 480bp) and that in *C. acetobutylicum* (CAC2942, 477 bp) with BLASTP indicated that, out of 158 amino acids, there exists 74% identity (117 amino acids) and 84% positive (133 amino acids) matches. The expression of Cbei_1237 started at a high level during the acidogenesis phase, and then increased by 2.2-fold and peaked during the transition to solventogenesis (time point 2), after which the expression decreased gradually to lower levels (Figure 4.4C).

4.4.8 Expression of sugar transport genes

The phosphoenolpyruvate-dependent phosphotransferase system (PTS) is the predominant sugar uptake pathway in *C. beijerinckii* (Mitchell and Tangney, 2005). The first two PTS domains (general PTS proteins), referred as EI (enzyme I) and HPr (heat stable, histidine-phosphorylatable protein), phosphorylate the sugar in the cytoplasm, which has been transported across the cell membrane from the extracellular side by substrate-specific enzyme II (EII) domains. Phosphorylated sugar is then able to enter the glycolytic pathway (Mitchell and Tangney, 2005). In *C. beijerinckii*, EI and HPr are encoded by Cbei_0196 (*ptsI*) and Cbei_1219 (*ptsH*) respectively. The expression of both genes was high throughout the fermentation process with a 2- to 3-fold decrease after onset of solventogenesis (Figure 4.4C). This agrees well with the previous finding that ATP-dependent glucose phosphorylation was predominant and phosphoenolpyruvate-dependent glucose phosphorylation was repressed during solventogenic stage for both *C. beijerinckii* 8052 and its mutant *C. beijerinckii* BA101 strains, although the latter exhibited a nearly 2-fold-greater ATP-dependent phosphorylation rate during solventogenic stage (Lee and Blaschek, 2001).

Previous competition studies indicated that glucose and mannose were assimilated by the same PTS in *C. beijerinckii* (Lee and Blaschek, 2001; Mitchell et al., 1991). The mannose family PTS transporters are commonly found in gammaproteobacteria and firmicutes, including *C. acetobutylicum* and *C. beijerinckii*, and exhibit broad substrate specificity for glucose, mannose, sorbose, fructose, and a variety of other sugars (Barabote and Saier, 2005; Shi and Blaschek, 2008). In the gene cluster encoding for the mannose family PTS

transporters, *manIIAB* encodes a membrane fusion protein of the IIA and IIB subunits involved in sugar phosphorylation, *manIIC* encodes a substrate-specific permease, and *manIID* encodes a mannose family-specific auxiliary protein (essential but of unknown function) (Barabote and Saier, 2005). The RNA-Seq data revealed that these three genes were organized in the same operon (Cbei_0711-0713) in *C. beijerinckii* 8052 (Wang et al., 2011), and the expression of *manIIAB* (Cbei_0711), *manIIC* (Cbei_0712) and *manIID* (Cbei_0713) was highly coordinated and unexpectedly up-regulated by 11.6, 20.4 and 31.6-fold respectively after onset of solventogenesis (Figure 4.4C).

There are 13 PTS domains identified in *C. acetobutylicum*, out of which six belongs to the glucose-glucoside (Glc) family (Mitchell and Tangney, 2005). Homologs were identified in *C. beijerinckii* for all the gene clusters encoding the Glc family PTS transporters in *C. acetobutylicum*. Among them, only Cbei_0751, Cbei_3273, Cbei_4532, Cbei_4533 and Cbei_4705 were actively expressed during the fermentation process, and Cbei_4533 (encoding N-acetylglucosamine-specific IIA subunit) and Cbei_4532 (encoding N-acetylglucosamine-specific IIBC subunit) were especially strongly expressed throughout the batch fermentation process (Figure 4.4C). By comparing the expression dynamics of gene clusters encoding for the Glc family PTS transporters and those for the mannose family PTS transporters, it could be inferred that the Glc family PTS transporters (especially those encoded by Cbei_4532 and _4533) play important roles during exponential and early stationary phases, while the mannose family PTS transporters may be more closely related to stationary and sporulation events.

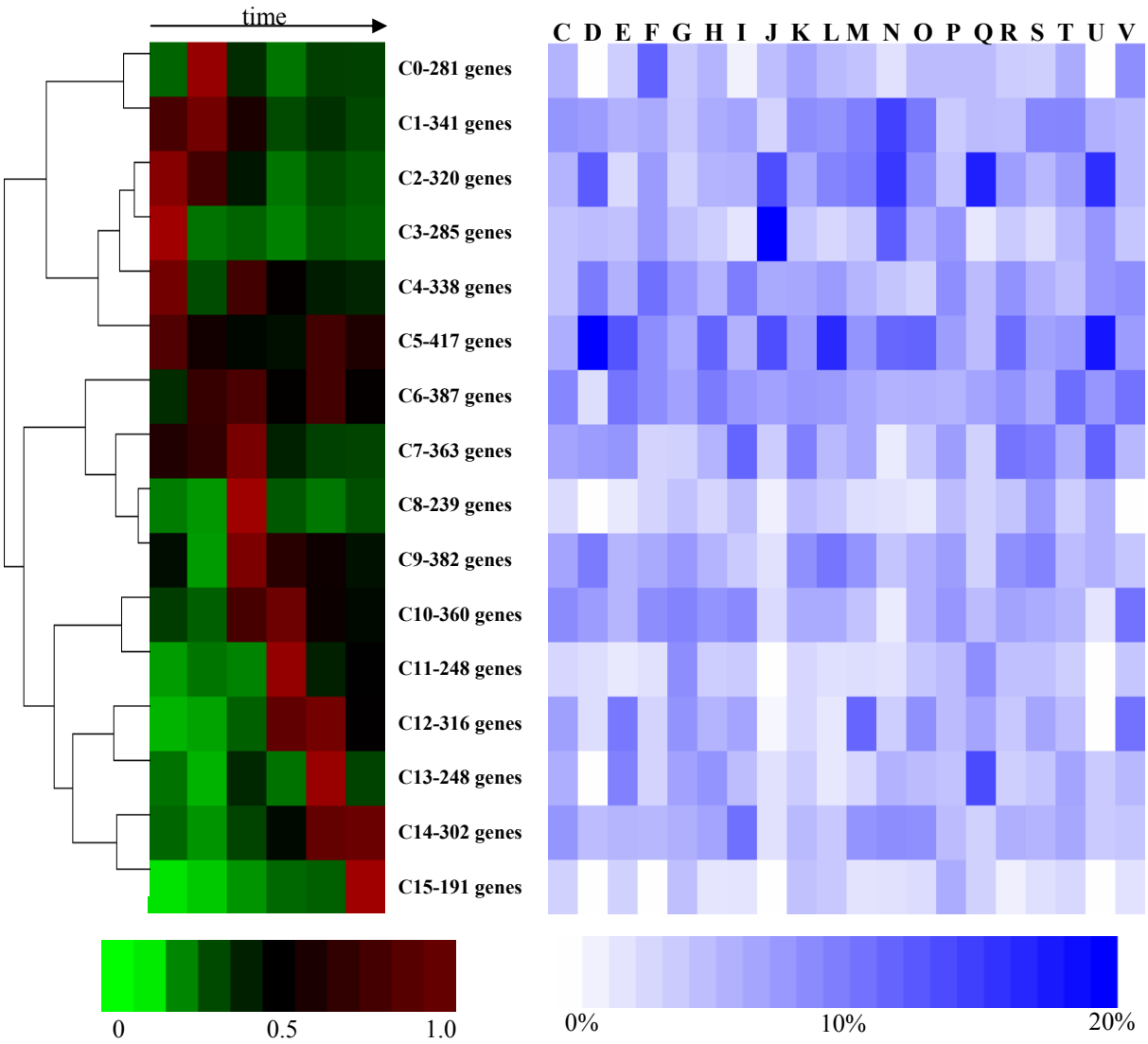


Figure 4.5 SOM clusters of gene expression profiles during the batch fermentation. SOM clusters are grouped by average-linkage hierarchical clustering based on expression pattern similarities. Distribution of COG categories across the clusters is represented colorimetrically on the right. COG functional groups are from the *C. beijerinckii* 8052 genome annotation defined by the COG database (Tatusov et al., 2000).

4.4.9 Self-organizing maps (SOM) analysis of functional groups

SOM analysis was employed to investigate the global expression patterns, especially for genes belonging to specific COG categories across the genome (Figure 4.5) (Alsaker and Papoutsakis, 2005). The expression of genes in clusters C2 and C3 was up-regulated during early exponential-growth phase and decreased to lower levels afterwards. While these two clusters jointly contain 605 genes (12.1%, out of totally 5018 in the analysis), they are

enriched with 49.2% of the genes related to translation (COG J), and 28.1% of genes related to cell motility (COG N). Cell motility genes were also over-represented by 14.8% in cluster C1, where the expression levels of genes were lower in the stationary phase than that in the exponential phase.

Carbohydrate transport and metabolism (COG G) Nine of the glycolysis genes discussed above (Figure 4.4A) falls into clusters C1 and C2, indicating more active glycolytic activities during the exponential-growth phase. On the other hand, solventogenic clostridia initiate granulose production at the early stationary phase; Granulose is a glycogen-like polymer used by bacteria for energy storage under severe environments (Reysenbach et al., 1986). Operon prediction for *C. acetobutylicum* indicated that *glgC* (CAC2237, encoding glucose-1-phosphate adenylyltransferase), *glgD* (CAC2238, encoding ADP-glucose pyrophosphorylase), *glgA* (CAC2239, encoding glycogen synthase) and CAC2240 (encoding cyclomaltodextrin glucanotransferase domain-containing protein) express as a single transcript (Paredes et al., 2004). In addition, *glgP* encodes a granulose phosphorylase (CAC1664, functions to depolymerize granulose), and *csrA* (CAC2209) encodes a putative carbon storage regulator, which was reported to repress the expression of both *glgA* and *glgP* in *Escherichia coli* (Alsaker and Papoutsakis, 2005; Yang et al., 1996). In *C. beijerinckii* 8052, it was observed that *glgC* (Cbei_4905) and *glgD* (Cbei_4904) were organized in an operon, while the gene encoding 1,4-alpha-glucan branching enzyme (Cbei_4909), *glgA* (Cbei_4908), *glgP* (Cbei_4907) and the gene encoding the catalytic region of alpha amylase (Cbei_4906) were organized in another adjacent operon (Wang et al., 2011). Cbei_4295 was annotated as the *csrA* gene. The five putative granulose formation genes (Cbei_4905-4909) in *C. beijerinckii* 8052 were grouped into cluster C6 based on their expression profiles (Figure 4.5). Expression of most granulose formation genes were up-regulated by 2- to 4-fold at transition phase from acidogenesis to solventogenesis (time point 2) and maintained at high expression levels over the stationary phase (point 3-5) (Figure 4.4C). Notwithstanding, the expression of *csrA* (Cbei_4295) was only detectable at samples 1, 3 and 6, which was much lower than those of the other granulose formation genes (Figure 4.4C). It is not yet clear whether the *csrA* gene in *C. beijerinckii* functions similarly to that in *E. coli*.

Translation genes (COG J) There are 181 translation-related genes included in the SOM analysis. The majority were grouped in the clusters where the gene expression is active during exponential phase (78% enriched in C1-C5). There are 58 genes in *C. beijerinckii*

genome encoding ribosomal proteins. Out of the 56 genes encoding ribosomal proteins included in the SOM analysis, 46 were clustered in C3, indicating that the protein synthesis machinery in the cell is mostly active during the early exponential phase (Figure 4.5). Similar results were reported for *C. acetobutylicum*, where most of the ribosomal protein genes were down-regulated at least 3-fold in stationary phase (Alsaker and Papoutsakis, 2005).

Signal transduction mechanisms genes (COG T) There are 371 genes (around 7.4% of all genes) annotated as signal transduction related genes in *C. beijerinckii*. SOM analysis did not reveal any specific enriched clusters for these genes. However, they are slightly over-represented in cluster C0 and C1 (jointly 16% of category T), where the expression levels are higher during the exponential phase (Figure 4.5). Among them, there are important cell chemotaxis and motility associated sensory transduction genes (Cbei_0279, _0665, _0755, _0804, _4012, _4161, _4307, _4309, _4310, _4466, _4604 and _4832). The expression of genes in clusters C8-C10 was momentarily up-regulated at onset of stationary phase and decreased quickly afterwards (Figure 4.5). The signal transduction genes in clusters C8-C10 may represent important signal transducers for solventogenesis and sporulation events. Among them, there are genes encoding response regulator receiver proteins, integral membrane sensor signal transduction histidine kinase and two-component transcriptional regulators (Table B4.4 in Appendix B) (Stock et al., 2000; West and Stock, 2001).

4.4.10 Correlations of RNA-Seq and qRT-PCR

The expression measurement with RNA-Seq approach was validated using qRT-PCR for selected genes. For all the selected genes, the dynamic gene expression profiles over time measured with RNA-Seq have very similar patterns as those measured with qRT-PCR (Figure B4.1 in Appendix B). It has been previously showed that good correlation exists between the overall gene expression quantification with RNA-Seq and that obtained with microarray or qRT-PCR (Wang et al., 2011). These results demonstrate that RNA-Seq is an effective approach for quantification of gene expression.

Previously, Shi and Blaschek reported the transcriptional analysis of *C. beijerinckii* using a ca. 500-gene set DNA microarray where only the expression of genes orthologous to those previously found to be important to the physiology of *C. acetobutylicum* 824 was examined (Shi and Blaschek, 2008). In this study, employing RNA-Seq technology, for the first time

the genome-wide dynamic transcriptional profiling of *C. beijerinckii* correlated with the physiological activities over the course of fermentation process has been revealed with a much larger dynamic range. In addition, single-nucleotide resolution RNA-Seq data also revealed the transcriptome structural organization of *C. beijerinckii* 8052 in depth (Wang et al., 2011). The transcriptome structural information together with the genome-wide dynamic transcriptional profiling can provide essential references for further improvement of *C. beijerinckii* strains by employing system biology-based strategies and metabolic engineering approaches. Furthermore, the genome-wide RNA-Seq analysis also disclosed the different expression dynamics of many allelic genes in the genome of *C. beijerinckii* 8052, which provided important insight for future studies on the properties of potential isoenzymes and their associated functions during specific fermentation stages.

4.5 Conclusions

By employing single-base resolution RNA-Seq technology, the genome-wide transcriptional dynamics of *C. beijerinckii* 8052 throughout the course of a batch fermentation was revealed in great depth. RNA-Seq technology can not only quantify the differential transcription of specific genes over time, but is also able to compare transcription levels of different genes in the same sample. Overall dynamic transcription profiles demonstrated obvious temporal variation trend throughout the fermentation, corresponding to the physiological state change of the cells. Glycolysis genes demonstrated higher expression during acidogenesis phase. The expression of acid formation genes declined at the onset of solvent formation, in accordance with the metabolic pathway shift from acidogenesis to solventogenesis. The *sol* operon, including *adc* as a part, was up-regulated at the onset of solventogenesis and maintained at high levels through the stationary phase. Most sporulation genes in *C. beijerinckii* 8052 demonstrated similar temporal expression patterns to those observed in *B. subtilis* and *C. acetobutylicum*, while sporulation sigma factor genes *sigE* and *sigG* exhibited accelerated and stronger expression in *C. beijerinckii* 8052, consistent with the more rapid forespore and endspore development in this strain. The PTS transport system encoding genes were highly expressed during acidogenesis and declined 2- to 3-fold after entering solventogenesis phase, agreeing well with the report that ATP-dependent glucose phosphorylation was predominant and phosphoenolpyruvate-dependent PTS was repressed during solventogenic stage for *C. beijerinckii*. Finally, SOM and clustering analysis revealed

that genes of specific COG functional classes demonstrated global gene expression patterns corresponding to the cell physiological change and fermentation pathway switch.

4.6 References

Alsaker, K. V., and E. T. Papoutsakis. 2005. Transcriptional program of early sporulation and stationary-phase events in *Clostridium acetobutylicum*. *Journal of Bacteriology* 187(20):7103-7118.

Alsaker, K. V., C. Paredes, and E. T. Papoutsakis. 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*. *Biotechnology and Bioengineering* 105(6):1131-1147.

Alsaker, K. V., C. J. Paredes, and E. T. Papoutsakis. 2005. Design, optimization and validation of genomic DNA microarrays for examining the *Clostridium acetobutylicum* transcriptome. *Biotechnology and Bioprocess Engineering* 10(5):432-443.

Alsaker, K. V., T. R. Spitzer, and E. T. Papoutsakis. 2004. Transcriptional analysis of *spo0A* overexpression in *Clostridium acetobutylicum* and its effect on the cell's response to butanol stress. *Journal of Bacteriology* 186(7):1959-1971.

Anders, S., and W. Huber. 2010. Differential expression analysis for sequence count data. *Genome Biology* 11(10):R106.

Annous, B. A., and H. P. Blaschek. 1991. Isolation and characterization of *Clostridium acetobutylicum* mutants with enhanced amylolytic activity. *Applied and Environmental Microbiology* 57(9):2544-2548.

Arigoni, F., L. Duncan, S. Alper, R. Losick, and P. Stragier. 1996. SpoIIE governs the phosphorylation state of a protein regulating transcription factor σ^F during sporulation in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences of the United States of America* 93(8):3238-3242.

Barabote, R. D., and M. H. Saier. 2005. Comparative genomic analyses of the bacterial phosphotransferase system. *Microbiology and Molecular Biology Reviews* 69(4):608-634.

Benedito, V. A., I. Torres-Jerez, J. D. Murray, A. Andriankaja, S. Allen, K. Kakar, M. Wandrey, J. Verdier, H. Zuber, T. Ott, S. Moreau, A. Niebel, T. Frickey, G. Weiller, J. He, X. Dai, P. X. Zhao, Y. Tang, and M. K. Udvardi. 2008. A gene expression atlas of the model legume *Medicago truncatula*. *Plant Journal* 55(3):504-513.

Chen, J.-S. 1995. Alcohol dehydrogenase: multiplicity and relatedness in the solvent-producing clostridia. *Fems Microbiology Reviews* 17(3):263-273.

Cornillot, E., R. V. Nair, E. T. Papoutsakis, and P. Soucaille. 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. *Journal of Bacteriology* 179(17):5442-5447.

Cutting, S., A. Driks, R. Schmidt, B. Kunkel, and R. Losick. 1991. Forespore-specific transcription of a gene in the signal transduction pathway that governs Pro-σ^K processing in *Bacillus subtilis*. *Genes & Development* 5(3):456-466.

Cutting, S., S. Panzer, and R. Losick. 1989. Regulatory studies on the promoter for a gene governing synthesis and assembly of the spore coat in *Bacillus subtilis*. *Journal of Molecular Biology* 207(2):393-404.

Driks, A., and R. Losick. 1991. Compartmentalized expression of a gene under the control of sporulation transcription factor σ^E in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences of the United States of America* 88(22):9934-9938.

Dürre, P. 2007. Biobutanol: An attractive biofuel. *Biotechnology Journal* 2(12):1525-1534.

Dürre, P., and C. Hollergschwandner. 2004. Initiation of endospore formation in *Clostridium acetobutylicum*. *Anaerobe* 10(2):69-74.

Eisen, M. B., P. T. Spellman, P. O. Brown, and D. Botstein. 1998. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America* 95(25):14863-14868.

Ezeji, T., and H. P. Blaschek. 2008. Fermentation of dried distillers' grains and solubles (DDGS) hydrolysates to solvents and value-added products by solventogenic clostridia. *Bioresource Technology* 99(12):5232-5242.

Feustel, L., S. Nakotte, and P. Dürre. 2004. Characterization and development of two reporter gene systems for *Clostridium acetobutylicum*. *Applied and Environmental Microbiology* 70(2):798-803.

Gentleman, R., V. Carey, D. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Yang, and J. Zhang. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* 5(10):R80.

Gerischer, U., and P. Dürre. 1990. Cloning, sequencing, and molecular analysis of the acetoacetate decarboxylase gene region from *Clostridium acetobutylicum*. *Journal of Bacteriology* 172(12):6907-6918.

Gerischer, U., and P. Dürre. 1992. mRNA analysis of the *adc* gene region of *Clostridium acetobutylicum* during the shift to solventogenesis. *Journal of Bacteriology* 174(2):426-433.

Green, D. H., and S. M. Cutting. 2000. Membrane topology of the *Bacillus subtilis* Pro-σ^K processing complex. *Journal of Bacteriology* 182(2):278-285.

Grimmler, C., H. Janssen, D. Krausse, R. J. Fischer, H. Bahl, P. Dürre, W. Liebl, and A. Ehrenreich. 2011. Genome-wide gene expression analysis of the switch between acidogenesis and solventogenesis in continuous cultures of *Clostridium acetobutylicum*. *Journal of Molecular Microbiology and Biotechnology* 20(1):1-15.

Halberg, R., and L. Kroos. 1994. Sporulation regulatory protein SpoIIID from *Bacillus subtilis* activates and represses transcription by both mother-cell-specific forms of RNA polymerase. *Journal of Molecular Biology* 243(3):425-436.

Harris, L. M., N. E. Welker, and E. T. Papoutsakis. 2002. Northern, morphological, and fermentation analysis of *spo0A* inactivation and overexpression in *Clostridium acetobutylicum* ATCC 824. *Journal of Bacteriology* 184(13):3586-3597.

Heap, J. T., S. A. Kuehne, M. Ehsaan, S. T. Cartman, C. M. Cooksley, J. C. Scott, and N. P. Minton. 2010. The ClosTron: Mutagenesis in *Clostridium* refined and streamlined. *Journal of Microbiological Methods* 80(1):49-55.

Hill, M. O., and H. G. Gauch. 1980. Detrended correspondence analysis: an improved ordination technique. *Vegetatio* 42(1-3):47-58.

Janssen, H., C. Doring, A. Ehrenreich, B. Voigt, M. Hecker, H. Bahl, and R. J. Fischer. 2010. A proteomic and transcriptional view of acidogenic and solventogenic steady-state cells of *Clostridium acetobutylicum* in a chemostat culture. *Applied Microbiology and Biotechnology* 87(6):2209-2226.

Jiang, M., W. L. Shao, M. Perego, and J. A. Hoch. 2000. Multiple histidine kinases regulate entry into stationary phase and sporulation in *Bacillus subtilis*. *Molecular Microbiology* 38(3):535-542.

Jones, S. W., C. J. Paredes, B. Tracy, N. Cheng, R. Sillers, R. S. Senger, and E. T. Papoutsakis. 2008. The transcriptional program underlying the physiology of clostridial sporulation. *Genome Biology* 9(7):R114.

Kenney, T. J., and C. P. Moran, Jr. 1987. Organization and regulation of an operon that encodes a sporulation-essential sigma factor in *Bacillus subtilis*. *Journal of Bacteriology* 169(7):3329-3339.

Kunkel, B., L. Kroos, H. Poth, P. Youngman, and R. Losick. 1989. Temporal and spatial control of the mother-cell regulatory gene *spoIIID* of *Bacillus subtilis*. *Genes & Development* 3(11):1735-1744.

Kunst, F., N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo, M. G. Bertero, P. Bessieres, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S. C. Brignell, S. Bron, S. Brouillet, C. V. Bruschi, B. Caldwell, V. Capuano, N. M. Carter, S. K. Choi, J. J. Codani, I. F. Connerton, N. J. Cummings, R. A. Daniel, F. Denizot, K. M. Devine, A. Dusterhoft, S. D. Ehrlich, P. T. Emmerson, K. D. Entian, J. Errington, C. Fabret, E. Ferrari, D. Foulger, C. Fritz, M. Fujita, Y. Fujita, S. Fuma, A. Galizzi, N. Galleron, S. Y. Ghim, P. Glaser, A. Goffeau, E. J. Golightly, G. Grandi, G. Guiseppi, B. J. Guy, K. Haga, J. Haiech, C. R. Harwood, A. Henaut, H. Hilbert, S. Holsappel, S. Hosono, M. F. Hullo, M. Itaya, L. Jones, B. Joris, D. Karamata, Y. Kasahara, M. KlaerrBlanchard, C. Klein, Y. Kobayashi, P. Koetter, G. Koningstein, S. Krogh, M. Kumano, K. Kurita, A. Lapidus, S. Lardinois, J. Lauber, V. Lazarevic, S. M. Lee, A. Levine, H. Liu, S. Masuda, C. Mauel, C. Medigue, N. Medina, R. P. Mellado, M. Mizuno, D. Moestl, S. Nakai, M. Noback, D. Noone, M. Oreilly, K. Ogawa, A. Ogiwara, B. Oudega, S. H. Park, V. Parro, T. M. Pohl, D. Portetelle, S. Porwollik, A. M. Prescott, E. Presecan, P. Pujic, B. Purnelle, G. Rapoport, M. Rey, S. Reynolds, M. Rieger, C. Rivolta, E. Rocha, B. Roche, M. Rose, Y. Sadaie, T. Sato, E. Scanlan, S. Schleich, R. Schroeter, F. Scoffone, J. Sekiguchi, A. Sekowska, S. J. Seror, P. Serror, B. S. Shin, B. Soldo, A. Sorokin, E. Tacconi, T. Takagi, H. Takahashi, K. Takemaru, M. Takeuchi, A. Tamakoshi, T. Tanaka, P. Terpstra, A. Tognoni, V. Tosato, S. Uchiyama, M. Vandenbol, F. Vannier, A. Vassarotti, A. Viari, R. Wambutt, E. Wedler, H. Wedler, T. Weitzenegger, P. Winters, A. Wipat, H. Yamamoto, K. Yamane, K. Yasumoto, K. Yata, K. Yoshida, H. F. Yoshikawa, E. Zumstein, H. Yoshikawa, and A. Danchin. 1997. The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* 390(6657):249-256.

Lee, J., and H. P. Blaschek. 2001. Glucose uptake in *Clostridium beijerinckii* NCIMB 8052 and the solvent-hyperproducing mutant BA101. *Applied and Environmental Microbiology* 67(11):5025-5031.

Legendre, M., S. Audic, O. Poirot, P. Hingamp, V. Seltzer, D. Byrne, A. Lartigue, M. Lescot, A. Bernadac, J. Poulain, C. Abergel, and J.-M. Claverie. 2010. mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus. *Genome Research* 20(5):664-674.

Li, H., J. Ruan, and R. Durbin. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Research* 18(11):1851-1858.

Losick, R., and P. Stragier. 1992. Crisscross regulation of cell-type-specific gene expression during development in *B. subtilis*. *Nature* 355(6361):601-604.

Lu, S., R. Halberg, and L. Kroos. 1990. Processing of the mother-cell σ factor, σ^K, may depend on events occurring in the forespore during *Bacillus subtilis* development. *Proceedings of the National Academy of Sciences* 87(24):9722-9726.

Mitchell, W. J., J. E. Shaw, and L. Andrews. 1991. Properties of the glucose phosphotransferase system of *Clostridium acetobutylicum* NCIB 8052. *Applied and Environmental Microbiology* 57(9):2534-2539.

Mitchell, W. J., and M. Tangney. 2005. Carbohydrate uptake by the phosphotransferase system and other mechanisms. In *Handbook on clostridia*, 155-175. P. Dürre, ed. London: CRC press.

Molle, V., M. Fujita, S. T. Jensen, P. Eichenberger, J. E. Gonzalez-Pastor, J. S. Liu, and R. Losick. 2003. The Spo0A regulon of *Bacillus subtilis*. *Molecular Microbiology* 50(5):1683-1701.

Mortazavi, A., B. A. Williams, K. McCue, L. Schaeffer, and B. Wold. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* 5(7):621-628.

Nagalakshmi, U., Z. Wang, K. Waern, C. Shou, D. Raha, M. Gerstein, and M. Snyder. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320:1344-1349.

Nolling, J., G. Breton, M. V. Omelchenko, K. S. Makarova, Q. D. Zeng, R. Gibson, H. M. Lee, J. Dubois, D. Y. Qiu, J. Hitti, Y. I. Wolf, R. L. Tatusov, F. Sabathe, L. Doucette-Stamm, P. Soucaille, M. J. Daly, G. N. Bennett, E. V. Koonin, and D. R. Smith. 2001. Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *Journal of Bacteriology* 183(16):4823-4838.

Ohtani, K., H. Hayashi, and T. Shimizu. 2002. The *luxS* gene is involved in cell-cell signalling for toxin production in *Clostridium perfringens*. *Molecular Microbiology* 44(1):171-179.

Paredes, C. J., K. V. Alsaker, and E. T. Papoutsakis. 2005. A comparative genomic view of clostridial sporulation and physiology. *Nature Reviews Microbiology* 3(12):969-978.

Paredes, C. J., I. Rigoutsos, and E. T. Papoutsakis. 2004. Transcriptional organization of the *Clostridium acetobutylicum* genome. *Nucleic Acids Research* 32(6):1973-1981.

Passalacqua, K. D., A. Varadarajan, B. D. Ondov, D. T. Okou, M. E. Zwick, and N. H. Bergman. 2009. Structure and complexity of a bacterial transcriptome. *Journal of Bacteriology* 191(10):3203-3211.

Perkins, T. T., R. A. Kingsley, M. C. Fookes, P. P. Gardner, K. D. James, L. Yu, S. A. Assefa, M. He, N. J. Croucher, D. J. Pickard, D. J. Maskell, J. Parkhill, J. Choudhary, N. R. Thomson, and G. Dougan. 2009. A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus *Salmonella typhi*. *Plos Genetics* 5(7):e1000569.

Petersen, D. J., J. W. Cary, J. Vanderleyden, and G. N. Bennett. 1993. Sequence and arrangement of genes encoding enzymes of the acetone-production pathway of *Clostridium acetobutylicum* ATCC 824. *Gene* 123(1):93-97.

Predich, M., G. Nair, and I. Smith. 1992. *Bacillus subtilis* early sporulation genes *kinA*, *spo0F*, and *spo0A* are transcribed by the RNA polymerase containing σ^H. *Journal of Bacteriology* 174(9):2771-2778.

Qureshi, N., A. Lolas, and H. P. Biaschek. 2001. Soy molasses as fermentation substrate for production of butanol using *Clostridium beijerinckii* BA101. *Journal of Industrial Microbiology & Biotechnology* 26(5):290-295.

Ravagnani, A., K. C. B. Jennert, E. Steiner, R. Grunberg, J. R. Jefferies, S. R. Wilkinson, D. I. Young, E. C. Tidswell, D. P. Brown, P. Youngman, J. G. Morris, and M. Young. 2000. Spo0A directly controls the switch from acid to solvent production in solvent-forming clostridia. *Molecular Microbiology* 37(5):1172-1185.

Reich, M., T. Liefeld, J. Gould, J. Lerner, P. Tamayo, and J. P. Mesirov. 2006. GenePattern 2.0. *Nature Genetics* 38(5):500-501.

Reysenbach, A. L., N. Ravenscroft, S. Long, D. T. Jones, and D. R. Woods. 1986. Characterization, biosynthesis, and regulation of granulose in *Clostridium acetobutylicum*. *Applied and Environmental Microbiology* 52(1):185-190.

Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139-140.

Sauer, U., and P. Dürre. 1995. Differential induction of genes related to solvent formation during the shift from acidogenesis to solventogenesis in continuous culture of *Clostridium acetobutylicum*. *Fems Microbiology Letters* 125(1):115-120.

Scotcher, M. C., F. B. Rudolph, and G. N. Bennett. 2005. Expression of *abrB310* and *sinR*, and effects of decreased *abrB310* expression on the transition from acidogenesis to solventogenesis, in *Clostridium acetobutylicum* ATCC 824. *Applied and Environmental Microbiology* 71(4):1987-1995.

Severin, A., J. Woody, Y.-T. Bolon, B. Joseph, B. Diers, A. Farmer, G. Muehlbauer, R. Nelson, D. Grant, J. Specht, M. Graham, S. Cannon, G. May, C. Vance, and R. Shoemaker. 2010. RNA-Seq Atlas of *Glycine max*: A guide to the soybean transcriptome. *BMC Plant Biology* 10(1):160.

Shi, Z., and H. P. Blaschek. 2008. Transcriptional analysis of *Clostridium beijerinckii* NCIMB 8052 and the hyper-butanol-producing mutant BA101 during the shift from acidogenesis to solventogenesis. *Applied and Environmental Microbiology* 74(24):7709-7714.

Steil, L., M. Serrano, A. O. Henriques, and U. Volker. 2005. Genome-wide analysis of temporally regulated and compartment-specific gene expression in sporulating cells of *Bacillus subtilis*. *Microbiology-Sgm* 151:399-420.

Steiner, E., A. E. Dago, D. I. Young, J. T. Heap, N. P. Minton, J. A. Hoch, and M. Young. 2011. Multiple orphan histidine kinases interact directly with Spo0A to control the initiation of endospore formation in *Clostridium acetobutylicum*. *Molecular Microbiology* 80(3):641-654.

Stephenson, K., and J. A. Hoch. 2002. Evolution of signalling in the sporulation phosphorelay. *Molecular Microbiology* 46(2):297-304.

Stock, A. M., V. L. Robinson, and P. N. Goudreau. 2000. Two-component signal transduction. *Annual Review of Biochemistry* 69(1):183-215.

Stragier, P. 2002. A gene odyssey: exploring the genomes of endospore-forming bacteria. In *Bacillus subtilis and its closest relatives: from genes to cells*, 519-525. A. Sonenshein, J. Hoch, and R. Losick, eds. Washington, D.C.: ASM Press.

Sun, D. X., R. M. Cabrera-Martinez, and P. Setlow. 1991. Control of transcription of the *Bacillus subtilis spoIIIG* gene, which codes for the forespore-specific transcription factor σ^G. *Journal of Bacteriology* 173(9):2977-2984.

Tatusov, R. L., M. Y. Galperin, D. A. Natale, and E. V. Koonin. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research* 28(1):33-36.

Tomas, C. A., K. V. Alsaker, H. P. J. Bonarius, W. T. Hendriksen, H. Yang, J. A. Beamish, C. J. Paredes, and E. T. Papoutsakis. 2003. DNA array-based transcriptional analysis of asporogenous, nonsolventogenic *Clostridium acetobutylicum* strains SKO1 and M5. *Journal of Bacteriology* 185(15):4539-4547.

Tomas, C. A., J. Beamish, and E. T. Papoutsakis. 2004. Transcriptional analysis of butanol stress and tolerance in *Clostridium acetobutylicum*. *Journal of Bacteriology* 186(7):2006-2018.

Tovar-Rojo, F., M. Chander, B. Setlow, and P. Setlow. 2002. The products of the *spoVA* operon are involved in dipicolinic acid uptake into developing spores of *Bacillus subtilis*. *Journal of Bacteriology* 184(2):584-587.

van Vliet, A. H. M. 2010. Next generation sequencing of microbial transcriptomes: challenges and opportunities. *Fems Microbiology Letters* 302(1):1-7.

Wadhams, G. H., and J. P. Armitage. 2004. Making sense of it all: bacterial chemotaxis. *Nature Reviews Molecular Cell Biology* 5(12):1024-1037.

Walter, K. A., G. N. Bennett, and E. T. Papoutsakis. 1992. Molecular characterization of two *Clostridium acetobutylicum* ATCC 824 butanol dehydrogenase isozyme genes. *Journal of Bacteriology* 174(22):7149-7158.

Wang, S. T., B. Setlow, E. M. Conlon, J. L. Lyon, D. Imamura, T. Sato, P. Setlow, R. Losick, and P. Eichenberger. 2006. The forespore line of gene expression in *Bacillus subtilis*. *Journal of Molecular Biology* 358(1):16-37.

Wang, Y., X. Li, Y. Mao, and H. Blaschek. 2011. Single-nucleotide resolution analysis of the transcriptome structure of *Clostridium beijerinckii* NCIMB 8052 using RNA-Seq. *Bmc Genomics* 12:479.

Wang, Z., M. Gerstein, and M. Snyder. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10(1):57-63.

Weir, J., M. Predich, E. Dubnau, G. Nair, and I. Smith. 1991. Regulation of *spo0H*, a gene coding for the *Bacillus subtilis* σ^H factor. *Journal of Bacteriology* 173(2):521-529.

Welch, R. W., F. B. Rudolph, and E. T. Papoutsakis. 1989. Purification and characterization of the NADH-dependent butanol dehydrogenase from *Clostridium acetobutylicum* (ATCC 824). *Archives of Biochemistry and Biophysics* 273(2):309-318.

West, A. H., and A. M. Stock. 2001. Histidine kinases and response regulator proteins in two-component signaling systems. *Trends in biochemical sciences* 26(6):369-376.

Wilkinson, S. R., D. I. Young, J. Gareth Morris, and M. Young. 1995. Molecular genetics and the initiation of solventogenesis in *Clostridium beijerinckii* (formerly *Clostridium acetobutylicum*) NCIMB 8052. *Fems Microbiology Reviews* 17(3):275-285.

Winzer, K., K. Lorenz, B. Zickner, and P. Dürre. 2000. Differential regulation of two thiolase genes from *Clostridium acetobutylicum* DSM 792. *Journal of Molecular Microbiology and Biotechnology* 2(4):531-541.

Yang, H. H., M. Y. Liu, and T. Romeo. 1996. Coordinate genetic regulation of glycogen catabolism and biosynthesis in *Escherichia coli* via the CsrA gene product. *Journal of Bacteriology* 178(4):1012-1017.

York, K., T. J. Kenney, S. Satola, C. P. Moran, Jr, H. Poth, and P. Youngman. 1992. Spo0A controls the σ^A-dependent activation of *Bacillus subtilis* sporulation-specific transcription unit *spoIIE*. *Journal of Bacteriology* 174(8):2648-2658.

CHAPTER 5. EFFECTS OF SUPPLEMENTARY BUTYRATE ON SOLVENT PRODUCTION WITH *CLOSTRIDIUM BEIJERINCKII* 8052 AND GENOME-WIDE TRANSCRIPTIONAL ANALYSIS WITH RNA-SEQ

5.1 Abstract

Butyrate has long been hypothesized to be associated with solventogenesis switch in the ABE fermentation with solventogenic clostridia. In this study, sodium butyrate was added to the chemically defined MP2 medium for ABE fermentation with *Clostridium beijerinckii* 8052 in order to investigate the effects of butyrate on solvent production and metabolic pathway shift from acidogenesis to solventogenesis. The fermentation results indicated that with butyrate addition, the solvent production was triggered early in the mid-exponential phase and finished rapidly in < 50 h from the beginning of innoculation. While in the control fermentation, solventogenesis started from late exponential phase and took > 90 h to reach the maximum butanol level. Butyrate supplementation led to a 31% improvement in the final butanol titer, 58% improvement in butanol yield and 133% improvement in butanol productivity, comparing to the control without butyrate addition. The butanol/acetone ratio increased to 2.4 (molar ratio) versus 1.8 in the control, indicating a metabolism shift towards butanol production. The addition of butyrate also triggered much earlier and stronger sporulation during the fermentation. Genome-wide transcriptional analysis was performed for the culture in the fermentation with and without butyrate addition (in reactors R1 and R2 respectively) with RNA-Seq. In R1, the gene expression in the central metabolic pathway related to solventogenesis was induced about 10 hours earlier and accelerated, and the maximum expression levels of butyryl-CoA and solvent formation genes were elevated comparing to the control, which may correlate with the faster and higher solvent production. In both reactors, the sporulation genes were induced at the onset of solventogenesis, which was much earlier in R1, and went through similar temporal expression patterns but with much more rapidly down-regulation to final lower levels in stationary phase in R1. The motility genes were generally down-regulated to lower levels before stationary phase in both reactors. However, it took much longer in R2 and the gene expression maintained at comparatively higher levels in R2 after entering the stationary phase. In addition, the general PTS protein genes (*ptsI* and *ptsH*) were depressed earlier and more dramatically in the fermentation with added butyrate, which may be associated with the quicker and higher solvent production in

R1. Supplemented butyrate provided feedback inhibition to the butyrate formation pathway and may be re-assimilated through the reversed butyrate formation pathway, and thus result in an elevated level of intracellular butyryl phosphate, which may act as a phosphate donor to Spo0A and then triggered the solventogenesis and sporulation events.

5.2 Introduction

Biobutanol produced through fermentation with solventogenic clostridia has recently attracted a lot of attention due to its high values as a promising biofuel and chemical feedstock (Dürre, 2007). *C. beijerinckii* is an important solvent-producing microbe featuring a typical biphasic acetone-butanol-ethanol (ABE) fermentation. In a typical batch fermentation, acetic and butyric acids are produced during the active exponential growth phase. As growth slows down, acids are re-assimilated by the culture and acetone, butanol and a small amount of ethanol are produced (Jones and Woods, 1986). The switch from acidogenesis to solventogenesis is usually associated with induction of solvent formation gene transcription and decreased cell growth and motility. However, the mechanism that triggers the metabolic switch is still elusive.

The level of intracellular undissociated butyric acid has long been suggested as a controlling factor for the shift from acid production to solvent production in the metabolic pathway (Jones and Woods, 1986). Terracciano and Kashket even determined that solvent production began in *C. acetobutylicum* when the undissociated butyric acid level reached 13-18 mM (Terracciano and Kashket, 1986). While Husemann and Papoutsakis suggested that 6–13 mM of undissociated butyric acid was required for the initiation of solventogenesis when the external pH was between 3.7 and 5.0 (Husemann and Papoutsakis, 1988). Recently, Lee et al. reported that the initial addition of butyrate into the culture medium for ABE fermentation with *C. beijerinckii* 8052 enhanced solvent production (Lee et al., 2008). But in their study, detailed fermentation dynamics were not included, and no transcriptional analysis or enzyme activity assays were preformed, thus making it impossible to draw conclusions regarding the effects of butyrate on the metabolic pathway switch from the gene transcription level.

Transcriptional profiling of the fermentation culture with external butyrate addition may provide biological evidences for the effects of butyrate on solventogenesis metabolism. As we have already demonstrated, with RNA-Seq technology, genome-wide gene transcription

can be quantified and profiled to an unprecedented resolution and depth (Nagalakshmi et al., 2008; Perkins et al., 2009; Severin et al., 2010; Wang et al., 2012), with various advantages over traditional array-based transcriptional analysis methods (van Vliet, 2010; Wang et al., 2009). Therefore, the objective of this study was to examine the effect of supplemented butyrate in the growth medium on solvent production and the metabolic pathway switch. Along with high-resolution genome-wide transcriptional analysis by RNA-Seq, further insights concerning the biochemical effects of butyrate on solventogenesis related events are expected to come to light at the gene regulation level.

5.3 Materials and methods

5.3.1 Bacterial culture and fermentation experiment

Laboratory stocks of *C. beijerinckii* 8052 spores were stored in sterile H_2O at 4 ^oC (Annous and Blaschek, 1991). Spores were heat-shocked at 80 °C for 10 min, followed by cooling on ice for 5 min. The heat-shocked spores were inoculated at a 1% inoculum level into tryptone–glucose–yeast extract (TGY) medium containing 30 g L^{-1} tryptone, 20 g L^{-1} glucose, 10 g L^{-1} yeast extract and 1 g L^{-1} L-cysteine. The TGY culture was incubated at 35±1 ^oC for 12-14 h in an anaerobic chamber under N_2:CO_2:H_2 (volume ratio of 85:10:5) atmosphere. Subsequently, actively growing culture was inoculated with a 5% ratio into two BioFlo® 310 benchtop bioreactors (New Brunswick Scientific Co., Enfield, CT) both containing sterilized solution with 60 g L^{-1} glucose, 1 g L^{-1} yeast extract and chemically modified P2 medium (MP2) (Chen and Blaschek, 1999). In the first reactor (R1), 40 mM (3.52 g L^{-1}) sodium butyrate was also included in the medium to test the effect of butyrate on solvent production comparing to the control without butyrate addition in the second reactor (R2). MP2 medium, which was modified from P2 medium (Qureshi et al., 2001), contained following compounds (in g L^{-1}): KH_2PO_4, 0.5; K_2HPO_4, 0.5; $(NH_4)_2SO_4$, 2; $MgSO_4 \cdot 7H_2O$, 0.2; $MnSO_4 \cdot H_2O$, 0.01; $FeSO_4 \cdot 7H_2O$, 0.01; NaCl, 0.01; *p*-Aminobenzoic Acid, 0.001; Thiamin-HCl, 0.001; Biotin, 0.00001. 2-(N-Morpholino)ethanesulfonic acid (MES) (100 mM) (Sigma-Aldrich Co. LLC, St. Louis, MO) was added to the fermentation medium to prevent overacidification (Chen and Blaschek, 1999). Before inoculation, the pH of the fermentation medium was adjusted to 5.6 in both reactors with filter-sterilized 2 M NaOH. Oxygen-free nitrogen was flushed through the broth to initiate anaerobiosis until the culture initiated its

own gas production (CO_2 and H_2). Temperature was controlled at 35±1 ^oC. A stirring at 50 rpm was employed for mixing. Cell density and product concentration were monitored through the course of fermentation. For sequencing purpose, samples were taken over the exponential and stationary phases from both R1 and R2 (Figure 5.1 and Table 5.1).

5.3.2 Culture growth and fermentation products analysis

Culture growth was measured by following optical density (OD) in the fermentation broth at A_{600} using a Varian's Cary 300 Bio UV-visible spectrophotometer (Agilent Technologies, Inc., Santa Clara, CA). ABE, acetic acid, and butyric acid concentrations were quantified using gas chromatography (GC) system as previously described (Wang et al., 2011). Butanol yield (g g^{-1}) was defined as total butanol produced divided by the total sugar utilized, and butanol productivity was calculated as total butanol produced (in g L^{-1}) divided by the fermentation time and is expressed in g L^{-1} h^{-1}.

5.3.3 RNA isolation, library construction and sequencing

In preparation for RNA isolation, 10 ml cultures were harvested at each time point, and centrifuged at 4,000 × g for 10 min at 4 °C. Total RNA was extracted from the cell pellet using Trizol reagent based on manufacturer's protocol (Invitrogen, Carlsbad, CA) and further purified using RNeasy minikit (Qiagen, Valencia, CA). DNA was removed using a DNA-*free*™ kit (Ambion Inc., Austin, TX). RNA quality was assessed using a nanochip on a model 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA). RNA concentration was determined with a nanodrop (Biotek Instruments, Winooski, VT). Bacterial 16S and 23S ribosomal RNAs were removed with Ribo-Zero™ rRNA Removal Kits from Epicentre (Madison, WI) following the manufacturer's protocol. The enriched mRNA was converted to a RNA-Seq library using the mRNA-Seq library construction kit (Illumina Inc., San Diego, CA) following manufacturer's protocols. For sequencing, six samples were pooled and sequenced on one single lane of an eight-lane flow cell with the HiSeq 2000 system (Illumina Inc., San Diego, CA). The derived sequence reads were 100 nt long. The overall error rate of the control DNA was 0.84%. The total number of reads generated from each library is summarized in Table 5.1.

5.3.4 RNA-Seq data analysis

The generated 100-nt reads were mapped to the *C. beijerinckii* 8052 genome using MAQ, and those that did not align uniquely to the genome were discarded (Li et al., 2008). The quality parameter (-q) used in MAQ pileup was set to 30. Each base was assigned a value based on the number of mapped sequence coverage. The quantitative gene expression value, RPKM (reads/Kb/Million), was calculated using custom Perl scripts by normalizing the sequence coverage over the gene length and total unambiguously mapped reads in each library (Mortazavi et al., 2008; Severin et al., 2010).

The gene expression results were visualized colorimetrically using heatmap plots. R language and the open source software Bioconductor based on R were used for the data analysis and plots (Anders and Huber, 2010; Gentleman et al., 2004; Robinson et al., 2010). Time course RPKM values were first transformed to $\log_2$-scale. The $\log_2$-transfromed RPKM values were then properly centered for better representation of the data by the heatmap plots. More details about this method were summarized in the Additional file C5.2 in Appendix C.

5.3.5 RNA-Seq sequencing data access

The RNA-Seq sequencing data have been included in the Additional file C5.1 in Appendix C.

5.4 Results and discussion

5.4.1 Growth kinetics and ABE fermentation

The cell growth response in both reactors is illustrated in Figure 5.1. The cell culture in the fermentation with butyrate addition in R1 grew rapidly to the exponential phase following inoculation, while the control culture in R2 exhibited an obvious lag phase. Both cultures grew to similar final maximum cell densities, although the one in R2 took around 10 hours longer. Sporulation in R1 was triggered much earlier and stronger than that in R2. As shown in Figure 5.2, in R1, production of solvents was detected at 7-10 h after the start of fermentation, which is around the middle of the exponential growth phase (A_{600} = 0.5); while in R2 the solvent production started at approximately 17-20 h which corresponded to the late exponential growth phase (A_{600} = 1.1) (Figures 5.1 & 5.2). In R1, the supplemented butyrate was assimilated corresponding to the solvent accumulation, while the level of acetic acid increased even after the initiation of solvent production for about 6 hours before it reached

the maximum level. The solvent production achieved maximum soon after the cell culture reached stationary phase at around 46 h. Comparatively, the fermentation in R2 experienced a switch from acidogenesis to solventogenesis at around 17-20 h. Acetic and butyric acids were the main products during exponential growth phase, although the levels were very low. Butanol continued to increase throughout the stationary phase.

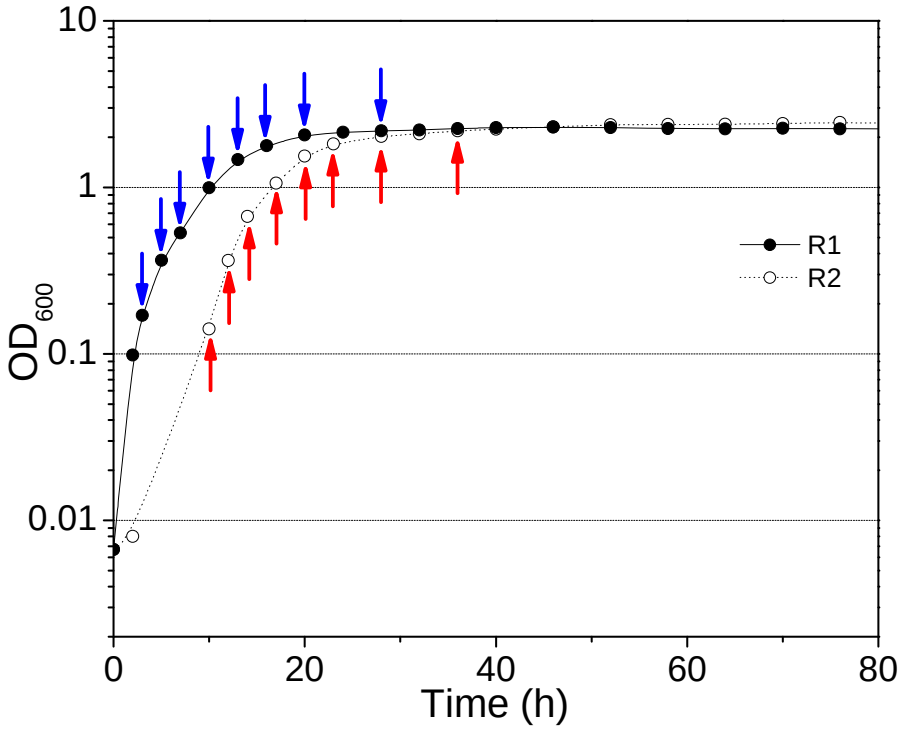


Figure 5.1 Cell growth kinetics of *C. beijerinckii* 8052 in batch fermentation with 40 mM sodium butyrate addition in R1 and no sodium butyrate addition in R2. Sampling points for RNA-Seq indicated by arrows.

With butyrate supplemented into the medium for ABE fermentation with *C. beijerinckii* 8052, the solvent production was initiated about 10 h earlier and finished much faster than the control. The butyrate supplementation also led to a higher final butanol titer (10.2 vs 7.8 g L^{-1}), yield (0.41 vs 0.26 g g^{-1}) and productivity (0.21 vs 0.09 g L^{-1} h^{-1}). In addition, the butanol/acetone ratio in the fermentation products increased to 2.4 (molar ratio) compared to 1.8 in the control, indicating a shift in the metabolism towards butanol production. Supplementation of butyrate in R1 triggered the solvent production from as early as mid-log phase. Even after the initiation of solventogenesis, acid accumulation was still observed in R1 until the late exponential growth phase. This does not conform to the classical phenomenological model that acids normally accumulate during the early acidogenesis phase prior to the solvent production. Butyrate addition may have provided feedback inhibition to

the butyrate production pathway, and thereby led to a declined butyrate formation flux. Since the butyrate formation pathway is also responsible for energy (ATP) generation for the cell culture, the culture needs to find an alternate ATP-generation pathway to make up the energy requirement. One option is the enhancement of the acetate formation pathway and increase in acetate accumulation and higher maximum level in R1 (Harris et al., 2000). So, this mechanism led to stronger acetate accumulation and higher peak level in R1. In addition, the supplemented butyrate provided a larger feedstock pool for conversion to a higher final total concentration of solvents, and resulted in a metabolism pathway shift towards butanol production and a higher butanol/acetone ratio. It is worthwhile to point out that the butanol yield in R1 (0.41 g g^{-1} glucose) was approximate to 100% of the maximum theoretical yield of butanol production from glucose (Shen et al., 2011), indicating that the supplemented butyrate was re-assimilated for butanol production.

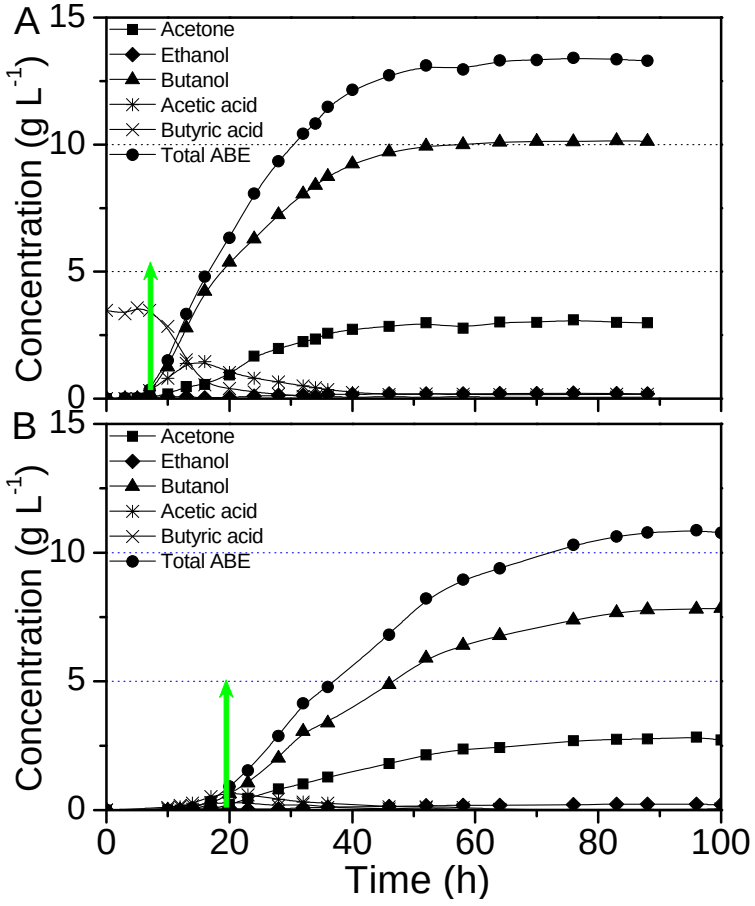


Figure 5.2 Solvents and acids production over time with the onset of solvent production indicated by green arrows. (A) Reactor R1: 40 mM sodium butyrate was added to the growth medium. (B) Reactor R2: no sodium butyrate was added to the growth medium.

Table 5.1 Summary of RNA-Seq sequencing and data analysis results

Samples from R1	R11	R12	R13	R14	R15	R16	R17	R18	---	Total
Time collected (h)	3	5	7	10	13	16	20	28	---	
Total number of reads	32934519	29367899	15075431	30451579	37230096	26184184	36117022	32444284	---	241287505
No. of reads mapped	31098649	27046332	14016757	29094448	34907591	24836210	34255629	30754172	---	227240126
No. of reads unambiguously mapped	29844200	26282300	13592284	28326600	34233900	24012800	33747700	30065800	---	221417584
No. of genes with detectable expression[1]	5070	5056	5032	5075	5084	5062	5087	5085	---	5088
Range in expression levels (RPKM)	7.4×10^{-3} ~ 2.0×10^{4}	1.1×10^{-2} ~ 1.8×10^{5}	5.7×10^{-3} ~ 8.6×10^{4}	2.3×10^{-2} ~ 5.0×10^{4}	1.6×10^{-3} ~ 9.2×10^{4}	1.5×10^{-3} ~ 1.1×10^{5}	2.3×10^{-2} ~ 2.2×10^{5}	2.3×10^{-2} ~ 8.8×10^{4}	---	1.5×10^{-3} ~ 2.2×10^{5}

Samples from R2	R21	R22	R23	R24	R25	R26	R27	R28	R23-2[2]	Total[3]
Time collected (h)	10	12	14	17	20	23	28	36	14	
Total number of reads	25350444	38025575	31352222	33877714	26099516	42485421	38176862	33624379	33916811	268992133
No. of reads mapped	23995017	35773620	29542453	32664767	24890605	40877908	36397654	31652246	32399849	255794270
No. of reads unambiguously mapped	22098900	31363000	27534800	31668100	24198000	39820500	35617800	30760200	31451300	243061300
No. of genes with detectable expression[1]	5059	5063	5058	5077	5084	5088	5088	5084	5081	5088
Range in expression levels (RPKM)	2.3×10^{-2} ~ 1.1×10^{5}	6.2×10^{-3} ~ 1.0×10^{5}	2.4×10^{-2} ~ 1.3×10^{5}	1.6×10^{-3} ~ 1.5×10^{5}	2.8×10^{-2} ~ 4.6×10^{4}	4.7×10^{-2} ~ 6.7×10^{4}	2.6×10^{-2} ~ 1.3×10^{5}	3.6×10^{-2} ~ 1.2×10^{5}	2.9×10^{-2} ~ 1.1×10^{5}	1.6×10^{-3} ~ 1.5×10^{5}

[1]Pseudogenes included;

[2]R23-2 is a biological replicate for sample R23; it was taken at the same fermentation time as for R23, but from a replicate reactor operated under the identical conditions;

[3]R23-2 is excluded.

5.4.2 RNA-Seq data analysis and overall gene transcription dynamics

The 100-nt sequence reads were mapped to the *C. beijerinckii* 8052 genome. Only those reads that mapped unambiguously to the genome were used for further analyses (Table 5.1). Based on the sequence data, 5008 out of 5020 (99.8%) protein-coding genes had detectable expression in all 16 samples, while 12 demonstrated no transcripts over all sampling time points, and are likely silent (Table C5.1 in Appendix C). In Chapter 3, 78 putative silent genes were identified during the batch fermentation with regular P2 medium (Table A3.5 in Appendix A). The silent genes may be associated with different fermentation conditions and the identification of silent genes is likely biased by the sequencing depth in different RNA-Seq runs. However, it is interesting that all of the 12 putative silent genes identified in this experiment are included in the previously identified list of 78 genes. In addition, except for one that codes for RNA-directed DNA polymerase (Cbei_3670), 11 out of the 12 putative silent genes are transposase encoding genes. As discussed earlier, although these transposases may have played important functions during the evolution of *C. beijerinckii*, they seem not functional any more during the ABE fermentation process.

A biological replicate was included in this study to further confirm the effectiveness of expression measurement with RNA-Seq approach. The biological replicate sample was taken at 14 h from a batch fermentation under the identical conditions as for R2. As illustrated in Figure 5.3, a high degree of correlation (R = 0.969) was observed for the $\log_2$-transformed RPKM values between sample R23 (taken at 14 h from R2) and the biological replicate sample R23-2 (Table 5.1). These results further demonstrated that RNA-Seq is an effective approach for quantification of gene expression.

Detrended Correspondence Analysis (DCA) was conducted as in Chapter 4 on the RNA-Seq data to compare the overall difference in dynamic transcription profiles throughout batch fermentation in both reactors. DCA is an ordination technique that uses detrending to remove the 'arch effect' usually existing in correspondence analysis (Hill and Gauch, 1980). As shown in Figure 5.4, when they were depicted using the first two detrended correspondences, the overall gene transcriptional dynamic profiles in R1 and R2 both demonstrated clear but very different temporal variation trends, which may be an indication of different cell physiology and metabolism under different fermentation conditions.

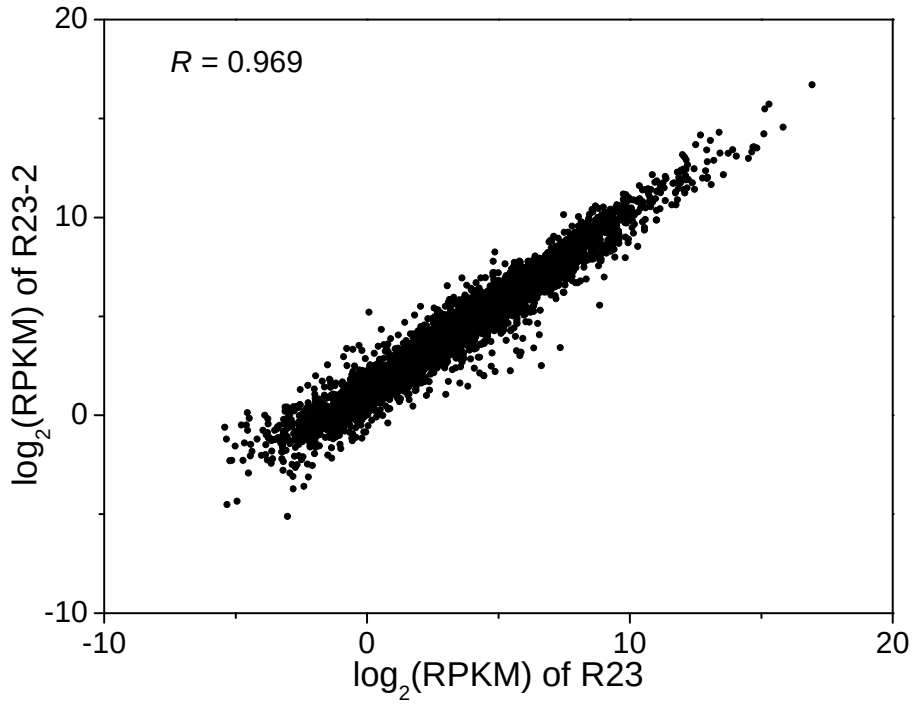


Figure 5.3 Correlation of gene expression in sample R23 (took at 14 h from R2) with its replicate sample R23-2 (took at 14 h from a fermentation operated under identical conditions as for R2).

5.4.3 Expression of glycolysis genes

As shown in Figure 5.5A, the expression of the glycolysis genes in both reactors showed very similar patterns, except that the corresponding gene transcription in R2 is about 10 h later than those in R1. Similar as what found in Chapter 4, the expression of most of these genes down-regulated by 2- to 6-fold in R1 after 16 h and by 2- to 3-fold in R2 after 28 h. However, *glcK* (Cbei_4895) in R2 demonstrated a very different expression pattern, which was up-regulated by 2.5-fold after 20 h and maintained at high levels during stationary phase. The reason for this unexpected expression pattern of *glcK* in R2 warrants further investigation. The three 6-phosphofructokinase (pfk) encoding genes (Cbei_0584, Cbei_0998 and Cbei_4852) demonstrated similar expression pattern in both fermentations. While Cbei_0584 was down-regulated after the culture entering late exponential phase (10 h for R1 and 20 h for R2 respectively), the expression of Cbei_0998 and Cbei_4852 kept at high levels until the start of stationary phase. Among the three pyruvate kinase (pyk) encoding genes (Cbei_0485, Cbei_1412 and Cbei_4851), in both reactors, the expression level of Cbei_0485 and Cbei_4851 decreased by 2- to 6-fold after entering stationary phase, whereas Cbei_1412

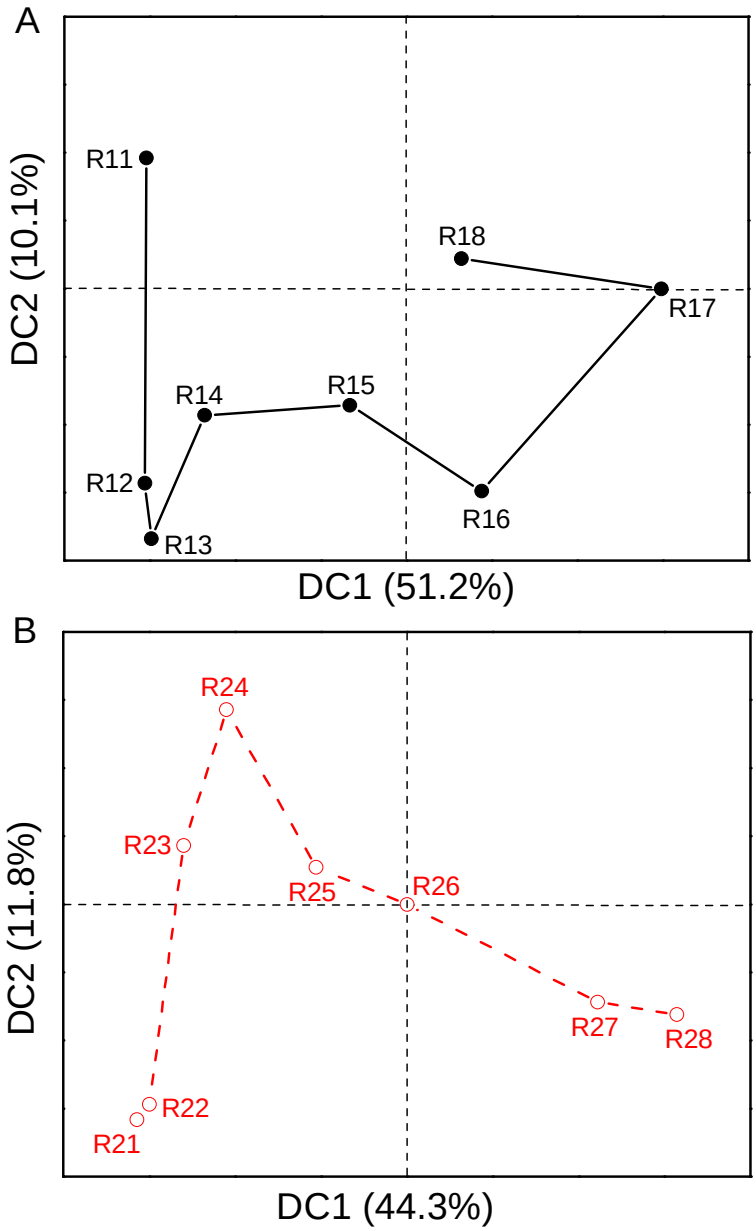


Figure 5.4 Detrended Correspondence Analysis (DCA) indicating different overall gene transcriptional dynamics in R1 and R2.

was expressed at relatively higher level throughout the late exponential phase and stationary phase with very low expression during early exponential phase. These results suggested that different allelic genes may play very different roles during different growth phases in the fermentation. The dynamic expression patterns of glycolysis genes observed here under both conditions with and without added butyrate are generally similar to those observed in Chapter 4.

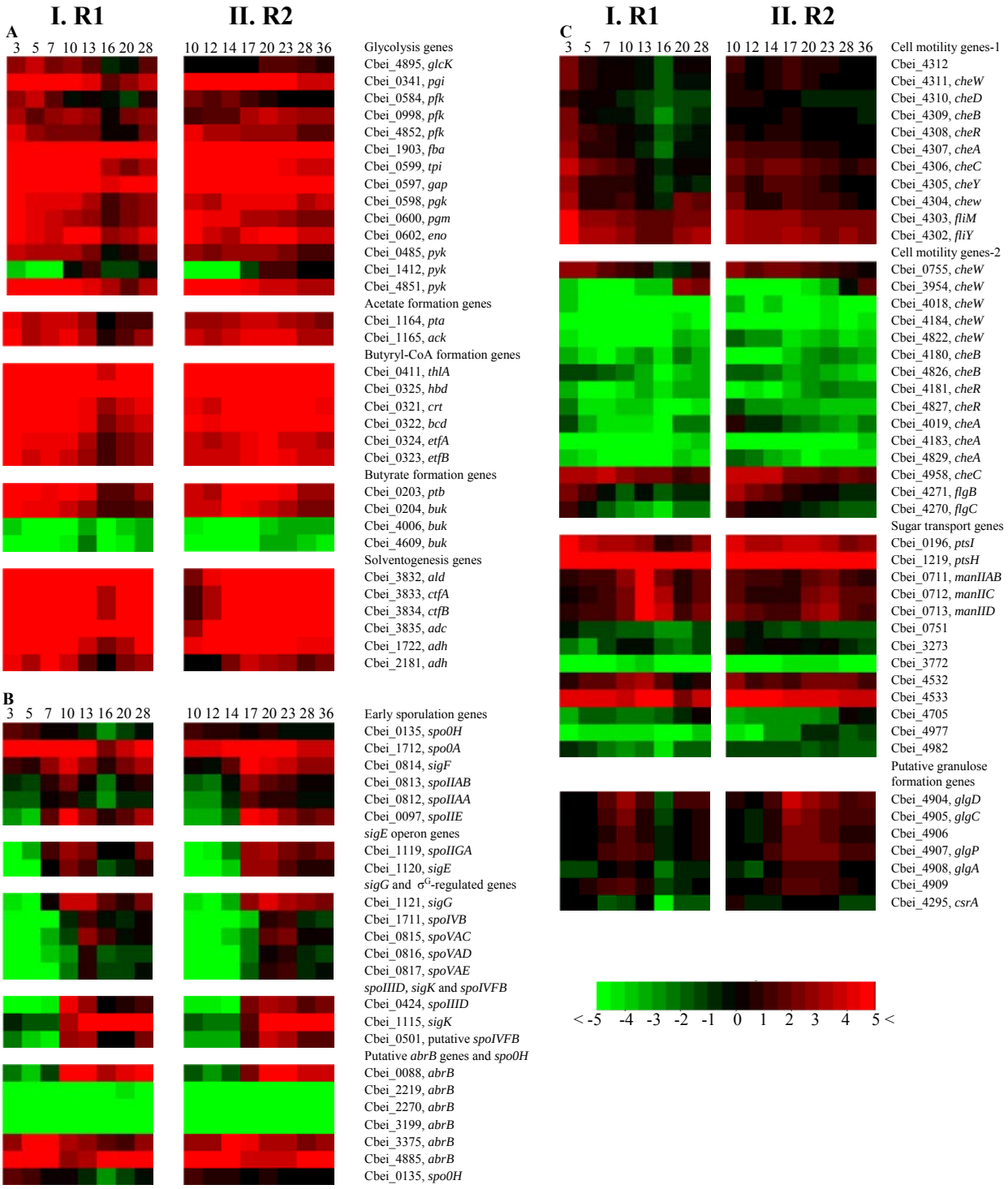


Figure 5.5 Comparative time course expression profiles of important genes during the batch fermentation in R1 and R2. Each corresponding sampling time point (in h) was indicated above the heatmap. (A) Genes related to glycolysis, acidogenesis and solventogenesis. (B) Sporulation genes, putative *abrB* genes and putative sensory kinase genes. (C) Genes related to cell motility, sugar transport and granulose formation.

5.4.4 Expression of acidogenesis and solventogenesis genes

In the fermentation with butyrate supplementation in R1, the genes involved in acid formation pathways, including *pta* (Cbei_1164) and *ack* (Cbei_1165) for acetic acid formation, *ptb* (Cbei_0203) and *buk* (Cbei_0204) for butyrate formation, and the genes related to acetoacetyl-CoA and butyryl-CoA synthesis were up-regulated from the very beginning of the fermentation, and declined 3- to 8-fold at 16 h during the late exponential phase corresponding to the time point when the accumulation of acetic acid reached maximum (Figures 5.5A and 5.2A). In the control fermentation in R2, the acid formation pathway genes were highly expressed and up-regulated from the middle of exponential phase and maintained at the highest expression levels until the onset of stationary phase (Figure 5.5A). The addition of butyrate triggered the solvent production in R1 as early as the middle exponential phase. Even after the onset of solvent production, acetic acid continued to accumulate, as a consequence of active acetic acid synthesis activity. Correspondingly, the expression of the acetic acid formation genes was maintained at high levels until the late exponential phase. Previously, a metabolic flux analysis of the acid formation pathway indicated that the butyrate uptake fluxes in *C. acetobutylicum* were not correlated with acetone production and thus suggested that butyrate synthesis enzymes might be also responsible for butyrate re-assimilation (Desai et al., 1999). Lately, further experimental evidence was provided that butyrate is not re-assimilated via the adc/ctfAB-dependent pathway as for acetate uptake in *C. acetobutylicum* because distinct uptake of butyrate has been observed in the fermentation with both adc-negative and ctfA-negative mutants (Lehmann et al., 2012). In this study, although butyric acid accumulation was not observed in R1 after initiation of the fermentation, the butyrate synthesis genes were up-regulated from the very beginning. This may suggest from a different point of view that butyrate formation genes are also responsible for butyrate re-assimilation in *C. beijerinckii*. In the control fermentation, the expression of the acid formation genes was slightly elevated from the middle of exponential phase until the onset of stationary phase, which also corresponded to the accumulation of acids in R2. In contrast to the fermentation in R1 with the supplementation of high level butyrate, the fermentation in R2 took longer to accumulate acids and then switch over to solventogenesis.

Based on the RNA-Seq sequencing data, a *sol* operon organized in the order of *ald* (Cbei_3832, encoding aldehyde dehydrogenase)-*ctfA* (Cbei_3833)-*ctfB* (Cbei_3834)-*adc*

(Cbei_3835) was revealed in *C. beijerinckii* 8052 (Wang et al., 2011). Highly coordinated expression patterns were observed for the *sol* operon genes in both reactors. In R1, corresponding to the early initiation of solvent production in exponential phase, the expression of these genes was high from the very beginning. It was further up-regulated by 3- to 5-fold at 7 h and kept at very high levels throughout the fermentation process; while in R2, the expression was up-regulated by 20- to 50-fold right before the onset of solvent production at 14 h and maintained at the highest levels afterwards (Figure 5.5A).

Two iron-containing alcohol dehydrogenase genes (*adh*, Cbei_1722 and _2181) were inferred to play key roles during primary alcohol production in *C. beijerinckii* 8052 by comparing them with their previously characterized orthologs in *C. beijerinckii* NRRL B592 (Chen, 1995; Shi and Blaschek, 2008; Wang et al., 2012). In R1, both genes were induced to high levels from the very beginning, and decreased by 2- to 3-fold before entering the stationary phase. In R2, their expression was up-regulated by 3- to 5-fold at 14 h and maintained at highest levels afterwards with a slight decrease after entering stationary phase (Figure 5.5A). Comparatively, in R1 both *adh* genes were transiently up-regulated for less than 10 hours from the beginning of the fermentation, while in R2 the expression of both genes kept at very high levels throughout the fermentation after they were induced, corresponding to the continuous solvent accumulation in R2. In addition, Cbei_1722 was comparatively higher regulated than Cbei_2181 in both reactors (2- to 6-fold higher in R1 and 4- to 22-fold higher in R2, respectively). It warrants further investigation whether this implies different significance or functions of these two genes for primary alcohol production in *C. beijerinckii*.

There are 20 annotated alcohol dehydrogenase genes in *C. beijerinckii* 8052, implying the great potential for solvent production of this strain (Wang et al., 2012). Many of these genes were expressed at high levels corresponding to the solvent production in both reactors. Besides Cbei_1722 and _2181, several other active alcohol dehydrogenase genes in both fermentations include Cbei_0685, _0869, _1932, _2243, _2421, _3864, _3890 and _4354 (Table C5.2 in Appendix C).

5.4.5 Expression of sporulation genes

In *B. subtilis*, the gene *spo0H* encodes σ^H, the earliest acting sigma factor associated with sporulation, which regulates the early sporulation genes, including *spo0A* and *sigF* operon

(Predich et al., 1992). In this study, the expression of *spo0H* (Cbei_0135) in R1 was induced during early log phase, and was down-regulated 2.5-fold soon at 7 h and down-regulated further later during the fermentation. In R2, a similar expression pattern was observed for *spo0H* where it was down-regulated slightly after 17 h (Figure 5.5B). The expression of *spo0A* (Cbei_1712) in R1 was induced early during the fermentation, and elevated by 2.4-fold at 7 h; afterwards the expression was down-regulated 5-fold after 13 h. In R2 the transcription of *spo0A* was rather constant at high levels prior to 14 h, and further up-regulated by 3.7-fold at 17 h and maintained at highest levels until 28 h, after which it was down-regulated by 7.5-fold. It has been reported that in *C. acetobutylicum*, *spo0H* and *spo0A* were both constitutively expressed at constant levels throughout the growth cycle, and the amount of *spo0H* transcript was much lower than that of *spo0A* (Dürre and Hollergschwandner, 2004). This was not consistent with the results for *C. beijerinckii* 8052 obtained in this study, where in both reactors the expression of *spo0H* tended to decline slightly corresponding to the initiation of solvent production, while *spo0A* was up-regulated corresponding to the solvent production and down-regulated before entering the stationary phase. However, in all cases the transcription level of *spo0A* was much higher than that of *spo0H* (Figure 5.5B).

The *sigF* operon (*spoIIAA-spoIIAB-sigF*) encodes the anti-anti-sigma factor (SpoIIAA), the anti-sigma factor (SpoIIAB) and the sigma factor σ^F. In R1, the *sigF* operon genes were up-regulated coordinately by around 4-fold at the onset of solvent production and soon down-regulated after 16 h. In R2, the expression of the *sigF* operon genes increased by 5- to 9-fold before the onset of solvent production (17 h) and declined by 5- to 6-fold before entering the stationary phase (28 h) (Figure 5.5B). SpoIIE is a phosphatase required for regulating the dephosphorylation of SpoIIAA, which helps to release the σ^F from SpoIIAB by forming an ADP-SpoIIAB-SpoIIAA complex (Arigoni et al., 1996). In R1, high expression of *spoIIE* (Cbei_0097) was induced at 7 h (53-fold up-regulated) corresponding to the onset of solvent production, and then peaked at 10 h after which it decreased gradually to a lower level. A similar expression pattern was observed for *spoIIE* in R2 where its expression was elevated 80-fold at 17 h and then declined slowly to lower levels, but was still 10-fold higher than that at 14 h (Figure 5.5B).

σ^E is the first mother cell-specific sigma factor in *B. subtilis* (Kenney and Moran, 1987). In *C. acetobutylicum*, the expression of *sigE* operon (*spoIIGA-sigE*, CAC1694 and CAC1695)

throughout the fermentation does not exceed 1.3-fold higher than that at the initiation of fermentation (Alsaker and Papoutsakis, 2005). In this study with *C. beijerinckii*, the expression of *sigE* operon in R1 (*spoIIGA-sigE*, Cbei_1119 and Cbei_1120) was up-regulated around 27-fold at 7 h, and decreased by 3- to 8-fold after 16 h. In R2, the expression of *sigE* operon was up-regulated by 32- and 90-fold for *spoIIGA* and *sigE* respectively at 17 h, and decreased slowly to lower levels after 23 h (Figure 5.5B). The expression patterns of *sigE* operon in this study further confirmed the observation that forespore and endspore developed more rapidly in *C. beijerinckii* than in *C. acetobutylicum*, as previously discussed (Shi and Blaschek, 2008; Wang et al., 2012).

σ^G is a forespore-specific sigma factor and it is active after complete engulfment of the forespore by the mother cell (Cutting et al., 1989). In the current study, the expression of *sigG* in R1 was induced at 7 h and further up-regulated by 13-fold at 10 h, and decreased to lower levels after 16 h although the expression was still high throughout the fermentation. Similarly, the expression of *sigG* in R2 was up-regulated by 78-fold at 17 h and peaked at 20 h after which the expression declined to lower levels (Figure 5.5B). More than 80 genes in *B. subtilis* related to spore protection and maturation were identified to be regulated by σ^G (Steil et al., 2005; Wang et al., 2006). In *C. beijerinckii* in this study, among the genes regulated by σ^G, the expression of *spoIVB* and the *spoVA* operon in both reactors was induced after the onset of high expression of *sigG*, and soon decreased to lower levels afterwards. The expression patterns of *sigG* and the putative σ^G–regulated genes well demonstrated their regulation manner in *C. beijerinckii* (Figure 5.5B), as was also discussed in Chapter 4.

In *B. subtilis*, *sigK* gene is expressed from a σ^E-dependent promoter, and regulated by another transcription factor *spoIIID* (Driks and Losick, 1991; Halberg and Kroos, 1994; Kunkel et al., 1989). The *sigK* gene is expressed as an inactive pro-σ^k factor first, and then cleaved to be activated by a membrane-localized protease, SpoIVFB, whose activity is induced by SpoIVB, the signaling protein produced in the forespore (Green and Cutting, 2000; Lu et al., 1990). In this study, in R1 the expression of *spoIIID* (Cbei_0424) and putative *spoIVFB* (Cbei_0501) was induced at 10 h, and then down-regulated to lower levels after 16 h (7.3- and 6.4-fold lower for *spoIIID* and *spoIVFB*, respectively), while *sigK* (Cbei_1115) was induced at 10 h and further up-regulated by 5.1-fold at 13 h, and maintained at highest levels throughout the fermentation. In R2, all three genes were induced at the onset of solvent production at 17 h, and further up-regulated at 20 h. Thereafter, the expression of

spoIIID and putative *spoIVFB* was slightly down-regulated to lower levels while the expression of *sigK* was maintained at highest levels throughout the fermentation (Figure 5.5B).

In *B. subtilis*, *abrB* encodes a transition state regulator, whose transcription depression by Spo0A leads to a burst of σ^H synthesis at the initiation of sporulation (Weir et al., 1991). Maximal expression of *abrB* in *B. subtilis* was observed 2 h before the onset of sporulation (Jiang et al., 2000). In *C. beijerinckii* 8052, six loci were annotated as genes encoding AbrB family transcriptional regulator. Similar to that observed in Chapter 4, the expression of three (Cbei_2219, Cbei_2270 and Cbei_3199) is at very low levels most of the time in both R1 and R2. Expression of Cbei_0088 was up-regulated at 10 h and 17 h for R1 and R2 respectively, and maintained at high levels afterwards. The expression of Cbei_3375 and Cbei_4885 was up-regulated before the onset of solvent production, and then decreased to lower levels in both reactors. So, as discussed in Chapter 4, none of the *abrB* genes in *C. beijerinckii* 8052 was observed to display an exact antagonistic expression pattern to that of *spo0H*.

Generally, the sporulation genes in both reactors went through a similar temporal expression patterns with much faster down-regulation in R1. In addition, after entering stationary phase, most of the sporulation genes demonstrated lower expression levels in R1 than in R2, corresponding to the much faster solventogenesis and sporulation events in R1.

5.4.6 Expression of cell motility genes

The cell motility-related genes encode products associated with chemotactic responses and flagellar assembly (Wadhams and Armitage, 2004). It was confirmed that in *C. beijerinckii* 8052 a flagellar/chemotaxis multi-gene cluster (Cbei_4312-4302) exists similar to that in *C. acetobutylicum* (CAC2225-2215) (Paredes et al., 2004; Wang et al., 2011). In R1, most genes in the flagellar/chemotaxis cluster were down-regulated by 2- to 4-fold at 5 h and then decreased gradually to even lower levels. In R2, although low expression levels were observed as early as 12 h for Cbei_4312, _4311, _4309 and _4308, generally the expression of the genes in the flagellar/chemotaxis cluster declined after 20 h when noticeable solvents were detected. Comparatively, *fliM* (Cbei_4303) and *fliY* (Cbei_4302) had much higher expression levels than the others in the cluster, although they were also down-regulated coordinately at the same time frame (Figure 5.5C).

There are several other genes in *C. beijerinckii* 8052 annotated as motility-related, although most of them are not actively expressed during the fermentation process (Figure 5.5C). Among them, Cbei_0755 (*cheW*) in both reactors gradually declined to lower levels after the onset of solvent production, with much lower final levels in R1. Another *cheW* gene (Cbei_3954) exhibited an unexpected antagonistic pattern, which was highly expressed after the cell culture entered stationary phase and reached the maximum level finally in both reactors. This observation was the same as in Chapter 4. In both reactors the gene Cbei_4958 (*cheC*) maintained at high expression levels throughout the fermentation although was down-regulated slightly after solvents were accumulated. The flagellar motility related genes *flgB* (Cbei_4271) and *flgC* (Cbei_4270) were strongly expressed during early growth phases (before 7 h and 17 h in R1 and R2 respectively), and down-regulated to very low levels afterwards (Figure 5.5C).

5.4.7 Expression of sugar transport genes

The phosphoenolpyruvate-dependent phosphotransferase system (PTS) is the predominant sugar uptake pathway in *C. beijerinckii* (Mitchell and Tangney, 2005). The general PTS proteins, referred as EI (enzyme I) and HPr (heat stable, histidine-phosphorylatable protein), phosphorylate the sugar in the cytoplasm, which has been transported across the cell membrane from the extracellular side by substrate-specific enzyme II (EII) domains. In *C. beijerinckii*, EI and HPr are encoded by Cbei_0196 (*ptsI*) and Cbei_1219 (*ptsH*) respectively. In R1, the expression of both genes was at high levels before 13 h, and decreased afterwards by 5- and 6.5-fold for *ptsI* and *ptsH* respectively, while in R2, the expression of *ptsI* and *ptsH* was maintained at high levels, and only decreased by 1.5- to 2-fold after 23 h. This observation was similar to that in Chapter 4, and agreed well with the previous report that ATP-dependent glucose phosphorylation was predominant and phosphoenolpyruvate-dependent glucose phosphorylation was repressed during solventogenic stage for both *C. beijerinckii* 8052 and its mutant *C. beijerinckii* BA101 strains (Lee and Blaschek, 2001). In addition, comparatively, the general PTS protein genes (*ptsI* and *ptsH*) were depressed earlier and more dramatically in the fermentation with butyrate addition, which is probably associated with the faster and higher solvent production in R1.

Previous studies indicated that glucose and mannose were assimilated by the same PTS system in *C. beijerinckii* (Lee and Blaschek, 2001; Mitchell et al., 1991). The RNA-Seq data

indicated that in *C. beijerinckii* 8052, the following genes encoding for the mannose family PTS transporters are organized in the same operon: *manIIAB* encoding for a membrane fusion protein of the IIA and IIB subunits involved in sugar phosphorylation, *manIIC* encoding for a substrate-specific permease, and *manIID* encoding for a mannose family-specific auxiliary protein (essential but of unknown function) (Barabote and Saier, 2005; Wang et al., 2011). In R1, the expression of the mannose family PTS transporter gene cluster was highly coordinated and gradually increased from the very beginning to a maximum at 13 h and then decreased to lower levels. In R2, the expression was rather constant from the beginning and up-regulated by 3-fold at 20 h after which the expression decreased slightly to lower levels (Figure 5.5C). It is interesting that in both reactors the expression of the PTS transporter genes reached peak levels at around the transition phases from exponential to stationary phases. Similar results were observed in Chapter 4. This observation may be associated with the cell activity for energy storage (for example, via the formation of granulose as discussed later) right before entering the more severe environments during stationary phase.

Homologs of glucose-glucoside (Glc) family PTS domains were identified in *C. beijerinckii* for all the gene clusters encoding the Glc family PTS transporters in *C. acetobutylicum* (Mitchell and Tangney, 2005; Wang et al., 2012). Among the encoding genes, Cbei_4533 (encoding N-acetylglucosamine-specific IIA subunit) and Cbei_4532 (encoding N-acetylglucosamine-specific IIBC subunit) were strongly expressed throughout the batch fermentation process (Figure 5.5C), which was similar to the observation in Chapter 4.

5.4.8 Expression of putative granulose formation genes

Granulose is a glycogen-like polymer used by bacteria for energy storage under severe environments (Reysenbach et al., 1986). In *C. beijerinckii* 8052, the RNA-Seq data indicated that *glgC* (Cbei_4905, encoding glucose-1-phosphate adenylyltransferase) and *glgD* (Cbei_4904, encoding ADP-glucose pyrophosphorylase) were organized in an operon, while the gene encoding 1,4-alpha-glucan branching enzyme (Cbei_4909), *glgA* (Cbei_4908, encoding glycogen synthase), *glgP* (Cbei_4907, encoding a granulose phosphorylase and functioning to depolymerize granulose) and the gene encoding the catalytic region of alpha amylase (Cbei_4906) were organized in another adjacent operon (Wang et al., 2011). The putative granulose formation genes (Cbei_4904-4909) in *C. beijerinckii* 8052 were highly coordinately expressed in both reactors; they were up-regulated at the onset of solvent

production, and were maintained at highest levels until the transition from exponential phase to stationary phase, corresponding to peak levels of the expression for PTS transporter genes as discussed above. The expression level of *csrA* (Cbei_4295, encoding carbon storage regulator) was much lower than that of the other granulose formation genes (Figure 5.5C).

With butyrate addition in R1, the expression of the genes in central metabolism related to solventogenesis events were induced earlier and accelerated. In addition, the culture in R1 showed elevated maximum expression of butyryl-CoA and solvent formation genes (although their maximum expression levels corresponded to very different time frames than in R2), which may correlate with the faster and higher solvent production. In both reactors, the sporulation genes were induced at the initiation of solvent production, and went through a similar temporal expression patterns to those observed in *B. subtilis* and *C. acetobutylicum*, although usually were down-regulated faster in R1 than the control in R2, corresponding to the faster and stronger sporulation in R1. The motility related genes were generally down-regulated to lower levels before the culture entered the stationary phase in both reactors. However, in R2, it took much longer and the gene expression was maintained at comparatively higher levels after entering stationary phase. This observation was consistent with the continuous solvent accumulation during stationary phase and lower final solvent titer in R2. Spo0A has been reported to negatively regulate the expression of chemotaxis/motility genes (Alsaker et al., 2004; Dürre and Hollergschwandner, 2004; Ravagnani et al., 2000). In R1, stronger induction of *spo0A* was observed with the onset of solvent production, which may coincide with the more rapid down-regulation and lower final expression levels of the motility genes.

The concentration of undissociated butyric acid has been suggested as the controlling factor for the switch to solvent production in solventogenic clostridia (Jones and Woods, 1986). Monot et al. reported that, when the concentration of undissociated butyric acid reached a level of 0.5 to 1.5 g L^{-1}, solvent production was initiated (Monot et al., 1983). Husemann and Papoutsakis determined that 6–13 mM of undissociated butyric acid was required for the initiation of solventogenesis when the external pH was between 3.7 and 5.0 (Husemann and Papoutsakis, 1988). However, Harris et al. later suggested that butyrate may be not involved in the induction of solventogenesis in a study where they obtained high solvent production from the fermentation with a butyrate kinase inactivated mutant of *C.*

acetobutylicum (Harris et al., 2000). Zhao et al. further indicated that butyryl phosphate (but not acetyl phosphate) correlated with (and possibly regulated) the initiation of solvent formation through a sensitive measurement of butyryl phosphate and acetyl phosphate levels in the wild type of *C. acetobutylicum* ATCC 824 and several mutant strains; transcriptional analysis further demonstrated that high butyryl phosphate levels also corresponded to down-regulation of chemotaxis and motility genes and up-regulation of solvent formation genes (Zhao et al., 2005). In this study, since the pH of the fermentation was known at each sampling point, the undissociated butyric acid level could be calculated using the Henderson–Hasselbalch equation (Harris et al., 2000; Husemann and Papoutsakis, 1988). Based on the assumption that undissociated acids can freely cross the cell membrane, the intracellular and extracellular undissociated acid concentrations are the same (Harris et al., 2000). As shown in Figure 5.6, the peaks of undissociated butyric acid levels in both reactors corresponded well to the onsets of solvent production. However, the peak level in the control (1 mM) was about six-time lower than that in the fermentation with butyrate supplementation (7 mM), and thus it is unlikely that there is a specific threshold level of undissociated butyric acid for the induction of solventogenesis in *C. beijerinckii*, although the latter falls well into the range of 6-13 mM as suggested by the established phenomenological models for initiation of solvent production in *C. acetobutylicum* (Husemann and Papoutsakis, 1988; Terracciano and Kashket, 1986). Instead, butyryl phosphate might be associated with the initiation of solvent production as suggested by Zhao et al. (Zhao et al., 2005). With the exogenous butyrate supplementation in R1, the high level of intracellular butyrate may result in elevated concentration of butyryl phosphate either through feedback inhibition to the butyrate formation pathway or through re-assimilation of butyrate converting to butyryl phosphate through the reversed butyrate formation pathway (Lehmann et al., 2012). While in R2, ATP were generated through the acid (both acetic and butyric) formation pathway during acidogenesis phase, thus leading to a much higher ATP level (or ATP/ADP ratio) comparing

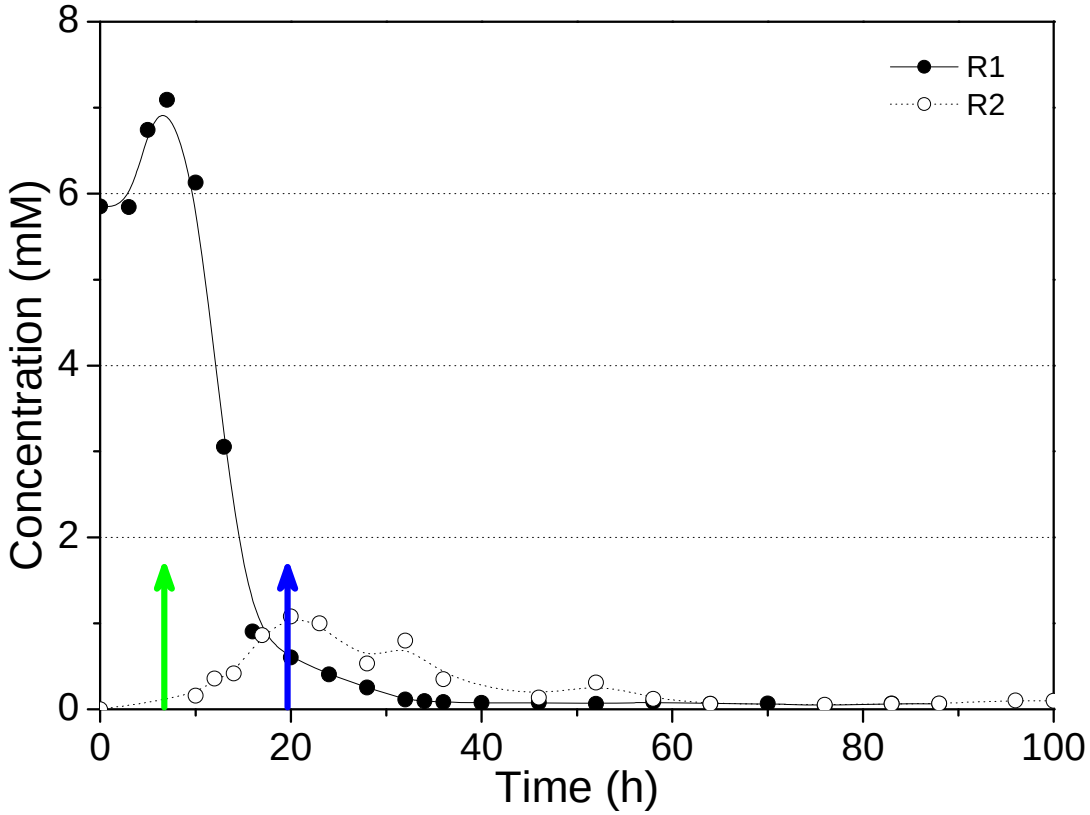


Figure 5.6 Calculated undissociated butyric acid levels over time in R1 and R2 based on the measured butyrate concentration. The onset of solvent production was indicated by green arrow (for R1) and blue arrow (for R2), respectively.

to that in R1. Therefore, although the level of undissociated butyric acid was much lower in R2, through the equilibrium of the reaction for the conversion between butyrate and butyryl phosphate, at the onset of solvent production the intracellular butyryl phosphate level (the "threshold" level) in R2 might be actually approximate to that in R1. Spo0A has been indicated as the master regulator for solventogenesis and sporulation events in clostridia and bacilli, and the phosphorylated Spo0A has been found to regulate the expression of various target genes (Alsaker et al., 2004; Dürre and Hollergschwandner, 2004; Ravagnani et al., 2000). The elevated level of butyryl phosphate beyond a "threshold level" may donate its phosphate group to Spo0A, and thus trigger solventogenesis. In addition, the phosphorylated Spo0A may also induce sporulation and cell motility reduction.

Actually, butyrate kinase from *C. acetobutylicum* has been reported to exhibit reversible activities, and the discrepancies in specific activities in the two directions suggest that there might be multiple forms of the enzyme (Hartmanis, 1987; Hartmanis and Gatenbeck, 1984; Jones and Woods, 1986). In *C. beijerinckii*, there are three genes (Cbei_0204, _4006 and

_4609) annotated to encode butyrate kinase. Besides Cbei_0204 which is co-operonic with *ptb* and highly regulated during the fermentation, Cbei_4006 and _4609 had comparatively only negligible expression throughout the fermentation in both reactors (Figure 5.5A). It still warrants further study in *C. beijerinckii* whether butyrate is re-assimilated through the reversible butyrate formation pathway and whether Cbei_0204 is responsible for this reversible reaction. In addition, further study is also required to elucidate the detailed molecular mechanism for solvent production in ABE fermentation. Development of advanced analytical methods for precise measurement of the intracellular fermentation intermediates, and generation of mutant strains with altered product profiles through metabolic engineering would be helpful to address this problem.

5.5 Conclusions

With exogenous butyrate supplementation, solvent production was triggered early in the mid-exponential phase and finished quickly in < 50 h; while in the control, solventogenesis started from late exponential phase and took > 90 h to reach the maximum butanol concentration. Butyrate supplementation led to a 31% improvement in final butanol titer, 58% improvement in butanol yield and 133% improvement in butanol productivity. The butanol/acetone ratio increased to 2.4 versus 1.8 in the control, indicating a metabolism shift towards butanol production. The addition of butyrate also triggered much earlier and stronger sporulation during the fermentation. Genome-wide transcriptional analysis with RNA-Seq indicated that, with butyrate addtion in R1, comparing to the control the gene expression in central metabolic pathway related to solventogenesis was induced about 10 hours earlier and accelerated, and the maximum expression levels of butyryl-CoA and solvent formation genes were elevated, which may correlate with the faster and higher solvent production. In both reactors, the sporulation genes were induced at the onset of solventogenesis, which was much earlier in R1, and went through a similar temporal expression patterns but with much faster down-regulation to final lower levels in stationary phase in R1. The motility related genes were generally down-regulated to lower levels before stationary phase in both reactors. However, in R2 it took much longer and the gene expression was maintained at comparatively higher levels after entering stationary phase. In addition, the general PTS protein genes were depressed earlier and more dramatically in the fermentation with butyrate addition, which may be associated with the more rapid and higher solvent production in R1.

Supplemented butyrate provided feedback inhibition to the butyrate formation pathway and might be re-assimilated through the reversed butyrate formation pathway, thereby resulting in an elevated level of intracellular butyryl phosphate, which might have acted as a phosphate donor to Spo0A and thus triggered the solventogenesis and related events.

5.6 References

Alsaker, K. V., and E. T. Papoutsakis. 2005. Transcriptional program of early sporulation and stationary-phase events in *Clostridium acetobutylicum*. *Journal of Bacteriology* 187(20):7103-7118.

Alsaker, K. V., T. R. Spitzer, and E. T. Papoutsakis. 2004. Transcriptional analysis of *spo0A* overexpression in *Clostridium acetobutylicum* and its effect on the cell's response to butanol stress. *Journal of Bacteriology* 186(7):1959-1971.

Anders, S., and W. Huber. 2010. Differential expression analysis for sequence count data. *Genome Biology* 11(10):R106.

Annous, B. A., and H. P. Blaschek. 1991. Isolation and characterization of *Clostridium acetobutylicum* mutants with enhanced amylolytic activity. *Applied and Environmental Microbiology* 57(9):2544-2548.

Arigoni, F., L. Duncan, S. Alper, R. Losick, and P. Stragier. 1996. SpoIIE governs the phosphorylation state of a protein regulating transcription factor σ^F during sporulation in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences of the United States of America* 93(8):3238-3242.

Barabote, R. D., and M. H. Saier. 2005. Comparative genomic analyses of the bacterial phosphotransferase system. *Microbiology and Molecular Biology Reviews* 69(4):608-634.

Chen, C. K., and H. P. Blaschek. 1999. Effect of acetate on molecular and physiological aspects of *Clostridium beijerinckii* NCIMB 8052 solvent production and strain degeneration. *Applied and Environmental Microbiology* 65(2):499-505.

Chen, J.-S. 1995. Alcohol dehydrogenase: multiplicity and relatedness in the solvent-producing clostridia. *Fems Microbiology Reviews* 17(3):263-273.

Cutting, S., S. Panzer, and R. Losick. 1989. Regulatory studies on the promoter for a gene governing synthesis and assembly of the spore coat in *Bacillus subtilis*. *Journal of Molecular Biology* 207(2):393-404.

Desai, R. P., L. M. Harris, N. E. Welker, and E. T. Papoutsakis. 1999. Metabolic flux analysis elucidates the importance of the acid-formation pathways in regulating solvent production by *Clostridium acetobutylicum*. *Metabolic Engineering* 1(3):206-213.

Driks, A., and R. Losick. 1991. Compartmentalized expression of a gene under the control of sporulation transcription factor σ^E in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences of the United States of America* 88(22):9934-9938.

Dürre, P. 2007. Biobutanol: An attractive biofuel. *Biotechnology Journal* 2(12):1525-1534.

Dürre, P., and C. Hollergschwandner. 2004. Initiation of endospore formation in *Clostridium acetobutylicum*. *Anaerobe* 10(2):69-74.

Gentleman, R., V. Carey, D. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Yang, and J. Zhang. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* 5(10):R80.

Green, D. H., and S. M. Cutting. 2000. Membrane topology of the *Bacillus subtilis* Pro-σ^K processing complex. *Journal of Bacteriology* 182(2):278-285.

Halberg, R., and L. Kroos. 1994. Sporulation regulatory protein SpoIIID from *Bacillus subtilis* activates and represses transcription by both mother-cell-specific forms of RNA polymerase. *Journal of Molecular Biology* 243(3):425-436.

Harris, L. M., R. P. Desai, N. E. Welker, and E. T. Papoutsakis. 2000. Characterization of recombinant strains of the *Clostridium acetobutylicum* butyrate kinase inactivation mutant: Need for new phenomenological models for solventogenesis and butanol inhibition? *Biotechnology and Bioengineering* 67(1):1-11.

Hartmanis, M. G. 1987. Butyrate kinase from *Clostridium acetobutylicum*. *Journal of Biological Chemistry* 262(2):617-621.

Hartmanis, M. G. N., and S. Gatenbeck. 1984. Intermediary metabolism in *Clostridium acetobutylicum*: levels of enzymes involved in the formation of acetate and butyrate. *Applied and Environmental Microbiology* 47(6):1277-1283.

Hill, M. O., and H. G. Gauch. 1980. Detrended correspondence analysis: an improved ordination technique. *Vegetatio* 42(1-3):47-58.

Husemann, M. H. W., and E. T. Papoutsakis. 1988. Solventogenesis in *Clostridium acetobutylicum* fermentation related to carboxylic acid and proton concentration. *Biotechnology and Bioengineering* 32(7):843-852.

Jiang, M., W. L. Shao, M. Perego, and J. A. Hoch. 2000. Multiple histidine kinases regulate entry into stationary phase and sporulation in *Bacillus subtilis*. *Molecular Microbiology* 38(3):535-542.

Jones, D. T., and D. R. Woods. 1986. Acetone-butanol fermentation revisited. *Microbiological Reviews* 50(4):484-524.

Kenney, T. J., and C. P. Moran, Jr. 1987. Organization and regulation of an operon that encodes a sporulation-essential sigma factor in *Bacillus subtilis*. *Journal of Bacteriology* 169(7):3329-3339.

Kunkel, B., L. Kroos, H. Poth, P. Youngman, and R. Losick. 1989. Temporal and spatial control of the mother-cell regulatory gene *spoIIID* of *Bacillus subtilis*. *Genes & Development* 3(11):1735-1744.

Lee, J., and H. P. Blaschek. 2001. Glucose uptake in *Clostridium beijerinckii* NCIMB 8052 and the solvent-hyperproducing mutant BA101. *Applied and Environmental Microbiology* 67(11):5025-5031.

Lee, S. M., M. O. Cho, C. H. Park, Y. C. Chung, J. H. Kim, B. I. Sang, and Y. Um. 2008. Continuous butanol production using suspended and immobilized *Clostridium beijerinckii* NCIMB 8052 with supplementary butyrate. *Energy & Fuels* 22(5):3459-3464.

Lehmann, D., D. Hönicke, A. Ehrenreich, M. Schmidt, D. Weuster-Botz, H. Bahl, and T. Lütke-Eversloh. 2012. Modifying the product pattern of *Clostridium acetobutylicum*. *Applied Microbiology and Biotechnology* 94:1-12.

Li, H., J. Ruan, and R. Durbin. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Research* 18(11):1851-1858.

Lu, S., R. Halberg, and L. Kroos. 1990. Processing of the mother-cell σ factor, σ^K, may depend on events occurring in the forespore during *Bacillus subtilis* development. *Proceedings of the National Academy of Sciences* 87(24):9722-9726.

Mitchell, W. J., J. E. Shaw, and L. Andrews. 1991. Properties of the glucose phosphotransferase system of *Clostridium acetobutylicum* NCIB 8052. *Applied and Environmental Microbiology* 57(9):2534-2539.

Mitchell, W. J., and M. Tangney. 2005. Carbohydrate uptake by the phosphotransferase system and other mechanisms. In *Handbook on clostridia*, 155-175. P. Dürre, ed. London: CRC press.

Monot, F., J. M. Engasser, and H. Petitdemange. 1983. Regulation of acetone butanol production in batch and continuous cultures of *Clostridium acetobutylicum*. *Biotechnology and Bioengineering* 25(S13):207-216.

Mortazavi, A., B. A. Williams, K. McCue, L. Schaeffer, and B. Wold. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* 5(7):621-628.

Nagalakshmi, U., Z. Wang, K. Waern, C. Shou, D. Raha, M. Gerstein, and M. Snyder. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320:1344-1349.

Paredes, C. J., I. Rigoutsos, and E. T. Papoutsakis. 2004. Transcriptional organization of the *Clostridium acetobutylicum* genome. *Nucleic Acids Research* 32(6):1973-1981.

Perkins, T. T., R. A. Kingsley, M. C. Fookes, P. P. Gardner, K. D. James, L. Yu, S. A. Assefa, M. He, N. J. Croucher, D. J. Pickard, D. J. Maskell, J. Parkhill, J. Choudhary, N. R. Thomson, and G. Dougan. 2009. A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus *Salmonella typhi*. *Plos Genetics* 5(7):e1000569.

Predich, M., G. Nair, and I. Smith. 1992. *Bacillus subtilis* early sporulation genes *kinA*, *spo0F*, and *spo0A* are transcribed by the RNA polymerase containing σ^H. *Journal of Bacteriology* 174(9):2771-2778.

Qureshi, N., A. Lolas, and H. P. Biaschek. 2001. Soy molasses as fermentation substrate for production of butanol using *Clostridium beijerinckii* BA101. *Journal of Industrial Microbiology & Biotechnology* 26(5):290-295.

Ravagnani, A., K. C. B. Jennert, E. Steiner, R. Grunberg, J. R. Jefferies, S. R. Wilkinson, D. I. Young, E. C. Tidswell, D. P. Brown, P. Youngman, J. G. Morris, and M. Young. 2000. Spo0A directly controls the switch from acid to solvent production in solvent-forming clostridia. *Molecular Microbiology* 37(5):1172-1185.

Reysenbach, A. L., N. Ravenscroft, S. Long, D. T. Jones, and D. R. Woods. 1986. Characterization, biosynthesis, and regulation of granulose in *Clostridium acetobutylicum*. *Applied and Environmental Microbiology* 52(1):185-190.

Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139-140.

Severin, A., J. Woody, Y.-T. Bolon, B. Joseph, B. Diers, A. Farmer, G. Muehlbauer, R. Nelson, D. Grant, J. Specht, M. Graham, S. Cannon, G. May, C. Vance, and R. Shoemaker. 2010. RNA-Seq Atlas of *Glycine max*: A guide to the soybean transcriptome. *BMC Plant Biology* 10(1):160.

Shen, C. R., E. I. Lan, Y. Dekishima, A. Baez, K. M. Cho, and J. C. Liao. 2011. Driving forces enable high-titer anaerobic 1-butanol synthesis in *Escherichia coli*. *Applied and Environmental Microbiology* 77:2905-2915.

Shi, Z., and H. P. Blaschek. 2008. Transcriptional analysis of *Clostridium beijerinckii* NCIMB 8052 and the hyper-butanol-producing mutant BA101 during the shift from acidogenesis to solventogenesis. *Applied and Environmental Microbiology* 74(24):7709-7714.

Steil, L., M. Serrano, A. O. Henriques, and U. Volker. 2005. Genome-wide analysis of temporally regulated and compartment-specific gene expression in sporulating cells of *Bacillus subtilis*. *Microbiology-Sgm* 151:399-420.

Terracciano, J. S., and E. R. Kashket. 1986. Intracellular conditions required for initiation of solvent production by *Clostridium acetobutylicum*. *Applied and Environmental Microbiology* 52(1):86-91.

van Vliet, A. H. M. 2010. Next generation sequencing of microbial transcriptomes: challenges and opportunities. *Fems Microbiology Letters* 302(1):1-7.

Wadhams, G. H., and J. P. Armitage. 2004. Making sense of it all: bacterial chemotaxis. *Nature Reviews Molecular Cell Biology* 5(12):1024-1037.

Wang, S. T., B. Setlow, E. M. Conlon, J. L. Lyon, D. Imamura, T. Sato, P. Setlow, R. Losick, and P. Eichenberger. 2006. The forespore line of gene expression in *Bacillus subtilis*. *Journal of Molecular Biology* 358(1):16-37.

Wang, Y., X. Li, Y. Mao, and H. Blaschek. 2011. Single-nucleotide resolution analysis of the transcriptome structure of *Clostridium beijerinckii* NCIMB 8052 using RNA-Seq. *Bmc Genomics* 12(1):479.

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

Wang, Z., M. Gerstein, and M. Snyder. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10(1):57-63.

Weir, J., M. Predich, E. Dubnau, G. Nair, and I. Smith. 1991. Regulation of *spo0H*, a gene coding for the *Bacillus subtilis* σ^H factor. *Journal of Bacteriology* 173(2):521-529.

Zhao, Y. S., C. A. Tomas, F. B. Rudolph, E. T. Papoutsakis, and G. N. Bennett. 2005. Intracellular butyryl phosphate and acetyl phosphate concentrations in *Clostridium acetobutylicum* and their implications for solvent formation. *Applied and Environmental Microbiology* 71(1):530-537.

CHAPTER 6. SUMMARY AND FUTURE WORK

6.1 Summary

In this study, with the high-throughput RNA-Seq technology, the genome-wide transcriptome structural organization and the dynamic transcriptional profiling in *C. beijerinckii* 8052 during a batch fermentation were first investigated. Some essential transcriptome structures in *C. beijerinckii* 8052 including genome-wide transcript boundaries and operon linkage were identified, and temporal transcriptional dynamic profiles corresponding to culture physiology and metabolism during a regular batch fermentation were demonstrated. In addition, transcriptional analysis was further performed using RNA-Seq for *C. beijerinckii* 8052 under butyrate supplemented conditions. Important gene transcriptional patterns associated with the effects of butyrate on solvent production and metabolic pathway switch were revealed. Several specific conclusions could be drawn from this study.

1. The transcription start sites and operon structure throughout the genome in *C. beijerinckii* 8052 were identified. The structure of important gene operons involved in metabolic pathways for acid and solvent production, including *pta-ack*, *ptb-buk*, *hbd-etfA-etfB-crt* (*bcs*) and *ald-ctfA-ctfB-adc* (*sol*) operons has been confirmed; important operons related to chemotaxis/motility, transcriptional regulation, stress response and fatty acids biosynthesis were defined.

2. In the current *C. beijerinckii* 8052 genome annotation, there are 20 previously non-annotated regions with significant transcriptional activities and 15 genes whose translation start codons were likely mis-annotated. Hence, the accuracy of existing genome annotation was significantly enhanced.

3. Transcriptome structural profiling provided important supplemental information on the accuracy of genome annotation, and revealed additional gene functions and regulation in *C. beijerinckii.*

4. The glycolysis genes were highly expressed throughout the fermentation, with comparatively more active expression during acidogenesis phase.

5. The expression of acid formation genes was down-regulated at the onset of solvent formation, in accordance with the metabolic pathway shift from acidogenesis to solventogenesis.

6. The *sol* operon genes exhibited highly-coordinated expression. Out of the > 20 genes encoding alcohol dehydrogenase in *C. beijerinckii*, Cbei_1722 and Cbei_2181 were highly up-regulated at the onset of solventogenesis, corresponding to their key roles in primary alcohol production.

7. Most sporulation genes in *C. beijerinckii* 8052 demonstrated similar temporal expression patterns to those observed in *B. subtilis* and *C. acetobutylicum*, while sporulation sigma factor genes *sigE* and *sigG* exhibited accelerated and stronger expression in *C. beijerinckii* 8052, which is consistent with the more rapid forespore and endspore development in this strain.

8. With 40mM sodium butyrate supplementation, the solvent production was triggered early in the mid-exponential phase and finished quickly in < 50 h from the beginning of innoculation. While in the control fermentation, solventogenesis started from late exponential phase and took > 70 h to reach the plateau of butanol production.

9. Comparing to the control without butyrate addition, butyrate supplementation led to a 31% improvement in final butanol titer, 58% improvement in butanol yield (sugar-based) and 133% improvement in butanol productivity. It also led to the butanol/acetone ratio increased to 2.4 versus 1.8 in the control, indicating a metabolism shift towards butanol production.

10. With butyrate addition, the gene expression in central metabolic pathway related to solventogenesis were induced about 10 hours earlier and accelerated, and the maximum expression levels of butyryl-CoA and solvent formation genes were elevated comparing to the control, which might correlate to the faster and higher solvent production.

11. In both reactors with or without butyrate addition, the sporulation genes were induced at the onset of solventogenesis, which was much earlier in the fermentation with butyrate addition, and went through a similar temporal expression patterns with much faster down-regulation in the reactor with butyrate addition.

12. The motility related genes were generally down-regulated to lower levels before stationary phase with or without butyrate addition. However, in the control it took much longer and the gene expression kept at comparatively higher levels after entering stationary phase.

13. Supplemented butyrate provided feedback inhibition to butyrate formation pathway and might be re-assimilated through the reversed butyrate formation pathway, and thus

resulted in an elevated level of intracellular butyryl phosphate, which might have acted as a phosphate donor to Spo0A and thus triggered the solventogenesis and sporulation events.

6.2 Recommendations for future work

Based on the results and conclusions from this study, some meaningful future work could be done in order to better understand the fundamental physiology and metabolism of *C. beijerinckii*, and further improve the strains for enhanced butanol production.

1. The RNA-Seq results from this study should be integrated and re-mined with system biology methods. The detailed gene regulation network in *C. beijerinckii* could be revealed and targets for metabolic engineering could be further identified.

2. Genome-wide transcriptional analysis should be performed for the hyper–butanol-producing mutant *C. beijerinckii* BA101 strain, the butanol tolerant strains SA-1 and SA-2 (Lin and Blaschek, 1983), comparing to the transcriptional profiling of wild type strain *C. beijerinckii* 8052. Along with the re-sequencing of the mutant strains, mechanisms for elevated butanol production, enhanced butanol tolerance could be identified. This can provide further insights for strain improvement with metabolic engineering approaches.

3. Further experiments or bioinformatic analyses should be carried out to confirm and characterize the putative new genes and the mis-annotated genes discovered in this study. They might be related to important metabolism of the strain.

4. RNA-Seq experiments should be carried out with specific protocols to further investigate the transcriptome and regulation elements, such as the small RNAs.

5. Metabolic engineering should be carried to better understand the solvent production metabolism and obtain mutant strains with better merits for butanol production.

For example,

(1) The putative transition state regulator genes (*abrB*), the putative sensory kinase genes, and the important signal transduction mechanism genes could be knocked out in order to further investigate their specific functions during the metabolic switch for solventogenesis and sporulation.

(2) The acid production pathways could be eliminated by knocking out the key acidogenesis genes, in order to improve the solvent production by leading more carbon flux to solventogenesis pathway.

(3) The acetone formation pathway genes could be knocked out in order to depress/eliminate acetone formation and thus improve butanol production.

(4) There are 20 annotated alcohol dehydrogenase genes in *C. beijerinckii* 8052. While their specific functions need to be further characterized, several of them that demonstrated high expression levels during the fermentation might be important ones related to primary alcohol production. Overexpression should be performed for these genes in *C. beijerinckii* to find out whether the solvent production could be enhanced.

6. Further experiments should be carried out to elucidate the effects of acids (both acetic acid and butyric acid) on solvent production and metabolic switch in *C. beijerinckii*. Development of advanced analytical methods for precise measurement of the intracellular fermentation intermediates, and utilization of isotope-labeled chemicals for the fermentation would be helpful to address this problem.

6.3 References

Lin, Y. L., and H. P. Blaschek. 1983. Butanol production by a butanol-tolerant strain of *Clostridium acetobutylicum* in extruded corn broth. *Applied and Environmental Microbiology* 45(3):966-973.

Appendix A: ADDITIONAL FILES TO CHAPTER 3

Additional files to Chapter 3 have been included in Appendix A as described following.

Additional file A3.1: The RNA-Seq sequencing data

Additional file A3.2: Refined genome annotation (in GenBank format) based on the findings from this work and the current *C. beijerinckii* 8052 genome annotation in NCBI. The GenBank file can also be downloaded from https://netfiles.uiuc.edu/blaschek/www/Wang-BMC2011.

Figure A3.1 Circular plots of the reads from all six samples mapping to the *C. beijerinckii* 8052 genome. The outermost and second outermost circles represent CDS on the forward and reverse strands respectively, both of which are colored according to Clusters of Orthologous Groups (COG) functional classification assigned to *C. beijerinckii* 8052 annotation. The gold peak and shading area represents greater than the average and lower (in purple). The COG functional classes and corresponding color-coding are as follows (with RGB color model values in the parentheses): Class J, black (0 0 0); Class K, blue (0 0 255); Class L, brown (165 42 42); Class B, dark blue (0 0 139); Class D, chocolate (210 105 30); Class V, cyan (0 255 255); Class T, red (255 0 0); Class M, yellow (255 255 0); Class N, dark green (0 100 0); Class U, grey (128 128 128); Class O, gold (255 215 0); Class C, orange (255 165 0); Class G, light grey (170 170 170); Class E, mid red (255 63 63); Class F, pink (255 192 203); Class H, purple (128 0 128); Class I, violet (238 130 138); Class P, skyblue (135 206 235); Class Q, tan (210 180 140); Class R, darkgrey (100 100 100); Class S, darkred (139 0 0); Not in COGs, green (0 255 0). If one gene belongs to more than one COG classes, the color for that gene was defined by the first class it belongs to as the above order.

Figure A3.2 Images of end-point RT-PCR products for 11 selected co-operonic gene pairs on a 1.5% agarose gel. For each gene pair (1-11, see details in Table A3.8) as labeled on the top, the band on the left lane is PCR product using cDNA reverse-transcribed from RNA samples as template, on the right is PCR product using genomic DNA as template.

Table A3.1 TSS and operon structure in *C. beijerinckii* 8052 chromosome

Table A3.2 Genes with 5'-UTR length ≥ 100 nt

Table A3.3 Putative riboswitches identified using RibEx among the 5'-UTRs over 100 nt

Table A3.4 Potential new genes predicted in non-annotated regions with significant transcriptional activities using GeneMark

Table A3.5 Putative silent genes over all the six sampling time points

Table A3.6 Putative housekeeping genes (HKGs)

Table A3.7 Transcription of the pseudogenes

Table A3.8 Operon structure validation with end-point RT-PCR

Table A3.9 Genes and primer sequences for qRT-PCR test

Table A3.10 The corresponding Gene ID and ortholog gene in *C. acetobutylicum* ATCC 824 for the genes in *C. beijerinckii* 8052 mentioned in the paper (three-letter short names were used in the main text)

Table A3.11 Summary of the supplementary information provided by RNA-Seq to the current *C. beijerinckii* 8052 genome annotation

Supplemental texts

For testing and enhancing the current *C. beijerinckii* 8052 genome annotation.

Non-annotated regions with significant transcriptional activity: Cbei_0576-0577, Cbei_1425-1426, Cbei_1460-1461, Cbei_1511-1512, Cbei_1572-1573, Cbei_1607-1608, Cbei_1748-1749, Cbei_2042-2043, Cbei_2151-2152, Cbei_2344-2345, Cbei_2395-2396, Cbei_2480-2481, Cbei_3091-3092, Cbei_3110-3111, Cbei_3202-3203, Cbei_3285-3286, Cbei_3362-3363, Cbei_3591-3592, Cbei_4206-4207 and Cbei_4988-4989.

These regions were tested for potential new genes with GeneMark (http://exon.gatech.edu/genemark/). See Table A3.4 for details.

Genes with potential incorrect annotated translation start sites (potential mis-annotations): Cbei_0427, Cbei_0469, Cbei_0574, Cbei_0889, Cbei_0933, Cbei_1291, Cbei_1377, Cbei_1404, Cbei_1407, Cbei_1716, Cbei_1787, Cbei_3528, Cbei_3767, Cbei_4000, Cbei_5042.

Appendix B: ADDITIONAL FILES TO CHAPTER 4

Additional files to Chapter 4 have been included in Appendix B as described following.

Additional file B4.1: RNA-Seq data analysis

Figure B4.1 The qRT-PCR verification of RNA-Seq quantification results for selected genes. The fold change of each gene's expression along time course measured by either qRT-PCR (in black) or RNA-Seq (in red) was compared to that of the same gene at first time point (time point 1 in Figure 4.1A). (A) Cbei_4852, *pfk*. (B) Cbei_1903, *fba*. (C) Cbei_4851, *pyk*. (D) Cbei_0203, *ptb*. (E) Cbei_0204, *buk*. (F) Cbei_3833, *ctfA*. (G) Cbei_3834, *ctfB*. (H) Cbei_0411, *thlA*. (I) Cbei_0325, *hbd*.

Table B4.1 The genes with Z-score > 1.7 in each sample

Table B4.2 The top 100 highest expressed genes in samples 2 and 4

Table B4.3 The expression of the putative alcohol dehydrogenase genes in *C. beijerinckii* 8052

Table B4.4 The putative signal transduction genes in SOM clusters C8-C10 in *C. beijerinckii* 8052

Table B4.5 Genes and primer sequences for qRT-PCR test

Appendix C: ADDITIONAL FILES TO CHAPTER 5

Additional files to Chapter 5 have been included in Appendix C as described following.

Additional file C5.1: The RNA-Seq sequencing data

Additional file C5.2: RNA-Seq data analysis

Table C5.1 Putative silent genes over all the sampling time points

Table C5.2 The expression of the putative alcohol dehydrogenase genes in *C. beijerinckii* 8052