The Dissertation Committee for Seung Hee Cho Certifies that this is the approved version of the following dissertation:

Exploring novel regulatory noncoding RNAs associated with ethanol tolerance in *Zymomonas mobilis* for strain engineering

Committee:

Lydia M. Contreras, Supervisor

Hal S. Alper

Scott W. Stevens

Jennifer A. Maynard

Edward M. Marcotte

Exploring novel regulatory noncoding RNAs associated with ethanol tolerance in *Zymomonas mobilis* for strain engineering

by

Seung Hee Cho, B.S.; M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2016

Dedication

I dedicate this work to my beloved family for their endless love and support

Acknowledgements

I would like to give my deepest appreciation to my advisor, Dr. Lydia M. Contreras for her great support, encouragement and patience all through my PhD work. I am really thankful for her excellent guidance and considerate personality to develop my professional capabilities and to enable me to enjoy the long journey of PhD study. Also, I am particularly grateful to have been advised by Dr. Hal Alper for his thoughtful comments and invaluable advice. I also want to thank Dr. Scott Stevens, Dr. Jennifer Maynard and Dr. Edward Marcotte for their helpful discussions and advice.

Many thanks go to Katie Haning for providing great collaborations and helpful discussions throughout entire studies. I also want to thank my undergraduate students, Cameron Blome, Roy Lei, Sang Hyun Ju and Trey Henninger, who helped my experiments for this work. All the members in the Contreras group, Kevin V., Juan, Jorge, Chen, Kevin B., Paul, Grant, Steve, Jo, Abby, Mia, Angela and Mark, deserve my heartful thanks for their friendship and help. I want to thank my Korean friends in Austin and South Korea for being my emotional supporters.

I would like to express my appreciation for my husband, Tae Hyun, who strongly supported my study and life in Austin. I give my love and thanks to my precious children, Lillian and Michael, who made my life to be filled with joy and happiness. I would like to thank my father and sister for their sacrifice and encouragement. I also would like to thank my father-in-law for his support and assistance. I will remember my mother-in-law's love in my memory. I want to give thanks to my sister-in-law and brother-in-law, who cheered me up and helped me a lot when I first came to Austin. Lastly, but most importantly, I am truly grateful for my mother, Hyun Sook, who is the reason of my life. This work would not have been possible without her sacrifice, support and love.

Exploring novel regulatory noncoding RNAs associated with ethanol tolerance in *Zymomonas mobilis* for strain engineering

Publication No._____

Seung Hee Cho, Ph.D.

The University of Texas at Austin, 2016

Supervisor: Lydia M. Contreras

Zymomonas mobilis has been identified as a promising cellular factory for biofuels due to its efficient, natural production of and tolerance to ethanol. With rising demands of efficiency and sustainability, the use of microbes as chemical factories is increasingly attractive. In bacteria, small noncoding RNAs have been highlighted as powerful tools due to their regulatory roles in cellular pathways and as such are being increasingly exploited for engineering purposes. As global controllers of gene expression, noncoding regulatory RNAs represent powerful tools for engineering complex phenotypes. However, mechanistic analysis of these regulators in bacteria lags far behind their high-throughput search and discovery; this makes it difficult to understand how to efficiently identify noncoding RNAs that could be used to engineer a phenotype of interest. This dissertation describes discovery of novel global regulatory small noncoding RNAs that impact ethanol tolerance in *Z. mobilis* using a forward systems approach (Chapter 2). The effect of the bioinformatically predicted candidates were experimentally characterized by building overexpression and deletion strains. This led to the successful

uncovering of several sRNAs that could be manipulated to enhance ethanol tolerance as well as ethanol production in *Z. mobilis*. Using these ethanol-related sRNAs, we then performed traditional genetic and biochemical approaches as well as transcriptomics and proteomics methods to identify a pool of mRNA targets and pathways that were being regulated by these sRNAs under conditions of enhanced ethanol tolerance (Chaper 3). A second major part of my PhD work, also addressed in this dissertation is the search and exploitation of control elements (especially untranslated regions (UTRs)) that regulate gene expression at the local level associated with ethanol as well as other metabolic stresses (Chapter 4) in *Z. mobilis*. Specifically, ethanol responsive-UTR element identified in this work regulates the expression of Hfq and affects cell growth under ethanol stress condition. This work represents the first application of a *de novo* sRNA engineering strategy in a non-model organism, *Z. mobilis,* which is of relevance to biofuel technologies. Furthermore, our development and application of a novel bioinformatics pipeline for this work has demonstrated the ability to use formal computational approaches (in conjuction with systems-level methods) to accelerate the discovery of specific pathways that could be further optimized to enhance a given complex phenotype. In addition to my major work in *Z. mobilis*, I have also worked on a small side project that has examined the use of a ribosomal protein to enhance translational yields. My work has resulted in 4 publications, listed below.

1. Synthetic chimeras with orthogonal ribosomal proteins increase translation yields by promoting mRNA associations with active ribosomes, *Biotechnology Progress,* 32 (2): 285-93 (2016) <u>CHO SH</u>, Ju SH, Contreras LM.

2. Strain Engineering via regulatory noncoding RNA mechanisms: not a one-blueprint fits all, *Current Opinion in Chemical Engineering*, 10:25-34 (2015) <u>CHO SH</u>*, Haning K*, Contreras LM. (* equal contribution)

3. Discovery of ethanol responsive small RNAs in Zymomonas mobilis, *Applied and Environmental Microbiology*, 80 (14): 4189-98 (2014) <u>CHO SH</u>, Lei R, Henninger TD, Contreras LM.

4. Small RNAs in Mycobacteria: an unfolding story, *Frontiers in cellular infection and microbiology*, 4(96):1-11 (2014) Haning K, <u>CHO SH</u>, and Contreras LM.

Table of Contents

List of Tables

List of Figures

(33.4kDa) and RIIIa (23.2kDa) band from pDuet-L4H2-RIIIa with an arrow, respectively. Lane 2, 4, 6, 8 showed un-induced sample from pRIIIa, pL29-RIIIa, pL4H2-RIIIa and pDuet-L4H2-RIIIa, respectively. The identity of the bands was confirmed by Western blotting using anti-His antibody. Anti-RecA was used as the loading control. (C) Band intensity from Western blot analysis was quantified using ImageQuant TL 8.1 for the expression of RIIIa from pRIIIa and pDuet-L4H2-RIIIa and L4H2-RIIIa from pL4H2-RIIIa. The intensity of each target is normalized by the intensity of RIIIa. Error bars were calculated as SEM. 125

Chapter 1

Introduction and Background

1.1 INTRODUCTION

With rising demands of efficiency, environmental care, and sustainability, the use of native or engineered microbes as biofuel organisms is increasingly attractive. *Zymomonas mobilis*, a natural ethanol-producing microbe, has been extensively studied at a fundamental level as well for industrial applications (Doelle et al, 1993; Jang et al, 2012; Seo et al, 2005; Yang et al, 2009b). Recent advances in synthetic and systems biology has enabled the identification of regulatory RNAs associated with metabolic, physiological, and pathogenic pathways (Romby & Charpentier, 2010). As regulatory noncoding RNAs (ncRNAs) continue to be discovered in organisms of traditionally high relevance to biotechnology, understanding of and exploitation of natural cellular regulation to help achieve production and efficiency goals has expanded. These ncRNAs (especially small RNAs in bacteria) are ~50-300 nt transcripts that act as regulators of mRNA and protein expression, typically by blocking translation or changing stability (Storz et al, 2011). In bacteria, ncRNAs have been highlighted as powerful tools due to their regulatory roles in cellular pathways (Chappell et al, 2013; Kang et al, 2014; Qi & Arkin, 2014; Vazquez-Anderson & Contreras, 2013b). Based on the literature, we expected engineering efforts involving regulatory ncRNAs to significantly contribute to the strain improvement for advanced biofuel production in any organism.

1.2 BACKGROUND

1.2.1 Microbial biofuel tolerance

Biofuel has been emphasized as a potential alternative energy source due to its sustainability and due to concerns about global environment contamination over several decades. Biofuel has mainly focused on bioethanol which can be produced naturally from microorganisms as first generation of biofuel (Fortman et al, 2008). Recent developments in metabolic engineering has enabled the production of other forms of potential biofuels such as biodiesel, butanol, longer-chain alcohols, fatty acid-derived fuels, cyclic isoprenoids, and short-branched chain alkanes in microorganisms (Kalscheuer et al, 2006; Lee et al, 2008b; Rude & Schirmer, 2009). As such, alcohol producing microorganisms such as *Escherichia coli*, *Clostridium acetobutylicum*, *Saccharomyces cerevisiae* and *Zymomonas mobilis* have been studied and engineered for high tolerance and to produce to produce industrial-scale biofuels (Dunlop, 2011; Ingram et al, 1987; Lee et al, 2008a; Lee et al, 2008b; Paredes et al, 2005; Stephanopoulos, 2007).

E. coli has been extensively engineered for the efficient production of biofuels as it can utilize both pentose and hexose sugars. Additionally, *E. coli* is well-characterized and easy to manipulate for its genetic features as well as a suitable host for the production of valuable metabolites (Kim et al, 2007; Martin et al, 2003). In *E.coli*, ethanol fermentation is not a primary pathway and, as result, maximum yields reach only 50% compared to that of primary ethanol producing pathway (Jarboe et al, 2007). However, by introducing the *pdc* (pyruvate decarboxylase) and *adhB* (alcohol dehydrogenase II) genes from *Z. mobilis*, engineered *E.coli* strain can generate ethanol with enhanced ethanol producing ability (Ohta et al, 1991). Furthermore, the application of engineering efforts via mutagenesis, direct-evolution of *E. coli* allowed development of improved ethanologenic *E. coli* strains (Yomano et al, 1998; Yomano et al, 2008).

Butanol has better properties than ethanol as a biofuel due to its energy density, corrosiveness, low volatility and its suitability for a substitute of gasoline (Jin et al, 2011). Butanol is naturally synthesized from acetone-butanol-ethanol (ABE) fermentation in *Clostridium acetobutylicum*. Even though *Clostridium acetobutylicum* has complex genetics, comprehensive research on better understanding of metabolic pathways has lead to enhanced butanol producing strains. For example, disruption of the acetoacetate decarboxylase gene increased butanol production in a way that engineering a *C. acetobutylicum* M5 strain facilitates the formation of precursor for ethanol and butanol (Jiang et al, 2009; Lee et al, 2009). Furthermore, advances in omics approaches and synthetic biology approaches have allowed development of robust strains that are more efficient at overcoming low butanol yield and toxicity (Xue et al, 2013). Due to its limited genetic tools and complex physiology relative to *E. coli*, genes associated with butanol fermentation pathways from *C. acetobutylicum* have also been trasferred to *E. coli* to produce engineered *E.coli* strains for 1-butanol production (Atsumi et al, 2008).

Zymomonas mobilis has been a major focus for ethanol production due to its ability to produce 95% theoretical yield (Rogers et al, 2007). *Z. mobilis* utilizes its unique Entner Doudoroff (ED) pathway to produce ethanol with high ethanol yields and high specific sugar uptake (Doelle et al, 1993). Extensive physiological and genetic studies of *Z. mobilis* have further helped this organism to be more promising for biofuel. For instance, transcriptomic analysis of *Z. mobilis* has contributed to the improvement of strain for industrial applications (He et al, 2012; Yang et al, 2009b). However, one of the disadvantages of *Z. mobilis* is that it cannot ferment pentose sugars to produce ethanol. Due to the limitations in its sugar usage, genes for xylose utilization have been imported to *Z. mobilis* to enhance its capability as a biofuel microorganism (Zhang M et al, 2007).

3

A key challenge with the development of biofuels from microorganisms is that the chemicals or molecules produced during biomass pretreatment and hydrolysis as well as the final product itself is often toxic and therefore inhibitory to cell growth. Biomass represents renewable resources for the production of biofuels and biologically produced chemicals such as carboxylic acids, alcohols, hydrocarbons, biodiesel (Tomes et al, 2011; Zheng et al, 2008). It most commonly refers to a plant-based material, lignocellulosic biomass. To be used as a carbon source for biofuel producing microorganisms, biomass should be treated to release hexose or pentose sugar via pretreatment process (Ingram et al, 1987; Lee et al, 2008b; Stephanopoulos, 2007). During biomass pretreatment process, by-products (acetic acids, carboxylic acids andphenylic compunds depending on pro cessing methods) and their accumulation often affect cell viability, resulting in cell death. Therefore, how engineering metabolic pathways to effectively overcome these toxicity effects, by engineering strain tolerance, is the key to achieve high biofuel yields from microorganisms.

Recently, many studies have focused on developing engineered strains with improved biofuel tolerance by decreasing product toxicity. It is importatnt to note that toxicity levels vary across different types of biofuels and the types of solvents (Dunlop, 2011); for instance, long-chain alcohols are more toxic than short-chain alcohols. Furthermore, each microorganism can tolerate different amount of stress. For example, engineered *E. coli* strain can tolerate 60g/L ethanol, but *Z. mobilis* can tolerate ethanol up to 120g/L (Rogers et al, 2007). Due to the dependence on the tocixity properties on the microorganisms, engineering strategies to overcome challenges for biofuel tolerance may vary and the effectiveness of strategies may also be different. Biofuel tolerance is complicated and often linked to general stress responses. Although, in general, biofuel tolerance is often correlated with membrane permeability (which can allow the release of

4

ATP, ions, phospholipids, RNA, and proteins etc.), membrane stability and energy transduction, it can be difficult to predict the effect of a given tolerance strategy (Dunlop et al, 2011).

Among various strategies, several promising methods have been applied to different organisms for strain engineering purpose. First of all, overexpresion of heat shock proteins in response to solvent stress contribute to the improvement of biofuel tolerance. Heat shock proteins (RpoH, groESL, gnaKJ, hsp18, and hsp90) function as chaperones under stress condition and their expression is upregulated in many organisms such as *E. coli* and *C. acetobutylicum* under stress (Brynildsen & Liao, 2009; Tomas et al, 2004). In *C. acetobutylicum*, overexpression of GroESL increases butanol tolerance as well as butanol yields (Tomas et al, 2003). Overexpression of heat shock proteins was also tried in *E. coli* and and *Lactobacillus plantarum* and demonstrated to show improved butanol tolerance (Fiocco et al, 2007; Reyes et al, 2011). Another strategy of increasing biofuel tolerance is to modify membranes to decrease membrane permeability as biofuel products often disrupt membranes and increase permeability, causing membrane structure disruption and ultimately cell death. By transforming trans-fatty acids to cis-fatty acids in membranes which are catalyzed by the *cis/trans* isomerase (*cti*), membranes whave been shown to be stabilized and as a result solvent tolerance in *Pseudomonas syringae* has been observed (Junker & Ramos, 1999; Kiran et al, 2004). Other approach to target membrane transport systems (pump) has been studied to increase biofuel tolerance in *E.coli*. After screening efflux pumps that were known to export toxins, libraries of efflux pump have been expressed heterologously to test their contributions to tolerance in *E. coli*. These overexpressed pumps have also been shown to increase biofuel tolerance in *E. coli* (Dunlop et al, 2011).

5

With advances in genome engineering approaches, recent studies have increasingly focused on regulators such as transcription factors or genes related to stress response mechanism. Regulators for controlling stress response are systematically identified as targets for the improvement of biofuel tolerance or switches to turn on and off the expression of gene tolerance mechanisms. In *E. coli*, extensive studies on increasing tolerance to ethanol, butanol, isobutanol, toluene and isooctane have demonstrated the correlation of the regulators and tolerance mechanisms (Chong et al, 2013; Kang & Chang, 2012; Xu et al, 2015). Remarkably, engineering of global regulator cAMP receptor protein (*crp*) in *E. coli* affected increase in tolerance to various stresses such as ethanol, butanol, isobutanol and toluene (Basak et al, 2012; Chong et al, 2014; Lee et al, 2011; Zhang et al, 2012). Besides *crp*, sigma factor rpoD was identified to increase ethanol tolerance up to 70g/L ethanol (Alper & Stephanopoulos, 2007) in *E. coli*. *C. acetobutylicum* was also engineered to increase butanol tolerance with the generation of knockout strains for solR and overexpression strains of spo0A which are two genes associated with sporulation and soventogenesis pathway (Alsaker et al, 2004; Harris et al, 2001). Table 1 shows a summary of regulators that have been engineered for the improvement of tolerance.

1.2.2 Roles of regulatory RNAs associated with stress responses

It has been discovered that RNA transcripts act as important regulators of gene expression at the post transcriptional level in both prokaryotes (Wassarman, 2002) and eukaryotes (Lee et al, 1993). These regulators, referred to as small regulatory noncoding RNAs (ncRNA) in prokaryotes, are relatively short around 50-300 nucleotides and are not translated. ncRNAs have diverse functions including synthesizing proteins, splicing, editing RNA, modifying rRNA and catalyzing ncRNAs (Eddy, 2001). Initially, ncRNAs

were identified as a kind of RNAs that were not transfer, ribosomal or messenger RNAs. To date, ncRNAs have been discovered and validated using many computational methods and then, confirmed by experimental strategies in various organisms (Vogel & Sharma, 2005).

Stress	Strain	Regulator	Engineering Methods	Improvement
Ethanol	*E. coli*	*crp*	mutant library	up to 62g/L
		rpoD	gTME	up to 70g/L
Butanol	*E.coli*	*crp*	mutant library	1.2% (v/v)
		rpoA	mutant library	0.9% (v/v)
		entC	overexpression	32.8% increase
		feoA	overexpression	49.1% increase
		astE	Knockout	48.7% increase
		ygiH	Knockout	48.4% increase
	C.acetobutylicum	*cac0003*	overexpression	13% increase
		cac1869	overexpression	81% increase
		solR	Knockout	25% more production
		spo0A	overexpression	Increased tolerance under 0.6% butanol
Toluene	*E. coli*	*crp*	mutant library	Increased tolerance under 0.23% (v/v)
Isobutanol	*E. coli*	*crp*	mutant library	Increased tolerance under 1.2% (v/v)

Table 1.1: Summary of regulators engineered for the improvement of tolerance.

Regulatory ncRNAs are divided into different sub-groups, depending on their genomic locations with respect to their mRNA target(s). (1) ncRNAs which are encoded on the same gene with a target mRNA but in opposite direction are called an antisense RNA or cis-acting ncRNAs. Antisense RNAs affect translation and mRNA instability of the complementary target gene (Georg & Hess, 2011a). (2) A second class of ncRNAs that are referred to as trans-acting ncRNAs regulates mRNAs by imperfect base pairing with distal mRNA targets. Trans-acting ncRNAs are also called small RNAs (sRNAs) in

bacteria. Generally, small ncRNAs inhibit translation by base pairing interactions with a target mRNA around the ribosome binding site so that they can modulate gene expression post-transcriptionally (Waters & Storz, 2009). ncRNAs can be both activators and repressors for gene expression depending on what part of the mRNA molecule they base-pair with. ncRNAs can activate mRNA translation through an anti-antisense mechanism where sRNA base-pairing with a target mRNA disrupt a secondary structure, sequestering the ribosome-binding site. As a result, the ribosome-binding site is liberated and free to bind ribosomes for translation. In contrast, repressor ncRNAs negatively regulate gene expression by binding to the 5' UTR often near the ribosome binding site. The binding inhibits translation by prohibiting ribosome binding and/or target the mRNA for degradation by RNases (often RNase E). The functions of trans-acting ncRNAs depend on Hfq in Gram-negative bacteria due to their weak interactions with target mRNA. Hfq is known as an RNA chaperone that binds to RNA and regulates the level of translation and/or RNA stability (Vogel & Luisi, 2011). Hfq functions as the core component of a global post-transcriptional network by facilitating interactions between small regulatory RNAs and target mRNAs. (3) Less common, known mechanism of ncRNA regulation includes regulation of protein targets by direct interactions (Suzuki et al, 2006). In this case, ncRNAs regulate an RNA-binding protein that contains a specific protein recognition site. Through this mechanism, ncRNAs can inhibit or activate proteins that have enzymatic activity (Willkomm & Hartmann, 2005). However, many important mechanisms for ncRNAs regulating protein activity still remain unknown and poorly described.

It has been known that small regulatory RNAs are also involved in environmental stresses such as nutrient stress, pH or temperature change, membrane stress, oxidative stress, quorum sensing and SOS response to DNA damage (Gottesman et al, 2006)

(Georg & Hess, 2011b). When cells encounter environmental changes, regulatory sRNAs help to modulate gene expression by optimizing cellular metabolism for survival. In response to stress conditions, ncRNAs act as powerful controllers by interacting complementarily with multiple target mRNAs or by switching the expression of genes in the stress response network (Wassarman, 2002). For example, *OxyS* RNA has been characterized as a small regulatory RNA that can be induced in response to oxidative stress in *E.coli. OxyS* RNA act with Hfq to regulate the translation of its target genes: *fhlA* and *rpoS*, which is a transcriptional activator of hyp operon (required for the synthesis of three hydrogenase isoenzymes) and a stationary phase sigma factor, respectively (Altuvia et al, 1997). Furthermore, it has been elucidated that sigma factor RpoS, a major stress regulator, which controls cellular response to various stresses, is regulated at the post-trascriptional level by a few sRNAs in *E.coli* (Battesti et al, 2011; Repoila et al, 2003). Another example is that while natural ncRNAs have been exploited to increase acid tolerance in *E. coli*, AR1 has been shown to depend on the RpoS sigma factor that directly or indirectly regulates about 500 genes (Venkataramanan et al, 2013). Given this dependence on RpoS, the AR1 system has been engineered in *E. coli* to manipulate RpoS levels. Interestingly, simple overexpression of *rpoS* by removal of its natural 5'-UTR does not significantly improve acid tolerance and is not sufficient to fully induce the genes it regulates (Battesti et al, 2011). The inability to use conventional overexpression strategies to increase levels of RpoS is not surprising given that expression of this protein is tightly regulated at multiple levels by a variety of ncRNAs (e.g. DsrA, RprA, and ArcZ) that stimulate rpoS translation through its 5'-UTR (Battesti et al, 2011). This complex phenotype presents an engineering challenge because strategies using localized regulators (i.e. riboswitches and promoters) are limited to the activation of a set of genes within a regulon.

9

The limited exploitation of ncRNAs stems from the challenge that much remains unknown about the networks of mRNAs, proteins, and transcriptional factors that are regulated in response to environmental changes. Importantly, the engineering of ncRNAs to improve survival under acid stress sets a precedent for using natural ncRNA regulators to tune expression of entire sets of pathways. However, unlike this demonstration, the vast majority of ncRNAs have not been characterized, leaving their mRNA targets and mechanisms of action unknown, particularly beyond *E. coli*. A challenge with overexpression approach is the risk of expressing ncRNAs at non-native levels given the variety of pathways that can be negatively impacted. To date, knockouts (or knockdowns) of natural ncRNAs are not widely used as engineering strategies, but have been performed with the motivation of characterizing the functions of ncRNAs as they are discovered. The lack of ncRNA knockouts for strain engineering is not surprising as the nature of ncRNA regulation is dynamic with respect to cellular stresses and growth phases; as such, simple knockouts may not be beneficial in all conditions throughout cellular growth. Additionally, the necessary genetic tools for ncRNA deletions are more complex than plasmid overexpression approaches, limiting the screening of combinatorial effects involving multiple ncRNAs.

1.2.3 Introduction of ethanol producing microorganism *Z. mobilis*

Zymomonas mobilis is a gram-negative bacterium that can efficiently produce ethanol from several carbon sources such as glucose, fructose and sucrose via its special Entner-Doudoroff pathway. *Z. mobilis* grows anaerobically and does not require the controlled addition of oxygen to maintain cell viability at high ethanol concentration (Yang et al, 2009b). Previous reports have indicated that the presence of oxygen during fermentation affects ethanol production by resulting in low ethanol yield. Due to an

increase in inhibitors such as acetaldehyde and acetate under aerobic conditions, anaerobic *Z. mobilis* fermentations facilitate glucose rapidly and grow with increase in ethanol productivity and yield (Lee et al, 2010).

Z. mobilis has a number of desirable characteristics as a biofuel organism (Widiastuti et al, 2011). For instance, *Z. mobilis* can efficiently produce ethanol up to 12% (w/v) from sugar at a faster rate and higher yield than yeast due to different carbohydrate metabolism. *Z. mobilis* utilizes Entner-Doudoroff pathway as a specific mechanism to produce ethanol from carbohydrate compared to the glycolytic pathway for *Saccharomyces cerevisiae*. In addition to high ethanol producing capability with relatively low biomass, its sugar uptake rates and processing are high. Other advantages include: (1) *Z. mobilis* can tolerate up to 16% (w/v) of ethanol (Rogers et al, 2007), (2) *Z. mobilis* is relatively easy to handle for genetic manipulation and as such amenable for developing recombinant strains to enhance ethanol productivity (Thiebaut et al, 2012), and (3) the complete genome sequence of *Z. mobilis* is available for metabolic engineering (Carey & Ingram, 1983).

Recently, *Z. mobilis* has been extensively studied further as a biofuel producing microbe. Despite these beneficial characteristics of *Z. mobilis* for ethanol production, industrial use of *Z. mobilis* has been limited due to its selective use of carbon sources such as glucose, fructose and sucrose (De Graaf et al, 1999). Many studies have focused on engineering *Z. mobilis* strains for utilization of a wider range of carbon sources to produce ethanol during fermentation (Gao et al, 2002). To develop xylose fermenting strain, xylA/B operon and tal and tkt genes from *E. coli* were transferred into *Z. mobilis* (Zhang et al, 1995). Through genome integration and metabolic engineering of xylose fermenting strains, improved strain has been developed to ferment xylose, arabinose and glucose to produce ethanol (Mohagheghi et al, 2014; Zhang M et al, 2007).

To develop robust industrial biofuel strains, utilization of biomass after pretreatment or hydrolysis process is essential. However, products during the process of pretreatment or hydrolysis and final products are often toxic to the cells or inhibitory to the cell growth. Therefore, research strategies to overcome the toxic effect of inhibitory compounds such as acetic acids, furfural, hydroxymethyfurfural, formic acid and phenolic aldehydes have been carried out in *Z. mobilis*. Several strategies have been performed to overcome the impact of toxic products. First of all, strains with improved tolerance to furfural (3g/L) and acetate (7g/L) were generated via adaptive laboratory evolution and mutagenesis approaches (Shui et al, 2015). Evolved strains not only showed higher acetate and furfural tolerance (under 7g/L acetic acid and 3g/L furfural stress condition), but also presented higher ethanol yield under furfural stress condition (95% of theoretical yield compared to 9% in WT strain (Shui et al, 2015). This feature may be achieved via upregulated *adh* and *pdc* gene expression. Other approaches to enhance tolerance and achieve high yield ethanol production are omics-based metabolic engineering methods. Utilizing transcriptomics and proteomics, genes affected by ethanol, acetate and furfural stress at a systems-level have been identified (He et al, 2012; Yang et al, 2014a; Yang et al, 2013). These genes including Hfq (ZMO0347), phenolic aldehyde responsive reductases (ZMO1116, ZMO1696, ZMO1885), himA (ZMO1122), nhaA (ZMO0117); it it thought that these genes represent useful potential targets for metabolic engineering. Taken together, omics approaches combined with proteomics and metabolomics have provided us insights into global stress responses and mechanisms in *Z. mobilis* that can contribute to the engineering of strains with enhanced tolerance to ethanol and improved ability to produce ethanol. Summary of strategies carried out for the improvement stress tolerance and ethanol production in *Z.mobilis* are shown in Table 1.2.

Even though extensive studies have been done to improve tolerance to various compounds, complete systematic investigation associated with carbon utilization, stress response and ethanol production has not been yet completed in this organism. As such, underlying mechanistic studies via genome scale omics approaches are still desirable for engineering and optimizing *Z. mobilis* as a high yield biofuel organism.

Engineering Methods	Genes	Improvement
Transformation	*xylA/B, tal, tkt*	Utilization of pentose sugar
Adaptation	(*adh, pdc*)	Increased acetate and furfural stress tolerance
Microarray/ overexpression, disruption	*hfq*	Increased acetate tolerance
Microarray	Phenolic aldehyde reductase	Increase tolerance to 4-hydroxybenaldehyde, vanillin
Mutagenesis	*nhaA*	Increased acetate tolerance
gTME	*rpoD*	Increased ethanol tolerance
Toluene	*E. coli*	Increased tolerance
Isobutanol	*E. coli*	Increased tolerance

Table 1.2: Summary of strategies for the identification of genes that enhance tolerance or production in *Z. mobilis*.

1.2.4 Exploring regulatory global and local RNAs for strain engineering in *Z. mobilis*

As we discussed above, advances in high-throughput sequencing technology has led to the discovery of various nocoding RNAs in response to environmental stress in different types of microorganisms. These regulatory RNAs are typically differentially expressed under stress conditions and control metabolic network to cope with stress environment. Recently, several computational genome-scale analyses suggest that stress response mechanism and adaptation involve ncRNAs for precise control and

simultaneous regulation of multiple genes (Wassarman, 2002; Widiastuti et al, 2011). As an ethanologenic microorganism, *Z. mobilis* has unique Entner-Doudoroff pathway and energy-uncoupled growth that may provide high ethanol production and tolerance. Heterologous expression of its unique genes, *adh* (alcohol dehydrogenase) and *pdc* (pyruvate decarboxylase) showed increased ethanol production in *E. coli* (Ohta et al, 1991). This can prove that natural metablic mechanisms in *Z. mobilis* can be successfully transferred to other microorganism.   Given that genome sequences for *Z. mobilis* were published and many genetic modification tools are available, *Z. mobilis* has potential to be a model organism in the context of metabolic engineering. Improving our understanding of how RNA regulators function in vivo and in the context of entire networks, combined with molecular tools for reprogramming their natural functions, will likely result in various useful applications. Therefore, we hypothesize that noncoding RNA-mediated regulatory mechanism have naturally evolved to coordinate efficient metabolic changes by up or down regulation of associated genes that function in an interdependent way in response in response to a specific cellular stress. Ultimately, in my work I have exploited stress induced complex metabolic network in *Z. mobilis* to achieve its ethanol tolerance (and production) capabilities. A major goal of my work has been to establish and understand the functional presence of ncRNAs in *Z. mobilis* to ultimately tune their expression for optimal ethanol producing phenotypes. Furthermore, this study will provide a new approach for genome wide engineering to target coordinated regulation of multiple pathways.

1.3 SUMMARY OF RESEARCH OBJECTIVES AND ACCOPLISHMENTS

Chapter 2 of this dissertation describes efforts to discover small RNAs (sRNAs) in response to ethanol utilizing high-throughput RNA sequencing in *Z. mobilis*. When I started my work, there was no reported or experimentally confirmed sRNAs in *Z. mobilis*. Such sRNAs may mediate global cellular regulation upon different levels of ethanol production condition. Aerobic and anaerobic growth conditions were used for the screening of sRNAs due to the different level of ethanol production. In this chapter we sought to identify differentially expressed novel sRNAs under ethanol stress condition. Zms6 and Zms18 showed different level of expression under ethanol stress compared to non-ethanol enriched conditions. Zms4 also showed differential expression under anaerobic condition compared to aerobic condition, where ethanol levels are naturally higher.

In Chapter 3, characterization of the identified sRNAs that were most relevant to our phenotype of interest was performed to confirm their physiological roles under ethanol stress and to identify potential targets. For this purpose, we utilized a combination of computational analysis and RIPseq (RNA immunoprecipitation-sequencing) and RaPID (RNA binding protein purification and identification) techniques (Cloonan et al, 2008; Slobodin & Gerst, 2010). Overexpression libraries and deletion strains were generated for selected sRNAs to confirm the specific association of these sRNAs to the ethanol stress response. Potential target mRNAs were predicted utilizing existing programs (IntaRNA and CopraRNA, (Wright et al, 2014)). These results were compared with the data from our trascriptomic and proteomic analysis.

Chapter 4 describes local regulatory RNAs that are associated with various stresses. Besides global regulation by RNAs, cis-regulatory elements were discovered. Utilizing transcriptomics, candidates were selected and narrowed down and then

examined experimentally. Using an in vivo screening system using GFP, candidate RNA elements were tested their responsiveness to ethanol, acetate and xylose stress. We expected that finding their roles in metabolic network responses to stress to contribute to the engineering of improved strain.

Additionally, Appendix A describes the demonstration of the use of a ribosomal protein to enhance translational yields of hard-to-express protein. In my work, we attributed this effect to channeling mRNAs to active ribosomes. These results parallel the way by which proximal channeling to metabolic enzymes has been used for increasing metabolic yields of target products. Moreover, our results demonstrate successful adaptation of secM-mediated ribosome stalling in vivo for highly targeted ribosome profiling. Overall, this article would be of strong interest to the protein expression and bioengineering communities as well as to the synthetic biology community that is working on identifying useful ribosomal parts for applications in biotechnologies.

Chapter 2

Identification of ethanol responsive small RNAs in *Zymomonas mobilis*

2.1 INTRODUCTION

High tolerance to ethanol is a desirable feature for ethanologenic strains used in industry. Given that ethanol is toxic to cells by inhibiting cell growth and metabolism, production of ethanol itself represents a bottleneck for the industrial use of biological systems (Osman & Ingram, 1985; Stanley et al, 2010). *Zymomonas mobilis* (*Z. mobilis*) is a gram-negative bacterium that can efficiently produces ethanol from several carbon sources that include glucose, fructose and sucrose, via the Entner-Doudoroff pathway (Rogers et al, 2007). In addition, *Z. mobilis* maintains cell viability anaerobically when yielding high levels of ethanol (Yang et al, 2009b). In fact, several reports have indicated that the presence of oxygen during fermentation affects ethanol production due to increased number of inhibitors (e. g. acetaldehyde and acetate) under aerobic conditions (Moreau et al, 1997; Swings & De Ley, 1977). On the other hand, anaerobic growth of *Z. mobilis* can facilitate rapid glucose consumption with increase in ethanol production relative to aerobic fermentation (Bringer et al, 1984; Moreau et al, 1997).

Z. mobilis has a number of desirable characteristics that make it attractive as a biofuel organism (Widiastuti et al, 2011). For instance, *Z. mobilis* can efficiently produce ethanol up to 12% (w/v) from carbohydrates at a faster rate and three to five fold higher yield than yeast (Jeffries, 2005).. In addition to high ethanol producing capability with

---

[1] Cho SH designed the study, performed experiments, analyzed the data and prepared the manuscript.

relatively low biomass, its rates of sugar uptake and processing are also high. Other advantages include: (1) *Z. mobilis* can tolerate up to 16% (w/v) of ethanol, (2) *Z. mobilis* is easy to handle for genetic manipulation and therefore amenable for developing recombinant strains with enhanced ethanol productivity, and (3) the complete genome sequence of *Z. mobilis* is available for metabolic engineering (Carey & Ingram, 1983; Rogers et al, 2007; Yang et al, 2009a; Zhang et al, 1995).

An intriguing aspect of *Z. mobilis* is the potential shifts in metabolism that likely occur as the organism transitions from high to low oxygen, where is the most efficient at accumulating ethanol. In this work, we wanted to examine the potential role of regulatory small RNAs (sRNAs) in this process. These regulators are relatively short in prokaryotes (~50-300 nucleotides) and are not translated (Livny & Waldor, 2007; Wassarman, 2002), although a possibility is that they produce small (functional or nonfunctional) peptides. As such, sRNAs represent a subset of non-coding RNAs that can be both activators and repressors for regulating proteins and mRNAs via a variety of mechanisms. For instance, (1) antisense sRNAs affect translation and mRNA stability of the complementary target gene, and (2) trans-acting sRNAs regulate mRNAs by imperfect base pairing with distal mRNA targets (Aiba, 2007; Gudapaty et al, 2001; Storz et al, 2011). sRNAs have been known to regulate various metabolic pathways under cellular stress conditions such as oxidative stress, ethanol, temperature or pH change (Altuvia et al, 1997; Georg & Hess, 2011b; Gottesman et al, 2006). When cells encounter environmental changes, regulatory sRNAs help to modulate gene expression by optimizing cellular metabolism for survival. Our motivation in this work is rooted by the ubiquitous discovery and validation of these regulatory elements in bacteria using many computational and experimental strategies (Altuvia, 2007; Livny & Waldor, 2007; Sridhar & Gunasekaran, 2013; Tsai et al, 2013). Interestingly, recent data have shown higher expression of Hfq under anaerobic

18

conditions in *Z. mobilis,* with higher ethanol production relative than to aerobic condition (Yang et al, 2009b). Hfq is a conserved bacterial Sm-like family of RNA-binding proteins particularly in Gram-negative bacteria, which can bind sRNAs and their target mRNAs to direct functionality (Vogel & Luisi, 2011) (Sittka et al, 2008). In addition, Hfq has been shown to play an important role in tolerance to multiple biomass pretreatment inhibitors such as acetate, vanillin and furfural (Yang et al, 2010b) in *Z. mobilis*. Collectively, these findings supported our initial hypothesis the possibility that sRNAs play important mechanistic roles under differential oxygen (and thereby ethanol) conditions in this bacterium.

The study of potential sRNA regulation in the context of bacterial strains that are capable of producing and tolerating high levels of biofuels (and precursors) dates back to previous studies. For instance, small RNAs have been confirmed in *Clostridium Acetobutylicum*, another important strain in the production of acetone and biobutanol from carbohydrates (Borden et al, 2010; Chen et al, 2011; Venkataramanan et al, 2013). In the case of *Z. mobilis*, although its genome has been completely sequenced (Seo et al, 2005), most research has focused on describing membrane composition, understanding patterns of gene expression and characterizing lipid composition. In this study, we focus on investigating the potential presence of regulatory sRNAs in *Z. mobilis.* We furthermore characterize the expression of these newly uncovered RNA elements under anaerobic and aerobic conditions (known to result in differential levels of ethanol accumulation).

2.2 MATERIALS AND METHODS

2.2.1 Strains and culture conditions

Zymomonas mobilis ZM4 (ATCC 31821) was cultured in RM media at 30°C (pH 6.0). A single colony was inoculated into 5 ml RM media (Glucose, 20.0 g/L; Yeast Extract, 10.0 g/L; KH_2PO_4, 2.0 g/L; pH 6.0) (Yang et al, 2009b) and cultured aerobically at 30°C overnight. A 1/100 of initial culture was added to 1 L of pre-warmed RM broth and then cultured overnight at 30°C with shaking at 150 rpm. The inoculum was added to each culture so that the initial OD 600nm was around 0.17. Each culture was grown aerobically or anaerobically and then collected at 13hrs (late exponential/early stationary phase) or 26hrs (late stationary phase) post-inoculation as pH was adjusted every 4 hr. The experiments were done in triplicates. For anaerobic culture, media was nitrogen-purged and tightly capped on a completely sealed flask. Cell density was measured at 600nm absorbance using a spectrophotometer (SmartSpec Plus Spectrophotometer, Bio-Rad).

2.2.2 Measurement of glucose and ethanol concentrations

Glucose concentrations were measured using YSI 7100 Multiparameter Bioanalytical System (YSI Life Sciences, Yellow Springs, OH). Ethanol concentrations were measured using the UV-based ethanol assay kit (R-Biopharm, Darmstadt, Germany) according to the manufacturer's protocol. Absorbance 340nm of reaction mixture with collected samples was measured using spectrophotometer.

2.2.3 Total RNA preparation

Total RNA was prepared according to a protocol previously published in (DiChiara et al, 2010) for all the growth conditions tested. Briefly, cells were grown aerobically or anaerobically and collected at 16 hrs after inoculation for Deep

Sequencing. All centrifugation was performed at 4°C. Cells were pelleted and resuspended in 1 ml TRIzol reagent (Invitrogen). Following pelleting, cells were transferred to screw cap tubes containing glass beads (Sigma) and incubated at 25°C for 5 min. Cells were lysed using a mini-beadbeater (BIOSPEC), with 100-s pulses three times. Cells were kept on ice for 10 min between each 100-s treatment. The beads and cellular debris were centrifuged at 4°C for 2 min. The supernatant was transferred to a clean siliconized 2 ml tube. After addition of 300 µl of chloroform: isoamyl alcohol mix (v/v 24:1), the samples were inverted for 15 s, and then incubated at 25°C for 3 min. Then, tubes were centrifuged at 13,000 rpm for 10 min, and the aqueous top phase transferred to a clean siliconized 1.5 ml tube.   Following this step, 270 µl of isopropanol and 270 µl of a mixture of 0.8 M sodium citrate and 1.2 M sodium chloride was added. The samples were mixed well, and then incubated on ice for 10 min. The RNA was pelleted by centrifugation at 13,000 rpm for 15 min. The pellet was washed with 1 ml 95% cold ethanol and centrifuged for 5 min. The pelleted RNA was allowed to air-dry for ∼5 min, and was resuspended in 30 µl RNase-free water (Ambion). RNA concentration was measured by NanoDrop ND-1000 spectrophotometer (Thermo). Samples were stored at –20°C. Total RNAs were validated on 10% urea gel to verify the quality of the RNAs and make sure RNAs did not undergo any degradation.

2.2.4 RNA deep sequencing and data processing

Prepared RNA was quantified and qualified using Bioanalyzer before sequencing. NEBNext® Small RNA Library Prep Set for Illumina® (New England Biolabs Inc.) was used for generating small RNA libraries. Sequencing was performed using Illumina® HiSeq technology with 2*100 run (Genomic Sequencing and Analysis Facility at the University of Texas at Austin). Prior to analyzing sequencing results, the adapter

sequences were trimmed to remove low quality bases at the end of the reads. Data was processed using BWA (Li & Durbin, 2010) and mapped onto *Z. mobilis* ZM4 complete genome (Genbank: NC_006526). The mapped sequencing reads were visualized in Integrated Genome Viewer (IGV) (Robinson et al, 2011).

2.2.5 Computational analysis of predicted sRNA by BLAST

Sequence conservation analysis of intergenic regions was implemented using WU-BLAST ((blastn 2.0MP-WashU (04 May 2006); W. Gish, personal communication)). WU BLAST output was filtered with a PERL script to a stringent threshold of at least 50% query sequence coverage with 50% identity in the conserved region. These parameters were selected according to search criteria that have been developed to analyze the conservation levels of protein-encoding sequences, where the expected level of conservation is much higher. We categorized with genus and outside genus for the data analysis.

2.2.6 Northen Blotting analysis

Small RNA Northern Blotting analysis was performed as described in (DiChiara et al, 2010). Briefly, Northern Blotting Analysis was performed to verify expression of potential sRNAs candidates that resulted from computational predictions and transcriptomic analysis. DNA oligonucleotide probes specific for each candidate sRNA (Table 2.2) were labeled using 20 pmoles of oligonucleotide in a 20 μl kinase reaction containing 25 μM γ-P^{32} ATP and 20 units T4 polynucleotide kinase (NEB) at 37°C for 1 h. Ladder (ΦX174 DNA/*Hinf I* (Promega)), was labeled in the same manner. Total RNA (50ug~100ug) was separated on a 10% denaturing polyacrylamide gel that was then was transferred to a positively charged membrane (Hybond N+, GE Life Sciences) for blotting. Hybridization was performed using Amersham Rapid-hyb buffer (GE

Healthcare), following their recommended protocol for oligonucleotide probes, with a 3hr incubation or overnight incubation at 42°C. After three washing step with washing buffer (5x SSC, 0.1% SDS for the 1st washing and 1X SSC, 0.1% SDS for the 2nd and 3rd washing step), membranes were exposed to a phosphor screen overnight and visualized with a phosphorimager (Typhoon Imager, Amersham Biosciences).

2.2.7 Deep 5' and 3' Rapid amplification of cDNA ends (RACE)

Deep RACE experiments were performed using total RNA samples from both aerobic and anaerobic cultures. 5' Deep RACE was performed using Ion Torrent 316 chip (Wadsworth Center Applied Genomic Technologies Core Facility) as previously described in (Beauregard et al, 2013; Tsai et al, 2013). Briefly, FirstChoice® RLM-RACE kit (Ambion) was used with minor modifications to the protocol. A total 8 ug RNA was treated with Tobacco Acid Pyrophospatase (TAP) at 37°C for 1hr, followed by ligation of the 5' RACE kit adapter at 37°C for 1hr. The resulting RNA was then reverse transcribed according to the manufacturer's protocol. PCR was performed on the resulting cDNA. All primer sequences used for Deep RACE are listed in Table 2.2. To increase the yield of some sRNAs, PCRs were re-amplified using the product from the original reaction as a template and the same primers. Resulting PCR products were purified using QIAquick PCR purification kit (Qiagen) and RNase-free water (Ambion) for final elution. All products were pooled together.

For 3' RACE, a published protocol (Beauregard et al, 2013) was followed, using miScript Reverse Transciption Kit (Qiagen) to perform reverse transcription. PCR was performed on the resulting cDNA. Resulting PCR products were purified using QIAquick PCR purification kit (Qiagen) and eluted in RNase-free water (Ambion). Sequences of all primers used were listed (Table 2.1). All products were pooled together and sequenced

using Ion Torrent 316 chip at the Wadsworth Center Applied Genomic Technologies Core Facility. Data analysis was done using public resources in Galaxy website (http://usegalaxy.org/). To analyze the sequencing results for the 5' and 3' RACE, adapter sequences were first removed for each sample and then sequences lacking 5' or 3' adapter sequences were removed. After analysis of the sequencing results, data was mapped onto the *Z. mobilis* ZM4 complete genome sequence (NC_006526) using Bowtie2 (Langmead & Salzberg, 2012) with default parameters and visualized with IGV (Robinson et al, 2011).

sRNA	Primer sequences
Zms1-5'RACE	CAAGCAGAAGACGGCATACGAACCTTGCCATTGCCGTT
Zms2-5'RACE	CAAGCAGAAGACGGCATACGAATGCCCGTTGTTTCGC
Zms3-5'RACE	CAAGCAGAAGACGGCATACGAATATGGCAATGTTC
Zms4-5'RACE	CAAGCAGAAGACGGCATACGAGCAGAAAACTTCTGA
Zms6-5'RACE	CAAGCAGAAGACGGCATACGATCAACCCCCTTG
Zms8-5'RACE	CAAGCAGAAGACGGCATACGAGGACAGCTCCCTTG
Zms9-5'RACE	CAAGCAGAAGACGGCATACGAAACAGAGCGTCTG
Zms10-5'RACE	CAAGCAGAAGACGGCATACGAGCTAGAATTTTGACT
Zms13-5'RACE	CAAGCAGAAGACGGCATACGATCAATCACGCCGGATG
Zms14-5'RACE	CAAGCAGAAGACGGCATACGAAATTTCTAAGCTGCCT
Zms16-5'RACE	CAAGCAGAAGACGGCATACGAACCGTCCGCCCGA
Zms18-5'RACE	CAAGCAGAAGACGGCATACGAACCACCGAAGCAGT
Zms20-5'RACE	CAAGCAGAAGACGGCATACGATTTCTCGCTTCCTTC
Zms24-5'RACE	CAAGCAGAAGACGGCATACGACAATGCTCATGTC
Zms1-3'RACE	CAAGCAGAAGACGGCATACGAATTTCCGGTAACGGCA
Zms2-3'RACE	CAAGCAGAAGACGGCATACGAATTATGACCGGCGAAA
Zms3-3'RACE	CAAGCAGAAGACGGCATACGAACGGGTGCGCTTGA
Zms4-3'RACE	CAAGCAGAAGACGGCATACGACACGAGCTCAGAAGT
Zms6-3'RACE	CAAGCAGAAGACGGCATACGATATATGTCGAGCAA
Zms8-3'RACE	CAAGCAGAAGACGGCATACGATAAATACATCCAAGGGA
Zms9-3'RACE	CAAGCAGAAGACGGCATACGAAAGACCATTCCAG
Zms10-3'RACE	CAAGCAGAAGACGGCATACGATTCAGAAAATAAAGTC
Zms13-3'RACE	CAAGCAGAAGACGGCATACGAACTTGATTGCCATCC
Zms14-3'RACE	CAAGCAGAAGACGGCATACGAACTGCCGCACAGG
Zms16-3'RACE	CAAGCAGAAGACGGCATACGAGCGATAGTGGAAG
Zms18-3'RACE	CAAGCAGAAGACGGCATACGAGATTTTTGAATGACTGC

24

Zms20-3'RACE	CAAGCAGAAGACGGCATACGAAAGAATAAAAAGAAGG
Zms24-3'RACE	CAAGCAGAAGACGGCATACGATTCAGAACCGGACA
Universal 5'RACE primer	AATGATACGGCGACCACCGAACACTGCGTTTGCTGGCTTTGATG
Universal 3'RACE primer	AATGATACGGCGACCACCGCATGCCGAGGTCGACTTCCTAG

Table 2.1: Primer sequences for 5' and 3' Deep RACE

2.3 RESULTS

2.3.1 Transciptome analysis of *Z. mobilis* for identifying putative small RNAs

A combination of computational and experimental methods was used in this work to identify novel sRNAs in *Z.mobilis*. First, we isolated total RNA from cells cultured under anaerobic conditions (as higher growth rates are observed under these conditions for *Z. mobilis*) and conducted a high-throughput transcriptome sequencing analysis using Illumina Hiseq. Prior to sequencing, RNA quantification and quality assessment was performed via Bioanalyzer. Following mapping of sequencing results to the *Z. mobilis* complete genome (NCBI Reference Sequence: NC_006526.2), all hits were visualized using Integrative Genomics Viewer (IGV, http://www.broadinstitute.org/igv/). The experimental search scheme is outlined in Fig 1A. Importantly, we identified a total of 95 candidates that mostly represented highly expressed transcripts (having over 100 mapped sequence reads; at least 10% of the average number of reads observed in tRNAs). Although we expected that lowly expressed sRNA candidates could also have an important role in regulation (Gottesman, 2005), our initial focus on highly expressed candidates stems from our interest in further confirming expression of these sRNAs via Northern Blotting analysis and in fully mapping the transcript ends. These sRNA characterization techniques are known to be more robust with higher sRNA quantities (Varallyay et al, 2008). It is worthwhile to note that in this first study, we narrowed our search to intergenic sRNA candidates. Our rationale for excluding sequences that even partially overlapped with known open reading frames is that intergenic candidates have

lower possibility of representing fragmented mRNA transcripts or other degradation products.

2.3.2 Computationally predicted sRNAs in *Z. mobilis*

As a complementary technique to sRNA identification in *Z. mobilis*, we used a combination of computational approaches that have proven successful in our previous work (Tsai et al, 2013). Our interest in complementing our experimental search with such approaches stems from the fact that, even though RNA sequencing is a powerful transcriptome analysis technique, it can only capture transcripts expressed during the particular experimental condition under which cells are collected for RNA preparation. It is therefore not surprising that computational predictions have also become widely used for the discovery of small regulatory RNAs in bacteria (Livny & Waldor, 2007; Sridhar & Gunasekaran, 2013). We performed two specific computational prediction approaches to identify novel sRNA candidates in *Z. mobilis*: (1) SIPHT (sRNA Identification Protocol using High-throughput Technologies) (Livny, 2012) and (2) a bioinformatics analysis recently developed in our lab based on the search of long and conserved intergenic regions (Tsai et al, 2015). Using SIPHT, we identified 4 novel sRNA candidates.   As a note, SIPHT predicts intergenic loci in any of the over 1500 bacterial replicons in the NCBI database guided by sequence conservation upstream of putative Rho-independent terminators (Livny, 2012).

In addition to using SIPHT to identify potential sRNA targets, we performed a genome-wide BLAST conservation and size analysis of all 1011 intergenic regions that have not been annotated to be gene encoding in *Z. mobilis* and predicted 20 additional candidates (see Materials and Methods for detailed description). These predictions take

26

advantage of sRNA enrichment trends that we have previously established in long and highly conserved regions of multiple bacterial genomes (Tsai et al, 2015). Results from all bioinformatics are not shown. Collectively, 106 sRNA candidates were identified from computational analysis and experimental approaches (Table 2.2).

#	Start coordinate	End coordinate	Prediction methods	Probes used for Northern blotting	Verified
1	12712	12864	Deep sequencing	F_GGAGGGGGTGAATGATAACAATATT R_CAAAATATTGTTATCATTCACCCC	
2	39493	39517	Deep sequencing	F_TTCAGAAAATAAAGTCAAAATTCTAGC **R_GCTAGAATTTTGACTTTATTTTCTGAA**	Zms10
3	89979	89998	Deep sequencing	F_TATGGCATGAAGTGTCGGACGAT R_ATCGTCCGACACTTCATGCCATA	
4	113035	113048	Deep sequencing	F_AGTCATTAGAGTTTTATAGACGATCTCG R_CGAGATCGTCTATAAAACTCTAATGACT	
5	113238	113285	Deep sequencing	F_TTGTGGACTGTGTTTTGGCCAT R_CCATCCGTCTGTTCACGACCTC	
6	135640	135767	Deep sequencing	F_TTGAAAACGGAGACCGGAATCTT R_TCCGTTTTCAAGTCACAGCACT	
7	138124	138143	Deep sequencing	F_TGCGCTTCTATCATCAGATG R_CATCTGATGATAGAAGCGCA	
8	138860	138909	Deep sequencing	F_AGGCGCTGTACTGCATGATAATG R_TATCTGTCGAAATTGATGCGAAACC	
9	139068	139086	Deep sequencing	F_TATCTTGTATAGACAGATGGAACC R_GGTTCCATCTGTCTATACAAGATA	
10	148289	148313	Deep sequencing	F_ATCCGAGTTGTTCAAGTGATTCGGT R_ACCGAATCACTTGAACAACTCGGAT	
11	149955	150224	Deep sequencing	F_AGGCGCTGTACTGCATGATAATG R_TATCTGTCGAAATTGATGCGAAACC	
12	150332	150457	Deep sequencing	F_AAATGTCAGTCGGGGTTCTGAAG R_CCTGTCTGCACCAGAAACCAAAAA	
13	155335	155386	Deep sequencing	F_CTTATCCGTCGGCCGAAAGCCTTTTTC R_GAAAAAGGCTTTCGGCCGACGGATAAG	
14	157399	157413	Deep sequencing	F_TATTTAGCTCGACAGTTAACGATG R_CATCGTTAACTGTCGAGCTAAATA	
15	157564	157717	Deep sequencing	**F_GGACAGCTCCCTTGGATGTATTTA** R_ACGTTTGGGCGTCAGTGGATATT	Zms8
16	214353	214403	Deep sequencing	F_AACGTCCCTTAATTTTGGGCGTTAT R_AGCCCAATAACGCCCAAAATTAAG	
17	238961	238989	Deep sequencing	F_AAAAGTTTAAGTCTTCAAGGGATTGGG R_CAATCCCTTGAAGACTTAAACTTTTAG	
18	258560	258585	Deep sequencing	F_AAGAATAAAAAGAAGGAAGCGAGAAA **R_TTTCTCGCTTCCTTCTTTTTATTCTT**	zms20
19	282505	282730	Deep sequencing	F_GCCGAGATAAACGTGATGGTTTTG R_GAGAGACATATACCCTGCTGCG	
20	297389	297458	Deep sequencing	F_GTTTTCTGGTGTCAGAGCCGTA R_GCGAAAGCGCTCTTTACGGCTCTGAC	

21	317294	317305	Deep sequencing	F_GCCTTTGATGTCTATTAGACGATCG
				R_CGATCGTCTAATAGACATCAAAGGC
22	340877	340943	Deep sequencing	F_TCTTCACATAGGCGTGCATCCAT
				R_CAAAACATCCACGTCCAGCAAATA

Table 2.2: List of all 106 sRNA candidates from experimental and computational methods and probes used for Northern blotting analysis. Verified sRNAs were detected by probes marked with bold letter.

#	Start coordinate	End coordinate	Prediction methods	Probes used for Northern blotting	Verified
23	354802	354909	Deep sequencing	F_TCAGGATCGATATAGGGTCGATAA R_TTATCGACCCTATATCGATCCTGA	
24	374730	375138	BLAST	F_AGAGGCCATCGGCTATTGGAATCGAA R_TCGATTCCAATAGCCGATGGCCTCTT	
25	405247	405300	Deep sequencing	F_TTTATCCGATGCATCAGGAATAGG R_GATGCATCGGATAAAATCAGATGCC	
26	414951	414992	Deep sequencing	F_AACTTGGACAATCTGCATCGTGGA R_CAAGTTCCGCCATCAGTAATGATG	
27	438333	438374	Deep sequencing	F_TAGAAAGACGAATCTGGCGATCTT R_GCAGGACAAGATCGCCAGATTC	
28	443735	443769	Deep sequencing	F_ATCAGTCTTGGGACAGCAGAATAA R_TTATTCTGCTGTCCCAAGACTGAT	
29	454961	455000	Deep sequencing	F_ACTATATGTCGAGCAAGGGGGTT **R_TCAACCCCCTTGCTCGACATATA**	Zms6
30	512975	513761	BLAST, Deep sequencing	F_ACGGGTGCGCTTGAACATTGCCATAT **R_ATATGGCAATGTTCAAGCGCACCCGT**	Zms3
31	517659	518391	BLAST, Deep sequencing	F_AGTCGCGTGAACAGTTTGAAGTGCGT R_TCAAACGCCTTGAGGCGAATGCGAAT	
32	520890	521168	BLAST, Deep sequencing	F_AATGGCCGCAAATTTGTGCCGGTACT R_AGTACCGGCACAAATTTGCGGCCATT	
33	540051	540556	BLAST	**F_TCAATCACGCCGGATGGCAATCAAGT** R_AGGTTTCCATTACGCAGCAACAGCGA	Zms13
34	558421	558461	Deep sequencing	F_GTTGTATGTCCCCAGCGAATCTC R_ATGGATTGTCGGTCTGCATGAGAT	
35	581581	581637	Deep sequencing	F_TGGATGTTCGTTACAATTCCG R_GTAACGAACATCCATATCGGAAG	
36	594215	594225	Deep sequencing	F_TCCGATTCGTTCTATCGAAAAAGAC R_GTCTTTTTCGATAGAACGAATCGGA	
37	656695	656843	Deep sequencing	F_AAGAAAAACAGGGATGACGGATAA R_TTATCCGTCATCCCTGTTTTTCTT	
38	662037	662105	Deep sequencing	F_ATGGTAAAGGAGTATTGTCATGGAC R_AACGTCCATGACAATACTCCTTT	
39	723950	723973	Deep sequencing	F_ATAGCGTATTTTATCTGTGAACTT R_AAGTTCACAGATAAAATACGCTAT	
40	739580	739603	Deep sequencing	F_TGTCTTAATTGAATGGAGGAGGTAG R_CTACCTCCTCCATTCAATTAAGACA	
41	754710	754734	Deep sequencing	F_CATCGCTCATGCCTTTCAGAACA R_TGTTCTGAAAGGCATGAGCGATG	
42	756943	756990	Deep sequencing	F_GAAATAAAAGAAAGACGAGAATGGC R_GCCATTCTCGTCTTTCTTTTATTTC	
43	818245	818326	Deep sequencing	F_TATAAACTGATTTGAGGGTTTTTAGG R_ATCTATCTCCTAAAAAACCCTCAAATC	

Table 2.2 (cont.): List of all 106 sRNA candidates from experimental and computational methods and probes used for Northern blotting analysis. Verified sRNAs were detected by probes marked with bold letter.

#	Start coordinate	End coordinate	Prediction methods	Probes used for Northern blotting	Verified
44	849470	849583	Deep sequencing	F_CGTCAATGCCTCGGAGGAGATGT R_ACATCTCCTCCGAGGCATTGACG	
45	868916	869028	Deep sequencing	F_GCGATAGTGGAAGTCGGGCGGACGGT **R_ACCGTCCGCCCGACTTCCACTATCGC**	Zms16
46	932692	932710	Deep sequencing	F_GACTGAGTATCTTTGCATTGTCGTAT R_GATACGACAATGCAAAGATACTCAGT	
47	971165	971267	Deep sequencing	F_TGGGCGCGAATAATCGTACTATCT R_CCTGACAGATGACTGGCAACATAT	
48	971281	971394	Deep sequencing	F_CCATTTCCCATAACGTTCATGCATAT R_GGTGTCTGGGACGCTCTTCTT	
49	1006014	1006076	Deep sequencing	F_CCCTGTTTAATAGAGCAAGAGT R_ACTCTTGCTCTATTAAACAGGG	
50	1047694	1047766	Deep sequencing	F_CTCCTGTCCGAAAACAGGAG R_CAAATAACCCTACCCTCCCTTGAAA	
51	1050324	1051409	BLAST,Deep sequencing	F_AATTCCCATGCGCGGAAATGCAGCA R_TTGCTCATGCTGCATTTCCGCGCAT	
52	1062459	1062496	Deep sequencing	F_AAGATGATAGCTGTCAGAAGAGAGTC R_TCTGACAGCTATCATCTTTTTCCTCA	
53	1079429	1080003	BLAST, SIPHT,Deep sequencing	**F_ACCTTGCCATTGCCGTTACCGGAAAT** R_ATTTCCGGTAACGGCAATGGCAAGGT	Zms1
54	1159449	1159486	Deep sequencing	F_GGACTATTAACGCTAGTTCTAAACC R_CTAGCGTTAATAGTCCTTGAGTTTTT	
55	1159549	1159582	Deep sequencing	F_CTAATCTGTGGTGCGCTCTTTTATA R_AGCGCACCACAGATTAGATCGATAT	
56	1223161	1223471	BLAST	F_ATGCCCGTTGTTTCGCCGGTCATAAT R_TTATGACCGGCGAAACAACGGGCATT	Zms2
57	1242629	124398	BLAST, SIPHT	F_ATACTTTGTTCACTGCCGCACAGGCA **R_AATTTCTAAGCTGCCTGTGCGGCAGT**	Zms14
58	1290141	1290169	Deep sequencing	F_ATTAAAAGTAATAACGCCGGAAAAACGTT R_AACGTTTTTCCGGCGTTATTACTTTTAAT	
59	1308714	1308797	Deep sequencing	F_CCAAGATAGATGCTCTCTACAGTGTG R_CACACTGTAGAGAGCATCTATCTTGG	
60	1313597	1313617	Deep sequencing	F_TAAGTGCTTCGGAGAATGTTGTG R_CACAACATTCTCCGAAGCACTTA	
61	1350990	1351016	Deep sequencing	F_CACAGAAAGCAGGGCAAGGAATT **R_GAATTCCTTGCCCTGCTTTCTG**	Zms4
62	1351023	1351047	Deep sequencing	F_GCAGAAAACTTCTGAGCTCGTG R_CTCCACGAGCTCAGAAGTTTTCT	
63	1393080	1393111	Deep sequencing	F_AATAGTGAAATTTTATGAAGGAAGAGA R_TCTCTTCCTTCATAAAATTTCACTATT	
64	1423490	1423533	Deep sequencing	F_AAGCAAAAGGTAGCATCCATGAAG R_CATGGATGCTACCTTTTGCTTTTTA	

Table 2.2 (cont.): List of all 106 sRNA candidates from experimental and computational methods and probes used for Northern blotting analysis. Verified sRNAs were detected by probes marked with bold letter.

#	Start coordinate	End coordinate	Prediction methods	Probes used for Northern blotting	Verified
65	1438391	1438447	Deep sequencing	F_TAAGTCAGGAGACGGTCATTATG R_CATAATGACCGTCTCCTGACTTA	
66	1448672	1449121	BLAST, Deep sequencing	F_TGGCAGGAAAGCGTTGTAGGGCTTAA R_TTAAGCCCTACAACGCTTTCCTGCCA	
67	1461441	1461478	Deep sequencing	F_AAAAGACGGTTCAAAATTCTGAACCG R_ACGGTTCAGAATTTTGAACCGTCTTT	
68	1503403	1503415	Deep sequencing	F_AGGTGATTGAGTAAGCCCCCTT R_AAGGGGGCTTACTCAATCACCT	
69	1510311	1510365	Deep sequencing	F_TCTTTGGAATAGACGAAATGAAATGG R_CCATTTCATTTCGTCTATTCCAAAGA	
70	1514939	1514985	Deep sequencing	F_AATGATATAGAGACCATGCCGAGTTCG R_CGAACTCGGCATGGTCTCTATATCATT	
71	1515822	1515899	Deep sequencing	F_TCAATATTCCCGAAGTCGCATGTGA R_ACTCCATTTCACATGCGACTTCG	
72	1548106	1548378	BLAST	F_TTCGATTCCAGAAATTGTTGATTGCCGTGC R_GCACGGCAATCAACAATTTCTGGAATCG	
73	1564730	1564827	Deep sequencing	F_ATTGTGGTTGCCTTCCTTTGTCAAC **R_AACAGAGCGTCTGGAATGGTCTT**	Zms9
74	1564915	1565010	Deep sequencing	F_GTATTCGTTTGGAAGAACACGCTCT R_AGGACACACTGGATGAGTGGGAAT	
75	1590369	1590908	BLAST,Deep sequencing	F_ACGATTTGCATGTTCACTGCCGCACA R_TGTGCGGCAGTGAACATGCAAATCGT	
76	1598200	1598215	Deep sequencing	F_GCGTTTAAAACCGGACTGTCG R_CGACAGTCCGGTTTTAAACGC	
77	1607580	1607625	Deep sequencing	F_ATTCAGAACCGGACATGAGCATTG **R_CAATGCTCATGTCCGGTTCTGAA**	Zms24
78	1644399	1644415	Deep sequencing	F_GTCAATCCATATAATCGGGGATAGA R_TCTATCCCCGATTATATGGATTGAC	
79	1660400	1660416	Deep sequencing	F_AGCTTAGTTGAAGACGGTCTAGA R_TCTAGACCGTCTTCAACTAAGCT	
80	1666725	1666996	BLAST,Deep sequencing	F_ATTAGATATGGCTGCCGTGTGTCGGT **R_ACCGACACACGGCAGCCATATCTAAT**	Zms15
81	1674564	1674853	BLAST	F_TGACCATAGCGCCTGATCCTGTTGA R_AGGTCAACAGGATCAGGCGCTATGGTCATT	
82	1706340	1706389	Deep sequencing	F_TTGAGGATATTAGAGGCGCCGGT R_TAATAAAGACTGAAGACGACCGG	
83	1728905	1729367	BLAST,Deep sequencing	F_ATCTCCTTAGAAAGGTGAACGGGCCA R_TGGCCCGTTCACCTTTCTAAGGAGAT	
84	1746286	1746307	Deep sequencing	F_AAAAAAGAACAAGAACGCACATAAAAGG R_CCTTTTATGTGCGTTCTTGTTCTTTTTT	
85	1753124	1753198	Deep sequencing	F_AGAAATCGGACTGTTTTATATAAAATCGG R_CCGATTTTATATAAAACAGTCCGATTTCT	

Table 2.2 (cont.): List of all 106 sRNA candidates from experimental and computational methods and probes used for Northern blotting analysis. Verified sRNAs were detected by probes marked with bold letter.

31

#	Start coordinate	End coordinate	Prediction methods	Probes used for Northern blotting	Verified
86	1764445	1764489	Deep sequencing	F_AATCTTAGACGACCATGGCCGC R_ATAAGATCGGGACGTAGAGCCGTT	
87	1765963	1766111	Deep sequencing	F_TCCTACCAGTTGGACGAATCGCAG R_CTGCGATTCGTCCAACTGGTAGGA	
88	1767313	1767351	Deep sequencing	F_ATTCCGAACTATTCAAAGAATAGTAAAAG R_TTACGACTGATCTTTTACTATTCTTTGAAT	
89	1776698	1776780	Deep sequencing	F_ACAATCTTAAAGCTATTCAATCAAAC R_GTTTGATTGAATAGCTTTAAGATTGT	
90	1776846	1776871	Deep sequencing	F_CATTGATCGAGAGGGACAATCATG R_GTCATGATTGTCCCTCTCGATCAA	
91	1840400	1840667	BLAST,Deep sequencing	F_AACTGCATGGCCACTATGGCAAATCC R_AGGATTTGCCATAGTGGCCATGCAGT	
92	1843483	1843493	Deep sequencing	F_TTATTATCTTTGATAAGAAGATTCTGTC R_GACAGAATCTTCTTATCAAAGATAATAA	
93	1846587	1846645	Deep sequencing	F_AAAGCTCTATCGCCCAAAGTAACTA R_GATGAATAGTTACTTTGGGCGATAGA	
94	1897313	1897377	Deep sequencing	F_AAAAGCCAAGCCTTCTAGAGTCCA R_TGGACTCTAGAAGGCTTGGCTTTT	
95	1897632	1897692	Deep sequencing	F_AACCAGCCTGAGCCAACCATCGCGC R_GCGCGATGGTTGGCTCAGGCTGGTT	
96	1901164	1901303	Deep sequencing	**F_ACCACCGAAGCAGTCATTCAAAAATC** R_GATTTTTGAATGACTGCTTCGGTGGT	Zms18
97	1911503	1911546	Deep sequencing	F_ACCAGCCTGAGCCAACCATCGCGC R_GCGCGATGGTTGGCTCAGGCTGGT	
98	1925820	1925844	Deep sequencing	F_CATGATTGATATGCGCCCTGAACCT R_AGGTTCAGGGCGCATATCAATCATG	
99	1935692	1935707	Deep sequencing	F_GGGTTTGGGTATACCAAGTCTCAA R_TTGAGACTTGGTATACCCAAACCC	
100	1982320	1982342	Deep sequencing	F_GCTGGAGATTTTAAATGGCAGCGA R_TCGCTGCCATTTAAAATCTCCAGC	
101	1984025	1984282	SIPHT	F_TGCTTTCCCAAATCTGACCCGGCTT R_AAAGCACCGGCATATGTTGAACTCGCCT	
102	1986644	1987548	BLAST	F_TGGCAGGCTTCGAGCATGAAGTCTTT R_AATTTGATCCACCGATGCCGGCATGT	
103	1991745	1992721	BLAST	F_ATTTCCGTGCATGTGCCTTCGTGCTT R_TCGCTTGCAAAGGCATTGAAGACCGT	
104	1995114	1995513	BLAST	F_ATGATCCTCGCTCTGCTCAACCAGAA R_AAAGCCGGTCTTTCGCTTCTGCGTTT	
105	2011148	2011173	Deep sequencing	F_CATAAATTTCATTTTAGCTTCGGCTG R_CAGCCGAAGCTAAAATGAAATTTATG	
106	2014378	2014388	Deep sequencing	F_CTGAGTATCTTTGCATTGTCGTATC R_TACGACAATGCAAAGATACTCAGTC	

Table 2.2 (cont.): List of all 106 sRNA candidates from experimental and computational methods and probes used for Northern blotting analysis. Verified sRNAs were detected by probes marked with bold letter.

When comparing results from our computational analysis, we found that 3 out of 4 candidates predicted by SIPHT method were also identified in our bioinformatics analysis. However, only 10 of the 85 candidates that were selected from the analysis of deep sequencing data were also predicted computationally. The combined experimental and computational scheme for selecting sRNA candidates is summarized in Figure 2.1A. Figure 2.1B shows the overlap in sRNA predictions from all methods used in this work. Strikingly, one only candidate was identified by all prediction methods; this further highlights the different pools of potential sRNA candidates that were tapped into by these methods.

A

Growth of *Z. mobilis*

↓

Total RNA extraction

↓

Library construction for Deep sequencing

↓

RNA Deep sequencing using Illumina Hiseq

↓

Data analysis and visulization in IGV

↓

95 candidates are selected
(at least ~100 reads)

B

Predicted by SIPHT
1

2

8

Predicted by Blast analysis

1

9

85
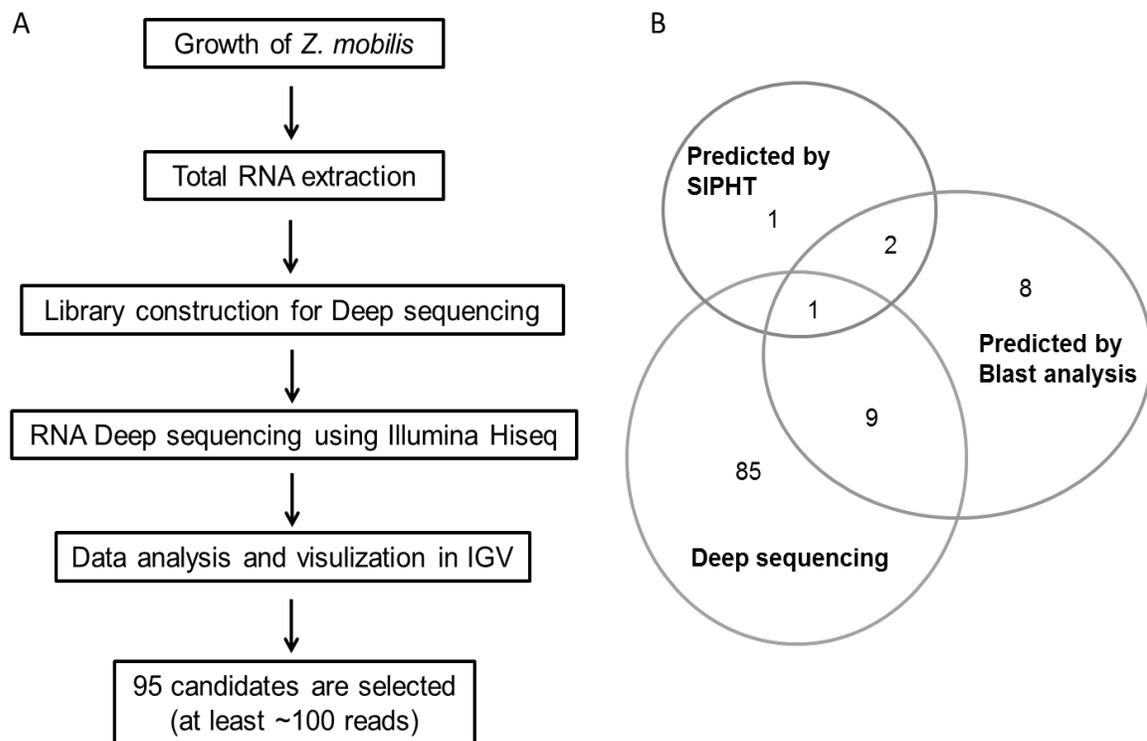
Deep sequencing

Figure 2.1: Experimental scheme for sRNA candidate selection. (A) This schematic shows the strategy for selecting sRNA candidates from deep sequencing methods. (B) Each number of candidate sRNAs from experimental and computational approaches is shown in Venn diagram.

2.3.3 High-throughput validation of sRNAs using Northern Blotting analysis

To validate sRNA expression from the pool of all candidates, we performed a large scale Northern Blotting analysis. Cells were grown anaerobically and collected for RNA extraction in stationary phase given that *Z. mobilis* Hfq has been shown to be more abundantly expressed in anaerobic, stationary phase relative to aerobic stationary phase (Yang et al, 2010b). As Hfq is known as a global sRNA regulator (Vogel & Sharma, 2005), we reasoned that there was a higher chance to identify (and experimentally validate) sRNAs under this condition. A list of all the probes used for Northern Blotting analysis is included in Table 2.1. Given that deep-sequencing data did not provide strand information, sRNAs were probed on both the plus strand and the minus strand. In addition, each candidate was probed with at least two different probes. Importantly, expression of a total of 15 candidates was confirmed with multiple probes, designed to bind different regions of the putative sRNA transcript.

Figure 2.2 and 2.3, shows a summary of all the confirmed sRNAs as well as an image of the positive signal obtained by Northern Blotting analysis using their corresponding probes. Confirmed sRNAs were originally enumerated with a designated "Zms" (*Zymomonas mobilis* sRNA) nomenclature, but they were then annotated according to a published system for bacterial sRNAs (Lamichhane et al, 2013). As indicated in Figure 2.2, 12 of the confirmed sRNAs were identified from the high-throughput sequencing analysis and 3 were identified computationally; 3 sRNAs were found from both prediction methods. It is worthwhile pointing out the presence of multiple bands in some of our samples; these could represent degradation products or several transcription products from the same region. Importantly, the same patterns were observed despite the use of different probes for each region.

34

	5' end	3'end	New nomenclature	Size (nt)	Prediction method	Category
Zms1	1080003	1079429	ncZMO11069Ac	1300	Computation (BLAST, SIPHT) Experiment	overlap with adjacent gene
ZMO1069 — ZMO1070						
Zms2	**1223382**	**1223136**	ncZMO1198Ac	72-118	Computation (BLAST)	overlap with adjacent gene
ZMO1197 — ZMO1198						
Zms3	512975	513362	ncZMO10513A	234-271	Computation (BLAST)	intergenic
ZMO0513 — ZMO2002						
Zms4	**1351044**	1350765	ncZMO11136Ac	271	Experiment	intergenic
ZMO1335 — ZMO1336						
Zms6	**454669**	**454972**	ncZMO10460A	271-310	Experiment	intergenic
ZMO0460 — ZMO0462						
Zms8	**157766**	**157687**	ncZMO0173Ac	200	Experiment	overlap with adjacent gene
ZMO0172 — ZMO0173						
Zms9	**1564777**	**1564963**	ncZMO115135A	118	Experiment	intergenic
ZMO1535 — ZMO1536						
Zms10	**39274**	**39493/ 39703**	ncZMO0037A	603/ 310	Experiment	overlap with adjacent gene
ZMO0037 — ZMO0038						
Zms13	540051	540556	ncZMO10546Ac	500	Computation (BLAST) Experiment	intergenic
ZMO0546 — ZMO0547						
Zms14	**1242778**	1242990	ncZMO11218A	~72/ 194-234	Computation (BLAST,SIPHT)	intergenic
ZMO2011 — ZMO1218						
Zms15	**1666899**	1666996	ncZMO11624A	550/ 1300	Computation (BLAST) Experiment	intergenic
ZMO1624 — ZMO1625						
Zms16	**868928**	**869052**	ncZMO10860A	72-118	Experiment	intergenic
ZMO0860 — ZMO0861						
Zms18	**1901203**	1900964	ncZMOr005Ac	310	Experiment	overlap with adjacent gene
ZMOr005 — ZMOt044						
Zms20	**258449**	**258569**	ncZMO10256A	115	Experiment	intergenic
ZMO0256 — ZMO0257						
Zms24	**1607606**	**1607830**	ncZMO11574A	194-234	Experiment	intergenic
ZMO1574 — ZMO1576						
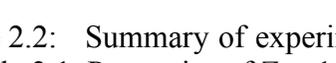
Figure 2.2: Summary of experimentally validated sRNA candidates using probes found in Table 2.1. Properties of Zms1-Zms24 are shown in this figure. The approximate sRNA size observed by Northern Blotting analysis which corresponded with 5' and 3' Deep RACE results. Coordinates with bold character mean that they are verified with 5' and 3'Deep RACE. Other coordinates are from predicted coordinates from computational search or calculated from Northern Blotting analysis. Arrows between coding genes are represented sRNAs and direction of arrows shows orientation of each sRNAs. All prediction methods are shown. Identified sRNAs are classified into two categories: entirely intergenic or overlap with adjacent genes.
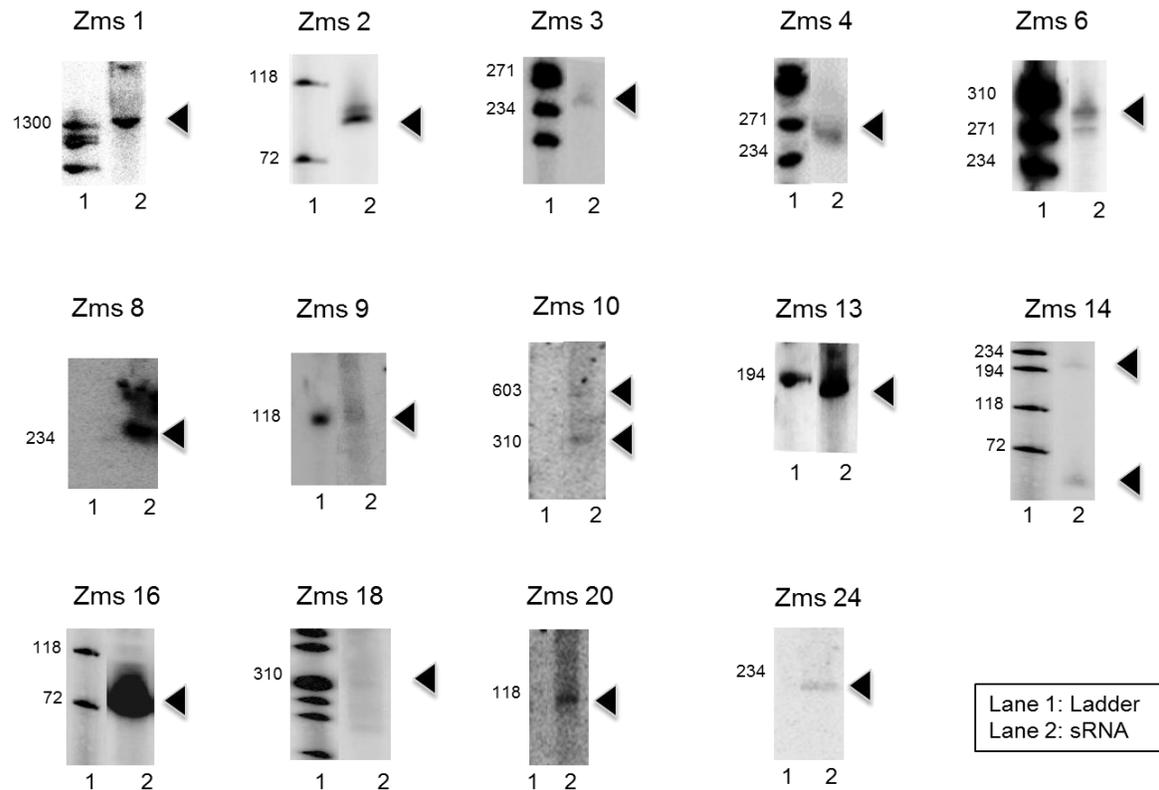
Figure 2.3: Representative Northern blots for confirmed sRNA in *Z. mobilis*. Northern Blotting analysis was performed to examine the expression of candidate sRNAs. Representative blots were confirmed with at least two different probes (Table 2.1). Black triangle indicates sRNA band for each candidate. Lanes are as follows: Lane 1: phiX 174 DNA-Hind III digested ladder, Lane 2: sRNAs

2.3.4 Mapping of transcription start and end site by 5' and 3' Deep RACE

Given that high-throughput transcriptome analysis data does not provide precise information of transcriptional start and end sites, we adapted Deep 5' and 3' RACE analysis for precise mapping of transcript ends and for further confirmation of sequencing results. This method combines conventional RACE technique with deep sequencing technology for the efficient verification of transcription start and end site in sRNA candidates (Beauregard et al, 2013; Olivarius et al, 2009). Coordinates for 5' and 3' end

of each sRNAs from Deep RACE analysis are shown in Figure 2.2. Fig 2.4 shows the data for mapping transcription start and end sites from 5' and 3' Deep RACE for all of sRNA candidates. When comparing the length of confirmed sRNAs, the results are in agreement with previous results confirmed by Northern Blotting analysis.
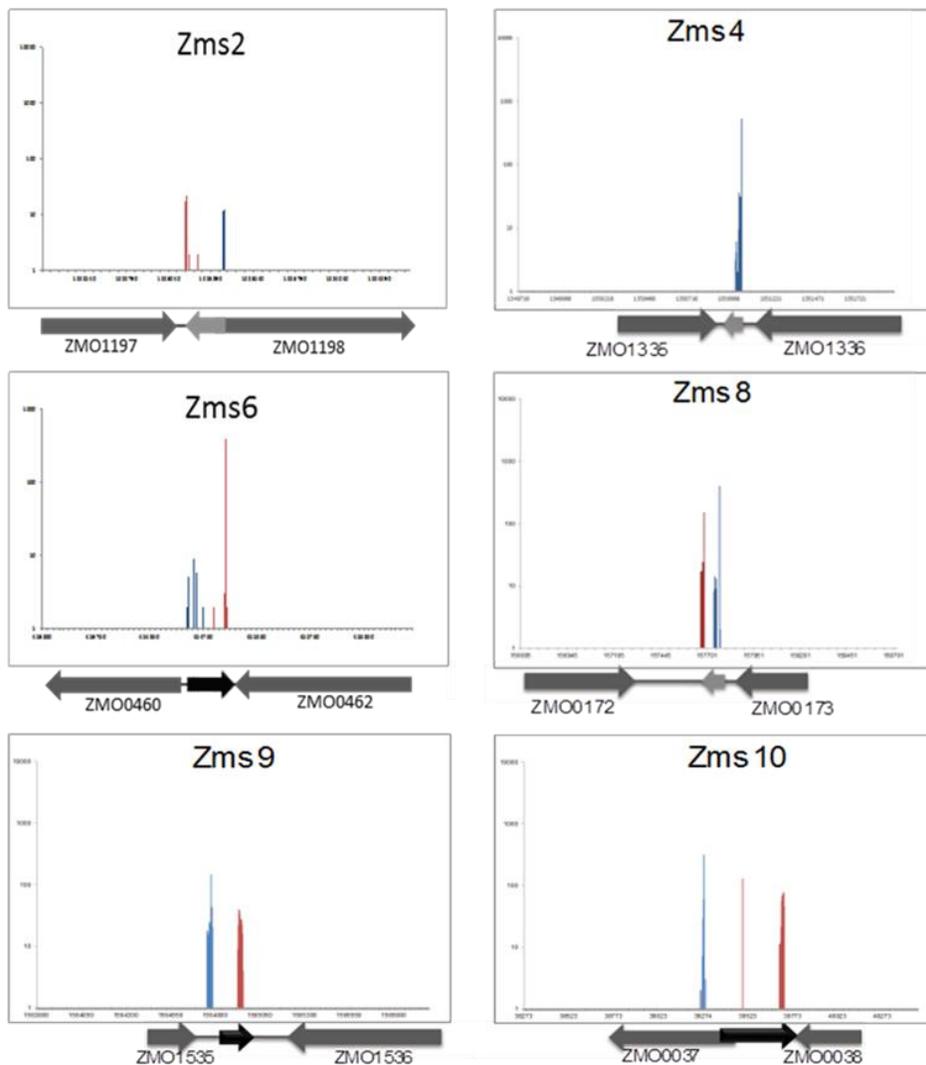
Figure 2.4: Mapping results of sRNAs by 5' and 3' Deep RACE with adjacent genes.
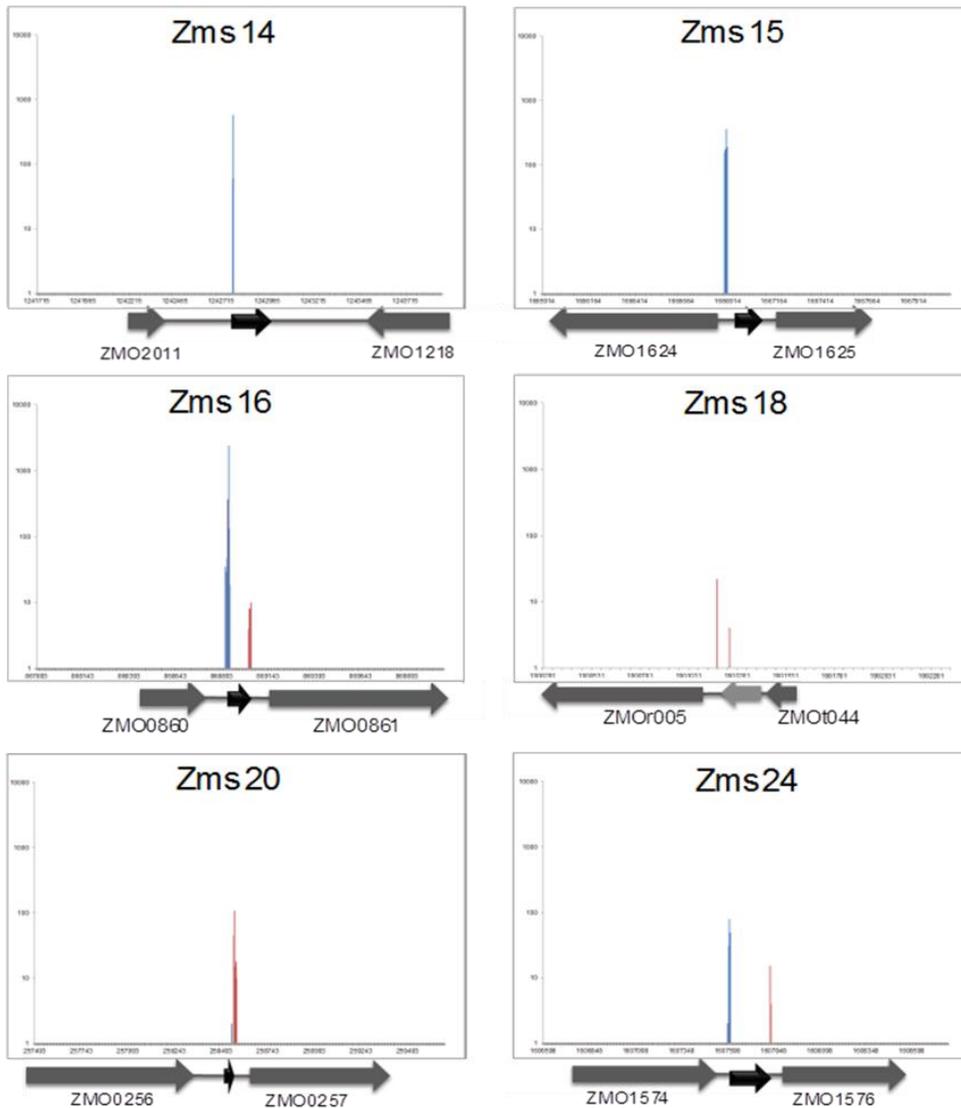
Figure 2.4 (cont.): Mapping results of sRNAs by 5' and 3' Deep RACE with adjacent genes. Blue lines show the number of 5′ RACE reads mapped to respective genome, while red lines show the number of 3′ RACE reads. The black arrow under the chart shows where the sRNA located and the grey arrows represented the adjacent annotated coding regions.

2.3.5 Differential expression of sRNA candidates under aerobic and anaerobic conditions

To understand how expression of the confirmed sRNAs could change under different conditions of ethanol accumulation, we cultured cells under anaerobic and

aerobic conditions. We pursued the analysis of all confirmed sRNAs under both aerobic and anaerobic conditions, as these could be important to basic cellular functions and to the regulation of ethanol production and/or tolerance, respectively. It has been known that the lack of oxygen positively affects glucose consumption, ethanol accumulation and growth in *Z. mobilis* (Yang et al, 2009b). To achieve conditions that show differential production of ethanol, *Z. mobilis* was grown aerobically and anaerobically. As shown in Figure 2.5, the maximal growth rate of *Z. mobilis* under aerobic and anaerobic conditions (estimated as 0.26 h^{-1} and 0.28 h^{-1}, respectively) did not show a significant difference. In addition, we verified established trends in glucose consumption and ethanol production under these conditions. After 26hrs of culture, 84.23mM and 169.74mM of ethanol were measured in aerobic and anaerobic conditions, respectively. As shown in Figure 2.6, glucose is consumed faster under the anaerobic conditions and the corresponding production of ethanol is more rapid under anaerobic conditions. These trends were also consistent with previous published reports (Yang et al, 2009b) and confirmed that the desired culturing conditions were achieved. After screening all confirmed sRNAs, one of the most interesting aspects of this work was the finding that 3 sRNAs (Zms2, Zms4 and Zms6) showed differential expression under aerobic and anaerobic culture conditions (Figure 2.7). Zms2 and Zms6 showed decrease 0.8-fold and 0.64-fold respectively in expression level under anaerobic culture condition. Inversely, Zms4 showed 1.5-fold increased expression in anaerobic culture condition. These results suggested the possibility that these sRNAs could be functionally associated to the metabolic regulation of ethanol production and/or ethanol tolerance.
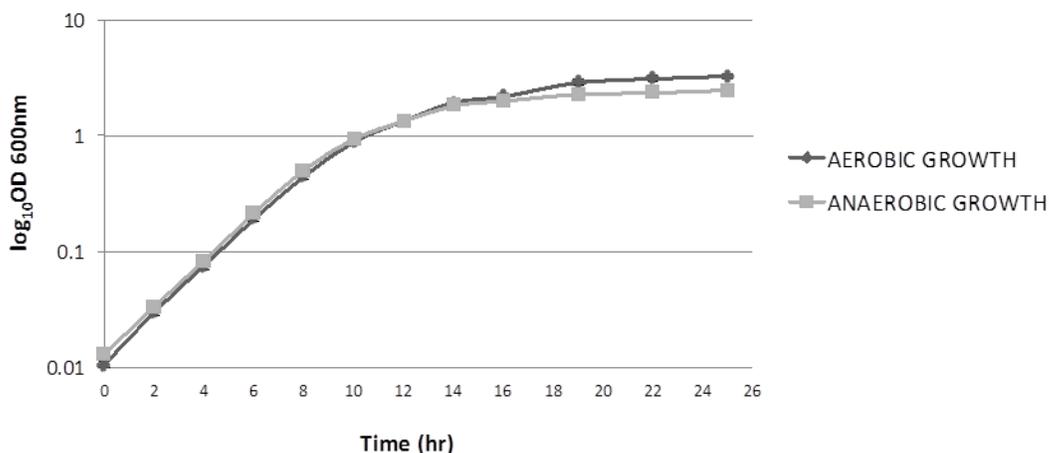
Figure 2.5: Growth of *Z. mobilis* and differential expression of sRNAs under anaerobic and aerobic conditions. Growth curve of each conditions are shown. Mean values for triplicate are shown for each condition ± standard deviation (bars). OD_{600} of cells were measured every 2hrs.
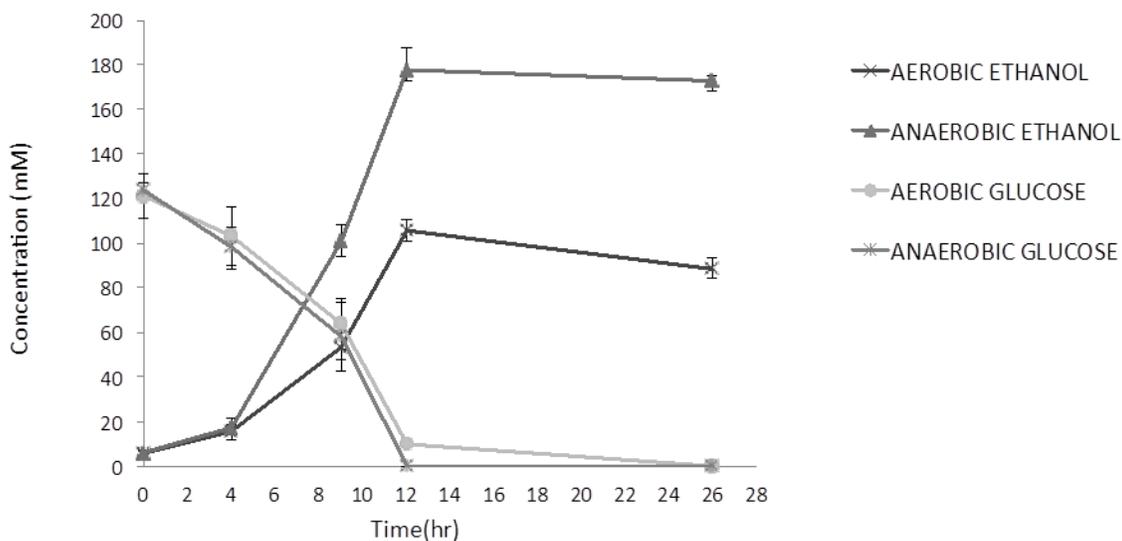
Figure 2.6: Growth of *Z. mobilis* and differential expression of sRNAs under anaerobic and aerobic conditions. Ethanol production and glucose consumption are shown. Mean values for triplicate are shown for each condition ± standard deviation (bars).
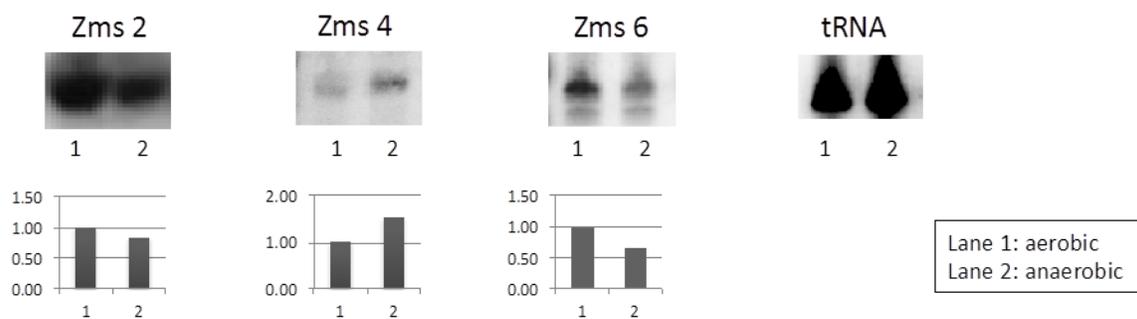
Figure 2.7: The expression of sRNA for each candidate is shown with corresponding intensity of each band detected by Northern blotting analysis. Zms2 (~90bp), Zms4 (~271bp) and Zms6 (~290bp) were shown differential expression between aerobic and anaerobic condition. Band intensities were normalized based on those of tRNA.

2.3.6 Differential expression of sRNA candidates is responsive to environmental growth conditions

After confirming differentially expressed sRNAs under different levels of ethanol, we next, tested directly the effect of ethanol stress on the expression of all identified sRNA. Previous work had shown that coordinated changes in expression of specific heat shock proteins and metabolic enzymes (e.g. alcohol dehydrogenase) are important under high ethanol stress (An et al, 1991; Thanonkeo et al, 2007). This supported the possibility that sRNAs could also be differentially expressed as potential post-transcriptional regulators during high ethanol stress conditions. Therefore, we systematically tested expression levels of all identified sRNA candidates under the ethanol stress conditions. We chose a 5% (v/v) ethanol supplement to the media as ethanol stress conditions given that 6% (v/v) ethanol was previously shown to affect cell viability dramatically (Yang et al, 2013). We confirmed that Zms2, Zms6 and Zms18 showed differential expression under ethanol stress conditions (Figure 2.8). Interestingly, Zms2 and Zms6 also exhibited differential expression between aerobic and anaerobic conditions. In contrast, Zms18

41

only showed differential expression between 0% ethanol supplemented growth conditions and 5% ethanol supplemented growth conditions, indicating its potential involvement in the regulation of the ethanol tolerance in *Z. mobilis*. In case of Zms4, even though it was observed to be expressed at higher levels under anaerobic conditions (relative to aerobic conditions), it was not observed to be differentially expressed between 0% and 5% ethanol stress conditions (data not shown). A plausible possibility is that Zms4 is more involved in managing oxygen stress.

Lastly, all experiments described above were done under conditions of late exponential phase. Given variations in gene expression levels that have been confirmed under different growth phases in Yersinia and Mycobacterium (DiChiara et al, 2010; Koo et al, 2011), we reasoned that functional sRNAs could also be differentially expressed under different growth phases in this bacterium. Interestingly, hfq that is known as RNA chaperone in bacteria (Sittka et al, 2008; Zhang et al, 2003) has been identified in *Z. mobilis* and showed greater expression in anaerobic stationary phase (Yang et al, 2009b). To test for differential sRNA expression as a result of different growth phases, we harvested total RNA samples from cells collected at 13hr post-inoculation (late exponential phase) and 26hr post-inoculation (late stationary phase). Importantly, Zms2 and Zms6 also showed differential expression between early and late stationary phase (Figure 2.9). Both sRNAs accumulate until late exponential phase and then decrease in late stationary phase.
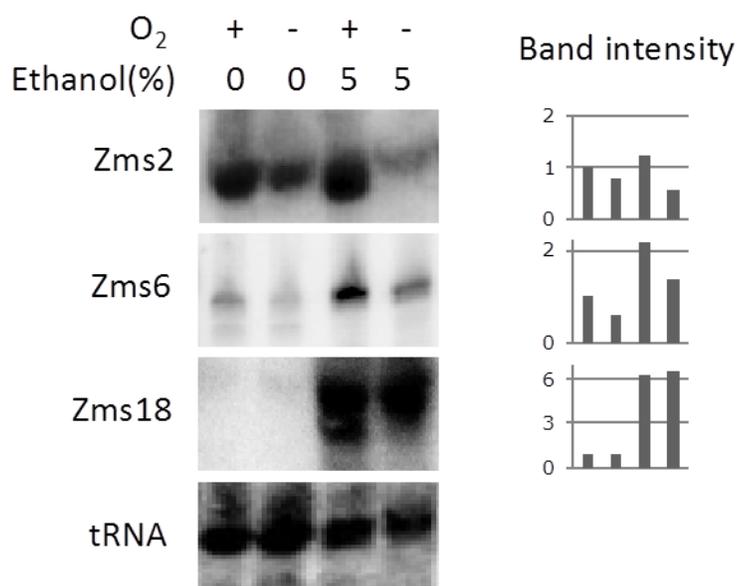
Figure 2.8: Expression patterns of sRNAs under ethanol stress. Addition of ethanol affects the expression levels of sRNAs. First two columns show results from no ethanol supplemented media and last two columns show sRNA expression under 5% ethanol supplemented growth media. All samples are collected under 13hrs (late exponential phase) after inoculation. Aerobic and anaerobic conditions are shown as O_2 +/-, respectively. Band intensities were normalized based on tRNA expression levels.
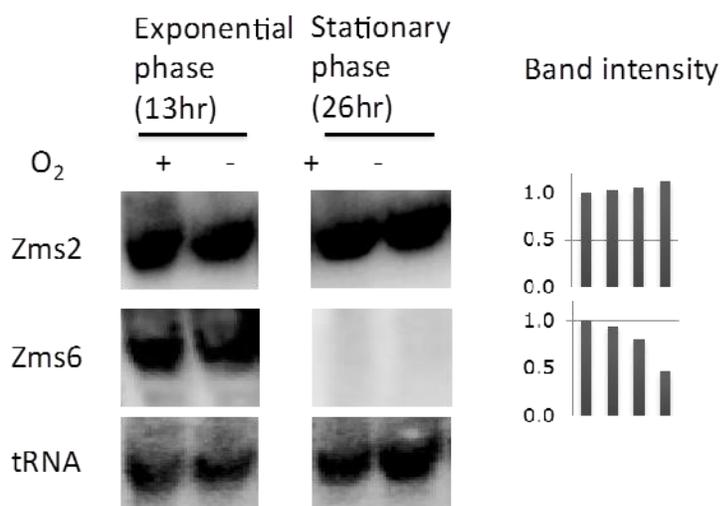
Figure 2.9: Expression patterns of sRNAs under various growth conditions. Different growth phase samples were collected 13hrs (late exponential phase) or 26hrs (late stationary phase) after inoculation for Northern Blotting analysis. Band intensities were normalized based on tRNA expression levels.

2.4 DISCUSSION

Recent research on *Z. mobilis* has unraveled changes in its transcriptomic and metabolic pathways associated with ethanol metabolism. In this study, we successfully discovered 15 novel sRNAs in *Z. mobilis* utilizing experimental and computational approaches. Although 106 candidates selected from our combinatorial methods were tested by Northern Blotting analysis, expression was only confirmed for 15 sRNAs. It is worth noting the possibility that many of candidates, identified by transcriptomic or computational analysis, were below the detectable threshold by Northern Blotting analysis under the experimental conditions used in this study. Compared to mapped reads of tRNAs in deep sequencing, which were mapped with an average of 1500 reads, identified sRNA candidates showed very low read numbers. This could partially explain why we only detected 14% of sRNAs by Northern Blotting analysis even though we used an excessive amount of total RNA (up to 100ug) for detection. Despite high levels of total RNA used for testing, the intensity of some detected sRNAs is very low; this attests to the limitation of Northern Blotting analysis as an experimental tool for sRNAs validation. Furthermore, total RNA, used for deep sequencing and Northern Blotting analysis, was extracted from cells only under limited physiological conditions. This limited number of growth conditions can also explain why some predicted sRNAs are not detected experimentally since they could still be transcribed under different conditions that we have not tested.

Besides the analysis of sRNA candidates from deep sequencing data, we also analyzed expression of mRNAs under aerobic and anaerobic conditions. When we compared our results with published transcriptomic analysis data (Yang et al, 2009b), most of up-regulated and down-regulated genes under aerobic conditions compared to anaerobic conditions were also identified in our study (Table 2.3 and Table 2.4,

respectively). We confirmed that ED pathway genes were more abundant in anaerobic conditions. We also observed several transcription and response regulators are up-regulated in aerobic conditions compared to anaerobic conditions. Additionally, we detected about 200 up-regulated genes and 62 down-regulated genes in aerobic conditions relative to anaerobic condition (Table 2.3 and Table 2.4). Most of newly found genes in our data are related to metabolism and cellular processes. Our analysis identified that alcohol dehydrogenase, which was down-regulate in ethanol treated condition (Yang et al, 2013), was less abundant in late exponential phase under aerobic conditions. We also confirmed that Hfq gene (ZMO0347) was more abundant in anaerobic conditions. Differences in our data relative to previously published microarray data (Yang et al, 2009b), (particularly in the levels of fold-changes detected) could be explained by the increase sensitivity of deep-sequencing methods and by the collection of samples under different growth phases.

It is also worthwhile to point out that although, we initially selected sRNA candidates from the intergenic regions (IGR), upon confirmation by Northern Blotting analysis, and some of sRNAs were detected to be longer than predicted. Our 5' and 3' Deep RACE results further confirmed that some sRNAs overlapped with 5' or 3' end of adjacent genes. Thus, we categorized our identified sRNAs into two groups based on their location: "intergenic sRNA" and "overlapping sRNAs". Intergenic sRNAs are transcribed from intergenic regions between adjacent genes. On the other hand, overlapping sRNAs can be located at the 5' UTRs of adjacent genes to function as riboswitches (Loh et al, 2009) or can also be generated form mRNA post-translational processing if encoded from 3'end of the adjacent gene. It has been known that sRNAs can be transcribed from independent promoter or derived from processing of mRNA UTRs (Kawano et al, 2005; Vogel et al, 2003). There are several evidences that the

sRNAs identified in this study are not fragments of mRNAs: (1) many sRNAs are transcribed in different orientations from adjacent genes and (2) our several Northern blots showed no larger bands that could correspond to pre-processed mRNAs. Even though it is unlikely that any of the sRNAs we identified in this study are generated by mRNA processing, it is well-established that regulatory sRNAs can be derived from processing of mRNA UTRs (Chao et al, 2012; Kawano et al, 2005).

To further characterize the uncovered sRNAs, we confirmed their expression levels under ethanol levels (5%) that have been reported to stress cell growth and decrease ethanol productivity (He et al, 2012). Three sRNAs (Zms2, Zms6 and Zms18) expressed differentially under ethanol stress, suggesting that they could be related to regulatory mechanisms of ethanol production or tolerance in *Z. mobilis*. Analysis of comprehensive comparison with transcriptomic and proteomic data under this condition might be the next step for defining targets of sRNAs to understand regulatory mechanisms. Likewise, we uncovered in our studies, two sRNAs (Zms2 and Zms6) that accumulate until late exponential phase and then decreased in late stationary phase. In *Yersinia*, some sRNAs showed the same pattern of expression under late exponential and stationary phase and these differential levels of sRNA expression correlated with Hfq expression (Koo et al, 2011). Therefore, we speculate that function of these sRNAs might be Hfq-dependent. Further analysis should be performed to understand the role of Hfq in *Zymomonas mobilis*. Lastly, it is noteworthy that preliminary target prediction analysis shows ABC transporter genes as a putative target of Zms2. However, there is no conservation in sequence of Zms2 in other organisms. Ongoing studies are focusing on elucidating the metabolic roles of Zms2 and other differentially expressed sRNAs under ethanol. As part of these future efforts, we are focusing on experimental validation of the targets identified by computational prediction methods.

Primary locus	Product	log2 (aerobic/anaerobic)	Published array
ZMO0311	Pyrroline-5-carboxylate reductase	1.44	2.00
ZMO1853	Dihydrodipicolinate synthase	1.50	1.40
ZMO1792	Dihydroxy-acid dehydratase	1.56	2.40
ZMO0251	Putative translation initiation inhibitor	2.61	1.60
ZMO0746	Aspartyl-tRNA synthetase	0.85	1.40
ZMO1358	Ribosomal protein S20	2.04	2.00
ZMO0209	Ribosomal protein L27	2.72	2.40
ZMO0249	Ribosomal protein L33	1.06	4.00
ZMO1887	Isochorismatase	1.28	1.10
ZMO1022	Cysteine desulfurase	0.91	1.70
ZMO0899	NAD+ synthetase	0.31	3.30
ZMO1879	Delta-aminolevulinic acid dehydratase	1.97	5.10
ZMO1861	Enoyl-(acyl-carrier-protein) reductase II	2.49	3.10
ZMO1489	3-deoxy-D-manno-octulosonate cytidylyltransferase	0.86	2.50
ZMO1088	Putative rare lipoprotein A	2.95	2.50
ZMO0754	SCP-2 sterol transfer family superfamily	3.64	3.10
ZMO1232	Glycosyltransferase	2.58	1.70
ZMO1460	Thiosulfate sulfurtransferase	3.60	1.30
ZMO1863	Putative phosphatase	1.24	4.70
ZMO1496	Phosphoenolpyruvate carboxylase	0.29	1.10
ZMO1347	Threonine aldolase	0.85	2.20
ZMO1851	Flavodoxin	4.49	2.30
ZMO1814	NADH:ubiquinone oxidoreductase subunit	1.98	3.30
ZMO1576	Short-chain dehydrogenase	1.97	2.20
ZMO1688	Glycine cleavage T protein	2.02	1.70
ZMO1287	LPS glycosyltransferase	1.90	4.40
ZMO1089	D-alanyl-D-alanine carboxypeptidase	0.45	1.50
ZMO1904	Metal-dependent protease	1.63	1.20
ZMO1097	Thioredoxin	3.76	1.90
ZMO1118	Glutathione S-transferase family protein	2.15	3.00
ZMO0433	Guanylate kinase	2.66	1.60
ZMO1732	Alkyl hydroperoxide reductase	6.09	4.20
ZMO0084	CheX protein	1.42	6.30
ZMO0641	Chemotaxis protein	2.31	1.70
ZMO0651	Flagellar hook-associated protein 2	0.83	3.20
ZMO0678	Nitroreductase	3.52	1.40
ZMO0964	Probable multidrug resistance lipoprotein	1.86	1.60
ZMO1121	Bacterial regulatory protein, MerR family	2.14	1.30

Table 2.3: List of up-regulated genes under aerobic condtition relative to anaerobic consition identified in this study

Primary locus	Product	log2 (aerobic/anaerobic)	Published array
ZMO1216	Two-component signal transduction histidine kinase	1.44	2.20
ZMO1720	DNA-directed RNA polymerase omega subunit	3.72	2.40
ZMO1063	Sigma 54-dependent transcription suppressor	3.05	2.40
ZMO1622	DNA primase	2.30	1.90
ZMO1387	Two-component response regulator	0.84	2.20
ZMO1216	Two-component signal transduction histidine kinase	1.44	2.20
ZMO0188	Ferric-pseudobactin M114 receptor precursor	2.50	2.60
ZMO0285	RND efflux system lipoprotein	1.15	2.00
ZMO1048	Phosphate ABC transporter permease	2.58	2.20
ZMO1430	Multidrug resistance efflux pump	1.45	1.20
ZMO1437	LysE family transporter	1.91	1.60
ZMO1541	Ferrous iron transport protein B	2.48	3.80
ZMO1463	Probable TonB-dependent receptor	2.28	1.10
ZMO1847	ABC Fe3+ transport system permease	2.80	1.80
ZMO1856	MFS subfamily transporter	1.43	5.10
ZMO1882	Putative transport protein	1.53	3.10
ZMO0112	Hypothetical protein	2.72	1.20
ZMO0319	WGR domain superfamily	0.82	3.10
ZMO0352	Hypothetical protein	0.67	2.90
ZMO0418	Uncharacterized ACR, COG1434 family	1.02	1.20
ZMO0557	Hypothetical protein	2.18	2.30
ZMO0621	Hypothetical protein	2.06	2.00
ZMO0681	Hypothetical protein	1.01	1.50
ZMO0683	Hypothetical protein	0.48	1.20
ZMO0763	TPR Domain domain protein	1.49	2.00
ZMO0786	Dehydrogenase subunit III, putative	0.57	1.30
ZMO0815	Hypothetical protein	1.49	3.20
ZMO1007	Uncharacterized protein family (UPF0187)	0.11	1.50
ZMO1062	Hypothetical protein	3.24	3.90
ZMO1080	Hypothetical protein	2.09	1.00
ZMO1129	Hypothetical protein	1.90	3.30
ZMO1170	Vng6254c	0.71	2.80
ZMO1464	Hypothetical protein	5.05	3.10
ZMO1660	Hypothetical protein	1.87	3.00
ZMO1671	Hypothetical protein	1.99	1.60
ZMO1850	Hypothetical protein	1.87	1.80
ZMO1860	Nodulin 21 like protein	0.35	2.00
ZMO1876	Hypothetical protein	2.59	6.40
ZMO1901	Hypothetical protein	1.84	1.70
ZMO1959	ATPase	1.14	3.40

Table 2.3 (cont.): List of up-regulated genes under aerobic condtition relative to anaerobic condition identified in this study.

Primary locus	Product	log2 (aerobic/anaerobic)	Published array
ZMO0585	Tryptophan synthase beta chain	-0.86	-1.9
ZMO0804	N-acetyl-gamma-glutamyl-phosphate reductase	-0.81	-1.8
ZMO1141	Ketol-acid reductoisomerase	-1.36	-6
ZMO1407	Aspartate-semialdehyde dehydrogenase	-0.20	-1.6
ZMO1891	Threonine synthase	-0.85	-1
ZMO0523	Ribosomal protein L16/L10E	-0.21	-3.8
ZMO0528	Ribosomal protein L5	-0.33	-4.4
ZMO0531	Ribosomal protein L6P/L9E	-0.67	-4.9
ZMO0539	Ribosomal protein S13	-0.29	-1.6
ZMO1910	Ribosomal protein L25	-0.87	-3.7
ZMO1321	Inosine-5-monophosphate dehydrogenase	-0.92	-1.1
ZMO0172	Thiamine biosynthesis protein	-0.33	-1.7
ZMO0889	Aldose 1-epimerase precursor	-0.98	-1.4
ZMO0239	ATP synthase alpha subunit	-1.97	-1.7
ZMO0241	ATP synthase beta subunit	-2.18	-1.8
ZMO1571	Cytochrome bd-type quinol oxidase subunit 1	-0.06	-1.6
ZMO1572	Cytochrome bd-type quinol oxidase subunit 2	-0.01	-1.5
ZMO0152	Pyruvate kinase	-0.31	-1.4
ZMO0367	Glucose-6-phosphate dehydrogenase	-1.67	-3.3
ZMO0369	Glucokinase	-1.42	-3.5
ZMO1240	Phosphoglycerate mutase	-0.08	-4.1
ZMO1478	6-phosphogluconolactonase	-0.05	-3.4
ZMO1608	Enolase	-1.10	-4.8
ZMO1649	Gluconolactonase	-1.60	-1.9
ZMO1719	Fructokinase	-1.90	-3.8
ZMO0347	RNA-binding protein Hfq	-0.39	-1.1
ZMO0732	DNA-directed RNA polymerase beta subunit	-0.40	-2.1
ZMO0366	Glucose facilitated diffusion protein	-1.99	-4.4
ZMO0165	Tol biopolymer transport system	-0.38	-2.3
ZMO0318	Oxidoreductase	-1.91	-3.3
ZMO0022	Fe-S oxidoreductase	-2.09	-2.1
ZMO1844	Probable oxidoreductase	-0.90	-1.9
ZMO1779	Hypothetical protein	-2.02	-6.3
ZMO1609	Hypothetical protein	-0.56	-2.9

Table 2.4: List of down-regulated genes under aerobic conditions relative to anaerobic consitions identified in this study.

This study reinforces the importance of sRNA-associated mechanism for engineering of microbes that are relevant to the production of biofuels. Interestingly, sRNA regulation could also be exploited in the metabolic synchronization of ethanologenic organisms within consortiums. This strategy is already being explored to increase levels of ethanol production involving co-cultures of bacteria and yeast (Zuroff et al, 2013).

Chapter 3

Comprehensive characterization of regulatory small RNAs in *Zymomonas mobilis*

3.1 INTRODUCTION

Current engineering efforts using ncRNAs focus primarily on designing synthetic transcripts to knock down expression of specific mRNA targets, typically by blocking their ribosome binding sites (RBS) (Cho et al, 2015). These targeted knockdowns are useful for optimizing individual pathways, but are limited in addressing complex phenotypes like stress tolerance like many natural ncRNAs are known to do (Wassarman, 2002). Natural ncRNA engineering strategies have been successful, but limited to well-characterized pathways in model organisms. For example, acid tolerance in *Escherichia coli* was improved by 8500-fold when ncRNAs RprA, ArcZ, and DsrA were overexpressed together on a plasmid (Gaida et al, 2013). Similarly, overexpression of ncRNA RyhB in *E. coli* increased production of 5-aminolevulinic acid by 16% (Li et al, 2014). In both cases, the mRNA targets and mechanisms of these ncRNAs had been remarkably well defined such that the effects of the overexpression strategy could be easily foreseen (Battesti et al, 2011; Masse et al, 2007).

Advances in high-throughput sequencing have enabled the discovery of hundreds of ncRNAs across bacteria (Gelderman & Contreras; Tsai et al, 2015), but characterization lags far behind, leaving the vast majority of ncRNAs with functions completely unknown. Mechanistic characterization of these ncRNAs requires low-throughput knockout and overexpression studies, which can be particularly challenging in non-model organisms (Mars et al, 2015; Modi et al, 2011; Papenfort et al, 2008). For metabolic engineers, approaching the large (and growing) pool of ncRNA regulators with

51

an engineering goal in mind has been impractical, not because the ncRNAs necessarily lack power, but because they lack foreseeable roles in producing phenotypes of interest. Importantly, the growth conditions and phenotypes documented in these studies connect strain performance with RNA and protein expression profiles.

We successfully discovered 15 ncRNAs in *Z. mobilis*, with 3 shown responsive in expression to anaerobic or ethanol stresses, representing potential regulators for engineering robustness to these stresses (Cho et al, 2014). In this section, we characterize the effect of sRNAs via overexpression libraries and deletion mutant strains and rank their predicted targets according to their potential to be relevant to ethanol stress using mRNA target prediction program through CopraRNA and IntaRNA (Wright et al, 2014) and existing proteomics data in 6% (v/v) ethanol (Yang et al, 2013). In addidtion, we will aim to exploit transcriptomics and proteomics datasets regarding RNAs and proteins pulling down with sRNAs to predict regulatory networks, characterize phenotypic impacts of predicted ncRNA regulators, and demonstrate the potential of exploiting ncRNAs for strain engineering.

3.2 MATERIALS AND METHODS

3.2.1 Plasmid constructions

To generate sRNA overexpression libraries, we utilized pBBR1MCS2-pgap vector (for constitutive expression) and pEZ_tet vector (for inducible expression). Each sequence confirmed small RNA fragment between NheI and Sall was synthesized from GenScript® and then cloned into pBBR1MCS2-pgap vector, resulting in pBBR1MCS2-pgap-sRNA. For pEZ-tet-Zms4/Zms6, primers containing NcoI and SalI were used for the amplification of Zms4 and Zms6 from pBBR1MCS2-pgap-Zms4/Zms6 constructs.

PCR products were digested and cloned into pEZ-tet, resulting in pEZ-tet-Zms4/Zms6. pBBR1MCS2-pgap and pEZ-tet vectors were used as empty vectors. To construct deletion constructs for Zms4 and Zms6, upstream and downstream fragments (each 1kb) of the target deletion gene were assembled with the spectinomycin gene *aadA* in the middle. LoxP sequences were added outside of *aadA* gene for removing spectinomycin resistance gene for the transcriptomic analysis (Figure 3.2). Upstream and downstream fragments of each target were amplified by PCR using genomic DNA as a template. 1ug of the purified PCR product was directly electroporated (200ohms, 25uF, 1.6kV) into the *Z. mobilis.* Electroporated cells were recovered for 6hrs and plated onto 200 ug/ml spectinomycin containing plates. Plated cells were incubated in anaerobic container with a BD GasPak™ for 3~4 days at 33 °C. Transformants appearing on RM plate with 200 ng/ml of spectinomycin were cultured and screened using PCRs. Colonies with correct PCR product sizes were selected as deletion candidates after sequencing confirmation using the Sanger sequencing. For 2MS2BD-Zms4/Zms6/control constructs, gBlock® (NEB) of 2MS2BD-Zms4/Zm6/control was used for cloning into pBBR1MCS2-pgap vector, resulting plasmids abbreviated 2MS2-Zms4/2MS2-Zms6/2MS2-control. All sequences of primers and constructs used in this study were listed in Table 4.1

3.2.2 Strains and culture conditions

Zymomonas mobilis 8b strain was used in this study (Zhang et al, 1995). *Z. mobilis* 8b strain was cultured in RM media (Glucose, 20.0 g/L; Yeast Extract, 10.0 g/L; KH_2PO_4, 2.0 g/L; pH 6.0) at 33 °C. *Escherichia coli* DH5α was used for plasmid construction and manipulation. Plasmids containing pBBR1MCS2-pgap-sRNA and 2MS2-Zms4/Zms6/control strains were cultured with 350 ug/ml of kanamycin for *Z. mobilis* 8b and with 50 ug/ml for *E. coli*. Overexpression strains containing pEZ-tet-

Zms4/Zms6 were grown with 200 ug/ml spectinomycin for *Z. mobilis* 8b and 50 ug/ml for *E.coli.* For the preparation of the samples for RNA sequencing, each overexpression, deletion, empty and wildtype strain was initially grown in 5ml culture overnight. Then, cells were transferred into 500ml to adjust starting OD_{600nm} at 0.1. Cells were grown at 33 °C for several hrs to reach OD_{600nm} around 0.4. 150ml of cells were collected at this time point for proteomics, transcriptomics and ethanol assay. Then, strains containing pEZ-tet (empty) / pEZ-tet-Zms4 / pEZ-tet-Zms6 were induced with 10 ug/ml tetracycline at $OD_{600nm}= 0.5$. When OD_{600nm} reached around 0.6, 150ml of cells were collected to compare the gene expression profile in the middle of exponential phase. By doing this, the effect of overexpressing Zms4 and Zms6 on gene expression profile can be confirmed by comparing the samples before and after induction. Final samples were collected during stationary phase. Pelleted cells were stored at -80 °C for the further processing.

3.2.3 RNA preparation

Total RNA was prepared according to a protocol previously published in (DiChiara et al, 2010) for all the growth conditions tested. Briefly, cells were grown anaerobically and collected at each time points for RNA Sequencing. All centrifugation was performed at 4°C. Cells were pelleted and resuspended in 1 ml TRIzol reagent (Invitrogen). Following pelleting, cells were transferred to screw cap tubes containing glass beads (Sigma) and incubated at 25°C for 5 min. Cells were lysed using a mini-beadbeater (BIOSPEC), with 100-s pulses three times. Cells were kept on ice for 10 min between each 100-s treatment. The beads and cellular debris were centrifuged at 4 °C for 2 min. The supernatant was transferred to a clean siliconized 2 ml tube. After addition of 300 µl of chloroform: isoamyl alcohol mix (v/v 24:1), the samples were inverted for 15 s, and then incubated at 25 °C for 3 min. Then, tubes were centrifuged at 13,000 rpm for 10

min, and the aqueous top phase transferred to a clean siliconized 1.5 ml tube. Following this step, 270 µl of isopropanol and 270 µl of a mixture of 0.8 M sodium citrate and 1.2 M sodium chloride was added. The samples were mixed well, and then incubated on ice for 10 min. The RNA was pelleted by centrifugation at 13,000 rpm for 15 min. The pellet was washed with 1 ml 95% cold ethanol and centrifuged for 5 min. The pelleted RNA was allowed to air-dry for 5 min, and was resuspended in 50 µl RNase-free water (Ambion). RNAs were digested with DNase I (RNase-free, ThermoScientific) for 1hr at 37 °C to prevent genomic DNA contamination. By adding 0.5mM EDTA to the reaction mixture, samples were heat inactivated at 75 °C for 10mins. Then, RNAs were incubated with isopropanol and GlycoBlue™ (ThermoScientific) at -20 °C overnight. After centrifugation, pelleted RNAs were washed with 95% cold ethanol and centrifuged. RNAs were resuspended in 50 µl RNase-free water (Ambion) and stored at -80 °C for sequencing.

3.2.4 Purification of MS2-MBP fusion proteins

For use as an affinity tag, MS2 coat protein fused with maltose binding protein (MS2-MBP) (Said et al, 2009) was expressed in *E. coli*. 100ml of cells were cultured and induced with 1mM IPTG at OD 0.5_{600nm} for 4 hrs. Cells were pelleted and resuspended in 10ml column buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, 10 mM β–mercaptoethanol pH7.4). 2mM PMSF (phenyl methylsulfonyl fluioride) was added to resuspended cells for preventing protein degradation. After the sonication on ice, DNase I was treated for 1 hr at 4 °C. Cell lysates were centrifuged at 15000 rpm and supernatants (MS2-MBP lysates) were collected. After vortexing and thoroughly suspending amylose magnetic beads (NEB), 200ul of aliquot was washed with 1ml column buffer twice. Entire MS2-MBP lysates were incubated with washed amylose magnetic beads for 2~3

hrs at 4 °C. Then, magnet was applied and supernatants were decanted. Beads were washed with 1ml wash buffer (column buffer + 0.1mM maltose) three times. 50ul of elution buffer (column buffer + 10mM maltose) was added to beads for the elution of MS2-MBP and incubated for 15 minutes at 4°C. By applying magnet, eluted MS2-MBP fusion protein was collected. To increase the yield, elution step was repeated with 50 ul of elution buffer. Purified proteins were confirmed by SDS-PAGE gel and the concentration was measured using Bradford assay.

3.2.5 Affinity purification

2ug of purified MS2-MBP proteins were incubated with 100ul of total RNAs (500ng/ul) extracted from the cells containing 2MS2BD-Zms4/Zms6/control for 1hr at 4 °C. Washed amylose beads were incubated with 2MS2BD-Zms4/Zms6/control+MS2-MBP complex for 2hrs at 4 °C. Supernatants were removed from the beads by applying the magnet. Beads were washed three times with wash buffer and incubated with 50 ul of elution buffer for 15 mins. Elution step was repeated so that total 100ul of elutions were collected. For the precipitation of RNA, equal volume of isopropanol and 10ul of GlycoBlue™ was added to elution sample and then, incubated overnight at -20 °C. RNAs were pelleted at 15,000 rpm for 15 mins at 4 °C and washed with 1ml ethanol. Air-dried RNA pellet was resuspended in 50ul RNase-free water. RNAs for sequencing were stored at -80 °C.

3.2.6 Protein sample preparation for proteomics analysis

Cell lysates containing 2MS2BD-Zms4/Zms6/control were incubated with 2ug of purified MS2-MBP proteins and purified according to affinity purification protocol. Then, 1ml Trizol was added to eluted samples for protein purification. 1.5ml of isopropanol was added to the phenol-chloroform layer and mxitures were incubated for

10 mins at room temperature and then, centrifuge at 12,000 g for 10mins at 4 °C to pellet the protein. Pelleted proteins were washed with 2ml of 0.3M guanidine hydrochloride in 95% ethanol and incubated for 20 mins at room temperature. Then, samples were centrifuged at 7500 g for 5 mins at 4 °C. Washing steps were repeated 2 more times. Then, 2ml of 100% ethanol was added to protein pellets and samples were centrifuged at 7500 g for 5 mins at 4°C. Air dried protein pellets were resuspended 3x SDS-loading buffer for SDS-PAGE gel loading for Mass spectroscopy.

3.2.7 Mass spectrometry

Proteins that co-purified with MS2-MBP were precipitated by adding two volumes of cold 20% trichloroacetic acid (ACS reagent, Sigma-Aldrich) and chilling at -20 °C overnight. After thawing and pelleting for 20 min at 15,000 rpm and 4 °C, the protein pellet was washed with 1 mL ice cold acetone (HPLC, Fisher). The proteins were pelleted by centrifugation as before and acetone evaporated at room temperature for 15 min. After resuspension and denaturing in sample loading buffer, proteins were loaded onto an SDS-PAGE mini-gel (5% stacking, 10% resolving) and moved ~3 mm into the resolving gel by electrophoresis at 70 V. The gel was Coomassie stained and total protein bands were excised, stored in destaining solution at 4 °C, and then digested with trypsin. To identify proteins, LC-MS/MS was performed using the Dionex Ultimate 3000 RSLCnano LC coupled to the Thermo Orbitrap Elite with a 2-hour run time at the ICMB Proteomics Facility using published methods. Proteins were searched with Sequest HT in Proteome Discoverer 1.4 using the Uniprot *Zymomonas mobilis* ATCC ZM4 database (date April 27, 2016). The identifications were validated with Scaffold v4.4.1 (Proteome Software) with greater than 99.0% probability with a minimum of two

peptides at 99.0% peptide probability are listed in Table 3.3. In Scaffold, peptide and protein false discovery rates were both calculated as 0.0%.

3.2.8 Transcriptomics data analysis

Prepared RNA was quantified and qualified using Bioanalyzer before sequencing. NEBNext® Multiplex RNA Library Prep Set for Illumina® (New England Biolabs Inc.) was used for generating RNA libraries. Sequencing was performed using Illumina® NextSeq technology with PE 2*150 run (Genomic Sequencing and Analysis Facility at the University of Texas at Austin). All sequenced libraries were mapped to the *Z. mobilis* 8b complete genome (pending publish) using bwa (0.7.12-r1039) (Li & Durbin, 2009).We used three replicate for each sample. Generated sam files were further analyzed using Cuffdiff (v2.2.1 (4237)) (Trapnell et al, 2012) to generate normalized count matrix. Analysis followed the procedures and steps described in the package documentation and unless stated otherwise default parameters were used.

3.2.9 Bioscreen analysis

Strains were grown in loosely capped 5 mL RMG seed cultures with appropriate antibiotic for selection at 33°C for 48 h. Cells were distributed into Bioscreen C (Growth Curves USA, NJ) plates with media such that the initial $OD_{600} = 0.05$ and there were triplicate wells for each combination of strain and medium. The Bioscreen C measured the turbidity with the wideband filter (420-580 nm) every 15 min for 48 h as the cultures grew without shaking at 33°C. The Bioscreen C was operated using EZ Experiment (Norden Logic Oy, Helsinki, Finland) and growth rates were calculated using MATLAB.

Primer name	Sequences
2MS2 gblock	GCTAGCCCTGAGGTAATTATAACCCGGGCCCTATATATGGATCCTAAGGTACCTAA TTGCCTAGAAAACATGAGGATCACCCATGTCTGCAGGTCGACTCCAGAAAACATGA GGATCACCCATGTCTGCAGTATTCCCGGGTTCATTAGATCTGCGCGCGATCTCTAGA
2MS2_fwd	ATGCCGCTAGCCCTGAGGTAATTATAACC
2MS2_rev	ATGTCGATATCGATCGCGCGCAGATCTAATGAACCCGGGAA
Zms2-SalI_rev	TCATGTCGACAAACGAAATTGTCTTTCTCTGAAATCG
NheI-Zms2_fwd	CGATGCTAGCACCAAACCGTAATTGGGGTCGG
Zms3-SalI_rev	TCATGTCGACTGACGGCCCTTACG
NheI-Zms3_fwd	CGATGCTAGCATGCCCATTTAAAACATCATGAATCCATGTC
Zms8-SalI_rev	TCATGTCGACAAATCTACCTTCCTGCCG
NheI-Zms8_fwd	CGATGCTAGCAAAATGGAGCAAGAGGAAAAAGAACGACTG
Zms9-SalI_rev	TCATGTCGACTTCTGATATTAATTACAATACTTAAAAAATTATTTTTACGG
NheI-Zms9_fwd	CGATGCTAGCGTATTGGATGTTTAATAAGCCGAAGCAGTTCAGG
Zms10-SalI_rev	TCATGTCGACAAGGCGATCATCCTCC
NheI-Zms10_fwd	CGATGCTAGCCTTTCAGGCGGACAAAAAAGCCG
Zms13-SalI_rev	TCATGTCGACATCCATCAGAAAAAGAGCCG
NheI-Zms13_fwd	CGATGCTAGCAAAGCCAGTTCAGTTTTGATTGATAAGCTAACAG
Zms14-SalI_rev	TCATGTCGACAGCTGCCTGTGCG
NheI-Zms14_fwd	CGATGCTAGCATATAAGGTCGCTCTTTTGAAGAGCGGC
Zms16-SalI_rev	TCATGTCGACAGAGAGGAAGAATCAACCATAGG
NheI-Zms16_fwd	CGATGCTAGCATAATTTTCCGTTGTTACGGGCTTGCA
Zms20-SalI_rev	TCATGTCGACCCCCTTCCAAAGTTGTTCG
NheI-Zms20_fwd	CGATGCTAGCGTCATACCAGCAAGGCCGCTTAT
UpZms4F	CATGGCATTACCAAACCTGAAATTTTGAAAGTGCG
UpZms4SpeR	ACTTGCTGACCTGCCAATTATTGGGCTGTCAGCTTTTTGGCC
SpeZms4F	TTGGCAGGTCAGCAAGTGCCTGCCCCGATG
SpeZms4DownR	CTTGGACAAAATAAATTACTGGAGCACAGGATGACGCCTAACAAT
DownZms4F	AGTAATTTATTTTGTCCAAGGTGGATTTTTAAACGC
DownZms4R	CCAGATTTTTATAAGGACGCTTTACAATCATACCAC
UpZms6F	GGCCTCTGGCCAGATAATGCCGTTATCGTC
UpZms6SpeR	CGTCATCCTGTGCTCCGTTCAGTTGCAGCAACGGGTTC
SpeZms6F	GGAGCACAGGATGACGCCTAACAATTCATTCAA
SpeZms6DownR	GTCAACCCCCTTGCTCGATTGGCAGGTCAGCAAGTGC
DownZms6F	TCGAGCAAGGGGGGTTGACGCTTACCGTC
DownZms6R	TACCGGATAGCGAAAGATCAAAAATCGCTCTTTT
northern_zms4	CACAGAAAGCAGGGCAAGGAATTCGGA
northern_zms6	CCCGAAAGAATCATAAAAAGACTTTAGTCTTTTAGACCAATCC

Table 3.1: Primers used in this section

59

3.3 RESULTS

3.3.1 Generation of overexpression small RNAs libraries

Previously we verified small RNAs experimentally in *Z. mibilis* (Cho et al, 2014). To confirm the physiological effect of experimentally confirmed sRNAs on cellular growth, we generated sRNA overexpression libraries: Ov_Zms3, Ov_Zms4, Ov_Zms6, Ov_Zms8, Ov_Zms9, Ov_Zms10, Ov_Zms13, Ov_Zms16, Ov_Zms18 and Ov_Zms20. sRNAs were cloned into downstream of pgap promoter which is natural Zymomonas promoter and resulting constructs were transformed into *Z. mobilis* 8b strain. Overexpression of sRNAs was confirmed by Northern blotting analysis and their expression level was compared with that of wild-type (WT) strain. We examined the effect of overexpression of sRNAs on cell growth and validated the ethanol tolerance for each strain under 8% ethanol. Figure 3.1 showed relative growth rate for each Ov_sRNA strain under no ethanol (RMG), 5% and 8 % ethanol stress condition. It is worth noting that Ov_Zms4, Ov_Zms6 and Ov_Zms16 showed slightly higher growth rate in no ethanol condition compared to 8b WT and empty vector containing strain. Interestingly, these strains also showed a lot more increased viability under 5% and 8% ethanol supplemented condition compared to other overexpression strains and WT strains. This data suggests that Zms4, Zms6 and Zms16 may contribute to *Z. mobilis* ethanol tolerance. Further investigation of the combinatorial effect of sRNAs on cell growth and ethanol tolerance will be needed to confirm their role in ethanol tolerance it they may exhibit accumulative effects on ethanol tolerance. Remarkably, Zms4 and Zms6 sRNA showed differential expression under aerobic/anaerobic culture condition or ethanol stress condition, respectively. Taken together, given the significant effect on cell growth in Ov_Zms4 and Ov_Zms6 strain, we chose Zms4 and Zms6 for further mechanistic study. We excluded Zms16 for the generation of the deletion strains, as it did not show

differential expression under any condition. However, overexpression of Zms16 affected on cell growth with high level under ethanol stress, further researches on Zms16 can be used useful for the mechanistic study regarding ethanol stress response.
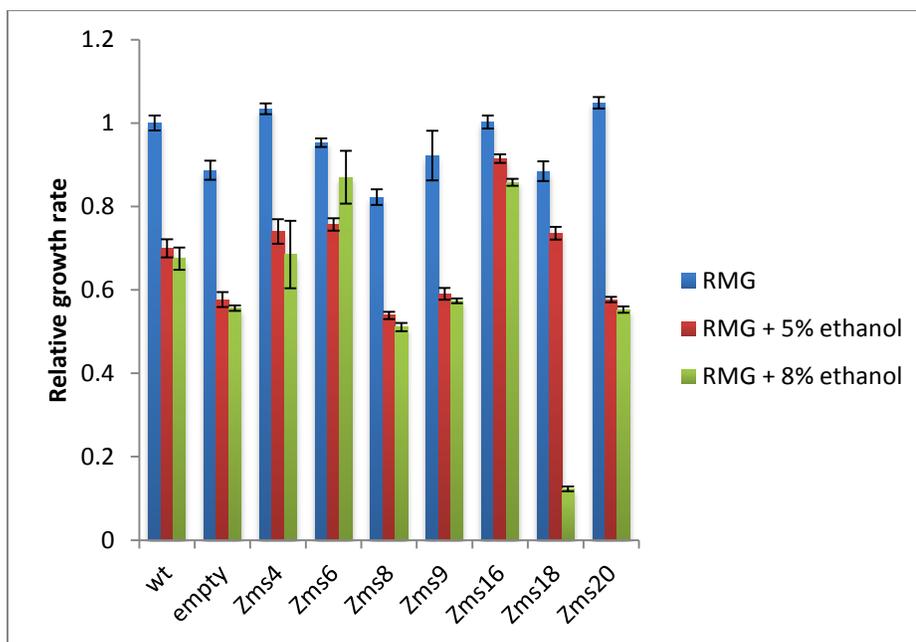
Figure 3.1:Effect of overexpression of sRNAs on growth in RMG media and 5% and 8% ethanol supplemented media. The growth of all strains was monitored by Bioscreen. Relative average growth rates were calculated compared to wt strain in RMG.

3.3.2 Construction of deletion strain for Zms4 and Zms6

To further characterize the direct effect of Zms4 and Zms6, we generated deletion constructs for replacing small RNA regions with spectinomycin antibiotic resistance genes. Deletion strategies were shown in Figure 3.2A. After we generated DelZms4Spe and DelZms6Spe, we activated cre to remove spectinomycin resistance gene between loxP sites resulting in generation of DelZms4ΔSpe / DelZms6ΔSpe strains. Upon successful screening of the deletion on the genome by PCR, Northern blotting analysis was performed to check that sRNA was not expressed (Figure 3.2). Bioscreen analysis

was used for the monitoring the cell growth of DelZms4 (ΔZms4) and DelZms6 (ΔZms6) compared to WT strain. We also compared growth under no ethanol, 5% ethanol and 8% ethanol supplemented conditions. Growth curve were shown in Figure 3.3. Figure 3.4 showed growth rate of each strain for Zms4 and Zms6 under 8% ethanol stress. ΔZms4 and ΔZms6 were showed decreased cell growth rate compared to WT as well as Ov_Zms4 and Ov_Zms6, respectively. Intriguingly, ΔZms6 and Ov_Zms6 strains showed significant decreased (average growth rate = 0.01 h^{-1}) and increased (average growth rate = 0.11 h^{-1}) cellular growth under 8% ethanol stress. This data suggests that Zms6 may play an important role in regulating ethanol tolerance to protect cells from extracellular stress factors.
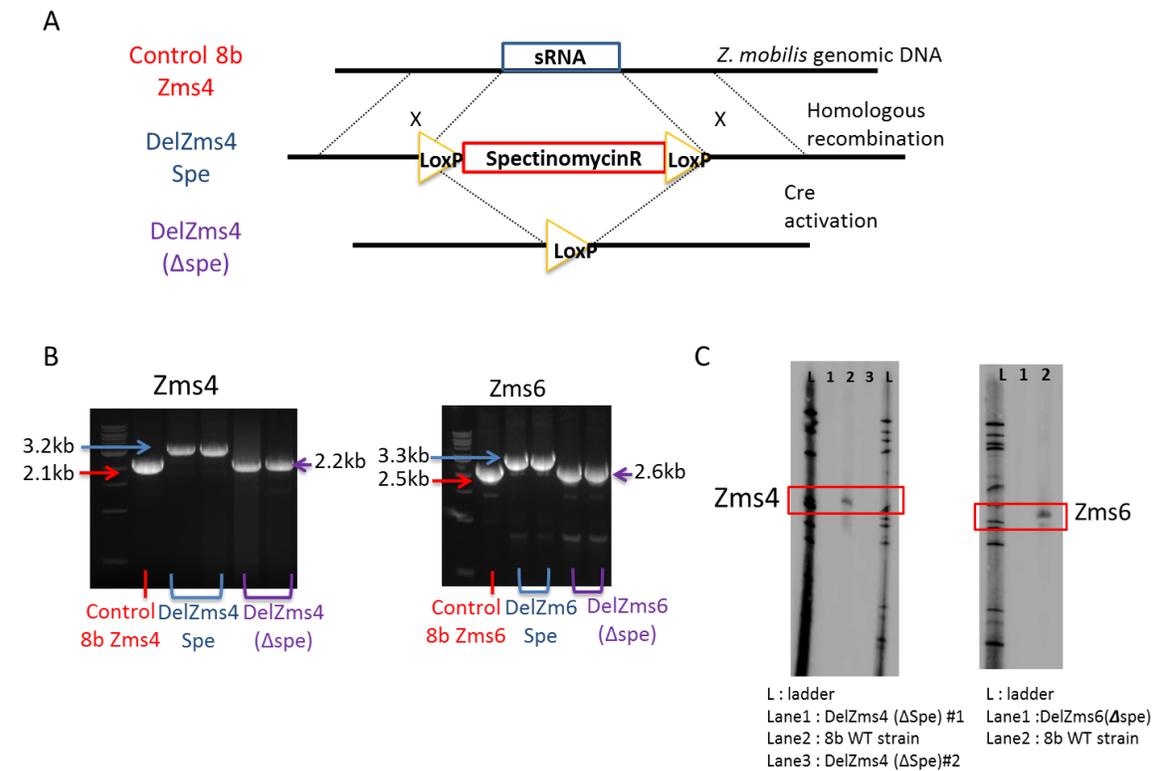
Figure 3.2: Generation of deletion constructs and confirmation. (A) Diagram of deletion strategy (B) deletion confirmation on the genome by PCR and by (C) Northern blotting analysis
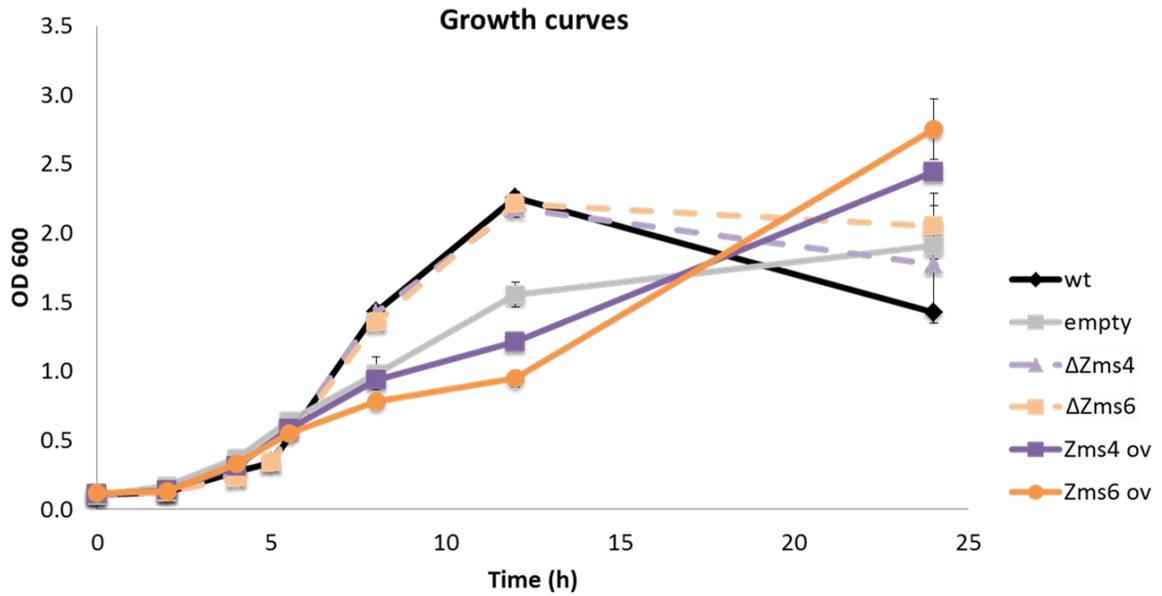
Figure 3.3: Generation of growth curve for deletion and overexpression strains for Zms4 and Zms6 compared with WT strain and empty vector strain. OD_{600nm} was measured at each time point using spectrophometer.
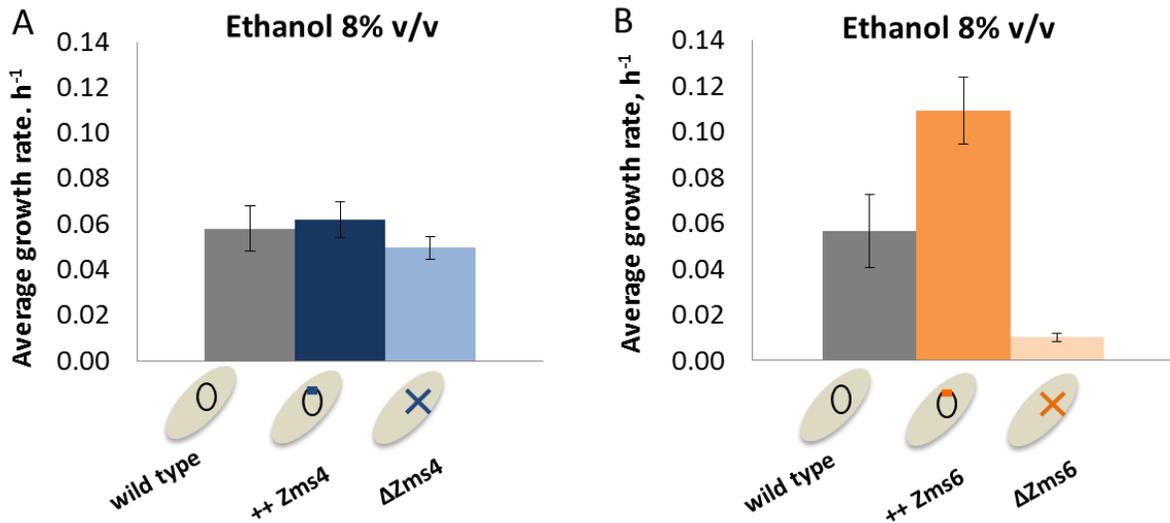
Figure 3.4: Comparison of growth rate of each strain under 8% ethanol stress. Growth of (A) Zms4 and (B) Zms6 strains was monitored by Bioscreen.

3.3.3 Effects of deletion and overexpression of sRNAs on ethanol production

As we validated the association of Zms4 and Zms6 with ethanol tolerance, we investigated ethanol production from mutant strains (overexpression as well as deletion strains). Measurement of ethanol production was performed at each time point (0, 4, 8, 12, 24 hr). For Zms4, Ov_Zms4 produced more ethanol than wt strain when it was in mid-exponential phase. On the contrary, ethanol production level of ΔZms4 was less than that of WT during lag and exponential phase but it was slightly more than that of WT and Ov_Zms4 during stationary phase (Figure 4). In case of Zms6, Ov_Zms6 showed a lot more ethanol production than WT and Ov_Zms4 when it reached around 8hrs of growth from the initial culture (mid-exponential phase). However, ΔZms6 produced less ethanol than WT at all times. This observation suggested that overexpression strains for Zms4 and Zms6 positively affected ethanol production and deletion of Zms6 definitely negativey affected ethanol production. Taken together, it is worth noting that the overexpression of Zms6 leading to an improved ethanol tolerance has also resulted in an increase in the ethanol production. This leads us to confirm Zms6 may play an important regulator of ethanol stress response. Although the efficiency of ethanol production in ethanol tolerant strains is important, most of the toxicity studies rely on the viability analysis of the strains in the presence of ethanol (Lewis et al, 2010; Yoshikawa et al, 2009).
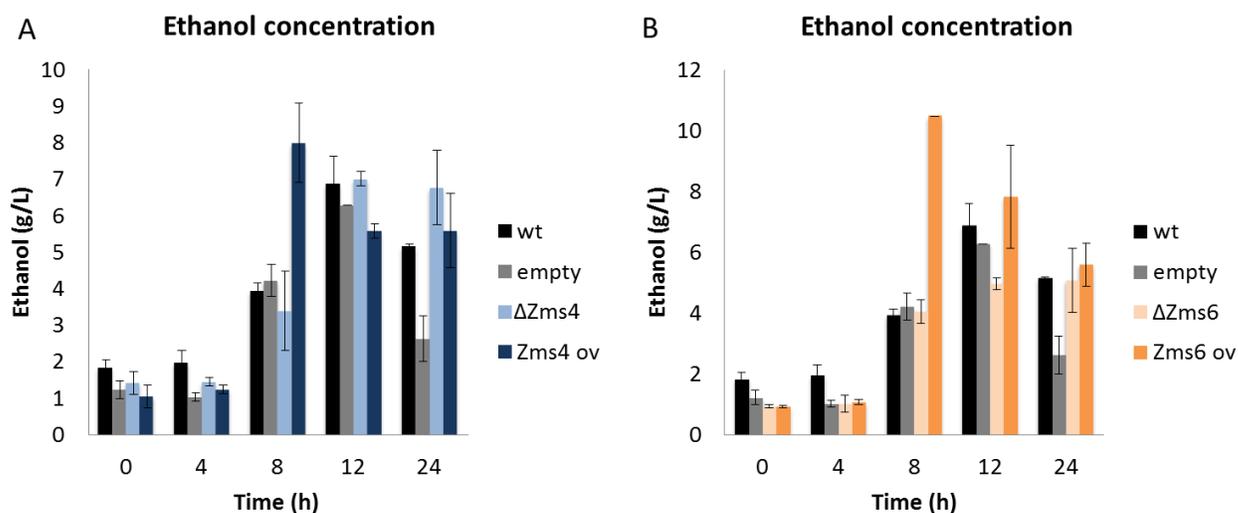
Figure 3.5: Measurement of ethanol production of each strain at the time point of 0, 4, 8, 12, 24hr from the initial culture. (A) Zms4 (B) Zms6

3.3.4 Transcriptomic and proteomic analysis utilizing aptamer-based affinity purification for the identification of sRNA binding targets

Even though we confirmed the role of Zms4 and Zms6 on ethanol production as well as tolerance, their direct mechanism about the control of ethanol in Z. mobilis was still missing part of this study. Therfore, we seek to find their targets for the identification of their metabolic mechanism associated with ethanol. When we recall the known mechanism of sRNAs, cis-encoded or trans-encoded sRNA paired with target mRNAs and resulted in repression or activation of the expression of target mRNAs. We employed RIPseq technique, specifically aptamer-based affinity purification to purify potential sRNA-target mRNA or sRNA-protein complex utilizing bacterio phage coat protein MS2 fused with MBP. sRNAs were tagged with 2MS2 aptamer sequences for binding to MS2-MBP protein. We generated 2MS2BD-Zms4, 2MS2BD-Zms6 and 2MS2-control and transformed into *Z. mobilis*. Purified samples were subjected to either RNA or protein precipitation. Purified RNAs were sent for RNA-sequencing and proteins were used for

mass spectrometry. First of all, we analyzed transcriptome data utilizing cuffdiff (Trapnell et al, 2012) to identify differentially expressed genes and selected genes that showed significant changes (at least more than 2 fold) compared to control. Figure 3.6 showed volcano plots for the transcriptome analysis. We further analyzed selected candidates with the result of target prediction. CopraRNA and IntaRNA, which are mRNA target prediction program, were used for the target prediction of Zms4 and Zms6 and then, predicted target mRNAs were ranked according to their predicted binding energy (Wright et al, 2014; Wright et al, 2013). Table 3.2 and Table 3.3 listed genes with significant changes in gene expression for Zms4 and Zms6, respectively. We excluded hypothetical gene from the table. These genes are candidate target mRNAs that are enriched in co-purified associations with 2MS2-Zms4 or 2MS2-Zms6 relative to 2MS2-control. Genes with bold character exhibited higher rank (top 100) with target prediction analysis.
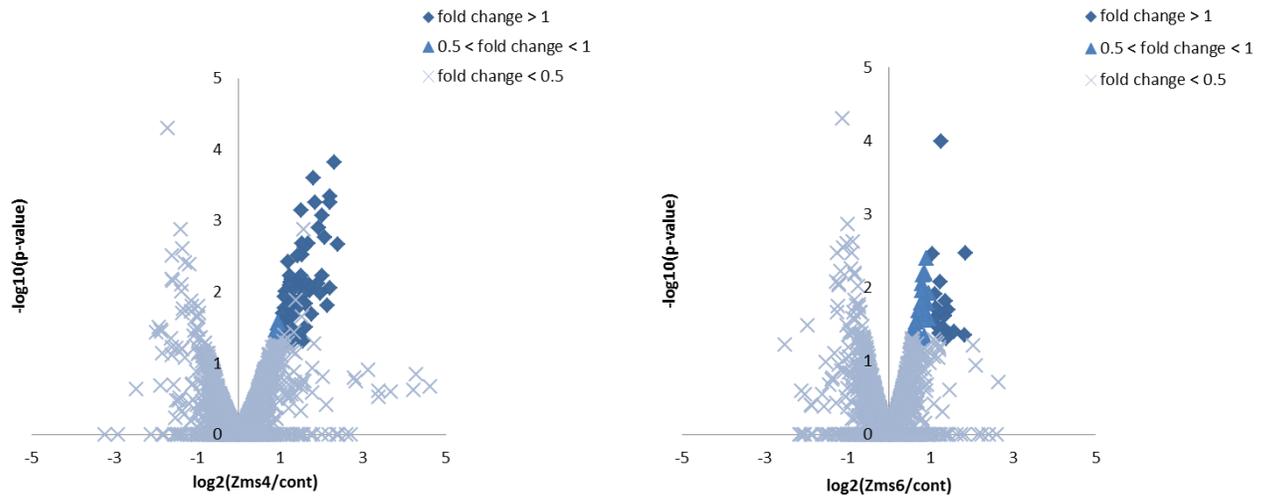
Figure 3.6: Volcano plots for transcriptome analysis

Zms4	Gene	log2 (fold_change)	target prediction rank
transcriptional regulator	ZMO1857	2.30	1221
two-component signal transduction histidine kinase	ZMO1162	2.20	1071
ABC-type nitrate/sulfonate/bicarbonate transportsystem permease component	ZMO1262	2.14	836
NTP pyrophosphohydrolase	**ZMO1041**	**2.07**	**43**
ABC-type transport system ATPase component	ZMO0275	1.97	1597
glutamine amidotransferase	ZMO1855	1.94	833
transposase	ZMO1864	1.90	
putative acetyltransferase	ZMO0733	1.81	549
lysine efflux permease	ZMO1973	1.72	454
Co/Zn/Cd efflux system component	ZMO0204	1.68	368
outer membrane protein	ZMO0798	1.64	1695
molecular chaperone	ZMO0989	1.62	590
signal transduction protein	ZMO0401	1.61	1338
Na+/H+ antiporter	ZMO0119	1.60	207
chemotaxis methyl-accepting protein	ZMO0085	1.59	289
TonB-dependent receptor	ZMO0561	1.56	1205
ABC-type cobalamin/Fe3+ transport systems	ZMO0230	1.56	1303
alginate lyase	ZMO1696	1.54	722
transglycosylase associated protein	ZMO1289	1.54	137
ABC-type multidrug transport system ATPase component	ZMO1029	1.51	277
dihydrodipicolinate synthase	ZMO1853	1.46	1615
mannose-6-phosphate isomerase	ZMO1233	1.43	790
putative phosphatase	ZMO1863	1.40	
1-acyl-sn-glycerol-3-phosphate acyltransferase	ZMO0099	1.40	1427
transcriptional regulator	ZMO1854	1.38	608
putative cation efflux pump	ZMO0214	1.32	1524
homoserine O-acetyltransferase	ZMO0225	1.29	1453
transcriptional regulatory protein	ZMO1574	1.28	945
oxidoreductase	**ZMO0318**	**1.26**	**97**
TonB-dependent receptor	**ZMO0979**	**1.26**	**17**
aminopeptidase N	ZMO1776	1.25	1600
aldose 1-epimerase precursor	ZMO0889	1.23	1249
TonB-dependent receptor	ZMO1298	1.23	781
small-conductance mechanosensitive channel	ZMO0996	1.22	1547
flavodoxin FldA	ZMO1851	1.19	909

Table 3.2: List of target mRNA candidates that are enriched in co-purification with 2MS2-Zms4 relative to 2MS2-control.

Zms4	Gene	log2 (fold_change)	target prediction rank
anthranilate/para-aminobenzoate synthase component I	ZMO0343	1.19	1277
thiamine biosynthesis protein ThiC	ZMO0172	1.18	1275
thioredoxin	ZMO1705	1.18	1529
polypeptide deformylase	ZMO0813	1.16	119
fumarate hydratase	ZMO1307	1.12	1546
LPS glycosyltransferase	ZMO1287	1.11	1539
transcriptional regulator	ZMO0471	1.11	1616
sphingosine kinase	ZMO1428	1.10	1414
multiple antibiotic transporter	ZMO0412	1.08	521
flagellin	ZMO0629	1.08	1353
NADH:ubiquinone oxidoreductase subunit	ZMO1813	1.07	760
DNA topoisomerase IV subunit A	ZMO1054	1.07	583
putative salt-induced outer membrane protein	ZMO1563	1.06	1641
TonB-dependent receptor	ZMO1463	1.06	1555
hemolysin	ZMO0297	1.04	1380

Table 3.2 (cont.): List of target mRNA candidates that that are enriched in co-purification with 2MS2-Zms4 relative to 2MS2-control.

Zms6	Gene	log2 (fold_change)	target prediction rank
1-acyl-sn-glycerol-3-phosphate acyltransferase	ZMO0099	1.85	384
chemotaxis protein	**ZMO0080**	**1.82**	**28**
putative undecaprenol kinase	ZMO1115	1.42	1411
membrane spanning export protein	ZMO0255	1.36	329
hypothetical protein	ZMO0513	1.34	858
multidrug resistance protein	ZMO0697	1.31	229
MFS permease	ZMO1457	1.26	1333
RTX toxin	ZMO0398	1.23	666
putative nicotinamide mononucleotide transporter	ZMO1564	1.21	175
Na+/H+ antiporter	ZMO0119	1.19	810
phytase	ZMO0061	1.13	607
putative 6-pyruvol tetrahydrobiopterin synthase	ZMO0818	1.11	1368
transcriptional regulator	ZMO0471	1.11	227
xanthine/uracil permease family protein	ZMO0969	1.06	1182

Table 3.3: List of target mRNA candidates that that are enriched in co-purification with 2MS2-Zms6 relative to 2MS2-control.

Next, we analyzed proteomics data utilizing RaPID to confirm if any proteins come out with Zms4 or Zms6. Even though protein targets for sRNA is less common, there is still possibility that Zms4 or Zms6 could bind to protein targets for the regulation. Another possibility is that additional proteins can help or bind to sRNA-mRNA complex. We also cannot exclude the possibility that proteins are being translated from mRNA while sRNA and mRNA complex has formed. Table 3.4 showed the analyzed data from proteomics that showed more than 2 fold changes in protein level.

log2 (Zms4/cont)	Gene	Protein name	Nucleotide binding domain
1.00	ZMO0856	Exodeoxyribonuclease 7 small subunit	-
1.08	ZMO0740	CsbD family protein	-
1.09	ZMO1876	Uncharacterized protein	-
1.09	ZMO0407	GcrA cell cycle regulator	-
1.22	ZMO1147	Outer membrane chaperone Skp (OmpH)	-
1.38	ZMO1305	Uncharacterized protein	-
1.42	ZMO0294	50S ribosomal protein L28	-
1.46	ZMO2031	50S ribosomal protein L32	-
1.56	ZMO1076	30S ribosomal protein S16	-
1.58	ZMO0178	**Phosphoglycerate kinase**	Yes
1.74	ZMO0994	Uncharacterized protein	-
1.85	ZMO1609	Uncharacterized protein	-
2.17	ZMO0542	50S ribosomal protein L17	-
2.46	ZMO1690	Chaperone DnaJ domain protein	-
2.58	ZMO2004	30S ribosomal protein S19	-
2.58	ZMO0531	50S ribosomal protein L6	-
2.91	ZMO1358	30S ribosomal protein S20	-
3.00	ZMO1721	Glyoxalase/bleomycin resistance protein/dioxygenase	-
3.17	ZMO0693	OsmC family protein	-
3.46	ZMO0753	Glutaredoxin 3	-
log2 (Zms6/cont)	Gene	Protein name	Nucleotide binding domain
0.62	ZMO0669	ATP synthase subunit b	-
0.89	ZMO0273	**S-adenosylmethionine synthase (AdoMet synthase)**	Yes
1.7	ZMO1657	Uncharacterized protein	-

Table 3.4: Protein candidates that may preferentially bind to Zms4 or Zms6.

GO SUMMARY for proteins co-purifying preferentially with Zms4		Count
structural constituent of ribosome	GO:0003735	7
translation	GO:0006412	7
ribosome	GO:0005840	5
rRNA binding	GO:0019843	3
cytoplasm	GO:0005737	2
phosphoglycerate kinase activity	GO:0004618	1
ATP binding	GO:0005524	1
cell	GO:0005623	1
glycolytic process	GO:0006096	1
DNA catabolic process	GO:0006308	1
protein folding	GO:0006457	1
response to oxidative stress	GO:0006979	1
exodeoxyribonuclease VII activity	GO:0008855	1
electron carrier activity	GO:0009055	1
exodeoxyribonuclease VII complex	GO:0009318	1
protein disulfide oxidoreductase activity	GO:0015035	1
large ribosomal subunit	GO:0015934	1
small ribosomal subunit	GO:0015935	1
cell redox homeostasis	GO:0045454	1
dioxygenase activity	GO:0051213	1

Table 3.5: GO term analysis for proteins copurified preferentially with Zms4.

Upon candidate proteins were selected, we analyzed proteins according to GO trems (Table 3.5). As shown in Table 3.4 and Table 3.5, many ribosomal proteins and structural component of ribosoemd were copurified with Zms4 as expected. Besides ribosomal protein, translation associated proteins, cytoplasmic proteins and chaperone (ZMO1690) were found. Interestingly, oxidative stress response related OsmC family protein (ZMO0693) was also found. Taken together, we can identify possible candidates by pulling down with each sRNA, even though further analysis will be needed for narrowing down the lists to find more reliable targets of sRNAs.

70

3.3.5 Target analysis using transcriptome data of deletion and overexpression strains

As we successfully identified candidate mRNA target lists from experimental and computational methods, we narrowed down of candidate target lists to identify direct target of each sRNA by searching gene expression level in deletion and overexpression strain. Utilizing deletion and overexpression of each sRNA transcriptome data, we searched each candidate target gene if their expression level is affected by sRNA or not. We compared expression level during mid-log phase and early-stationary phase.

We expected many genes to generically change with ethanol stress, and hypothesized that these strains would allow us to decouple effects that were specific to each of our sRNA under study. Specifically, we expected to see opposite trends in the directionality of change in the expression of a true sRNA target if studied in the deletion vs. overexpression strain. For example, if the expression of a potential mRNA target increased in the overexpression strain, we expected it to decrease in the deletion strain. This analysis focuses on the expression of all the genes that we determined to differentially co-purified with Zms4 and Zms6 based on the analysis above.

Table 3.5 and Table 3.6 showed differential expression of candidate target mRNA between deletion and overexpression of sRNA strains. Marked with bold text denoted interesting candidates that expression level was up or down regulated in deletion and overexpression strain in opposite directions. For example, molecular chaperone (ZMO0989) showed less expression in Ov_Zms4 and more expression in ΔZms4. Ton B receptor (ZMO0979) showed increased level in ΔZms4 and decreased in Ov_Zms4. It is note worthing that ZMO0979 ranked 17 in target prediction program, which means that binding energy of ZMO0979 and Zms4 is low so the possibility of binding to each other is high. Flavodoxin *fldA*, which is essential for superoxide response in *E. coli* (Zheng et al,

71

1999), is one of the interesting targets and showed the similar expression pattern with ZMO0979 in ΔZms4 and Ov_Zms4. *FldA* may contribute to oxidative stress response by restoring the redox balance in the cell. For Zms6, transcription regulator ZMO0471 showed differential expression. Gene expression of putative undecaprenol kinase (ZMO0818) was increase in Ov_Zms6 strain and decreased in ΔZms6, even though there is not much information about this gene. ZMO0471 is a transcriptional regulator as LysR family and contains nucleotide-binding domain. Taken together, browsing gene expression level in mutant strains help us to narrow down candidates so that we can revisit regulatory mechanism of sRNA on ethanol tolerance.

Zms4	Gene	Mid-exponential phase		early-stationary phase	
		log2 (Del/wt)	log2 (Ov/wt)	log2 (Del/wt)	log2 (Ov/wt)
1-acyl-sn-glycerol-3-phosphate acyltransferase	ZMO0099	-0.08	-0.13	0.44	0.04
ABC-type cobalamin/Fe3+ transport systems	ZMO0230	-0.23	0.20	0.56	0.34
ABC-type multidrug transport system ATPase component	ZMO1029	-0.54	-0.80	0.09	-0.62
ABC-type nitrate/sulfonate/bicarbonate transportsystem permease component	**ZMO1262**	**-2.68**	**2.46**	-0.21	-1.63
ABC-type transport system ATPase component	ZMO0275	0.72	0.72	0.37	0.32
aldose 1-epimerase precursor	ZMO0889	0.07	-0.58	-0.64	-1.27
alginate lyase	**ZMO1696**	**0.13**	**-0.28**	**0.94**	**-0.84**
aminopeptidase N	ZMO1776	-0.12	-0.22	-0.71	-0.37
anthranilate/para-aminobenzoate synthase component I	ZMO0343	0.18	0.35	0.07	-2.27
chemotaxis methyl-accepting protein	ZMO0085	-0.65	-0.38	-0.15	0.68
Co/Zn/Cd efflux system component	ZMO0204	-0.83	0.25	-0.25	-1.99
dihydrodipicolinate synthase	ZMO1853	0.34	0.55	0.20	-0.17
DNA topoisomerase IV subunit A	ZMO1054	-0.17	0.42	0.01	-0.17
flagellin	ZMO0629	-0.61	0.15	0.14	1.83
flavodoxin FldA	**ZMO1851**	**-0.86**	**0.65**	**0.34**	**-2.10**
fumarate hydratase	ZMO1307	-0.76	0.11	0.42	1.64
glutamine amidotransferase	ZMO1855	0.54	-0.02	0.13	-0.76

Table 3.6: Expression changes in target mRNA candidates as determined from transcriptomic analysis of ΔZms4 and Ov_Zms4 strains relative to wt strain. Each value was gradiently colored from green (increase in mutant strain) to red (decrease in mutant strain).

Zms4	Gene	Mid-exponential phase		early-stationary phase	
		log2 (Del/wt)	log2 (Ov/wt)	log2 (Del/wt)	log2 (Ov/wt)
hemolysin	ZMO0297	-0.43	-0.20	-0.09	0.59
homoserine O-acetyltransferase	ZMO0225	0.10	0.03	0.02	0.94
LPS glycosyltransferase	ZMO1287	-0.13	-0.68	0.39	0.34
lysine efflux permease	ZMO1973	0.93	-0.09	0.14	-0.72
mannose-6-phosphate isomerase	ZMO1233	-0.06	-0.01	-0.04	0.19
molecular chaperone	**ZMO0989**	**0.33**	**-0.51**	**-0.20**	**-6.15**
multiple antibiotic transporter	ZMO0412	0.23	-0.77	0.08	1.19
Na+/H+ antiporter	ZMO0119	0.14	0.46	-0.24	-2.33
NADH:ubiquinone oxidoreductase subunit	ZMO1813	0.22	0.02	0.10	0.49
NTP pyrophosphohydrolase	ZMO1041	-0.05	0.87	-0.44	0.18
outer membrane protein	ZMO0798	-0.16	-0.42	0.33	0.09
oxidoreductase	ZMO0318	0.13	0.48	-0.80	-0.94
polypeptide deformylase	ZMO0813	0.39	-0.71	-0.60	-1.08
putative acetyltransferase	**ZMO0733**	**-0.50**	**0.88**	-0.01	-1.22
putative cation efflux pump	**ZMO0214**	**0.27**	**-0.38**	0.10	-1.00
putative phosphatase	ZMO1863	0.87	0.53	-0.08	-1.94
putative salt-induced outer membrane protein	ZMO1563	-0.44	-0.13	-0.27	0.68
signal transduction protein	ZMO0401	-0.39	0.68	0.76	-1.47
small-conductance mechanosensitive channel	ZMO0996	-0.18	-0.72	0.12	0.38
sphingosine kinase	ZMO1428	0.66	0.06	0.09	-0.91
thiamine biosynthesis protein ThiC	ZMO0172	0.80	0.05	-0.41	-0.85
thioredoxin	ZMO1705	0.90	-0.63	-0.12	-4.80
TonB-dependent receptor	ZMO0561	-0.50	0.13	0.22	-0.88
TonB-dependent receptor	**ZMO0979**	0.47	0.34	**1.02**	**-1.05**
TonB-dependent receptor	ZMO1298	0.17	0.54	0.56	-1.68
TonB-dependent receptor	**ZMO1463**	**-1.37**	**2.44**	0.01	-0.79
transcriptional regulator	ZMO0471	0.40	0.09	0.17	-0.24
transcriptional regulator	ZMO1854	1.21	0.31	-0.01	-0.15
transcriptional regulator	ZMO1857	-0.40	0.35	-0.17	0.48
transcriptional regulatory protein	ZMO1574	0.46	0.72	0.09	-1.32
transglycosylase associated protein	ZMO1289	-0.16	0.76	-0.59	-0.95
transposase	ZMO1864	0.50	0.66	-0.33	-0.68
two-component signal transduction histidine kinase	ZMO1162	-0.06	-0.53	0.17	-2.02

Table 3.6 (cont.): Expression changes in target mRNA candidates as determined from transcriptomic analysis of ΔZms4 and Ov_Zms4 strains relative to wt strain. Each value was gradiently colored from green (increase in mutant strain) to red (decrease in mutant strain).

Zms6	Gene	Mid-exponential phase		early-stationary phase	
		log2 (Del/wt)	log2 (Ov/wt)	log2 (Del/wt)	log2 (Ov/wt)
1-acyl-sn-glycerol-3-phosphate acyltransferase	**ZMO0099**	-0.06	0.68	**0.59**	**-0.54**
chemotaxis protein	**ZMO0080**	**0.15**	**-0.13**	-2.26	-3.25
membrane spanning export protein	ZMO0255	0.83	0.45	0.22	-0.08
MFS permease	**ZMO1457**	-0.71	-1.56	**-0.32**	**0.19**
multidrug resistance protein	ZMO0697	0.14	0.69	-0.49	-0.03
Na+/H+ antiporter	ZMO0119	-0.06	0.41	-0.98	-1.51
phytase	ZMO0061	0.34	0.18	-0.24	0.72
putative 6-pyruvol tetrahydrobiopterin synthase	**ZMO0818**	0.52	0.72	**-0.46**	**0.95**
putative nicotinamide mononucleotide transporter	ZMO1564	-0.26	-1.37	-0.80	-0.88
putative undecaprenol kinase	**ZMO1115**	-0.74	-0.46	**-0.19**	**1.70**
RTX toxin	ZMO0398	-1.75	-0.23	-0.03	-0.70
transcriptional regulator	**ZMO0471**	**0.63**	**-0.41**	-0.41	-0.33
xanthine/uracil permease family protein	ZMO0969	-0.09	0.89	-0.75	-0.46

Table 3.7: Expression changes in target mRNA candidates as determined from transcriptomic analysis of ΔZms6 and Ov_Zms6 strains relative to wt strain. Each value was gradiently colored from green (increase in mutant strain) to red (decrease in mutant strain).

3.4 DISCUSSION

In this section, we successfully proved that Zms4 /Zms6 regulate ethanol tolerance. In addition, overexpression strains of Zms6 greatly improve ethanol production level in the cells. Identification of targets for small RNAs uncovers complex regulatory mechanism in various microorganisms. Our approaches to pull down mRNA or proteins with MS2 coat protein and following omics study revealed promising target candidates. Figure 3.7 showed sketches on mechanistic network of sRNAs utilizing omics data sets. According to schematic logic in Table 3.7, gene expression of candidate target mRNAs (Table 3.2 for Zms4, Table 3.3 for Zms6) was searched in transcriptome data of each

ΔZms4/6 and Ov_Zms4/6 strains. Genes listed in the Figure 3.7 showed more drastic changes (over 1.5-fold) and that may bind to Zms4. Interestingly, these genes are most likely associated with membrane transport. In addition, target mRNAs predicted by transcriptomic (RIPseq) and proteomic analysis (RaPID) were mostly related with membrane transporters, chaperones, reductases, energy metabolism and stress response (Table 3.2 and Table 3.4).

Computational predictions combined with omics studies provide clues for finding targets of sRNAs, which will help demonstration of regulatory mechanism for ethanol tolerance in *Z.mobilis*. Further studies will be followed to confirm direct interaction between sRNA and mRNA in the candidate lists so that final target mRNAs can be elucidated. Even though sRNA and protein binding is less common mechanism, this cannot be excluded. Our proteomics data suggests that list of proteins may preferentially bind to sRNA itself or sRNA-mRNA complex. This also helps to elucidate possible mechanism. Their physiological roles in the metabolic pathway give us insight into the regulatory network in response to stress in *Z. mobilis* and unlock new strategies for engineering robust industrial strains.
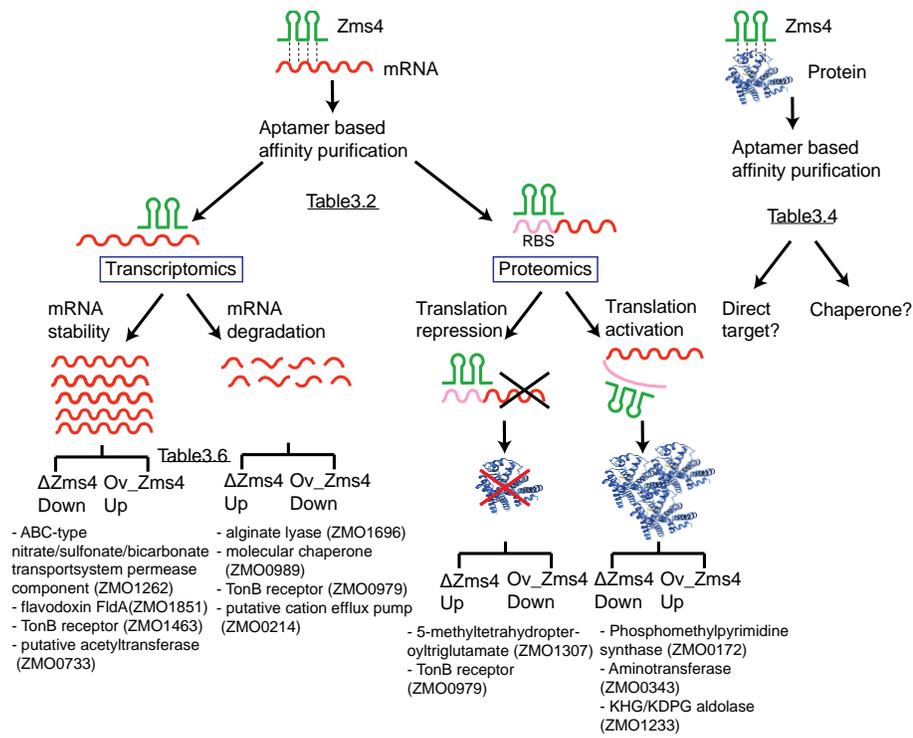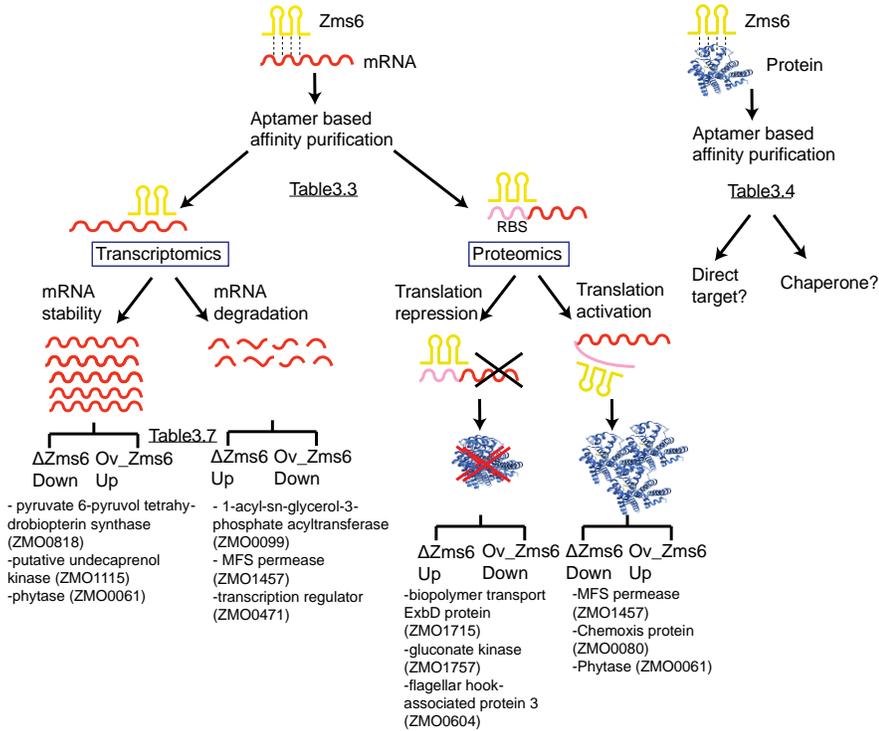
Figure 3.7: Schematic diagram of sRNA-mRNA/protein regulatory network for (A) Zms4 and (B) Zms6 utilizing omics data

Chapter 4

Genome-wide discovery of 5' untranslated regions (UTRs) that control gene expression in response to ethanol and other metabolic stresses in *Zymomonas mobilis*

4.1 INTRODUCTION

Zymomonas mobilis has sown its ability as a promising ethanologenic bacterium with more efficient ethanol production and higher ethanol tolerance (16% v/v) compared to yeast (Rogers et al, 2007). Transcriptomic and proteomic analysis of *Z. mobilis* in ethanol supplement conditions revealed that genes associated with DNA repair, membrane biogenesis, carbohydrate metabolism, transport, and transcriptional regulation have been up and down regulated in response to ethanol stress (He et al, 2012). This result showed that ethanol stress response is a complex phenotype impacting multiple pathways in vivo.

Fundamental studies of *Z. mobilis* have been focused on strain engineering for industrial strain development due to limited usage of carbon source. As wildtype *Z. mobilis* can only utilize glucose, sucrose, and fructose as a carbon source for the production of ethanol, but strains were engineered to utilize xylose and arabinose as carbon sources, which are abundant sugars in pretreated biomass feedstock (Morris & Mattick, 2014; Zhang M et al, 2007; Zhang et al, 1995). During biomass pretreatment which releases sugar monomers from cellulose, xylose (pentose sugar) is the most abundant sugar and acetate is one of major inhibitors produced (Mohagheghi et al, 2014) (Doran-Peterson et al, 2008). Therefore, besides ethanol, xylose and acetate are important stress factors to the physiology of *Z. mobilis*. Acetate toxicity negatively affects cell growth and ethanol production (Yang et al, 2010a). Upon co-utilization of xylose and

glucose, especially with inhibitors such as acetate or furfural, expression level of genes associated with redox mechanism, carbon and energy metabolism was dramatically changed to reduce the shock from the stress (Yang et al, 2014a). However, underlying direct molecular mechanisms involved in acetate tolerance with xylose utilization are still unclear. To uncover potential stress response mechanisms in response to stress in *Z. mobilis*, we focused on the studies of regulatory RNAs using transcriptomic analysis (Cho et al, 2014).

Regulatory RNAs include 5' and 3' untranslated regions (UTRs), riboswitches, cis-acting antisense RNAs and trans-acting small noncoding RNAs that regulate gene expression in response to external stress (Beisel & Storz, 2010). The function of untranslated regions has been elucidated recently in prokaryote as well as eukaryotes and found important in regulation of gene expression. 5' UTRs have been reported to modify gene regulation on the basis of the changes in temperature, pH or other metabolites (Gripenland et al, 2010). Thermo-sensing 5' UTRs control gene expression by temperature dependent conformational changes in pathogenic bacteria. For example, the 5' UTR in front of *prfA* mRNA immediately responds upon temperature changes in *Listeria monocytogenes*, which is critical for survival for pathogenic bacteria to in the host (Toledo-Arana et al, 2009). The 5' UTR of *alx* gene was reported as pH sensor in *E. coli* (Nechooshtan et al, 2009). Alkaline conditions triggered alternative structural changes in 5' UTRs, resulting in active translation of *alx* as well as other genes. Riboswitches represent one class of sensors in which metabolites control gene expression in various metabolic pathways. Upon sensing small molecule metabolites, riboswitches trigger structural changes to regulate transcription or translation of mRNAs. Riboswitches consist of two components; an aptamer and an expression platform. Aptamer domains are between 35-200 nucleotides and responsible for direct binding to

small molecule metabolites such as ions, nucleotides, amino acids or coenzymes (Breaker, 2008; Roth et al, 2007; Soukup & Soukup, 2004). Aptamer domains are highly structured and conserved among different species because of their specific binding to ligands present across many organisms. However, expression platforms can vary in sequence and structure and undergo structural changes in response to ligand binding to aptamer domains, resulting in controlling downstream gene expression; either activation or repression (Barrick & Breaker, 2007; Vazquez-Anderson & Contreras, 2013a; Winkler & Breaker, 2003). Based on the conservation of aptamer domains in riboswitches, bioinformatics and comparative genome analysis enables the discovery of new classes of riboswitch candidates (Corbino et al, 2005; Weinberg et al, 2007) as well as the identification of known riboswitches in diverse organisms (Nahvi et al, 2004; Rodionov et al, 2002). Due to challenges of the ligand identification for riboswitch candidates, there are "orphan" riboswitches such as *ykkC/yzkD, yybP/ykoY* and *pfl* RNAs, which, even though they contain the characteristics for riboswitches such as high sequence conservation and the motif associated with expression platforms, their corresponding ligand is still unknown (Meyer et al, 2011). As a new class of RNA elements, OLE (ornate, long and extremophile) RNA has been studied (Ko & Altman, 2007; Wallace et al, 2012). OLE RNAs, which highly expressed and stable unlike the characteristics of most RNA elements, interacted with ole-associated protein (OAP) to protect extremophiles in response to ethanol stress (Wallace et al, 2012).

To date, even though many different types of regulatory UTRs including riboswitches were discovered among various bacteria species, there were no experimentally confirmed regulatory UTRs in *Z. mobilis*. There were three annotated UTRs (riboswitches) in the genome by computational analysis using gene prediction CMfinder (Yao et al, 2006), which were not experimentally confirmed in *Z. mobilis*. To

understand how stress response is associated with regulatory UTRs in *Z. mobilis*, we discovered and characterized potential regulatory 5' UTRs in *Z. mobilis* utilizing available transcriptomic data. Recent advances in transcriptomics and proteomics combined with computational analysis help us understand comprehensive cellular regulatory networks related to stress responses. Here, we developed a bioinformatics approach to elucidate 5' UTRs in *Z. mobilis* and experimentally validated their regulatory roles under stress conditions utilizing an in vivo GFP reporter system.

4.2 MATERIALS AND METHODS

4.2.1 RT-PCR

To confirm the existence of 5' UTRs from the candidate lists experimentally, we used RT-PCR. Total RNA was prepared using Trizol reagent (Invitrogen) and resulting RNA was treated with DNase I (RNase free, ThermoScientific) to prevent of genomic DNA contamination as described in the manufacturer's protocol. Subsequently, RNAs were precipitated with isopropanol and then washed with ethanol. Rehydrated total RNAs with Nuclease-Free water (Ambion) were used as templates for reverse transcription. 200 ng of RNA was incubated for the first strand synthesis with 100ng of random hexamer and 10 mM dNTPs at 65 °C for 5 mins. According to the manufacturer's protocol, SuperScript™ III Reverse Transcriptase (Invitrogen) was added to RNA-primer mix with RNaseOUT™ RNase Inhibitor, 0.1 M DTT, 5x First-strand buffer and then incubated for 5 mins at room temperature. The final reaction mixture was incubated at 55 °C for 1 hr and then heat inactivated at 70 °C for 15 mins. cDNAs from first-strand synthesis were used as templates for the PCR reactions. Even though the Tm value for each primer was about 58°C - 60°C, minimum annealing temperature was 64°C to remove all erroneous,

non-specific PCR bands from negative controls. Three primers were designed for PCR steps: first forward primer (with reverse primer, set A) was located in the middle of 5' UTRs and second forward primer (with reverse primer, set B) was designed to bind in the front part of following mRNA regions. The reverse primer was designed to bind in the middle of mRNA regions (Figure 4.3A). All primers used for RT_PCR are listed in Table 4.2. Phusion® High-Fidelity DNA Polymerase (NEB) was used for PCR amplification. No reverse transcriptase was used for the negative control to exclude potential genomic DNA contamination. Primer set B was used for the positive control as it represents amplification of the mRNA coding region.

4.2.2 5' Rapid Amplification of cDNA Ends (RACE)

RACE experiments (Figure 3C) were performed using total RNA samples using FirstChoice® RLM-RACE kit (Ambion) according to manufacturer's protocol. Briefly, 8 ug RNA was treated with Tobacco Acid Pyrophospatase (TAP) at 37°C for 1hr, followed by ligation of the 5' RACE kit adapter at 37°C for 1hr. The resulting RNA was then reverse transcribed according to the manufacturer's protocol and PCR was performed on the resulting cDNA. All primer sequences used for RACE are listed in Table 4.4. Resulting PCR products were purified using QIAquick PCR purification kit (Qiagen) and RNase-free water (Ambion) for final elution. Final PCR products were sequenced via Sanger sequencing and results were compared with the genome. 5' RACE adapter sequences (5'-GCUGAUGGCGAUGAAUGAACACUGCGUUUGCUGGCUUUGAUGAAA-3') and adjacent mRNA sequences were used for the detection of TSS in the UTR.

4.2.3 Construction of GFP reporter plasmids with 5' UTRs

We utilized tetracycline inducible promoter with GFP construct (pEZ-tet-GFP). This plasmid contains an origin of replication with promoters for *E. coli* as well as *Zymomonas mobilis* so that it can be used as a shuttle vector in both organisms. Utilizing the pEZ-tet-GFP vector, we incorporated PheS counter-selection marker (Kast, 1994) in front of GFP gene flanked by BsmBI sites, which is one of the Type IIS restriction endonuclease to enable Golden Gate cloning for the efficient cloning of each UTR-containing GFP (Engler et al, 2008). We generated a parental plasmid containing PheS counter selection marker (Miyazaki, 2015) between BsmBI sites (Type IIS enzyme) in front of GFP gene coding region (Figure 4E). Then, we designed to clone identified 5'UTRs with first 90bps of mRNAs for each candidate right in front of the GFP gene in frame. Primers used for the amplification of UTR + 90bps are listed in Table 4.5. Each primer contains BsmBI enzyme site on the 5' end.

4.2.4 Bacterial strains and culture condition for GFP expression

5' UTR-GFP plasmids were transformed into *Zymomonas mobilis* 8b cells (Zhang et al, 1995). Cells were cultured in 5 ml RMG (Glucose, 20.0 g/L; Yeast Extract, 10.0 g/L; KH_2PO_4, 2.0 g/L; pH 6.0) (Yang et al, 2009b) overnight at 33 °C and then inoculated into 100 ml. Initial OD_{600nm} was around 0.05 and cells were induced with 10 ug/ml of tetracycline when OD_{600nm} reached ~0.4. We added 1 % (v/v), 3 % (v/v), 5 % (v/v) ethanol into RMG when we inoculated cells from the initial culture. We collected cells after 4hrs or 12hrs post induction for measuring GFP expression. We added 10 g/L of sodium acetate (NaAc) (Yang et al, 2014b) in RMGfor acetate stress. In case of xylose stress, 10 g/L glucose + 10 g/L xylose was used instead of 20 g/L glucose in the media.

4.2.5 Fluorescence measurements

Cells were analyzed by flow cytometry using the FACSCalibur™ (BD Biosciences) as described in the previous study (Gelderman et al, 2015). Collected cells were prepared for cytometry by resuspending the cells into phosphate buffered saline (PBS: 137 mM sodium chloride, 2.7 mM potassium chloride, and 10 mM phosphate buffer, pH 7.5) to a concentration on the order of 10^7 cells/mL. The cells were excited with the 488 nm argon laser and the cell population was determined from the forward scatter and side scatter distributions reported by the cytometer. Data was collected for at least 50,000 active cells, ensuring enough events to assume that the population distribution would be unaffected by rare events. Sample data were analyzed using CellQuest Pro (BD biosciences) with a user defined gate. We calculated the averages of median values for each sample from at least triplicates. Error bars were calculated as SEM.

4.2.6 Western blot analysis and quantification of protein expression level

Western blotting analysis was performed to detect GFP expression using Anti-GFP antibody (Roche 11814460001). Standard Western blotting protocols were used (Cho et al, 2016). Briefly, total cellular lysates were loaded onto a 12% denaturing SDS-PAGE gel. Gels were transferred to methanol activated PVDF membranes using the Trans-Blot® Semi-Dry Electrophoretic Transfer Cell (Bio-Rad) and run for 40 mins at 15V. Blocking with 5% dry milk in Tris-buffered saline (TBS) was done for 1 hr at room temperature. The proteins were detected with Anti-GFP antibody at 1:1000 dilutions. As a secondary antibody, we used Anti-Mouse IgG (H + L) HRP Conjugate (Promega #W4021) at a dilution of 1:2500. All images were developed using Clarity™ ECL Western blotting Substrate (BioRad, #170-5060) and the ChemiDoc™ MP Imaging System (BioRad). Bradford assay measurements were used to normalize the loading of

all protein analysis by total protein mass. Specific proteins were detected on the membrane by Western blot analysis for the accurate quantification utilizing ImageQuant TL 8.1. Each protein was detected using anti-GFP. The level of GFP expression was measured and then normalized using expression of recA as an internal control.

4.2.7 Construction of deletion of 5' UTRs

Upstream and downstream fragments (each about 1 kb) homologous to the target deletion gene were assembled with the spectinomycin gene *aadA* in the middle.   The assembled product was used as a template for PCR amplification. The purified PCR product was directly electroporated into the *Z. mobilis.* Transformants appearing on RM agar plate with 200 ug/ml of spectinomycin were cultured and screened using PCRs. Colonies with correct PCR product sizes were selected as deletion candidates after sequencing confirmation using the Sanger sequencing.

4.3 RESULTS

4.3.1 Systematic transcriptome analysis of genome enables to identify potential 5' UTRs in *Z. mobilis*

Previously, we have identified novel small RNAs from transcriptomic analysis in *Z. mobilis* and confirmed by Northern blotting analysis (Cho et al, 2014). The identification of novel sRNAs suggested that there would likely be other types of regulatory noncoding RNAs such as unannotated 5' UTRs among transcripts detected by RNA sequencing. We utilized these datasets to discover putative 5' UTR regions that could contribute to the gene regulation in *Z. mobilis*. Initially, we screened all transcripts for expression in the 5' UTR region of their adjacent coding regions from transcriptome datasets. 392 potential candidates were identified from the initial screening. However, we

84

needed to filter these 5' UTR candidates from large number and diversity of other types of transcripts.

We developed bioinformatics pipeline for the selection of final potential candidates from large number of initial candidates (Figure 4.1). First of all, we analyzed initial candidates that showed comparable level of expression compared to the adjacent genes. Then, we further filtered out all the transcripts less than 30 base pairs that was known as the shortest length of known UTR regulatory element (Roth et al, 2007). This filtering step led to a significant decrease in total number of potential candidates and ruled out any noncoding intergenic regions. Lastly, we analyzed potential candidates according to their function. As well-characterized metabolic enzymes are highly regulated by their 5' UTRs, we prioritized these types of candidates in the list. Finally, total 101 potential candidates were chosen for experimental analysis.
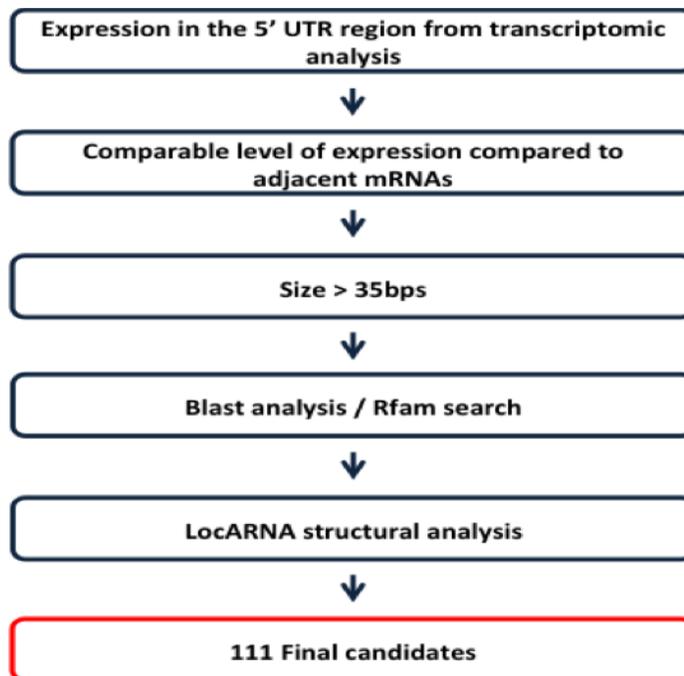
Figure 4.1: Pipeline for the selection of UTR candidates

85

4.3.2 Computational analysis supports the existence of 5' UTRs

To further verify the properties of 5' potential candidates, we performed computational analysis using Rfam. Rfam is known as a database of multiple sequence alignments, consensus secondary structures, and covariance models (CM) representing RNA families (Griffiths-Jones et al, 2003). In Rfam database, CM are used to describe each RNA family by modeling RNA sequence and the structure in an elegant and accurate way. There are three RNA categories in Rfam: non-coding RNAs, structured *cis*-regulatory elements and self-splicing RNAs (Nawrocki et al, 2015). We analyzed all potential candidates and confirmed 5 candidates were matched with a known riboswitch in Rfam. Previously annotated putative riboswitches (TPP and cobalamin riboswitches) in our candidate list were also found by Rfam analysis and marked in Table 1. These widely conserved riboswitches have been demonstrated in *Escherichia coli* to control the regulation of the downstream genes by direct binding to thiamine pyrophosphate and cobalamine, respectively (Nahvi et al, 2002; Winkler et al, 2002). Additionally, Rfam prediction identified the crcB RNA motif which is called Fluoride riboswitch (Baker et al, 2012) in front of the gene "chloride channel protein (ZMO0547)". The fluoride riboswitch is newly found regulator that changes structure in response to fluoride ions to regulate downstream gene expression. Genes encoding for this fluoride-specific subtypes of chloride channel proteins have been shown to be regulated by fluoride riboswitches in a variety of organisms (Stockbridge et al, 2012). Taken together, it is worthwhile to note that these results can not only support the presence of 5' UTRs before experimental confirmation but also validate our UTR prediction methods.

Given that structural conservation is closely associated with the regulatory roles of RNAs (Yang et al, 2010a), we performed structure based-conservational analysis utilizing LocARNA software. LocARNA is a global prediction tool for structural RNAs

86

based on sequence and structure similarities (Will et al, 2012). LocARNA differs from other sequence-based prediction tools in that it's algorithm considers structure as well as sequences (Will et al, 2012). LocARNA software requires sequences with homology from several different organisms for the analysis; therefore, we conducted conservation analysis using BLAST for each 5' UTR candidate before applying into LocARNA software (Camacho et al, 2009). A set of sequences which were collected with apparent homology (expect value $< 10^4$) was entered into LocARNA software for the prediction. After the analysis by LocARNA for each 5' UTR candidate, we successfully identified 28 candidates that contained structurally conserved motifs, many of which showed complexity. Figure 4.2 shows representative data of this analysis.

Lastly, we compared all final candidates with the genes which were differentially expressed under stress such as ethanol, acetate and xylose from previously published literature (He et al, 2012; Yang et al, 2014a; Yang et al, 2014b; Yang et al, 2013) as we hypothesized that differentially expressed gene could be regulated their 5' UTRs upon stress. We found that 17 candidate genes and 7 proteins were up or down regulated under stress (Table 4.1). Due to their association with stress response, these candidates will be further investigated with their mechanistic role in metabolic engineering field.

.

UTR_ZMO0172
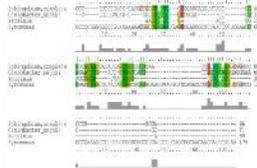
UTR_ZMO1137
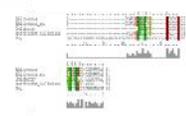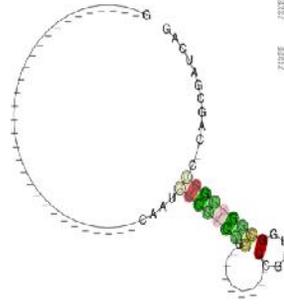
UTR_ZMO0347

UTR_ZMO0547
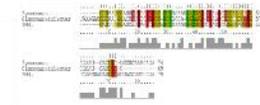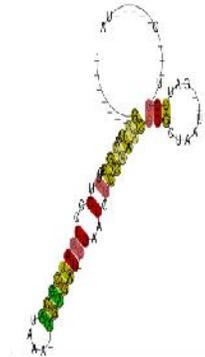
UTR_ZMO1113
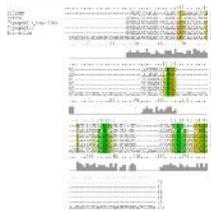
UTR_ZMO0709
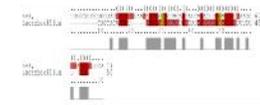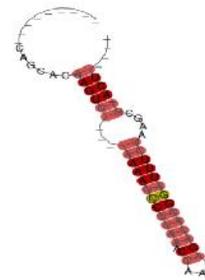
Figure 4.2: Structural analysis of UTR candidates using LocARNA

NCBI Gene ID	Gene	Rfam Prediction	Characteristics of known riboswitches in other organisms	Gene regulation under stress in the literature	Type of Stress induced differential GFP expression
ZMO0172	Thiamin biosynthesis protein/phosphomethylpyrimidine synthase	TPP	TPP riboswitch in E.coli	-	Acetate/Xylose
ZMO0979	TonB-dependent receptor	Cobalamin	AdoCbl	Down in Acetate stress Down in Xylose stress	-
ZMO1000	5-methyltetrahydropteroyltriglutamate	Cobalamin	metE in E.coli/Bacillus	Upregulated in Xylose	Acetate
ZMO0547	chloride channel core	crcB	Fluoride riboswitches	-	-
ZMO0056	glucosamine/fructose-6-phosphate aminotransferase	-	glmS ribozyme	-	-
ZMO0376	ATP-dependent protease La	-	-	-	-
ZMO0546	sulphate transporter	-	ABC transporter family	Up in 5% Ethanol stress UP in Xylose stress	Acetate
ZMO0660	dnaK molecular chaperone DnaK	-	-	-	-
ZMO1069	molecular chaperone DnaJ	-	-	-	-
ZMO1139	acetolactate synthase large subunit	-	-	Up in Xylose	-
ZMO1142	thioredoxin reductase	-	-	Up in Xylose stress Down in Ethanol	Ethanol
ZMO0709	phosphoribosylaminoimidazole synthetase	-	-	-	-
ZMO1137	phosphoserine phosphatase SerB	-	serC regulated by glycine riboswitch	-	-
ZMO0187	3-deoxy-7-phosphoheptulonate synthase	-	-	-	Xylose
ZMO0369	glucokinase	-	-	-	-
ZMO0405	ATP-dependent Clp protease ATP-binding subunit ClpA	-	-	Up in Xylose stress Down in Ethanol stress	-
ZMO0937	aromatic amino acid aminotransferase	-	-	Down in Acetate stress Down in Xylose stress	-
ZMO1179	(uracil-5)-methyltransferase	-	-	Up in Ethanol stress	-
ZMO1275	threonine dehydratase	-	-	-	-
ZMO0689	oxidoreductase domain-containing protein	-	-	Down in Xylose stress	-
ZMO0131	metallophosphoesterase	-	-	Down in Acetate stress Down in Xylose stress	-
ZMO0748	cysteine synthase	-	-	Up in Acetate stress Up in Xylose stress	-
ZMO1034	calcium-binding EF-hand-containing protein	-	-	Up in Acetate stress Up in Xylose stress	-
ZMO1113	FAD-dependent pyridine nucleotide-disulfide oxidoreductase	-	-	Up in Xylose	-
ZMO0347	RNA-binding protein Hfq	-	-	Down in Ethanol stress	Ethanol
ZMO1198	5-aminolevulinate synthase	-	-	Down in Xylose stress	-
ZMO1478	6-phosphogluconolactonase	-	-	-	Acetate
ZMO0275	ABC transporter	-	-	-	-
ZMO1399	fatty acid hydroxylase	-	-	-	-
ZMO1432	fusaric acid resistance protein	-	-	-	-
ZMO0367	glucose-6-phosphate 1-dehydrogenase	-	-	Up in Ethanol stress	-
ZMO1412	MucR family transcriptional regulator	-	-	Down in Ethanol stress	-
ZMO1048	phosphate ABC transporter inner membrane subunit PstC	-	-	-	-
ZMO0140	protein tyrosine phosphatase	-	-	-	-
ZMO0366	sugar transporter	-	-	-	-
ZMO1612	toluene tolerance family protein	-	-	-	-

Table 4.1: List of final 5' UTR candidate with their features

89

4.3.3 Final potential 5' UTR candidates were validated experimentally by RT-PCR

To confirm the expression of final candidates in the cell, we tested all final candidates for their expression by RT-PCR experimentally. We designed two sets of primers. Primer set A was aimed for the amplification of a long transcript from hypothetical 5' UTR regions to the middle of mRNA coding region. Primer set B was designed to amplify a relatively short transcript inside the adjacent mRNA coding region as a positive control representing the adjacent mRNA expression level (Figure 4.3A). Primers used for RT_PCR were listed in Table 4.2. As a negative control, reverse transcriptase was not added to the reaction confirm that genomic DNA contamination was not present. Representative data is illustrated in Figure 4.3B. PCR bands from both primer sets A and B proved the expression of transcript containing potential UTRs. However, if there is a band from primer set B and not from set A, then the UTR was not detected upstream of the mRNA. All RT-PCR results were shown in Figure 4.4. From the experimental analysis, we have confirmed 50 positive 5' UTR candidates that showed contiguous expression in the transcript along with the mRNA. However, we could not detect the expression for the rest of the candidates. Some of them did not even show mRNA expression. This could be due to unsuccessful PCRs or false positives in candidate selection from transcriptome data.

To understand the regulatory potentials of candidate 5' UTRs for the further analysis, we experimentally mapped the transcription start sites (TSS) for all candidates of interest by conventional RACE (Figure 4.3C). Table 4.4 showed primers used for 5' RACE. A total of 36 candidates were finally selected by 5' RACE for the further application. Summary of final candidates and their features were shown in Table 4.1. The precise 5' ends for each 5' UTR candidate sequences are shown in Table 4.3. We excluded 14 candidates from 50 experimentally confirmed candidates due to their

overlapped transcription with the adjacent gene (3' end of the adjacent gene was connected with potential 5' UTRs) or length less than 30 bp or no 5' UTRs experimentally detected (Figure 4.5).
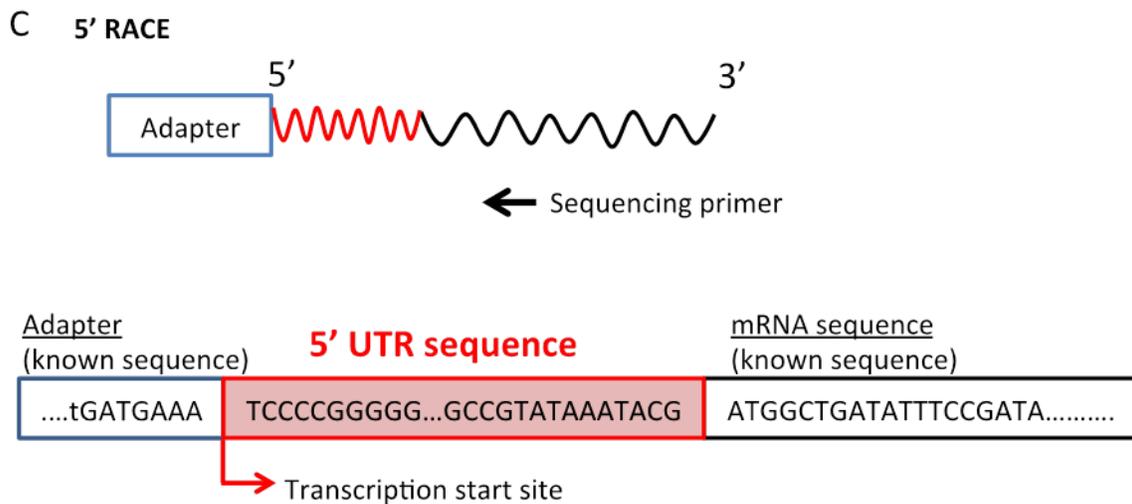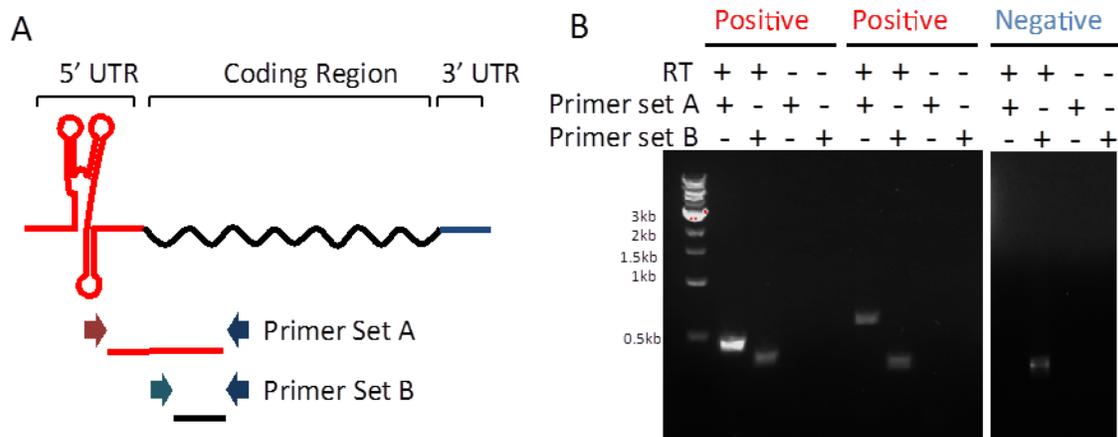
Figure 4.3: Experimental confirmation of UTR candidates by RT-PCR
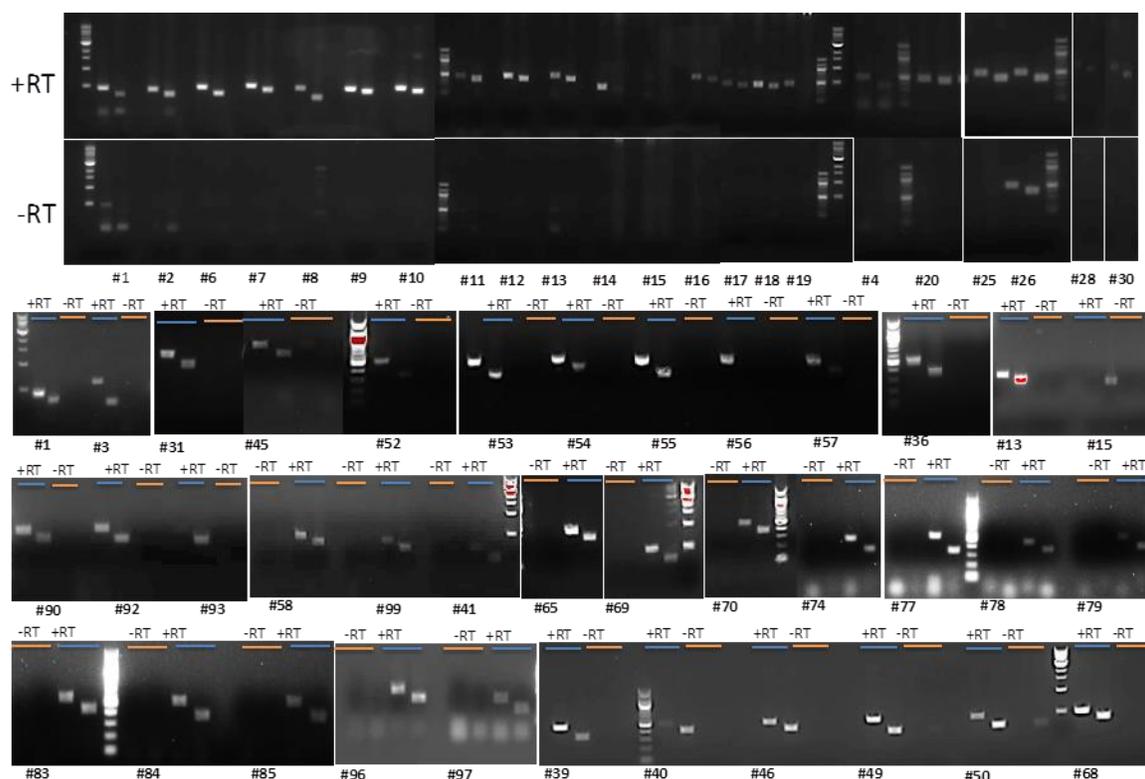
Figure 4.4: Experimental analysis of UTRs by RT_PCR.
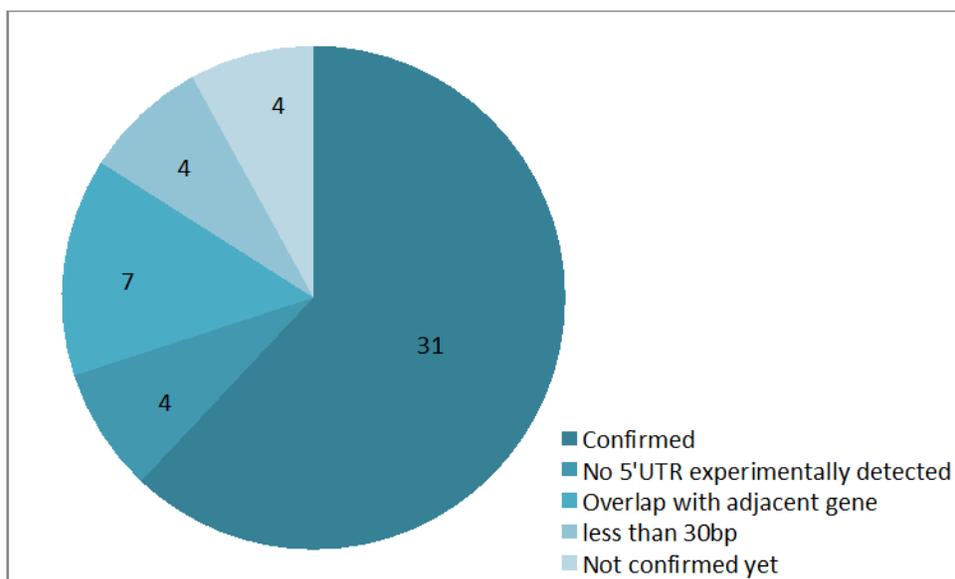
Figure 4.5: Summary for the results of 5' RACE

Target gene	Primer name	Sequences	Target gene	Primer name	Sequences
	zmo0172_5U_F1	gtagggagggaaggcatcag		zmo0031_5U_F1	cccgacagccttccaaag
ZMO0172	zmo0172_mRNAF2	cgaacgccattttttaccttc	ZMO0031	zmo0031_mRNAFw	aatactggccgttccatcag
	zmo0172_mRNARv	cacgcaaaggatgatggtt		zmo0031_mRNARv	tcaattttggtcgcattgaa
	zmo0979_5U_F1	attgggaacaaggtgcaaaa		zmo0131_5U_F1	aggttcacgaggaaataggg
ZMO0979	zmo0979_mRNAF2	cgactttggctttgacaaca	ZMO0131	zmo0131_mRNAFw	aacagcacctcgcctttatg
	zmo0979_mRNARv	actgttactccgggcgttct		zmo0131_mRNARv	tccagaataacgaaatgcaca
	zmo1000_5U_F1	gatcggagagcaatgaggaa		zmo0748_5U_F1	gcctagacaatcatttttaaggaca
ZMO1000	zmo1000_mRNAF2	tttgggtttcccaagaattg	ZMO0748	zmo0748_mRNAFw	cggctcaatcaaagaccgta
	zmo1000_mRNARv	tgttttcagcagaggtcgtg		zmo0748_mRNARv	ctgttgtggtggcatggata
	zmo0547_5U_F1	agacaaggtaatggaatctacctga		zmo0918_5U_F1	caatcaggtaagagggacgtt
ZMO0547	zmo0547_mRNAF2	ccttgattaaaggatgccaga	ZMO0918	zmo0918_mRNAFw	ccaaccctcctccaagatgt
	zmo0547_mRNARv	gccaccaaaaagatgggtta		zmo0918_mRNARv	ccatcttgggtgtcttttaaattc
	zmo0861_5U_F1	ccggcttccacgtcttaaa		zmo1029_5U_F1	gccaataaccctgacgtctt
ZMO0861	zmo0861_mRNAF2	atcgcatcctgaccaatca	ZMO1029	zmo1029_mRNAFw	gcaataacggtcagggaaaa
	zmo0861_mRNARv	atggttgattcttcctctctgg		zmo1029_mRNARv	atgtgccaaagccaatcttt
	zmo0056_5U_F1	aaagacagtccgcctttcaa		zmo1410_5U_F1	agggttgtaaaaagcaaaagg
ZMO0056	zmo0056_mRNAF2	caaggattaagacggctgga	ZMO1410	zmo1410_mRNAFw	gggccataaagaggtcattg
	zmo0056_mRNARv	tcagaaccttggaaacagca		zmo1410_mRNARv	ccgtctggctgttaagctg
	zmo0376_5U_F1	cgccccagtggagttgtg		zmo1034_5U_F1	cagacaatattcataaactgcctga
ZMO0376	zmo0376_mRNAF2	gcttttttgtcggtcgtgaa	ZMO1034	zmo1034_mRNAFw	gtctgtttttagcggcaacc
	zmo0376_mRNARv	actttgacagccaaattaacagc		zmo1034_mRNARv	tgatgatgaccacctttaccg
	zmo0546_5U_F1	cagccttgctaaatcccaaa		zmo1113_5U_F1	caaaccaccacgccaaag
ZMO0546	zmo0546_mRNAF2	cagcgcaaagagtccaaaag	ZMO1113	zmo1113_mRNAFw	gaacgtttcgggaaagagg
	zmo0546_mRNARv	tagataaccgcagcagacca		zmo1113_mRNARv	gcggctatccaaaaacagac
	zmo0660_5U_F1	attcctggggaaggaaggta		zmo0753_5U_F1	cagtaacatttagtcggttttctcc
ZMO0660	zmo0660_mRNAF2	acagctgcgttgctgttatg	ZMO0753	zmo0753_mRNAFw	tattgcaaacgggcaaaag
	zmo0660_mRNARv	tcttaccggcatctttggtt		zmo0753_mRNARv	ccactttgcgactgataccat
	zmo0732_5U_F1	tttgttgaccctttctgagg		zmo1639_5U_F1	tgagcggctttcttcacttt
ZMO0732	zmo0732_mRNAF2	tgccagtacagagcatttcg	ZMO1639	zmo1639_mRNAFw	agccggatactcgggtttat
	zmo0732_mRNARv	tcggtcagaagctggaattt		zmo1639_mRNARv	aacagccaactgaaattcacc
	zmo1069_5U_F1	aattttttcggtcggggggtat		zmo1350_5U_F1	aatatctgccgcatcagagg
ZMO1069	zmo1069_mRNAFw	gttggcgcggttataatga	ZMO1350	zmo1350_mRNAFw	gtctggacgcaccgaagat
	zmo1069_mRNARv	gattaagatcgggatgccatt		zmo1350_mRNARv	acaccggccatgataaaaat
	zmo1139_5U_F1	cgattggatcatgggggttt		zmo1576_5U_F1	cggtgcgaagaattttgtct
ZMO1139	zmo1139_mRNAFw	cgccattgttcttgaaacac	ZMO1576	zmo1576_mRNAFw	gtcagatttgggggcaact
	zmo1139_mRNARv	agatggaaggcctcgacaat		zmo1576_mRNARv	ggcaaccgaagagacattga
	zmo1142_5U_F1	tgccgtttttattctgctgt		zmo1705_5U_F1	taaatatcaaagaaacgggcaat
ZMO1142	zmo1142_mRNAFw	gctgatcctatatctacccgtgtt	ZMO1705	zmo1705_mRNAFw	gctgtattttcggccacat
	zmo1142_mRNARv	ttccacggaaaaagaaacca		zmo1705_mRNARv	cccaaaagagtaacggcttg
	zmo1215_5U_F1	aagatagattacctcggtaagactgaa		zmo1385_5U_F1	ccggaaataaaaggccagat
ZMO1215	zmo1215_mRNAFw	aatggcagcatgacagcac	ZMO1385	zmo1385_mRNAFw	cagcttgaagccaaggtgag
	zmo1215_mRNARv	tttgccaaaagggctagaag		zmo1385_mRNARv	ataatcgcggagcttgtttc
	zmo1324_5U_F1	cgcttttgactgtgattatctttg		zmo0347_5U_F1	tcttgagagactcttcacataggc
ZMO1324	zmo1324_mRNAFw	cggtggttcaaacgaatcat	ZMO0347	zmo0347_mRNAFw	gccgaaaaggtcaacaatct
	zmo1324_mRNARv	tcaattttttatcggtgtcgaag		zmo0347_mRNARv	agctgcgcgtcgtcatta
	zmo1480_5U_F1	cactaactccagcatgcttcata		zmo1756_5U_F1	aatgctggctttcccatagat
ZMO1480	zmo1480_mRNAFw	taatagtgcccgttaaacgtgtg	ZMO1756	zmo1756_mRNAFw	tgattgcaaagtggcgtcta
	zmo1480_mRNARv	ttttattatcggcaatcgtcaa		zmo1756_mRNARv	gctggcacgtcctaaggtaa
	zmo0077_5U_F1	tcggcttttcaactgtttgg		zmo1222_5U_F1	cagcccctcttctggttcta
ZMO0077	zmo0077_mRNAFw	ccgaccttctttaaagcatttttt	ZMO1222	zmo1222_mRNAFw	ctgttgtaacgggtgcctct
	zmo0077_mRNARv	aatttatcgcagccttcctg		zmo1222_mRNARv	gcagcataattggcttgtcc

Table 4.2: Primer sequences for RT_PCR

93

ZMO1360	zmo1360_5U_F1	ttcagacatagtgttttgaatatatgg	ZMO1198	zmo1198_5U_F1	tgaaatcgaccagattattttgttt
	zmo1360_mRNAFw	gagttatactgtcggtacctatttagc		zmo1198_mRNAFw	ggccgctatcgtgtgtttat
	zmo1360_mRNARv	atcgatttagccggagctt		zmo1198_mRNARv	tggttcaattcatcggaaaa
ZMO0017	zmo0017_5U_F1	accttcggttgtccattcg	ZMO1478	zmo1478_5U_F1	aaaaggcagaaatgccgata
	zmo0017_mRNAFw	gtcccaaagggaaagacagc		zmo1478_mRNAFw	tcaagcaagccattgagaaa
	zmo0017_mRNARv	taaacgccatgaaccagacg		zmo1478_mRNARv	aaattgggacctgggaaaat
ZMO0709	zmo0709_5U_F1	tccaagtctgatggtgtaggg	ZMO0275	zmo0275_5U_F1	tttccccgatatggatttga
	zmo0709_mRNAFw	aacagcacgcaaaaatacagtt		zmo0275_mRNAFw	atcgcgaaatcctacgatca
	zmo0709_mRNARv	gtaatcgccatcggcataga		zmo0275_mRNARv	gtctttcaccaccggaaagc
ZMO1184	zmo1184_5U_F1	ggatgattgttatataaaccataagga	ZMO0778	zmo0778_5U-_F1	tccgtatcaacgatctggaa
	zmo1184_mRNAFw	atgtcagaacgcgaatccat		zmo0778_mRNAFw	aaatcccgattgtggcagt
	zmo1184_mRNARv	ttttcgtcatacaaaatatctgctg		zmo0778_mRNARv	aaggtcattcagctgctgttc
ZMO1708	zmo1708_5U_F1	ccgcttaaaacaaggctcaa	ZMO1673	zmo1673_5U_F1	cagcaagaaaggattggagtc
	zmo1708_mRNAFw	aagaataaggcttggggtcaa		zmo1673_mRNAFw	tggtcgctttgggatatgat
	zmo1708_mRNARv	attcgataaccggcgcatt		zmo1673_mRNARv	tttttgctggaaataagggtgt
ZMO0027	zmo0027_5U_F1	cgcccagattagtcggaaa	ZMO1497	zmo1497_5U_F1	tcttatcgggggataatga
	zmo0027_mRNAFw	ttaaaaccgcactgctttcc		zmo1497_mRNAFw	ttttgtgttaacggcggaag
	zmo0027_mRNARv	gataatctttcgggtcggtca		zmo1497_mRNARv	gcagcgggaaggttgataat
ZMO0914	zmo0914_5U_F1	taatattttttgttgtaaggattgttc	ZMO1770	zmo1770_5U_F1	gtaatgacccagtcgggaga
	zmo0914_mRNAFw	gtaacgtcgttcctggcttg		zmo1770_mRNAFw	atattgccggttccaaagc
	zmo0914_mRNARv	accaaagcatttttgcctttt		zmo1770_mRNARv	gtaaagcccggcatcacg
ZMO1137	zmo1137_5U_F1	gcgatgaggagaactggatg	ZMO1107	zmo1107_5U_F1	tatctagccggaagattaaatagaag
	zmo1137_mRNAFw	atcaagctcggtcgcttct		zmo1107_mRNAFw	tgacttggcacgtaaagcag
	zmo1137_mRNARv	tgatccctggggttaaatga		zmo1107_mRNARv	attactctttggtatcaacaggaaca
ZMO1485	zmo1485_5U_F1	ttgaataaatctctttggaatagacg	ZMO1407	zmo1407_5U_F1	tatatcggtcattccatccttg
	zmo1485_mRNAFw	tcaagatgcagaccggatta		zmo1407_mRNAFw	gtgaaatgttggcgattctg
	zmo1485_mRNARv	atttcaagctgggtcagcaa		zmo1407_mRNARv	caacaacgcgtttgatcttt
ZMO0405	zmo0405_5U_F1	gcatcaggaataggggagtgg	ZMO1033	zmo1033_5U_F1	caagttaaagcatgtaaactctgacc
	zmo0405_mRNAFw	atgaacatgcttcgcaagtg		zmo1033_mRNAFw	ttatatgggatcgggtgagg
	zmo0405_mRNARv	gctgggtcttttcttcttcg		zmo1033_mRNARv	ccagcttggatgcgatgt
ZMO0443	zmo0443_5U_F1	cattttcagaatcggcagga	ZMO1008	zmo1008_5U_F1	gtctatgcgagcttcggttc
	zmo0443_mRNAFw	cagcagcttcattggcttc		zmo1008_mRNAFw	atgatttgcgccagctattg
	zmo0443_mRNARv	gctgccgacatcaaaggtt		zmo1008_mRNARv	tctttcttttcagtggcatcg
ZMO0820	zmo0820_5U_F1	tgggtgctgacagaaaggtt	ZMO1399	zmo1399_5U_F1	acttcggttccgattgtgac
	zmo0820_mRNAFw	gacaacgcctgaaaccaaat		zmo1399_mRNAFw	ttggacggatgacattgaaa
	zmo0820_mRNARv	tcccccgaagaatatgacac		zmo1399_mRNARv	acctccgaaaccccctataa
ZMO0903	zmo0903_5U_F1	cgttgttgttctggtgatcc	ZMO0179	zmo0179_5U_F1	ttaccttatatcccaagggaagg
	zmo0903_mRNAFw	gatttacccaatcggcagtg		zmo0179_mRNAFw	cgggacaaaggctaatcttg
	zmo0903_mRNARv	cacctgcgatagccagtttt		zmo0179_mRNARv	cgcaatatggcggaattctt
ZMO0923	zmo0923_5U_F1	gccagtctattgtgattgaagg	ZMO1235	zmo1235_5U_F1	atataatagaaggaaatctggcct
	zmo0923_mRNAFw	cgcaatgatctcagcctgt		zmo1235_mRNAFw	cgtgcctccgtttacaaag
	zmo0923_mRNARv	tggcattagccatatgttcc		zmo1235_mRNARv	aatgtgaacatttaccccgtattt
ZMO0937	zmo0937_5U_F1	tttatcgaggaatagaaagagggata	ZMO1411	zmo1411_5U_F1	caaaatttgacggagaaatgtt
	zmo0937_mRNAFw	tcgtgaagatactcgcgaaa		zmo1411_mRNAFw	tctctttacgaacgggcatc
	zmo0937_mRNARv	ccagtattggcaatgatctgaa		zmo1411_mRNARv	ggaatcgtcatcaaaataaccaa
ZMO1020	zmo1020_5U_F1	gtcttgccaccctaccctct	ZMO1432	zmo1432_5U_F1	tcttgctatgataggcgactg
	zmo1020_mRNAFw	ttgcgggaaagacgatgtat		zmo1432_mRNAFw	gggcgatgacgactgtctat
	zmo1020_mRNARv	cagggtcaatgccgaattt		zmo1432_mRNARv	tcgcaatcatgtgagaatctg
ZMO1052	zmo1052_5U_F1	attaagcgccttgcctttt	ZMO0367	zmo0367_5U-F1	aaggattcggcctctgtttt
	zmo1052_mRNAFw	ccggcaccattatccagtat		zmo0367_mRNAFw	gatctgcgtatcgtctgcac
	zmo1052_mRNARv	tgcatttcttcctgcgtagc		zmo0367_mRNARv	aatatgatcggaagaagcaaga

Table 4.2 (cont.): Primer sequences for RT_PCR

ZMO1179	zmo1179_5U_F1	cttggcatcaaagggagac		ZMO1556	zmo1566_5U_F1	tttggatagagggactatatcttgtg
	zmo1179_mRNAFw	cggttgccgtttacaatatg			zmo1566_mRNAFw	tggcacagagcatgaaaaac
	zmo1179_mRNARv	tcaattccttgatctgctaacg			zmo1566_mRNARv	tttaggcattaaaggcaggtc
ZMO1275	zmo1275_5U_F1	ttaaaagtaataacgccggaaaa		ZMO1599	zmo1599_5U_F1	tgtaacagctaaggcgcttg
	zmo1275_mRNAFw	tatcaaacaggcgcatcaac			zmo1599_mRNAFw	gctggtctctgcctatttcg
	zmo1275_mRNARv	agattgtcttttcagccaattc			zmo1599_mRNARv	atcttgaacatccggcaact
ZMO0247	zmo0247_5U_F1	cttcttgctggaaaaagattgt		ZMO1596	zmo1596_5U_F1	gttgttttcgggttgttgct
	zmo0247_mRNAFw	gtgctgttgctgtcgctatg			zmo1596_mRNAFw	aaaatgcgctgatcgtttct
	zmo0247_mRNARv	cgtcaagatgcccatctaaaaa			zmo1596_mRNARv	acgtgacggtcaacaatgg
ZMO0309	zmo0309_5U_F1	gcctaataggactttttacgacctg		ZMO1771	zmo1771_5U_F1	tacgcataagaaggcggaaa
	zmo0309_mRNAFw	gccaatcgggataaaagaacc			zmo1771_mRNAFw	cggttctagcgctcaaaaga
	zmo0309_mRNARv	ggcaaaaaccttatgaccttc			zmo1771_mRNARv	attgacttgcgggttaccac
ZMO0512	zmo0512_5U_F1	ttacaaaggaaagagaacttaactca		ZMO1233	zmo1233_5U_F1	ctcggctcccgacaatagta
	zmo0512_mRNAFw	ggtggtatctgcttgaattgg			zmo1233_mRNAFw	ggcgaaaatagcctgtttca
	zmo0512_mRNARv	gcttggccatcagaatgaaca			zmo1233_mRNARv	aaccgtaacctgtttcaggaga
ZMO0120	zmo0120_5U_F1	atcatttgggaggccagag		ZMO1412	zmo1412_5U_F1	tttaaagcttctctggaaagataga
	zmo0120_mRNAFw	tcgcagatattgtcgaaaagc			zmo1412_mRNAFw	tatcgtttcagcccatgtca
	zmo0120_mRNARv	cttggaaactgccgatcat			zmo1412_mRNARv	gccacatttgtggtaatgctt
ZMO0187	zmo0187_5U_F1	gcgcagcttgaaaggaataa		ZMO0178	zmo0178_5U_F1	ggaggctgtctccgttatca
	zmo0187_mRNAFw	gcgcgacaattaccgattta			zmo0178_mRNAFw	atcgcgttacggatgatacc
	zmo0187_mRNARv	aacagacgatctgggtcagg			zmo0178_mRNARv	ccgagaaagcatcattgaca
ZMO0215	zmo0215_5U_F1	tgaatattttctgtgtgcttggatg		ZMO1048	zmo1048_5U_F1	ggaaaaagatgatagctgtcagaa
	zmo0215_mRNAFw	tattcagccaacagcgtcat			zmo1048_mRNAFw	acgctgacccttgttgctat
	zmo0215_mRNARv	aaaaccgactaggggcgtta			zmo1048_mRNARv	cccgcaaccactctacctaa
ZMO0369	zmo0369_5U_F1	cctgttggggtagccttctga		ZMO0140	zmo0140_5U_F1	aacgatggatgaatccgttta
	zmo0369_mRNAFw	tggtcgggttctttctcttg			zmo0140_mRNAFw	gaaggggcttttcgtgatct
	zmo0369_mRNARv	cgaaataacggccttcagtc			zmo0140_mRNARv	ggttgagacatcttgccatgt
ZMO1181	zmo1181_5U_F1	tttgtgtttttcttaagcgtaaattg		ZMO1606	zmo1606_5U_F1	tcgatttatcctgtccgtca
	zmo1181_mRNAFw	tgatttgggaatcgacaaaag			zmo1606_mRNAFw	cacgatttgccaccgatac
	zmo1181_mRNARv	ttaggtcgcttttcaagcaag			zmo1606_mRNARv	ccccaacgataccatttcc
ZMO1682	zmo1682_5U_F1	tcgcttttttcctagatatttcaattag		ZMO0322	zmo0322_5U_F1	cctgtgttcttatgtaaatcgagac
	zmo1682_mRNAFw	aaacgcaggtcgtggtaatc			zmo0322_mRNAFw	catttaggccatcaggctgt
	zmo1682_mRNARv	cgaccatttcccgtaagaca			zmo0322_mRNARv	tctcgaatacgggtagaggaaa
ZMO0689	zmo0689_5U_F1	accacgacggatcttctctg		ZMO0366	zmo0366_5U_F1	agattaaggcgggagaggaa
	zmo0689_mRNAFw	tttcagcaacggatgacaac			zmo0366_mRNAFw	cattttattgcccctcgtca
	zmo0689_mRNARv	cagcgtcgatttttggatct			zmo0366_mRNARv	catctgctgaccagaaacca
ZMO1236	zmo1236_5U_F1	ccctatattcgcaagatgtatgtc		ZMO1097	zmo1097_5U_F1	aattcatatccccgtcaggag
	zmo1236_mRNAFw	ttacgccctctgaaatacgg			zmo1097_mRNAFw	cacctgctttgggtgaaatc
	zmo1236_mRNARv	caagaccatctggcactttg			zmo1097_mRNARv	ttaaagtgaactttcaatccatga
ZMO0976	zmo0976_5U_F1	tttaaaaaatgtggagagtttctg		ZMO0765	zmo0765_5U_F1	catctttgaaataaaagaaagacga
	zmo0976_mRNAFw	gggcgattgatcttggtatc			zmo0765_mRNAFw	gcccgttttaatggtcagat
	zmo0976_mRNARv	gcatagggcaggatgtcttt			zmo0765_mRNARv	atttttcgccctgagattcc
ZMO1522	zmo1522_5U_F1	tttggggacggtgcatatct		ZMO1612	zmo1612_5U_F1	gccttgaaaaccgacttctc
	zmo1522_mRNAFw	gcctgtttttgcacaggaa			zmo1612_mRNAFw	cattgcgcttgtctcatttg
	zmo1522_mRNARv	gcatcaaacgacgaccatta			zmo1612_mRNARv	tagccctgactggtgtgttc
ZMO0130	zmo0130_5U_F1	tcttatataacaaggaaagactgacga		ZMO0054	zmo0054_5U_F1	agacggcttgcgtgagac
	zmo0130_mRNAFw	tgaaccctcctcgacttatca			zmo0054_mRNAFw	gaattggatgcgtcttcgat
	zmo0130_mRNARv	agccttgtttcccagacct			zmo0054_mRNARv	cttttcggtaagggcaacc
ZMO1289	zmo1289_5U_F1	ggcgtttatacacaaatattgaca				
	zmo1289_mRNAFw	gatggctcgctggacttatt				
	zmo1289_mRNARv	tcagttgcgtgatttaaagaaa				

Table 4.2 (cont.): Primer sequences for RT_PCR

NCBI Gene ID	5'UTR SEQUENCE CONFIRMED by 5' RACE
ZMO0172	tccccggggggccgtataaatacggctgagaatgagctgattgctctaacccgtcgaacctgatccggcttaacaccggcgtagggagggaaggcatcagaccttgcttcttatccgtcggccgaaagcctttttccggtcggcctgatggttgacctgcaagaaggatgcataa
ZMO0979	ggaaattttttgcatagggtttccttcgagtgaaggaaaattgggaacaaggtgcaaaaccttggctgcccctgcaactgtaaacagttgaaacgccaaaaagccactgaatctattcgggaaggcggttgtttcgatgctgtgagccaggagaccgaccctatgtaatcgttccacgacaggacagggcgttcccgtcgaacgaggtaaattctcctcgcacggacgataatagtgaccttcaagccaatcggtttggaggctgttttgctgtgagtaacagaggggggatag
ZMO1000	aaggattaaggttctttgtcattggcaaagctaagagggaaactggtgcgaaagaattttcaaagccagtgctgcccccgcaactgtaaacggcgagcaaagatcaaaatgccactgatattattatcgggaaggctgatcggacgcggtgacccgtcaagtcaggagacctgccttaaaccaagtcatccactatcggttttgaaaactggaaacaggctttaaaactggtttctttcggggttgaagagaaaa
ZMO0547	gaatgaccatttccattatccactgacttgacttttggcattattttggccttaatatagacaaggtaatggaatctacctgatcctttct
ZMO0056	atctcaaagacagtccgcctttcaaaatagatatcacaaaatcgggaaacagaatt
ZMO0376	aaaaggtgtgagcattttgccgccccagtggagttgtg
ZMO0546	aattttggtcacattgtcggcaatattcctgatattttgacgatcgggctcagccttgctaaatcccaaaaccttgatcttgcccaaaaaactgaattgatatttcccatcgcagattttccttcgaggaagccggataaaatagagctaaaatctttcggtgaaattttgagctctgataaaagatggcaatctcaattcctagcatttttaaggatacccg
ZMO0660	gtagacttgttccaagggcaagctttattttttagaagagtaaattttgtcattcctggggaaggaaggtaggac
ZMO1069	tttgctggtggacagaacccggcttatagaccttctggcgtatcacctctgttgattttctttgtcatttgcctctttttcctaaagaaactatattcttggttttttccgactttggccaattttttcggtcgggggtattctcatacaggcagaacaagaaatatt
ZMO1139	cccttatcaaggctacaaactgaaccatgacaatcatcgcgattggatcatggggtttcgaaaaggaaataaagc
ZMO1142	tctgctgtagagtgacaatcttttatctcatttccctttttgaaggagcgctggtg
ZMO0709	acatgcaacctttgaacagttcagcttagaaatccaagtctgatggtgtagggctttttt
ZMO1137	ggagaactggatgtttttttaggagattcggat
ZMO0187	tttttgccagagcttttctctgacagaaagaacgctcctgccgaatagaaaactgatgaaccgcctttgcgcagcttgaaaggaataata
ZMO0369	Tttacctgttgggtagccttctgattttagaaaggaattatt
ZMO0405	agacgagcttatttaggcatntgattttatccg
ZMO0937	aatatccctctttctattcctcgataaaaagaggttaaacgcaatttaatccctgaaacaaggataaagcataaggcttttctgtatctataaaatgatagcacaggataataaatttacttttattatagagataatgtcttttctttatatccattttatagctatattgatatataattattttttgtgatctgtatcatataacggataataatataatttgttatcctctttaataaaaagaggcttccctatatcgaaatgtcacatttcgatgatattattggttttttcctagaatttttttaaggg
ZMO1179	gtaacagtcgcctcatgaaaaaaattatacgcttggcatcaaagggagacg
ZMO1275	aatactatctgttttagattaaaagtaataacgccggaaaaacgttattgcccagacgcaggataaagc
ZMO0689	ctggaaaagccggcgaatcagataacagttccgcacaggtgagaaccacgacggatcttctctgaattgttggttagttaagaaagaaacaaggatt
ZMO0131	ctggcctcatttctctcgaaatgaaaatttcaaaaaccgatattttgtaggttcacgaggaaatagggtagaatcc
ZMO0748	aagacgcgccctacatataaatgaacagattattctgcctagacaatcatttttaaggacaggaaaa
ZMO1034	aatcagacaatattcataaactgcctgatacccatcatcttagagagaaaatattgggatcatatcactgtaaaacagttttccacttctctcttggagtaaggtcaaaata
ZMO1113	aattgaggcggctcgcaaaccaccacgccaaaggggggctgggtcgcttggaagagaatagaggtttcaa
ZMO0347	atcccgatggcgtttttcttttctttacaattctatgataattgaataatatttcattgtcctttgtaataaaggatagaaaaatggattggcctagtaaattgtttctgtgacatttgaaattttaaactatattatcatctcttgagagactcttcacataggcgtgcatccatatttgctggacgtggatgttttgtaatgatttttgtccagccaacaaaccggagaaccggacaatggccgaaaaggtcaacaatcttcaggattttttccttaataccttgcgcaagacccgcacaccggtgacg
ZMO1198	attcttccttcttatagtcaggctgtcacaaaagcctctttcgtcaaaaggagtgtcatcct
ZMO1478	atattctcaatagtctggcaatgagtcgtcaaaatagtttgaaactgttctaagacgacgccgttctgtaagtgagatagataaacctgctcaaaaggcagaaatgccgatagtccttaatgtctggcaggtatcaatggagaagaag
ZMO0275	atccaaaaggttgcctgttttcaatcggttaaacgggctgcctctttttttttctgccttttcccccgatatggatttgaga
ZMO1399	aataaaagcaacaaccttntgaacttcggttccgattgtgacccgatttgacgggtgtctttcggatctgcgaggattatgacg
ZMO1432	ttttacaaataggcataaaacgcttttcttaacaggtgatgtctgcccgattatgtatgatatgaccgtttacactataaacattgcacaagggtaaaggacaggccactgcgcctatcgactcttgctatgataggcgactgcgatgct
ZMO0367	agaagttggaccagccagacc
ZMO1412	gtaaacgttgggagatcaaccgagatctttccaaaggtaaaagctttaaagcttctctggaaagatagaatattctcccacttttttcctaaaaaaggattgttct
ZMO1048	gtatattttggctgtttgttgtaatattctacatcagcctaaacagatatgaggaaaaagatgatagctgtcagaagagagtcggataaatttt
ZMO0140	agcgctacagaaaaataggaatttcaccgccgtttttcccaaaaaacagtgatgtatacccctcgctattcagggataaaaaatctaaaattaaacataacgatggatgaatccgtttaatcctttatccgatccgaaacaaagcctttttttgcaggagtcctaaacccg
ZMO0366	agtttatcgccaaggattcggtttgtgatgttattttttggtcagaaactaaaataagaccaatgtttaacattgccgatactcggcgattgtaagatttacagattaaggcgggagaggaatcgcc
ZMO1612	attgattcgcccgaataaccattagagaagaatctgcatatcaacaggccattntccttgaaggtgttttttaacattagccgaatcagcaaaaagccagaaaatctggatttagccttgaaaaccgacttctctgacaccgtccgggaaaggagaatt

Table 4.3: 5' UTR sequences for final candidates confirmed by 5' RACE

Name	Sequence	Name	Sequence
5RACE_RSE1	TGACATTGGCACCCGCTTTG	5RACE_RSE26	CATCCCAAGATGCCTTCAACC
5RACE_RSE2	GAAATCGCCATGCCCTGACTG	5RACE_RSE27	GGGATCGACGCCATATTCAGC
5RACE_RSE3	TTGGTGTCGAACCATTTGGTC	5RACE_RSE28	GATCATCGGCGGTGCGTTTG
5RACE_RSE4	CCACCAGCCGCCAATTTACC	5RACE_RSE29	GGGAATCCAAGCGCCTTCTG
5RACE_RSE5	CATCAAGAAGATGGGCGACAACC	5RACE_RSE30	TGTTCCCGACTAAATCCCAATTCC
5RACE_RSE6	TTTCGACGGAGGTATCCTCAAG	5RACE_RSE31	AAAGCCAATCTTTGCCGCATC
5RACE_RSE7	CATTAACAAATCCCGTCCGCACTG	5RACE_RSE32	ACCACCTTTACCGTCTTCTGGAG
5RACE_RSE8	CTGAAGGATGAAAGCCGAAATCTGG	5RACE_RSE33	CACCTGGCGTATTGAAATCATTGG
5RACE_RSE9	ATGCGGCTTGGCAGTGATTTC	5RACE_RSE34	CATTACGGGTCAACAAGACGCAG
5RACE_RSE10	TTTCTTCCGCTTGACGTTTGG	5RACE_RSE35	AATCGTGGACAAGGTCGCTTC
5RACE_RSE11	CTTGGTGCAATGGCGGGTAATG	5RACE_RSE36	GCCATCATTACCCATACCCAACC
5RACE_RSE12	GAAACCATCACAGGTAGCGCAAG	5RACE_RSE37	TGGAGTTCAAGGACGGAATCG
5RACE_RSE13	TCTACAGAATGGGCAGGATCAGG	5RACE_RSE38	GGATGACCAAAGCCGTCTGAAG
5RACE_RSE14	GAAGCAAAGCATCCCTGTGG	5RACE_RSE39	CGCTGGAAACGATTGGTCAGATG
5RACE_RSE15	CGTCTTTCGGGAAGCGCATC	5RACE_RSE40	CATGAAGCGCGGACAAGATCAC
5RACE_RSE16	TGCCAAGAGCGTGATGCAAC	5RACE_RSE41	GCCGACAGCATAACCGATACAG
5RACE_RSE17	TTTGGCTAATTCGGGATCGAGATG	5RACE_RSE42	CTACGCTAAGAAAGCCGATAACAG
5RACE_RSE18	ATTGCGCGGATTGTGATTCTTCG	5RACE_RSE43	AGAAGTTGGACCAGCCAGACC
5RACE_RSE19	AGCTTGGCAAAGAAACGCCATC	5RACE_RSE44	GCATTAAAGGCAGGTCGCTACG
5RACE_RSE20	TCCCGGACCAAGAATAGTGATAAC	5RACE_RSE45	CGGTATTGTTCCGGCGTCAG
5RACE_RSE21	TGACAGCATCCAATCGGCTC	5RACE_RSE46	AATACCGGCCAACAATTCAACC
5RACE_RSE22	CGGATTGCCTGCATGGATGAG	5RACE_RSE47	GGTCAGCTACCGATTTGCCTTC
5RACE_RSE23	TCTGTGCTAAAGATCGTGGAG	5RACE_RSE48	CACCGATACCTAAACCGGCAAG
5RACE_RSE24	CCATAAAGGACGATATTCGCACC	5RACE_RSE49	AGCAGCAGGGTCGGGATATTG
5RACE_RSE25	GCTTATCTTGGCTGTCGGCTATAC	5RACE_RSE50	ACAATACCGCTGCACCATCC

Table 4.4: Primer sequences for 5' RACE

4.3.4 High-throughput fluorescence-based screening system for 5' UTR

To test each 5' UTR candidate's ability to regulate downstream gene expression, we generated in vivo fluorescence-based screening systems in *Z. mobilis*. Previously, it has been demonstrated that GFP can be efficiently expressed in *Z. mobilis* (Douka et al, 2001) and we confirmed inducible expression of GFP in our plasmid under promoter ptet (Figure 4.6A) Figure 4.6C shows the fluorescence shift from this system when it is induced with tetracycline compared to the uninduced sample. Here, we have confirmed our fluorescence system is functional in *Z. mobilis*. Then, we selected the well-characterized theophylline synthetic riboswitch as a test case for establishing our screening system mediated by 5' UTRs. The theophylline riboswitch controls gene

expression at the translational level by binding to a small molecule, theophylline, which triggers structural changes by helix slippage to increase gene expression (Suess et al, 2004). The theophylline riboswitch was engineered as a synthetic riboswitch system to control gene expression in various bacterial species (Lynch et al, 2007; Topp et al, 2010). In this study, we cloned the theophylline riboswitch element in front of GFP gene in *Z. mobilis* (Figure 4.6B) and compared the level of GFP expression with 2mM theophylline compared to a DMSO control in *Z. mobilis*. Figures 4.6C and D showed that about a 2 fold change was observed. Even though this level of change in GFP was not high, it could be due to the pairing strength of region between RBS (ribosome binding site) and the aptamer in our version of theophylline switch, which determines the function of riboswitch (Lynch et al, 2007). Additionally, even with the same theophylline riboswitches, activation ratio was different depending on bacteria species (Topp et al, 2010). This could be optimized by engineering the region between the aptamer and RBS in the theophylline switch in *Z. mobilis* as a future study. Still, we successfully used the theophylline synthetic riboswitch as in vivo fluorescence-base screening system in *Z. mobilis* for the first time. This can be a very useful tool for screening the control of gene expression in metabolic pathway related with ethanol tolerance or other stress response.

Utilizing the pEZ-tet-GFP vector, we developed high-throughput cloning strategy for the efficient cloning of each UTR-containing GFP plasmid by the combination of Golden Gate assembly and PheS counter selection (Figure 4.6E). Given that nucleotides in the coding region may affect the structure of 5' UTRs for the regulation of the gene, each verified UTR sequence by 5' RACE along with 90 base pairs of the downstream coding region was used for the generation of UTR-GFP libraries. Primers used for the generation of UTR-GFP libraries were shown in Table 4.5.

98
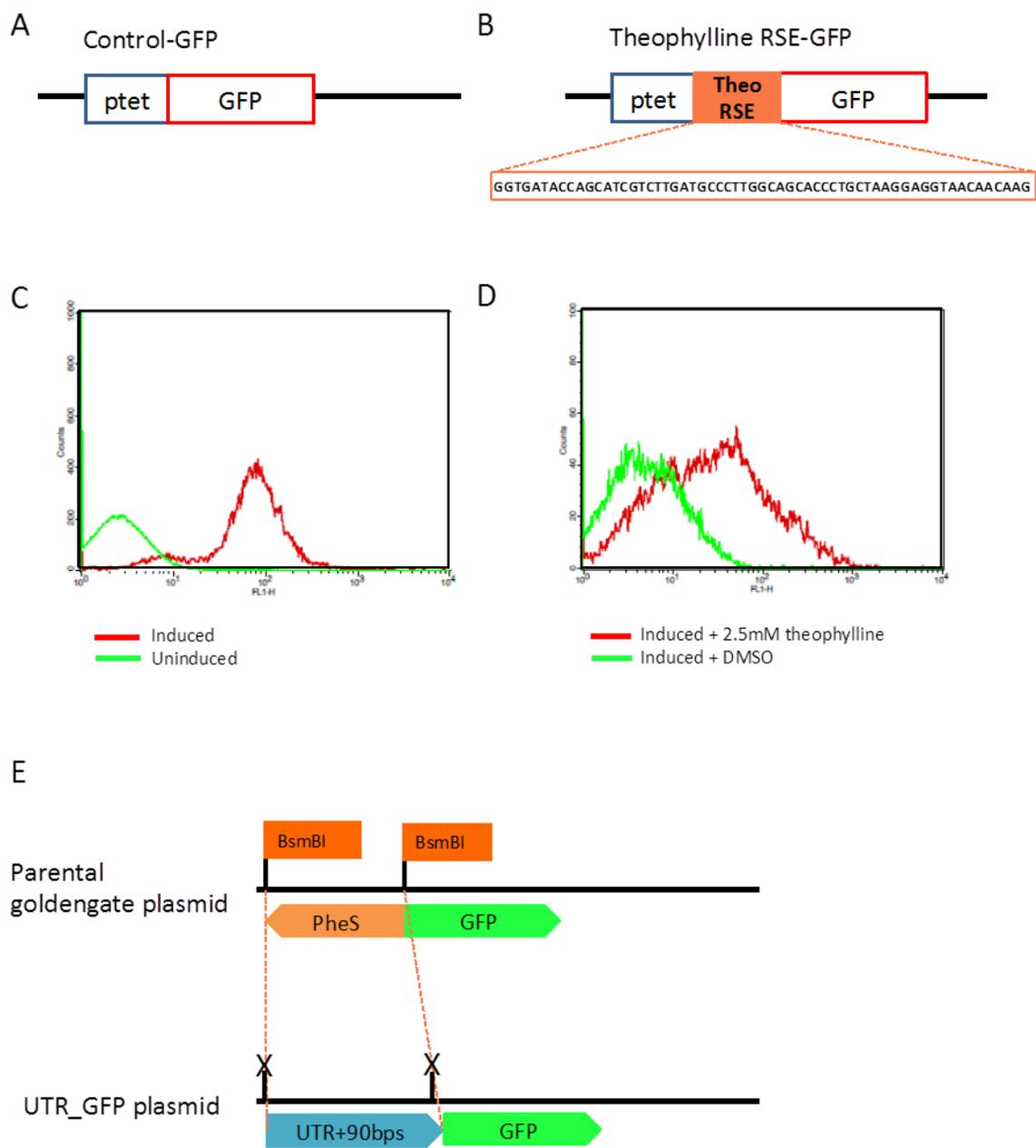
Figure 4.6: Establishing 5' UTR dependent GFP expression in *Z.mobilis*

Target Gene	Sequences	Target Gene	Sequences
ZMO0172	F- agatt CGTCTC TCTCCC tccccgggggggccgtataa R- agatt CGTCTC CACCAT atttttaccttcaacgtaaattttacgggaa	ZMO1275	F- agatt CGTCTC TCTCCC aatactatctgttttagatt R- agatt CGTCTC CACCAT tctaataatggcattct
ZMO0979	F- agatt CGTCTC TCTCCC ggaaattttttgcatagg R- agatt CGTCTC CACCAT ctgataggcttctttg	ZMO0689	F- agatt CGTCTC TCTCCC ctggaaaagccggcgaat R- agatt CGTCTC CACCAT acgggtcagatttggcgt
ZMO1000	F- agatt CGTCTC TCTCCC aaggattaaggttctttgtc R- agatt CGTCTC CACCAT acggccaccccagaaac	ZMO0131	F- agatt CGTCTC TCTCCC ctggcctcatttctctc R- agatt CGTCTC CACCAT aacaccggcataaaggc
ZMO0547	F- agatt CGTCTC TCTCCC gaatgaccatttccatt R- agatt CGTCTC CACCAT aagccgataaggtttgtat	ZMO0748	F- agatt CGTCTC TCTCCC aagacgcgccctacata R- agatt CGTCTC CACCAT tatccatacttcatggt
ZMO0056	F- agatt CGTCTC TCTCCC atctcaaagacagtccg R- agatt CGTCTC CACCATatcataaccgcgatattcc	ZMO1034	F- agatt CGTCTC TCTCCC aatcagacaatattcat R- agatt CGTCTC CACCAT ttgggcggaagcct
ZMO0376	F- agatt CGTCTC TCTCCC aaaaggtgtgagcattttg R- agatt CGTCTC CACCAT cgattttttcacgaccga	ZMO1113	F- agatt CGTCTC TCTCCC aattgaggcggctcgca R- agatt CGTCTC CACCAT gaaacgttcccccaaag
ZMO0546	F- agatt CGTCTC TCTCCC aatttggtcacattgt R- agatt CGTCTC CACCAT aataagcgcgaaggtcg	ZMO0347	F- agatt CGTCTC TCTCCC atcccgatggcgtttttc R- agatt CGTCTC CACCAT ctgtgactgaccatctc
ZMO0660	F- agatt CGTCTC TCTCCC gtagacttgttccaaggg R- agatt CGTCTC CACCAT tgcgtttcaataacctt	ZMO1198	F- agatt CGTCTC TCTCCC attcttccttcttatag R- agatt CGTCTC CACCAT tttattccgcaaaatat
ZMO1069	F- agatt CGTCTC TCTCCC tttgctggtggacagaa R- agatt CGTCTC CACCAT taatcttacttttctgggcagc	ZMO1478	F- agatt CGTCTC TCTCCC atattctcaatagtctg R- agatt CGTCTC CACCAT ggcttgcttgataataa
ZMO1139	F- agatt CGTCTC TCTCCC cccttatcaaggctacaa R- agatt CGTCTC CACCAT tgctccgccgggatagc	ZMO0275	F- agatt CGTCTC TCTCCC ataaaaagcctatcttg R- agatt CGTCTC CACCAT cgcgatagaaataatgg
ZMO1142	F- agatt CGTCTC TCTCCC tctgctgtagagtgaca R- agatt CGTCTC CACCAT tccggcacgagccg	ZMO1399	F- agatt CGTCTC TCTCCC aataaaagcaacaacctt R- agatt CGTCTC CACCAT caactctttcaatgtca
ZMO0709	F- agatt CGTCTC TCTCCC acatgcaacctttgaa R- agatt CGTCTC CACCAT gggtgcaatggctttga	ZMO1432	F- agatt CGTCTC TCTCCC ttttacaaataggcataa R- agatt CGTCTC CACCAT aagcccaatccgcatag
ZMO1137	F- agatt CGTCTC TCTCCC ggagaactggatgtttt R- agatt CGTCTC CACCAT tttaccgccctttgaag	ZMO0367	F- agatt CGTCTC TCTCCC ttaaacttgctttggctg R- agatt CGTCTC CACCAT atcaagaccataaagcg
ZMO0187	F- agatt CGTCTC TCTCCC tttttgccagagcttt R- agatt CGTCTC CACCAT agccttatccagtttatcg	ZMO1412	F- agatt CGTCTC TCTCCC gtaaacgttgggagatc R- agatt CGTCTC CACCAT aacagaaacgctgttgt
ZMO0369	F- agatt CGTCTC TCTCCC tttacctgttgggtagc R- agatt CGTCTC CACCAT tccaagagaaagaaccc	ZMO1048	F- agatt CGTCTC TCTCCC gtatattttggctgtttg R- agatt CGTCTC CACCAT gcacagggaacggaaaa
ZMO0405	F- agatt CGTCTC TCTCCC agacgagcttatttagg R- agatt CGTCTC CACCAT gacggctggcgatag	ZMO0140	F- agatt CGTCTC TCTCCC agcgctacagaaaaata R- agatt CGTCTC CACCAT gctgggcgagatcac
ZMO0937	F- agatt CGTCTC TCTCCC aatatccctctttctat R- agatt CGTCTC CACCAT tgttaccgatgaaagac	ZMO0366	F- agatt CGTCTC TCTCCC agtttatcgccaaggat R- agatt CGTCTC CACCAT aaccgctgaatcgtaac
ZMO1179	F- agatt CGTCTC TCTCCC gtaacagtcgcctcatg R- agatt CGTCTC CACCAT tatgatgcgccccgaat	ZMO1612	F- agatt CGTCTC TCTCCC attgattcgcccgaata R- agatt CGTCTC CACCAT tggggtctggttatcaat

Table 4.5: Primer sequences for the amplification of UTRs + 90 pbs of ORF

4.3.5 Ethanol stress-responsive regulatory 5' UTR was efficiently identified

As high tolerance to ethanol is one of the desirable features of *Z. mobilis*, we initially screened UTR-GFP libraries under an ethanol stress condition, which allows us to see any potential 5' UTR activation more clearly. For the efficient identification of ethanol stress responsive regulatory 5' UTRs, we tested 5% ethanol supplemented media

100

and compared fluorescence level with normal media (RM media). As a control, a strain with only GFP and no UTR (Control-GFP) was used. All the experiments were done in triplicate. After we confirmed signal difference was much higher at 10 h post induction, we collected fluorescence at 10 h post induction for the rest of the experiments. This is consistent with previous data that showed the maximum fluorescence in late exponential phase (Douka et al, 2001). We identified 2 candidates in which changes in fluorescence level under ethanol stress were measured by the median fluorescence values relative to control-GFP strain. These candidates corresponded to the 5' UTR of ZMO0347 (RNA binding protein Hfq, UTR_ZMO0347) and ZMO1142 (thioredoxin reductase, UTR_ZMO1142). To explore responsiveness of UTRs under different levels of ethanol, we further examined these two ethanol-responsive candidates under 1%, 3% and 5% (v/v) of ethanol (Figure 4.7). Importantly, we found that the fluorescence of the UTR_ZMO0347 (RNA binding protein Hfq) strain was decreased about the same level under different concentrations of ethanol. However, GFP expression of UTR_ZMO1142 was downregulated gradually depending on the concentration of ethanol. In contrast to the 1% ethanol stress, which induced almost no change in fluorescence, 5% ethanol stress on UTR_ZMO1142 induced about 60% decrease in fluorescence level (Figure 4.7). Compared to no UTR containing control GFP (Control-GFP) expression under ethanol stress, the fluorescence level of UTR_ZMO0347 and UTR_ZMO1142 showed significant changes: 40% and 60% decrease in fluorescence under 5% ethanol stress, respectively. Interestingly, it was reported that thioredoxin reductase (ZMO1142) protein was less abundant under ethanol stress, but the transcript level was increased (Yang et al, 2013). However, decreased transcript and increased protein for ZMO0347 were detected under ethanol stress (Yang et al, 2013). This data is also consistent with our own transcriptomic data (pending publish). GFP expression in the cell was also confirmed by Western blot

analysis, which was corresponded with fluorescence data (Figure 4.7). These observations indicated that UTRs of these genes are part of a biologically relevant mechanism by which gene expression is controlled to protect the cells from ethanol stress. Even though the complex regulatory network between ZMO0347 and ZMO1142 and ethanol needs to be further elucidated, it is worth noting that regulatory function of UTR_ZMO0347 and UTR_ZMO1142 is highly associated to the ethanol stress response in *Z. mobilis*.
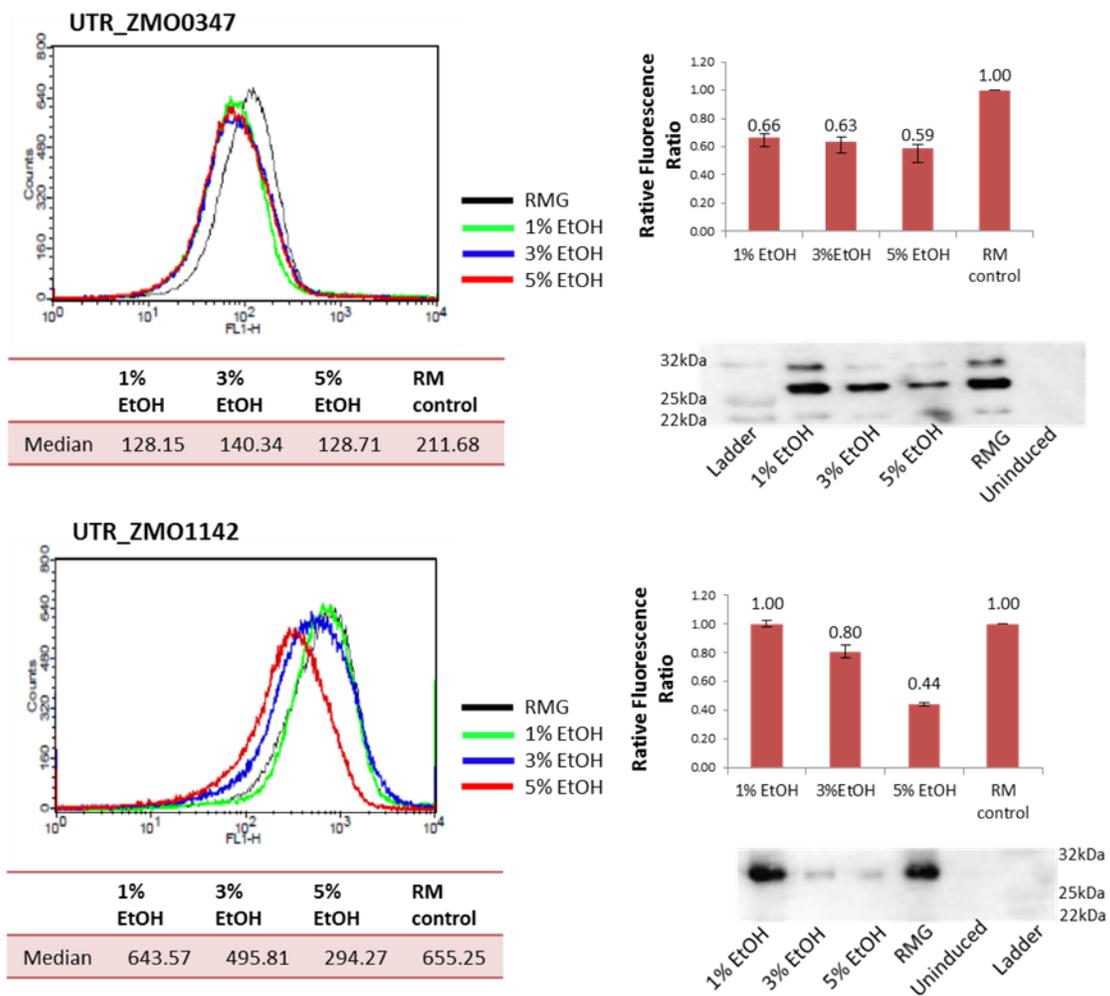
Figure 4.7: The effect of ethanol sress on 5' UTRs
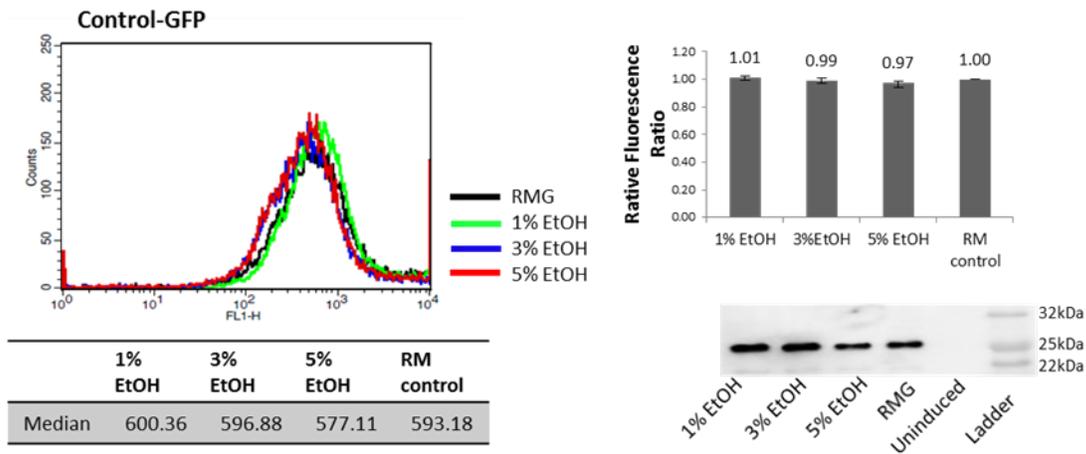
Figure 4.7 (cont.): The effect of ethanol sress on 5' UTRs

4.3.6 5' UTRs found responding to acetate and xylose stress conditions

Understanding of acetate tolerance and utilization of xylose are important for ethanol production in *Z. mobilis*. Therefore, in addition to ethanol stress, we evaluated the effect of acetate and xylose on 5' UTRs in *Z. mobilis*. We supplemented acetate (sodium acetate 10 g/L) and xylose (1%) into the media for the analysis. We identified 4 novel 5' UTRs under acetate stress: UTR_ZMO0172 (thiamine biosynthesis protein), UTR_ZMO1000 (5-methyltetrahydropteroyltriglutamate), UTR_ZMO0546 (sulphate transporter) and UTR_ZMO1478 (6-phosphogluconolactonase). UTR_ZMO0187 (thiamin biosynthesis protein) and UTR_ZMO0187 (3-deoxy-7-phosphoheptulonate synthase) showed 40% decrease and 400% increase in fluorescence under xylose stress, respectively (Figure 4.8). Interestingly, UTR_ZMO0172 (thiamine biosynthesis protein) was responsive to both acetate and xylose stress. This observation suggests that ZMO0172 is associated with both of metabolic pathways for xylose fermentation as well as controlling acetate stress.

GFP expression of UTR_ZMO0172 under acetate and xylose supplementation was decreased about 40% when compared to normal condition. However, GFP expression of UTR_ZMO1000, UTR_ZMO0546 and UTR_ZMO1478 was increased upon acetate stress. In contrast to UTR_ZMO0172, UTR_ZMO0187 activated GFP expression in response to xylose stress. Our data were consistent with the previous reports that these genes were shown to be up or down regulated under acetate/xylose supplemented conditions (Yang et al, 2014a). Further analysis will provide us clues for understanding the complete metabolic mechanism under stress.
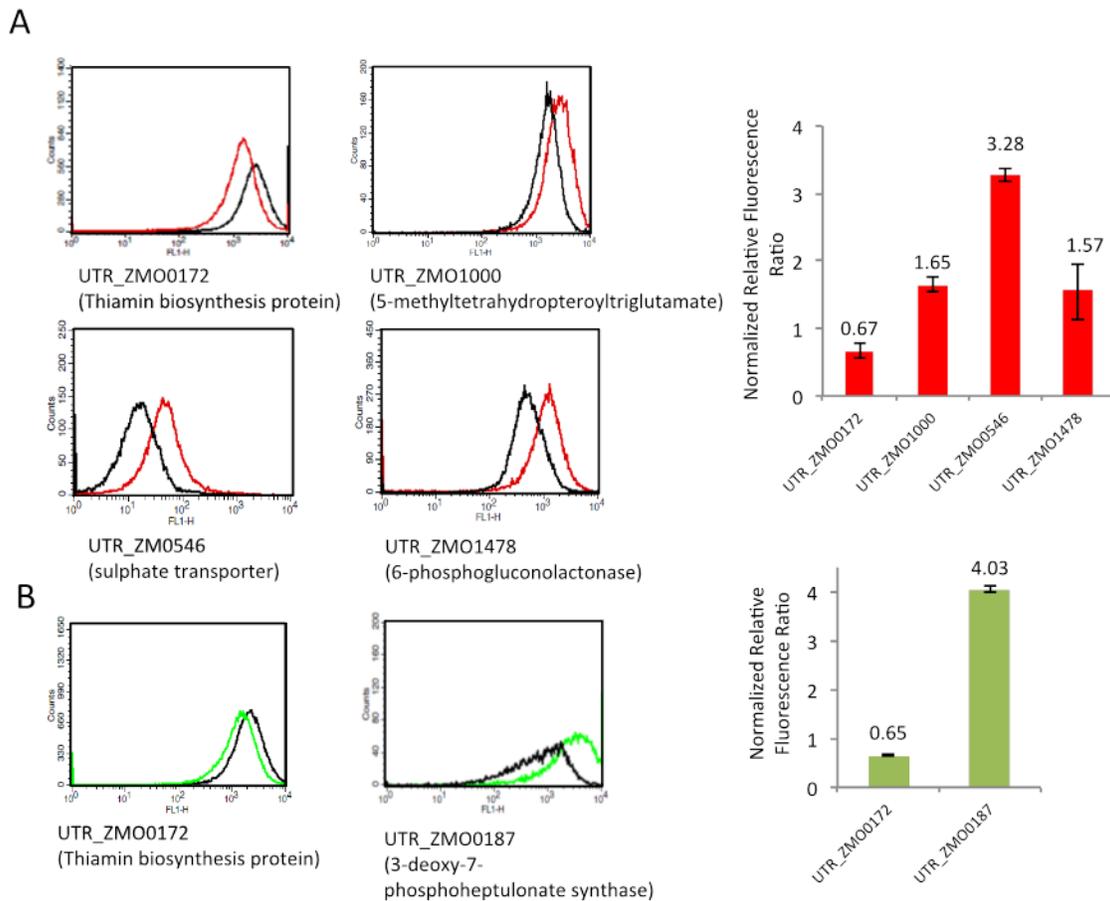
Figure 4.8: The effect of acetate (A) and xylose (B) on 5' UTRs

104

4.3.7 Genetic studies on the effect of verified 5' UTRs under stress conditions

To evaluate the physiological effects of ethanol stress-responsive regulatory 5' UTRs in *Z. mobilis*, we constructed a 5' UTR deletion strain for candidate UTR_ZMO0347 (ethanol stress). We selected this candidate that seemed to be more promising than other candidates depending on the feature of the genes and fold change in fluorescence. ZMO_0347 encodes Hfq that is highly associated with stress response in various organisms (Guisbert et al, 2007; Torres-Quesada et al, 2014). Furthermore, UTR of ZMO0347 repressed downstream gene expression under ethanol stress in this study. Another promising candidate, UTR_ZMO1142 seems to be very interesting but it is excluded for the construction of deletion strain because the size of the candidate was too short (only 57 bp) for the deletion, as we need to leave predicted promoter and RBS regions. We utilized a homologous recombination deletion technique to disrupt 5' UTR regions (except for promoter and RBS regions) of each gene in *Z. mobilis* 8b strain. Spectinomycin gene was used as a selection marker between about 1 kb up and down stream homology arms. The deletion of the target region was confirmed by PCR of genomic DNA (data not shown). Constructs for deletion of each 5' UTR are shown in Figure 4.9A. After the deletion was confirmed, we tested viability of WT and ΔUTR_ZMO0347 strain under stress conditions. First of all, ΔUTR_ZMO0347 strain was cultured under ethanol stress condition including 1%, 3%, 5% ethanol. Figure 4.9B showed growth curve for WT and ΔUTR_ZMO0347 under different stress condition and Table 4.6 showed specific growth rates. WT strain was used as a control. Remarkably, growth of ΔUTR_ZMO0347 strain was affected by ethanol level. Specific growth rate of ΔUTR_ZMO0347 was more than 2 fold decrease under 5% ethanol stress. Also, deletion strain showed extended lag phase under 5% ethanol stress. Compared to WT strain, growth of deletion strain under 1% and 3% ethanol slightly slow. Additionally, we tested

the effect of acetate stress on growth, but WT and mutant strain showed similar growth rate. Taken together, this data suggests that UTR of ZMO0347 negatively regulates the expression of ZMO0347 upon ethanol stress and regulation of ZMO0347 via UTR region may confer ethanol tolerance in *Z.mobilis*.
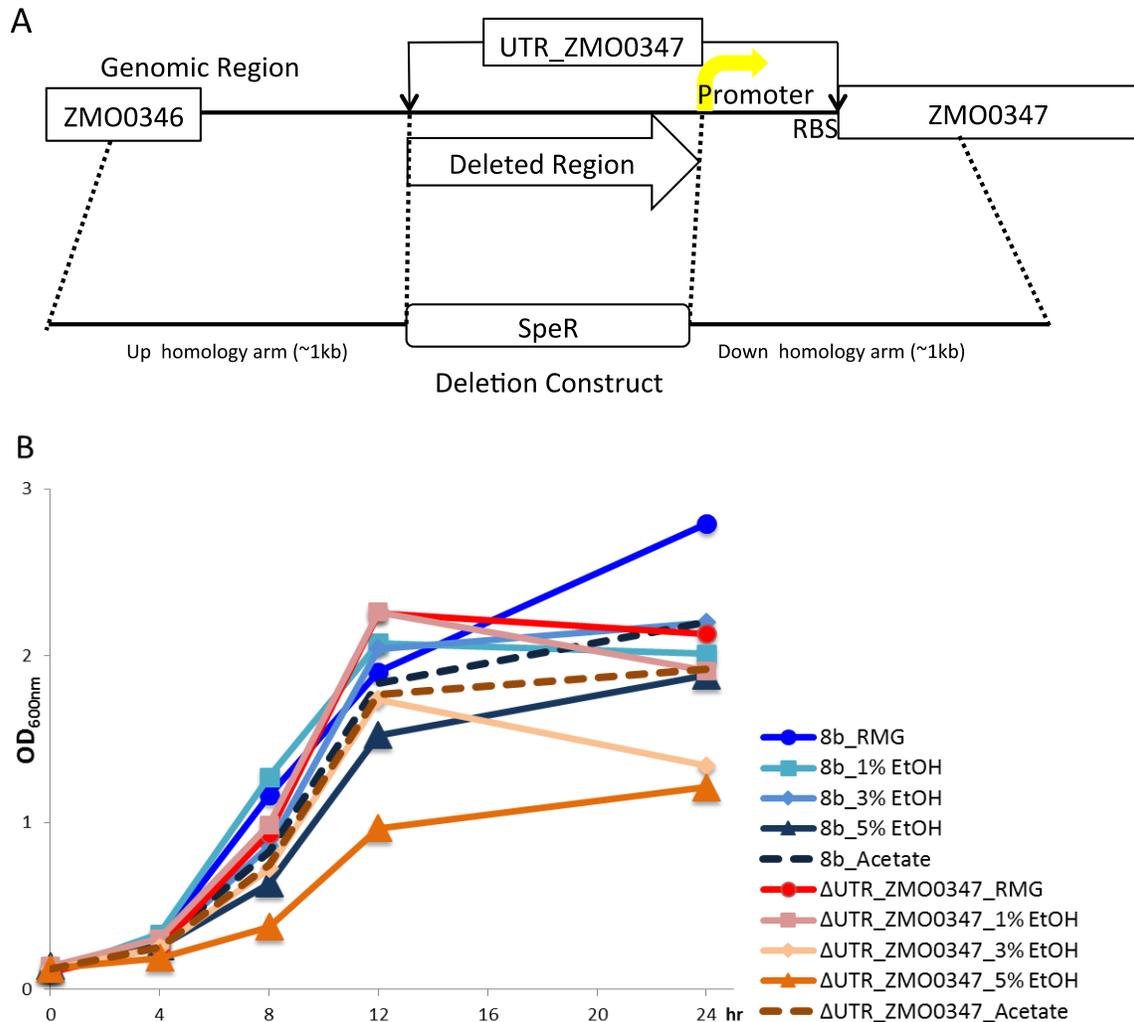
Figure 4.9: Deletion of UTR_ZMO0347 confirms the physiological role in *Z. mobilis* under ethanol stress. (A) Deletion construct for UTR_ZMO0347 and (B) Growth curve of WT and ΔUTR_ZMO0347 strain under 1%, 3%, 5% and Acetate (10g/L) compared with normal RMG media.

Strain	8b_WT					ΔUTR_ZMO0347				
Stress	1% RMG	3% EtOH	5% EtOH	10g/L EtOH	Acetate	1% RMG	3% EtOH	5% EtOH	10g/L EtOH	Acetate
Specific growth rate (h^{-1})	0.2	0.21	0.21	0.16	0.19	0.25	0.24	0.18	0.09	0.18

Table 4.6: Specific growth rate of WT and ΔUTR_ZMO0347 under ethanol and acetate stress
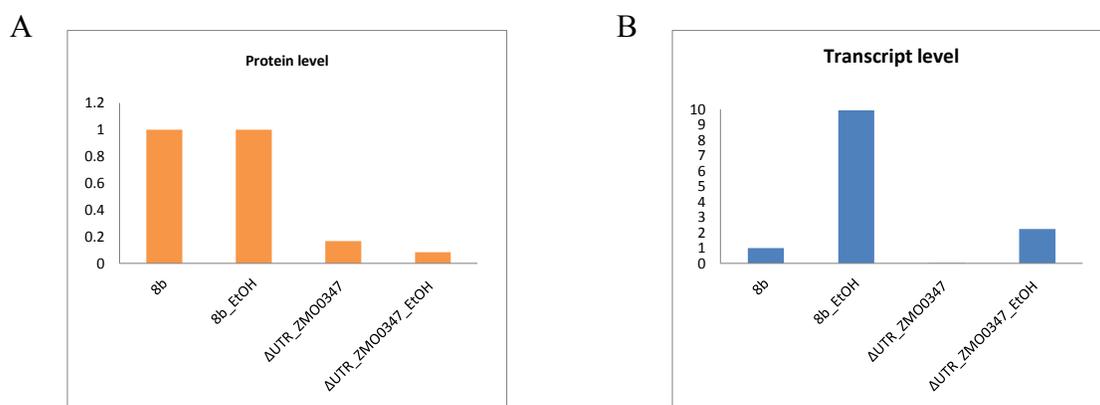
Figure 4.10: The level of proteins and transcripts were confirmed in WT and ΔUTR_ZMO0347 under normal and 5% ethanol stress condition. (A) Relative protein levels were calculated by proteomics analysis and (B) Relative transcript level was measured by qRT_PCR.

To confirm the level of protein expression in WT and ΔUTR_ZMO0347 strains, we performed proteomics analysis. Even though the detection level was low, Hfq protein was decreased about 6-fold in mutant strains compared to WT strains in both normal and ethanol stress conditions (Figure 4.10A). We also quantified the level of transcripts in WT and ΔUTR_ZMO0347 strains by qRT_PCR. Interestingly, the expression of transcripts in WT strain was 10-fold increased under ethanol stress. In ΔUTR_ZMO0347 strain, 20 times more transcripts were detected under ethanol stress compared to normal

condition even though overall expression level of ΔUTR_ZMO0347 strains was about 10 fold less than WT strains (Figure 4.10B). Therefore, we confirmed that Hfq protein was still expressed in ΔUTR_ZMO0347 strain and appropriate level of protein was important for the cell growth under ethanol stress condition. Consequently, regulation of transcript as well as protein level by UTR is essential for the cell growth in response to ethanol stress.

4.4 DISCUSSION

Bacteria utilize 5' UTR elements for rapid sensing of and responding to environmental changes so that they orchestrate a cascade of alterations in gene expression and protein activity (Oliva et al, 2015). Therefore, 5' UTR elements contribute to the comprehensive gene regulation under stress conditions (Oliva et al, 2015). Recent advances in transcriptome analysis provide useful information about regulatory 5' UTRs. Our approach appears to be effective as we identified all of the previously annotated in the genome, and, even though many of them were not experimentally verified, we found many of novel 5' UTR candidates. Thus, our findings highlight the utilization of transcriptome data in combination with computational analysis for identifying regulatory 5' UTRs. Conventionally, sequence-based conservation analysis such as Rfam is widely used to identify UTRs, but this kind of approach is limited to the known UTRs and to the identification of functional homologs from closely related species since it is based on alignments of UTRs across organisms. Therefore, using high-throughput transcriptomic data to identify novel 5' UTRs and functional homologs of known 5′ UTRs in less-studied bacteria could significantly improve the identification of UTRs.

Several recent studies have uncovered that the ligands and environmental signals that trigger 5′ UTR-mediated regulation are diverse (Caldelari et al, 2013; Shapiro & Cowen, 2012). As the identification of new classes of 5' UTRs is increasing via various approaches, finding their corresponding signals will become more difficult. The GFP-based reporter systems we have developed here for the validation of 5′ UTR-mediated regulation should be useful in addressing this challenge, providing an efficient way to screen a large number of candidate 5′ UTRs in a wide variety of conditions in *Z.mobilis*.

In this study, our finding suggests that 5' UTR elements in *Z. mobilis* have central regulatory roles, particularly under stress conditions. Particularly, UTR_ZMO0347 (RNA binding protein Hfq) and UTR_ZMO1142 (thioredoxin reductase) associated with ethanol stress response showed down-regulation of downstream gene expression upon stress in a dose-dependent way. The downstream gene of UTR_ZMO0347 is a homolog of RNA binding protein Hfq, which has been known as an RNA chaperone and regulator of the small RNA network. Hfq also mediates transcription anti-termination via binding to Rho factor for the control of gene expression at transcription level in *E. coli* (Rabhi et al, 2011). Previous studies indicated that Hfq homolog in *Z. mobilis* associated with stress responses (Yang et al, 2009b). Remarkably, previous transcriptomics studies indicated that levels of the same transcript (ZMO0347, an Hfq homolog in *Z. mobilis)* were associated with stress responses and naturally down regulated under ethanol stress (He et al, 2012; Yang et al, 2013; Yang et al, 2009b). This suggests that the discovered associated UTR could also act at the post-transcriptional level, perhaps by inducing transcript degradation. Interestingly, it was also reported that, post-translationally, levels of the encoded protein by ZMO0347 increases under ethanol stress. Although these experiments were one in a variant strain from that used in our studies (8b strain) and under stlightly different conditions (ZM4), these observations suggest that other

regulatory factors (besides the UTR_ZMO0347) that contribute to up regulating this protein under with ethanol stress. For instance, it is known that expression of the Hfq homologue in under bacterial species (like in *E.coli*) is also regulated by the global carbon storage regulator (csr) system. However, from the decreased growth effect that we find upon deletion of this UTR element, we conclude that this natural level of local transcript regulation is important during ethanol stress for basic organism survival. Further studies may be necessary for the elucidation of fundamental direct mechanism.

Besides UTR_ZMO0347 as ethanol stress related UTRs, UTR_ZMO1142 (thioredoxin reductase) seems to be responsive to high concentration of ethanol (1% vs 5%). Thioredoxin reductase catalyzes the reduction of thioredoxin coupled with NADPH. It is worth noting that thioredoxin plays a major role in defense mechanism for the oxidative stress via the reduction of disulfide bonds by thioredoxin reductase (Koharyova & Kolarova, 2008).

We also identified acetate and xylose responsive 5' UTR elements. Acetate toxicity and xylose utilization has been studied for strain improvement of *Z. mobilis*. A previous study confirmed that regulatory mechanisms in responding to acetate stress and xylose utilization mainly involved with carbon and energy metabolism to reduce the impact of stress on the cell (Yang et al, 2014a). Acetate stress induced down-regulation of genes associated with flagellar system, glycolysis and up-regulation of the genes related to stress responses and energy metabolism. In this study, we confirmed that UTR_ZMO1478 (6-phosphogluconolactonase) activated GFP expression under acetate stress. 6-phsphogluconolactonase catalyzes hydrolysis of the ester linkage of lactone resulting in production of 6-phophogluconate in pentose phosphate pathway. Transcriptome data confirmed that the level of ZMO1478 was up-regulated under acetate stress (Yang et al, 2014a). Taken together, up-regulation of ZMO1478 by UTR is closely

related with the regulation of carbon metabolism for adapting acetate stress. UTR_ZMO0172 (thiamine biosynthesis protein) showed down-regulation of GFP expression in both acetate and xylose stress. When xylose fermentation was processed, acetate (acetic acid) was produced as a byproduct. These observations indicate that UTR_ZMO0172 is involved in multiple responses to metabolic stressors. Further studies on protein expression level of ZMO0172 help us understand the underlying regulatory mechanism of UTR_ZMO0172. Overall, our study has provided insight into the regulatory role of UTR regions against ethanol tolerance in *Z. mobilis*. Ultimately, this could contribute to the enhanced ethanol tolerance phenotype for strain engineering purpose.

Chapter 5

Conclusion and Perspectives

The work presented in this dissertation describes the discovery and characterization of regulatory noncoding RNAs including small RNAs and 5' UTRs associated with the ethanol stress response utilizing transcriptomic analysis combined with computational prediction methods. Through this analysis, we have also identified potential targets of ncRNAs that could allow rewiring of associated metabolic pathways to contribute to strain engineering in *Z. mobilis,* an organism of high biotechnological interest given its high ethanol producing capabilities.

To identify novel sRNAs in *Z. mobilis*, RNA-sequencing was performed for aerobic and anaerobic condition, which showed different levels of ethanol production. Computational prediction of sRNAs and transcriptomic analysis between aerobic and anaerobic conditions provided the list of potential candidates, which were further confirmed experimentally. Among 15 sRNAs identified by Northern blot analysis, Zms2, Zms4 and Zms6 showed differential expression between aerobic and anaerobic conditions. Moreover, Zms2 and Zms6 were differentially expressed under 5% ethanol stress. These data suggest that their regulatory mechanism is associated with ethanol and emphasized the importance of the further characterization of these sRNAs. This work was the first demonstration of the presence of sRNAs in *Z. mobilis*.

Generation of sRNA overexpression libraries and deletion strains for Zms4 and Zms6 confirmed their regulatory effect on ethanol tolerance and, combined with target predictions through transcriptomics and proteomics analysis, elucidated their potential targets. This knowledge of potential regulatory networks between sRNAs and their targets could significantly contribute to the strain improvement for ethanol tolerance.

Regulatory RNA regions within a transcript, particularly in the 5' UTR, have been shown in a variety of organisms to control the expression levels of these mRNAs in response to various metabolites or environmental conditions. Therefore, transcriptomic data was searched in this work for regulatory 5' UTR candidates in *Z. mobilis*. After transcriptomic and computational analysis, the predicted 5' UTRs were experimentally verified. Under various stresses, 5' UTRs were tested via fluorescence based screening system, revealing UTR_ZMO0347 (hfq RNA binding protein) is responsive to ethanol. The deletion of UTR_ZMO0347 (except for promoter and RBS) significantly affected cell growth rate under 5 % ethanol stress, showing slow growth as well as long lag phase under 5% ethanol relative to the wild type strain, even though there was not much difference between this deletion strain and the wild type strain in normal RMG media. Therefore, UTR_ZMO0347 appears to have physiological roles in the ethanol stress response regulatory network in *Z. mobilis*.

We have demonstrated application of a novel bioinformatics process to accelerate the discovery of specific pathways and extract insight about regulatory mechanisms that could be further optimized to enhance a given complex phenotype. This work represents the first application of a de novo sRNA engineering strategy in non-model *Z. mobilis* that is of relevance to biofuel technologies. Overall, this study would be of strong interest to the regulatory non-coding RNA and microbiology communities. This study also serves a biotechnological community that continues to search for new metabolic strategies to engineer organisms of relevance to biofuel synthesis.

Appendix

Synthetic chimeras with orthogonal ribosomal proteins increase translation yields by recruiting mRNA for translation as measured by profiling active ribosomes

A1 INTRODUCTION

The use of natural or engineered "parts" from ribosomes has shown significant potential in expanding biotechnological capabilities. Highly versatile roles have been assigned to single ribosomal proteins, outside the translation role performed by the ribosome. Ribosomal proteins that are located within the ribosomal polypeptide exit tunnel have been particularly linked to enhancing translation of complex, non-native proteins in *E. coli* (Contreras-Martinez et al, 2012). Other ribosomal proteins have been associated with translational regulation and antibiotic resistance (Wilson et al, 2001). As effective use of these ribosomal parts continues to be demonstrated for various applications in biotechnology, efforts to engineer ribosomal components for enhanced synthesis of non-natural proteins continue to rise (Filipovska & Rackham, 2013; Jewett et al, 2013). For instance, ribosome-inspired small molecule machines have been recently suggested as a way to synthesize peptide sequences from amino acids (Lewandowski et al, 2013). Moreover, orthogonal ribosome-mRNAs pairs have enabled expansion of the genetic code (Weinberg et al, 2007) and regulation of mRNA affinity to ribosomes (e.g. by engineering ribosome binding sites) (Salis et al, 2009).

---

[2] Cho SH designed the study, performed experiments and prepared the manuscript.

In this work, we explore the contribution of an engineered version of ribosomal protein L29 (L4H2) to translation yield enhancement. This protein variant L4H2 contains 5 mutations (V11I, E13G, L21Q, Q39R and V46A) relative to the wild-type (wt) L29 and has been reported to contribute to the expression of complex heterologous proteins in bacteria (Contreras-Martinez et al, 2012). Specifically, since overexpression of L4H2 was demonstrated to augment expression of a number of nonnative proteins in *E. coli*, we reasoned that L4H2 could be used as a general protein expression-enhancing factor. A few key characteristics of the L29 protein is that it is located at the surface of the ribosomal exit tunnel, relatively small (10 kDa), non-essential, and poorly incorporated into ribosomes when overexpressed from a plasmid source. A central premise of this work is that synthetic fusions of target proteins to L4H2 enhance protein expression by promoting affinity (in the absence of assembly) to actively translating ribosomes.

In other studies, engineering of direct associations with ribosomes via tethering with a natural ribosomal protein L23, has been shown to have beneficial outcomes in protein expression (Kristensen & Gajhede, 2003; Sorensen et al, 2004); L23 exhibits strong affinity to ribosomes and represents another attractive anchor point for protein expression technologies given its C-terminal exposure to the cytosol. However, the fact that L23 and many ribosomal proteins are essential for translation (Wegrzyn et al, 2006) and compromise any gained benefits of enhanced protein synthesis.

Given the continual interest in developing recombinant protein technologies in bacterial hosts (Costa et al, 2014; Nettleship et al, 2010), several strategies have been developed to address enhancement of complex protein (e.g. human) production. A few traditional approaches have encompassed overexpression of cellular factors to alleviate cellular toxicity (Saida, 2007), protein modifications to improve intrinsic solubility, including the use of fusion partners (Sorensen & Mortensen, 2005), codon usage

optimization and optimization of culture conditions(Rosano & Ceccarelli, 2014). More recent, the explosion of synthetic biology techniques has enabled the use of tunable promoters and of synthetic ribosome binding sites (RBSs) to control protein expression (Alper et al, 2005; Ellis et al, 2009; Salis et al, 2009).

Here, we focus on addressing the mechanistic role of L4H2 in protein expression in the context of a human human Fc gamma receptor (FcγR) IIIa and two other model bacterial proteins that are difficult to express in *E. coli*. Nucleoside-diphosphate kinase and Aspartate-semialdehyde dehydrogenase from *Pyrobaculum aerophilum* were also selected for our test cases. In case of all the *Pyrobaculum aerophilum* proteins, the expression of these cytoplasmic proteins has been previously reported as undetectable due to their misfolding on the ribosome. (Waldo et al, 1999) FcγRIIIa has been known as highly difficult to express in the *E. coli* cytoplasm (Gruel et al, 2001; Jung et al, 2010; Maenaka et al, 2001). FcγRs are expressed on the cell surface as part of the immune response and are classified based on their differences in function, affinity to IgG, and expression in different cells (Ivan & Colovai, 2006). For these proteins, total cellular expression has been shown to be still low in bacterial system even after codon optimization (Jung et al, 2010).

A2 MATERIALS AND METHODS

A2.1 Bacterial strains and plasmids

All *E. coli* strains and plasmids used are listed in Table A1. Briefly, for L4H2 containing constructs, the L4H2 region was amplified via PCR, using F5'-cgcgcggaattcttaaagaggagaaa-3' and R5'-gccgcggtcgaccagatcctcttctg-3', and then cloned into pET21a (Novagen) between EcoRI and SalI sites to generate pET-21a-L4H2.   RIIa,

RIIb, and RIIIa were amplified from pET21a-FcγRIIa, pET21a-FcγRIIb, and pET21a-FcγRIIIa, respectively (three plasmids generously provided by Dr. George Georgiou (Jung et al, 2010)). Amplifications of the genes encoding the receptors were done using primers F5'-gccgcggtcgacatcgaaggtcgta-3' and R5-gcgcgaagcttttattagtgatgatg-3' for FcγRIIa, F5'-gccgcggtcgacatcgaaggtcgta-3' and R5'- cgcgcggaattcttaaagaggagaaa-3' for FcγRIIb, and F5'-gccgcggtcgacatcgaaggtcgta-3' and R5'- gcgcgaagcttttattagtgatgatg-3' for FcγRIIIa.   Each of these fragments was cloned into pET-21a-L4H2 between SalI and HindIII sites to generate pET-21a-L4H2-RIIa/IIb/IIIa. For construction of the dual expression plasmid, pETDuet™-1 (Novagen) was used. L4H2 fragment was cut using EcoRI and SalI site from pET-21a-L4H2-RIIIa and then cloned into the first MCS of pETDuet™-1, resulting in pETDuet™-L4H2. Then, the FcγRIIIa fragment was excised from pET21a-FcγRIIIa using NdeI and XhoI site and cloned into the second MCS of pETDuet™-L4H2 to yield pETDuet™-L4H2-RIIIa (pDuet-L4H2-RIIIa). For the proteins from *Pyrobaculum aerophilum*, primers F5'-ctcatggtcgacgtgcatgctataaatattgcttttttcgc-3' and R5'-acgcactcgagctctaaaaacctcctcttctcgaaac-3' for Nucleoside-diphosphate kinase, and F5'-ctcatgccatggatcgctataaggtatatatt-3', R5'-acgcactcgagaacgcggtttgctatcattaactccg-3' for Aspartate-semialdehyde dehydrogenase were used for amplification of each gene from each GFP plasmids(Waldo et al, 1999) and then cloned into pET21a for control (using NdeI and BamHI) and pET21a-L4H2 for L4H2 constructs (using SalI/NcoI and XhoI). All plasmid sequences were confirmed by Sanger sequencing in our university core facilities. All constructs were transformed into *E. coli* BL21 (DE3) for protein expression. For SecM constructs, we used plasmid pET-SecM17 (Contreras-Martinez & DeLisa, 2007) that encodes 17 amino acids of SecM stalled sequence (FSTPVWISQAQGIRAGP). L4H2-RIIIa and RIIIa sequences were amplified from pET21a-L4H2-RIIIa (pL4H2-RIIIa) and pET21a-FcγRIIIa (pRIIIa), respectively.

NcoI/XbaI and EcoRI sites were used for cloning into pET-SecM17 and primers F5'-agtcctctagattgtttaactttaagaaggagatat-3' and R5'-agtccgaattcgtggtggtggtggtggtgctcgagtgc-3' for RIIIa and primers F5'-aagtctctagattaaagaggagaaaggtcatgaaagca-3' and R5'-agtccgaattcgtggtggtggtggtggtgctcgagtgc-3' for L4H2-RIIIa. After ligation of digested plasmids and PCR products, pET-RIIIa-SecM (pRIIIa-SecM) and pET-L4H2-RIIIa-SecM (pL4H2-RIIIa-SecM) were constructed and sequenced. After verifying all intended sequences, these plasmids were transformed into BL21 (DE3) for protein expression.

A2.2 Cell growth and analysis of total cell lysate

Cells harboring each constructed plasmid (Table 1) were grown in 5 ml Luria-Bertani (LB) growth medium at 37°C with shaking overnight. 500 ul of saturated overnight culture was transferred to 50 ml LB. When OD_{600} reached 0.4-0.6, 1mM isopropyl β-D-thiogalactopyranoside (IPTG) was added to induce expression of target proteins. Non-induced cells were used as a control to compare protein expression levels. Ampicillin (100ug/ml) was added to each culture for selection. Cells were harvested at 4,000 rpm for 20 mins after 5 hrs of induction; final OD_{600} was measured to adjust volume of collected samples so that the total protein amount was the same for all samples. To process total cellular lysate, pelleted cells were resuspended in lysis buffer (50mM Tris (pH 8.0), 1mM EDTA) and boiled at 100°C for 5 mins. After boiling, samples were loaded and run on denaturing 12% SDS-PAGE at 80V for 2hrs. Samples were loaded with Colorplus^{TM} Prestained Protein marker (NEB) or PageRuler™ Unstained Broad Range Protein Ladder (Thermo Scientific #26630). The loading amounts of samples were normalized by final cell OD_{600}. Gels were stained with

Coomassie brilliant blue R-250 and destained with 40% methanol/10% glacial acetic acid.

A2.3 Western blotting analysis

Western blotting analysis was performed to detect protein expression from all target plasmids listed in Table 1 using Anti-His monoclonal antibody (Invitrogen #R930-25). Standard Western blotting protocols were used (Gelderman et al). Briefly, total cellular lysates were loaded onto a 12% denaturing SDS-PAGE gel. Gels were transferred to methanol activated PVDF membranes using the Trans-Blot® Semi-Dry Electrophoretic Transfer Cell (Bio-Rad) and run for 40 mins at 15V. Blocking with 5% dry milk in Tris-buffered saline (TBS) was done for 1hr at room T. The proteins were detected with Anti-His monoclonal antibody at 1:5000 dilution (Invitrogen #R930-25). As a secondary antibody, we used Anti-Mouse IgG (H + L) HRP Conjugate (Promega #W4021) at a dilution of 1:2500. All images were developed using Clarity™ ECL Western blotting Substrate (BioRad, #170-5060) and the ChemiDoc™ MP Imaging System (BioRad). Bradford assay measurements were used to normalize the loading of all protein analysis by total protein mass.

A2.4 Quantification of protein expression level

Specific proteins were detected on the membrane by Western blot analysis for the accurate quantification utilizing ImageQuant TL 8.1. Each protein was detected using anti-his. The level of protein expression was and then normalized using expression of recA as an internal control. Then, the expression of L4H2-target proteins was calculated and normalized by the expression of wt target proteins.

Plasmid	Plasmid description	Reference or source
pET21a-FcγRIIa (pRIIa)	FcγRIIa gene with C-terminal 6×His in pET21a	(Jung et al, 2010)
pET21a-FcγRIIb (pRIIb)	FcγRIIb gene with C-terminal 6×His in pET21a	(Jung et al, 2010)
pET21a-FcγRIIIa (pRIIIa)	FcγRIIIa gene with C-terminal 6×His in pET21a	(Jung et al, 2010)
pET21a-L4H2-RIIa (pL4H2-RIIa)	FcγRIIa gene with N-terminal L4H2 in pET21a	This study
pET21a-L4H2-RIIb (pL4H2-RIIb)	FcγRIIb gene with N-terminal L4H2 in pET21a	This study
pET21a-L4H2-RIIIa (pL4H2-RIIIa)	FcγRIIIa gene with N-terminal L4H2 in pET21a	This study
pETDuet™-L4H2-RIIIa (pDuet-L4H2-RIIIa)	L4H2 and FcγRIIIa gene were cloned into 1st and 2nd MCS of pETDuet™ and expressed individually from separate promoter	This study
pL29-RIIIa	FcγRIIIa gene with N-terminal L29 in pET21a	This study
pAspartate-semialdehyde dehydrogenase (ASD)	WT Aspartate-semialdehyde dehydrogenase gene from *Pyrobaculum aerophilum* in pET21a	This study
pL4H2-pAspartate-semialdehyde dehydrogenase	Aspartate-semialdehyde dehydrogenase gene from *Pyrobaculum aerophilum* with N-terminal L4H2 proteins in pET21a	This study
pNucleoside-diphosphate kinase (NDK)	WT Nucleoside-diphosphate kinase gene from *Pyrobaculum aerophilum* in pET21a	This study
pL4H2-Nucleoside-diphosphate kinase	Nucleoside-diphosphate kinase gene from *Pyrobaculum aerophilum* with N-terminal L4H2 proteins in pET21a	This study
pscFv	WT scFv in pET21a	This study
pL4H2-scFv	scFv gene with N-terminal L4H2 in pET21a	This study
pET-RIIIa-SecM (pRIIIa-SecM)	FcγRIIa gene with SecM17 stalling sites in pET28a	This study
pET-L4H2-RIIIa-SecM (pL4H2-RIIIa-SecM)	L4H2-RIIIa gene with SecM stalling sites in pET28a	This study

Table A1. *E. coli* plasmids used in this study

A2.5 Isolation of stalled ribosomes assaying transcripts associated with mRNAs undergoing active translation

We isolated ribosomes according to a procedure modified from (Contreras-Martinez & DeLisa, 2007). Specifically, 100 ml cultures were grown at 30 °C (as described above) and induced with 1 mM IPTG at OD_{600}=0.5. After an additional 30 mins of growth at 30 °C, two Buffer C (20 mM Tris–HCl (pH 7.5), 50 mM NH_4Cl, and 25 mM $MgCl_2$) ice cubes were added to each culture flask. Flasks were rapidly swirled for 1 min on ice and incubated on ice for an additional 30 mins. Next, cells were pelleted by centrifugation as described above and resuspended in 600 µl of cold Buffer C. Cells were lysed by five cycles of freeze–thawing in liquid nitrogen followed by the addition of three 30 µl aliquots of lysozyme (EMD Millipore #71110-4), where the stock lysozyme solution was diluted 50-fold in cold Buffer C and each lysozyme addition was followed by a 20 min incubation at 4 °C and three additional freeze–thawing cycles in liquid nitrogen. To reduce the viscosity of the lysates (due to cell debris), DNase I (Thermo Scientific #EN0521) were added and samples were rotated for 15 min at 4 °C after each dose of the enzyme. Samples were spun in a microcentrifuge for 20 mins at 13,000 rpm at 4 °C to pellet the debris. The collected supernatant was loaded onto a cold two-phase cushion was made of equal volumes of Buffer C, supplemented with a 5% (w/v) sucrose and Buffer B (20 mM Tris–HCl (pH 7.5), 500 mM NH_4Cl, 25 mM $MgCl_2$) supplemented with 37% (w/v) sucrose. Ribosomes were isolated by ultracentrifugation for 35 h at 24,000 rpm and 4 °C using a Beckman L8 ultracentrifuge with an SW28 rotor. The crude ribosome pellet was resuspended in cold 200 µl of Buffer C and ultracentrifuged in a 10% - 40% (w/v) sucrose gradient in Buffer A (20 nM Tris–HCl (pH 7.5), 100 mM NH_4Cl, 25 mM $MgCl_2$) for 17 h at 22,000 rpm and 4 °C in a SW41 rotor. Gradient fractionation was

performed manually by sequentially pipetting 250 μl from the top part of the gradient. All collected samples were stored at -20°C for further analysis.

A2.6 Quantitative RT-PCR

Total RNA samples were extracted from cells using RNeasy mini kit (NEB) according to manufacturer's protocol for reverse transcription (RT) analysis. RT was performed in 20 μl using SuperScript III Reverse Transcriptase (Invitrogen #18080-) with protocols provided from the manufacturer. 1 ul of RNaseOUT^{TM} Recombinant ribonuclease inhibitor (Invitrogen #10777) was added to each reaction. Target specific reverse primers were used to prime cDNA synthesis. The RT reactions were all performed at 55°C for 60 minutes. The reactions were inactivated at 70°C for 15 mins and then incubated on ice. RT reactions were treated with RNase H (NEB #R0297S) for 20 mins at 37°C. The cDNA product was used as a template in a 50 μl PCR reaction containing Power SYBR® Green PCR master mix (Invitrogen #4368577). Specific primers were used for each target (primers F5'-aaataccgcgctgcataaag-3'and R5'-tttcgctgctcacatttttg-3' for the amplification of RIIIa/L4H2-RIIIa and primers F5'-caagacatcatggcccttac-3'and R5'-acttcatggagtcgagttgc-3' for 16S rRNA, used as a gene endogenous control). The same primers for the analysis of *RIIIa/L4H2-RIIIa* transcripts were used. The temperature cycle used for the PCR reactions is as follows: 95 °C for 10 min, 40 cycles of 95 °C for 15 s and 60 °C for 60 s. The same procedure was used for extracting and analyzing total RNA samples associated with active translating *RIIIa and L4H2-RIIIa* mRNA. Relative quantification of RIIIa/L4H2-RIIIa RNAs was performed using Viia 7 Software (Life Technologies) following the comparative delta-delta threshold cycle ($\Delta\Delta C_T$) method. Samples were collected in biological triplicates.

A3 RESULTS

A3.1 Expression of human Fcγ receptors in *E. coli*

To identify a target protein that exhibited compromised synthesis in *E. coli*, we tested the heterologous expression of human FcγRIIa (RIIa), FcγRIIb (RIIb) and FcγRIIIa (RIIIa) in BL21 (DE3), a strain commonly used for protein overexpression. This set of proteins was selected given the challenges in their expression that have been previously reported (Jung et al, 2010). We performed expression analysis under a T7 promoter and RBS (AAAGAGGAGAAA) with a maximum strength and efficiency, selected according to the RBS calculator (Salis et al, 2009), to drive the high levels of synthesis that would further challenge protein processing in a bacterial host. Given that RIIIa was the most poorly expressed protein out of the three tested (Figure A1), we used it as a model.
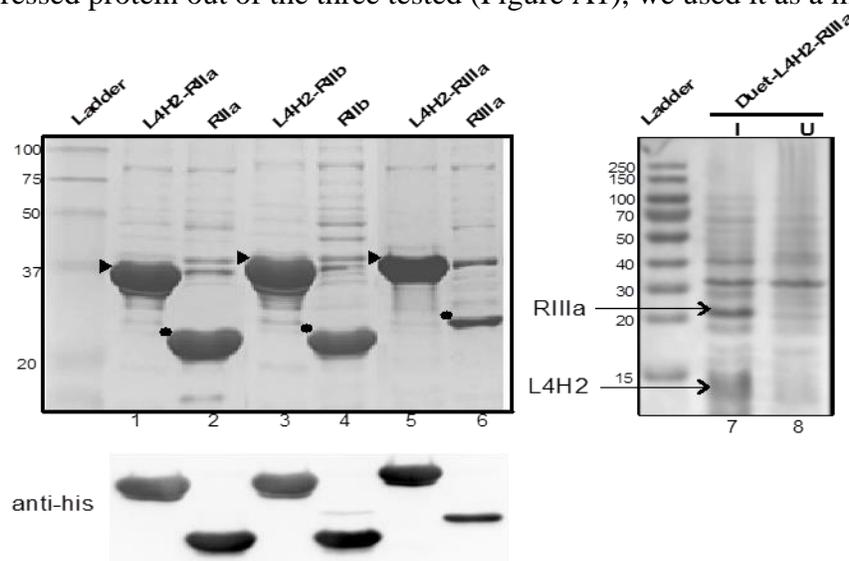
Figure A1: Expression profile of Fcγ receptors IIa, IIb, and IIIa in *E.coli* with or without L4H2. SDS-PAGE (top panel) for protein expression of RIIa, RIIb, and RIIIa (Lane 2, Lane 4 and Lane 6, respectively) compared to L4H2-RIIa (Lane 1), L4H2-RIIb (Lane 3) and L4H2-RIIIa (Lane 5). Triangle denotes L4H2-fusion proteins and circle denotes WT proteins. Specific bands were compared via Western blotting with an Anti-His antibody (bottom panel). In lane 7 and 8, the expression of pDuet-L4H2-RIIIa has been shown in a different gel for better visualization of small L4H2 (10 kDa) protein.

123

A3.2 Fusion to L4H2 enhances synthesis of RIIIa

Given previous demonstrations that overexpression of an engineered variant of the ribosomal protein L29 (L4H2) led to significant increases in protein synthesis for a variety of proteins (Contreras-Martinez et al, 2012), we investigated if L4H2 needed to be incorporated into 70S ribosomes to exert this effect. We observed that, when overexpressed, L29 is poorly incorporated into ribosomes (data not shown) particularly relative to other essential components at the surface of the exit tunnel (such as L23) (Kramer et al, 2002). This was not fully surprising since L29 is not essential for translation. However, this orthogonality offered an advantage to the use of L29 variants as they present minimal risk of interfering with cellular translation and thereby with cell viability.   However, our first attempt of overexpressing L4H2 and RIIIa individually from separate T7 promoters led to no increase in RIIIa synthesis (Figures A1, A2). Notably these results differ from previously published work (Contreras-Martinez et al, 2012) involving other proteins; we attribute this to the fact that RIIIa (a human protein) was likely more difficult to synthesize in *E. coli* relative to the suite of proteins previously tested; furthermore, previous studies overexpressed L4H2 from a different and weaker (pBAD) promoter.

We next fused FcγRIIIa to wt L29 and L4H2 by designing a construct where L29 or L4H2 was fused to the N-terminal of RIIIa, as shown in Figure 1A. After comparing protein expression levels of L4H2-RIIIa and RIIIa, we observed about 2.4-fold increase with L29 fusion and a 3.2-fold increase with L4H2 fusion as measured by Western blotting analysis (Figures A2). This increase was significantly higher than the one observed when RIIIa and L4H2-RIIIa were expressed individually from separate T7 promoters. This comparison also indicated that the 4H2 variant was more efficient for enhancing protein expression.

124

A

pRIIIa — T7 promoter — RBS | RIIIa | 6x His

pL29-RIIIa — T7 promoter — RBS | L29 | RIIIa | 6x His

pDuet-L4H2-RIIIa — T7 promoter — RBS | L4H2 | STOP — T7 promoter — RBS | RIIIa | 6x His

pL4H2-RIIIa — T7 promoter — RBS | L4H2 | RIIIa | 6x His

B

C
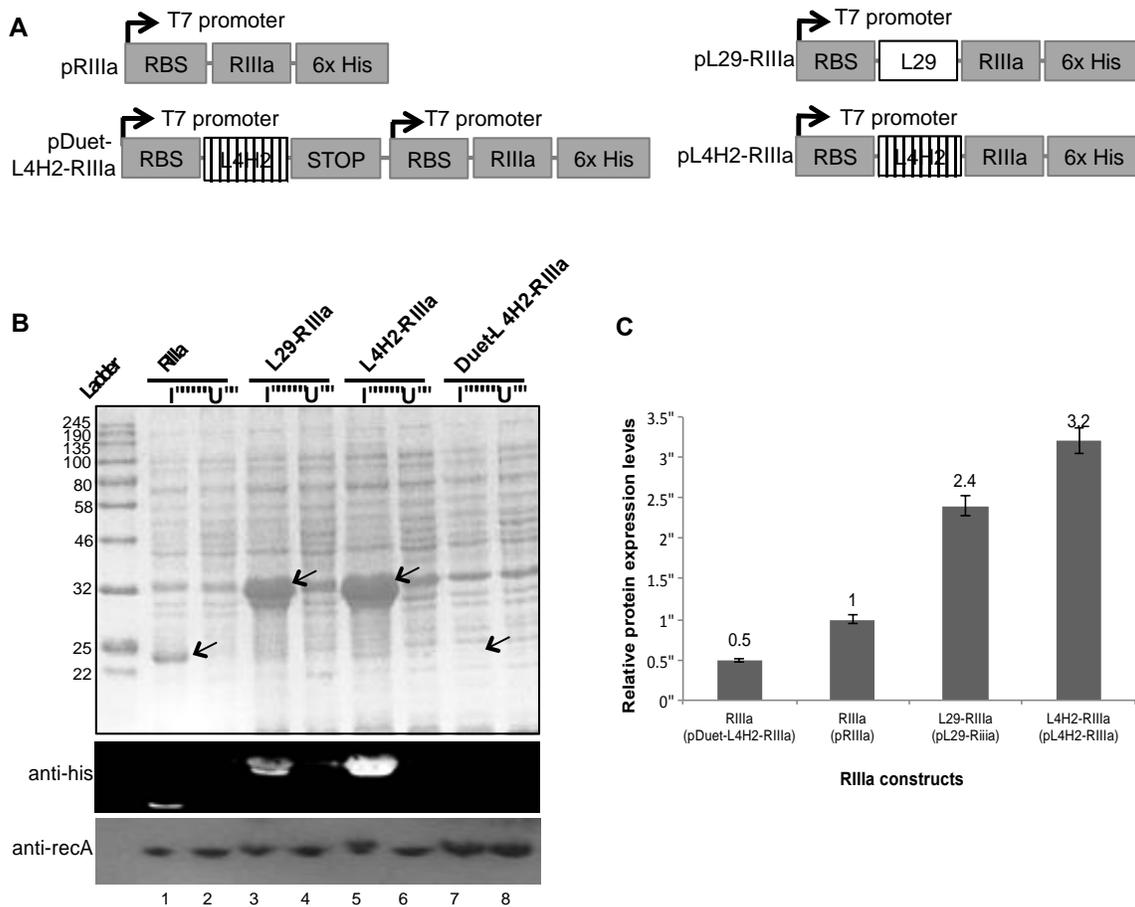
Figure A2: Expression profile of Fcγ RIIIa in *E.coli* with or without L29 or L4H2. (A) Constructs of RIIIa, L29-RIIIa, Duet-L4H2-RIIIa and L4H2-RIIIa plasmids. In the pET-Duet vector, L4H2 and RIIIa fragments are expressed from separate T7 promoters. L4H2 is fused in the 5' region of RIIIa in pET-L4H2-RIIIa. (B) SDS-PAGE gels for protein expression of RIIIa, L29-RIIIa, Duet-L4H2-RIIIa and L4H2-RIIIa. U denotes un-induced sample and I denotes induced sample with 1mM IPTG. Lane 1, 3, 5, 7 showed RIIIa (23.2kDa), L29-RIIIa (33.4 kDa), L4H2-RIIIa (33.4kDa) and RIIIa (23.2kDa) band from pDuet-L4H2-RIIIa with an arrow, respectively. Lane 2, 4, 6, 8 showed un-induced sample from pRIIIa, pL29-RIIIa, pL4H2-RIIIa and pDuet-L4H2-RIIIa, respectively. The identity of the bands was confirmed by Western blotting using anti-His antibody. Anti-RecA was used as the loading control. (C) Band intensity from Western blot analysis was quantified using ImageQuant TL 8.1 for the expression of RIIIa from pRIIIa and pDuet-L4H2-RIIIa and L4H2-RIIIa from pL4H2-RIIIa. The intensity of each target is normalized by the intensity of RIIIa. Error bars were calculated as SEM.

A3.3 Expression of FcγRIIIa in *E. coli* using several different fusion proteins

A traditional strategy to enhance complex protein production in bacteria is to use of fusion partners (Sorensen & Mortensen, 2005; Villaverde & Carrio, 2003). To test if increase in RIIIa synthesis upon fusion to L4H2 was due to the N-terminal fusion of any sequence that would change the 5' end mRNA structure, we compared synthesized levels of L4H2-RIIIa to TrxA-RIIIa and MBP-RIIIa (Figure A3). TrxA (Thioredoxin-1) and MBP (Maltose binding protein) represent native *E. coli* proteins that are typically employed as fusion tags in *E. coli* (Bach et al, 2001; McCoy & La Ville, 2001). As shown in Figure A3.B, RIIIa showed about 1.8-fold higher increase in the context of the L4H2 fusion relative to fusions to TrxA and MBP.
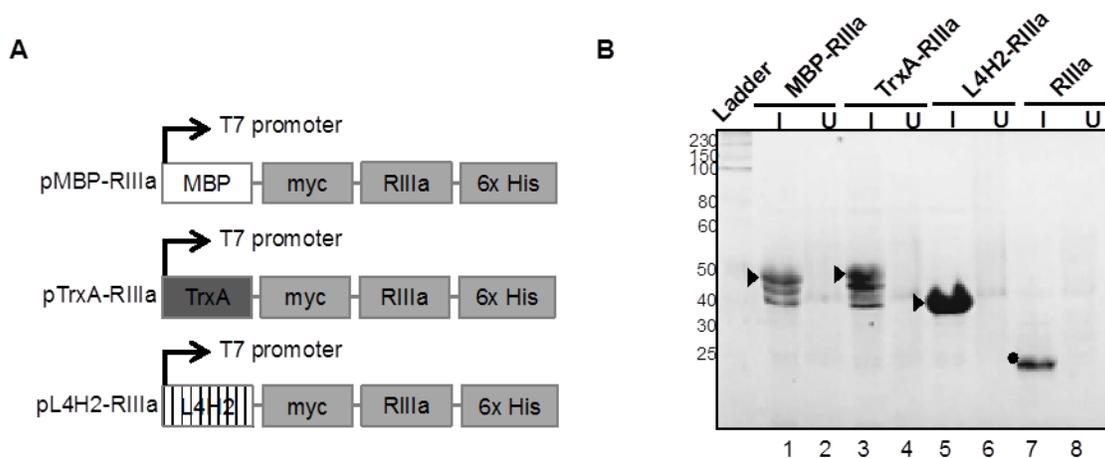
Figure A3: Expression profile of Fcγ RIIIa in *E.coli* with or without L29 or L4H2. (A) Constructs of MBP-RIIIa (pMBP-RIIIa), TrxA-RIIIa (pTrxA-RIIIa), L4H2-RIIIa (pL4H2-RIIIa) and RIIIa (pRIIIa) plasmids. (B) SDS-PAGE gel shows the expression of MBP-RIIIa (triangle, 46.5kDa, Lane 1), TrxA-RIIIa (triangle, 47kDa, Lane 3) and L4H2-RIIIa (triangle, 33.2kDa, Lane 5) compared to RIIIa (circle, 23kDa, Lane 7) Lane 2, 4, 6 and 8 showed un-induced sample from pMBP-RIIIa, pTrxA-RIIIa, pL4H2-RIIIa and pRIIIa, respectively. Gel was stained with Coomassie Blue.

A3.4 Expression of difficult-to-express proteins utilizing L4H2 fusion protein

To confirm that the effect of L4H2 fusion is not limited to the RIIIa protein, we selected two known difficult-to-express proteins (Nucleoside-diphosphate kinase and Aspartate-semialdehyde dehydrogenase) of *Pyrobaculum aerophilum*, previously shown to be highly insoluble in the context of GFP fusions. These two particular proteins were selected as they represent 2 of the most poorly expressed proteins of a set of 20 native *Pyrobaculum aerophilum* proteins previously characterized in *E. coli* (Waldo et al, 1999). Importantly, after fusion with L4H2, Aspartate-semialdehyde dehydrogenase showed a quantified~ 60% increase compared to wt Aspartate-semialdehyde dehydrogenase expression (Figure A4) as estimated by ImageQuant 8.1. In the case of Nucleoside-diphosphate kinase, fusion with the L4H2 tag allowed visualization of protein expression (whereas no expression of this protein was detectable in the absence of the 4H2 tag via SDS-PAGE and Western blot analysis. Therefore, this data demonstrated that the enhancements in protein expression due to L4H2 fusion are not limited to the expression of RIIIa and can contribute to increase protein expression of other difficult to express proteins. Notably expression of the *Pyrobaculum aerophilum* proteins tested was not significantly increased with fusions to other proteins (like GFP), as previously reported.

A3.5 L4H2 fusion increases levels of actively translating ribosomes synthesizing RIIIa mRNA but not of total levels of mRNA transcript

To further understand the enhancement of RIIIa synthesis in the context of L4H2, we first investigated whether observed increase in RIIIa production resulted from increases in *L4H2-RIIIa* mRNA levels relative to *RIIIa* mRNA. We focused on the RIIIa protein given that the changes in solubility of this protein were the most pronounced that we observed out of the proteins tested in this study.
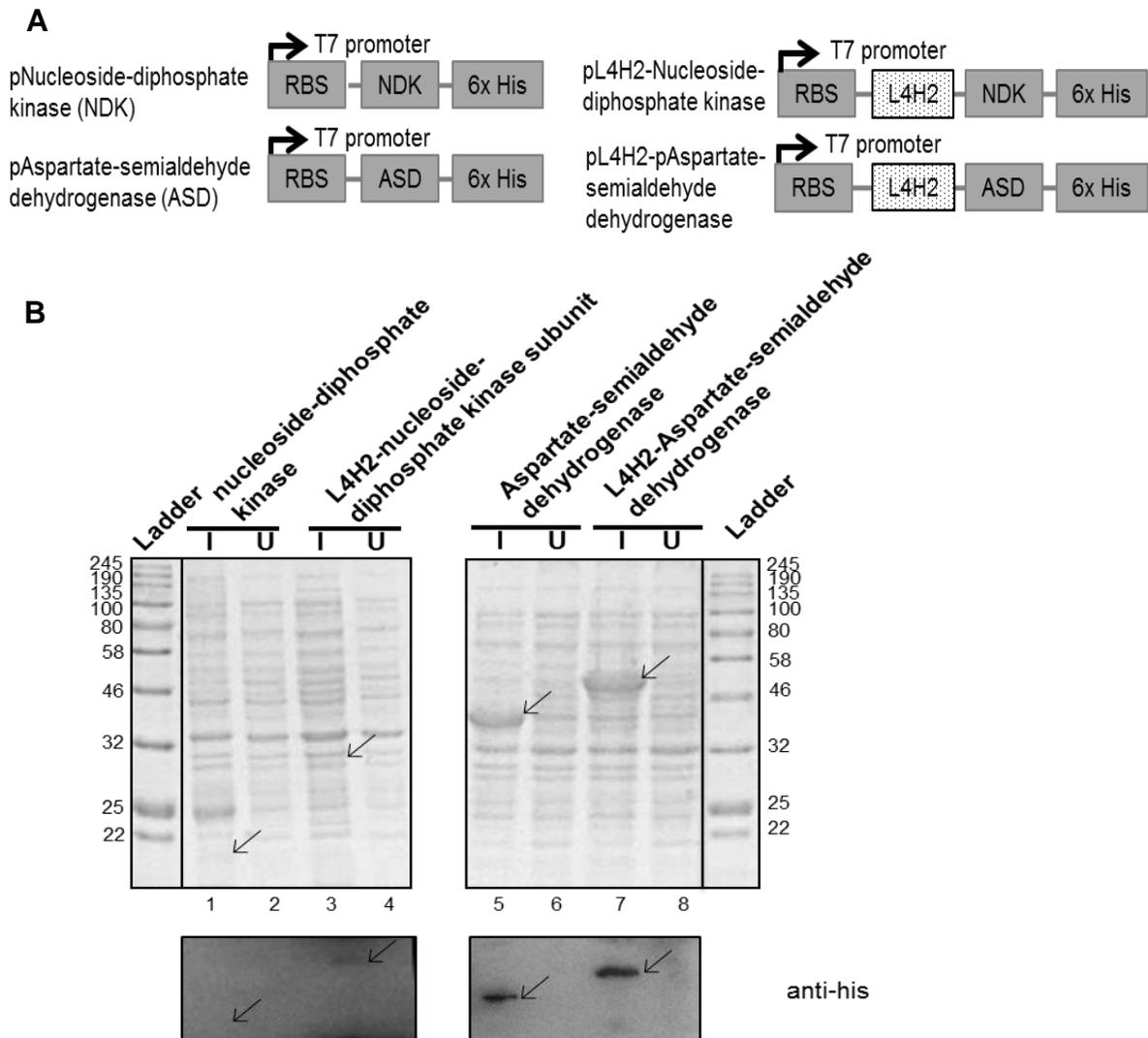
Figure A4: Expression profiles of difficult-to-express proteins of *Pyrobaculum aerophilum* with or without L4H2 in *E. coli* (A) Constructs of Nucleoside-diphosphate kinase and Aspartate-semialdehyde dehydrogenase plasmids with or without L4H2. (B) SDS-PAGE and Western blotting for expression of nucleoside-diphosphate kinase and aspartate-semialdehyde dehydrogenase (Lane 1 and 5, respectively) compared to L4H2-nucleoside-diphosphate kinase (Lane 3) and L4H2-aspartate-semialdehyde dehydrogenase (Lane 7). Lane 2, 4, 6, 8 showed un-induced samples per each protein.

We reasoned that increases in mRNA levels were plausible given the potential

impact on cellular abundance that changes introduced to the RIIIa (in both length and

sequence) could have on the mRNA secondary structure and/or cellular stability. To test this possibility, we analyzed the total cellular RNA levels in cells expressing RIIIa and L4H2-RIIIa. As shown in Figure A5, total levels of *RIIIa* transcripts were not significantly different than total levels of *L4H2-RIIIa*, as measured in three biological replicas. This result suggests that fusion to L4H2 does not affect overall mRNA levels of RIIIa in the cytoplasm. This result is consistent with previous findings reporting poor correlation between mRNA abundance and protein expression (Maier et al, 2009).
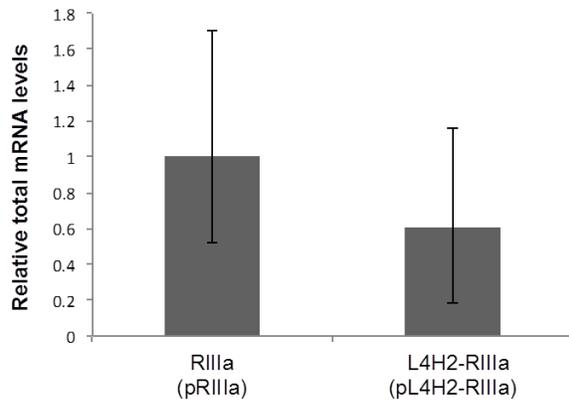
Figure A5: Total mRNA level of Wild type RIIIa and L4H2-RIIIa. Relative total mRNA levels of WT RIIIa and L4H2-RIIIa as measured by quantitative RT-PCR using total cellular RNA as template and normalized by endogenous levels of 16S rRNA. This data showed the mean value of 3 biological replicas with 3 experimental replicas. Error bars were calculated as SEM.

Since fusions to L4H2 did not affect total cellular mRNA levels, we next explored the possibility of L4H2 enhancing the recruitment of the mRNA transcripts to ribosomes. Specifically, we quantified and compared the levels of ribosomes that were actively translating each transcript (*L4H2-RIIIa* and *RIIIa*) by isolating stalled ribosomes. Our hypothesis that L4H2 could impact the number of active ribosomes engaged in RIIIa synthesis is justified by recent findings that ribosomal occupancy and density affect translation efficiency (Picard et al, 2012). To test if levels of actively translated mRNA

were higher in the context of L4H2, we employed SecM-mediated arrest, as used in previous works (Contreras-Martinez & DeLisa, 2007), to profile ribosomes that were actively engaged in translation of the specific transcripts of interest, *RIIIa* and *L4H2-RIIIa*. Briefly, in this highly targeted scheme (Figure A6), the C-terminal 17-aa stalling sequence of the *E. coli* secM protein was added at the 3'end of RIIIa and L4H2-RIIIa transcript. Since the secM17 stall sequence has been shown to induce translation arrest in the context of heterologous proteins (Contreras-Martinez & DeLisa, 2007), we expected both transcripts to remain associated with actively translating 70S ribosomes (i.e. unreleased) during synthesis of these proteins. This natural stalling mechanism has been well studied (Gumbart et al, 2012) and serves an important natural role in controlling translation of the secA protein. Importantly, after isolating RIIIa and L4H2-RIIIa stalled ribosomes and measuring their actively translated mRNA levels using qRT-PCR, we uncovered that the actively translated mRNA level of L4H2-RIIIa was about 3.8-fold higher than that of RIIIa (Figure A6); it is worth noting that the difference observed in actively translated mRNA transcript levels is highly similar with the difference observed in protein expression level (Figure A2). The difference in levels of actively translated transcripts observed between these constructs suggest that L4H2-RIIIa recruits more ribosomes than the RIIIa transcript and that this "mRNA channeling" to ribosomes can lead to improved protein synthesis. In addition, these results validated the novel application of in vivo ribosome stalling for examining synthesis of targeted mRNAs of interest. Note that elegant high-throughput methods (such as ribosome profiling (Ingolia, 2014)) were not convenient or economically feasible as we were only interested in relative levels of two mRNA transcripts.
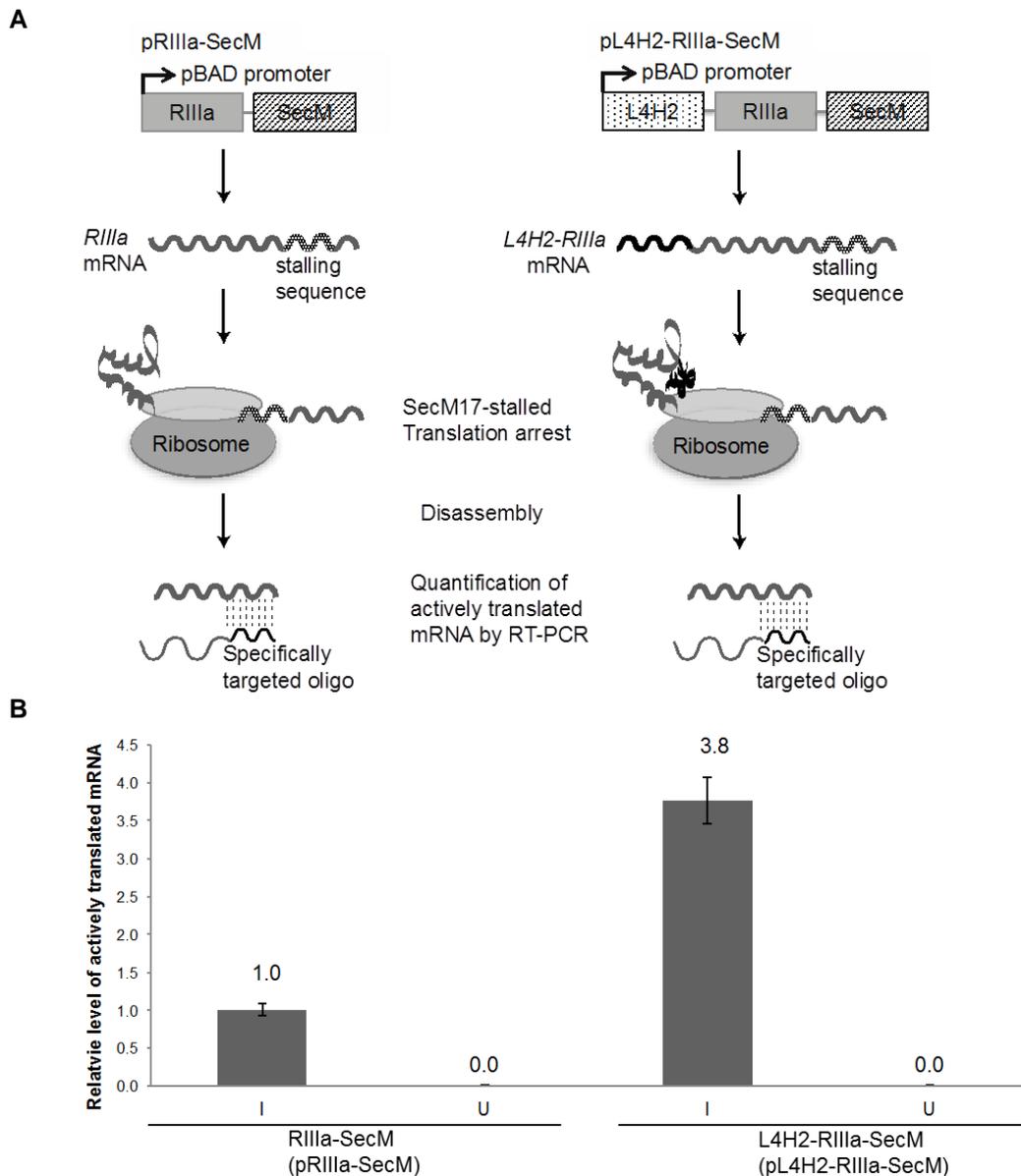
Figure A6: Relative levels of mRNAs undergoing active translation as measured in stalled ribosomes. (A) Schematic diagram showed how to isolate actively translated ribosomes. First translation is stalled in vivo using a 17aa stalling sequence (SecM17) and actively translated mRNAs were disassembled from ribosome. Measurement of mRNAs in SecM-stalled ribosomes enables quantification of a particular actively translated mRNA by qRT-PCR. (B) Relative level of actively translated mRNAs as measured by quantitative RT-PCR. U denotes un-induced sample and I denotes induced sample. This data showed the mean value of 3 biological replicas with 3 experimental replicas. Error bars represented SEM.

A4 DISCUSSION

In this study, we demonstrated that L4H2 could help improve protein synthesis of the human receptor FcγRIIIa as well as two other model cytoplasmic proteins from *Pyrobaculum aerophilum*, which show poor expression in bacterial systems. However, about 3.2-fold enhancement was only observed when L4H2 was fused to FcγRIIIa and not simply upon its coexpression. At least a partial mechanistic explanation of the observed improvement in protein synthesis is the recruitment of more ribosomes to the *L4H2-RIIIa* mRNA transcript relative to the *RIIIa* mRNA, as quantified by profiling stalled ribosomes. This result suggests the possibility that fusion of L4H2 could enhance proximity and affinity of the target mRNA to cellular ribosomes resulting in increased target protein expression; however, the details of this "ribosome channeling" mechanism need to be further investigated. Evidence of engineering physical proximity to increase enzymatic activity has been demonstrated in the context of bacterial cells to enhance yields of biosynthetic products (Conrado et al, 2012). For instance, the coupling of enzymes and substrates, using fusion proteins as scaffolds, has been shown to have significant improvements on product synthesis (Castellana et al, 2014; Dueber et al, 2009). It is therefore plausible that, similarly, the channeling of mRNA substrates to ribosomes could contribute to enhancing translation (to potentially already strong RBS-mediated interactions).

It is also worth noting that these same levels of protein increase were not observed for the other Fcγ receptors (RIIa and RIIb) that initially already exhibited higher levels of synthesis relative to RIIIa (Figure A1). This observation showed that the mechanisms by which L4H2-fusions enhance translation could be only relevant to proteins that are produced at very low (to non-detectable) levels. This was supported by the fact that very poorly expressed wt Nucleoside-diphosphate kinase and Aspartate-semialdehyde

dehydrogenase both exhibited significant enhances in their expression with L4H2, to expression levels that were readily detectable (Figure A4). Importantly, upon performing a secondary structure analysis of the *L4H2-RIIIa*, *L4H2-RIIa*, and *L4H2-RIIb* vs *RIIIa*, *RIIa* and *RIIb* transcripts using RNAstructure version 5.7 (Mathews et al, 2004), the predicted stability improvement of the mRNA transcript for RIIa and RIIb upon L4H2 fusions were 45% and 38%, respectively as calculated by the ΔG differences between the fused and wild-type transcripts; in contrast, L4H2 fusion contributed 57% stability increase to the *RIIIa* transcript. This could imply that fusion of L4H2 could also enhance the mRNA stability of RIIIa target protein; however, we suspect that this is a minor effect since an increase in total *RIIIa* mRNA levels was not detected by qRT-PCR upon fusion to L4H2.

Overall, a key implication of our findings is that orthogonal tags that enhance affinity to ribosomes can be powerful synthetic ways to enhance translation. Traditionally, codon optimization and manipulation around the 5' translation initiation region have been used as tools for increasing bacterial translation efficiency (Burgess-Brown et al, 2008; Seo et al, 2013). Therefore, it is plausible that the engineering of 5' regions with tags that further affect the affinity of the target gene (beyond the effect obtained by optimizing RBS strength) could be used as a strategy to further augment translation of very poorly expressed proteins. Future studies should be followed to find out the underlying mechanisms for broader usage of this approach. Lastly in this work, we have also shown a different application of stalling ribosomes for a specific transcript, where profiling actively translating ribosomes can be used to understand protein synthesis in vivo for targeted small number of transcripts.

Bibliography

Aiba H (2007) Mechanism of RNA silencing by Hfq-binding small RNAs. *Current opinion in microbiology* **10:** 134-139

Alper H, Fischer C, Nevoigt E, Stephanopoulos G (2005) Tuning genetic control through promoter engineering. *Proceedings of the National Academy of Sciences of the United States of America* **102:** 12678-12683

Alper H, Stephanopoulos G (2007) Global transcription machinery engineering: a new approach for improving cellular phenotype. *Metab Eng* **9:** 258-267

Alsaker KV, Spitzer TR, Papoutsakis ET (2004) Transcriptional analysis of spo0A overexpression in Clostridium acetobutylicum and its effect on the cell's response to butanol stress. *J Bacteriol* **186:** 1959-1971

Altuvia S (2007) Identification of bacterial small non-coding RNAs: experimental approaches. *Current opinion in microbiology* **10:** 257-261

Altuvia S, Weinstein-Fischer D, Zhang A, Postow L, Storz G (1997) A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. *Cell* **90:** 43-53

An H, Scopes RK, Rodriguez M, Keshav KF, Ingram LO (1991) Gel electrophoretic analysis of Zymomonas mobilis glycolytic and fermentative enzymes: identification of alcohol dehydrogenase II as a stress protein. *J Bacteriol* **173:** 5975-5982

Atsumi S, Cann AF, Connor MR, Shen CR, Smith KM, Brynildsen MP, Chou KJ, Hanai T, Liao JC (2008) Metabolic engineering of Escherichia coli for 1-butanol production. *Metab Eng* **10:** 305-311

Bach H, Mazor Y, Shaky S, Shoham-Lev A, Berdichevsky Y, Gutnick DL, Benhar I (2001) Escherichia coli maltose-binding protein as a molecular chaperone for recombinant intracellular cytoplasmic single-chain antibodies. *Journal of molecular biology* **312:** 79-93

Baker JL, Sudarsan N, Weinberg Z, Roth A, Stockbridge RB, Breaker RR (2012) Widespread genetic switches and toxicity resistance proteins for fluoride. *Science* **335:** 233-235

Barrick JE, Breaker RR (2007) The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome biology* **8:** R239

Basak S, Song H, Jiang RR (2012) Error-prone PCR of global transcription factor cyclic AMP receptor protein for enhanced organic solvent (toluene) tolerance. *Process Biochem* **47:** 2152-2158

Battesti A, Majdalani N, Gottesman S (2011) The RpoS-mediated general stress response in Escherichia coli. *Annu Rev Microbiol* **65:** 189-213

Beauregard A, Smith EA, Petrone BL, Singh N, Karch C, McDonough KA, Wade JT (2013) Identification and characterization of small RNAs in Yersinia pestis. *RNA biology* **10**

Beisel CL, Storz G (2010) Base pairing small RNAs and their roles in global regulatory networks. *Fems Microbiol Rev* **34:** 866-882

Borden JR, Jones SW, Indurthi D, Chen Y, Terry Papoutsakis E (2010) A genomic-library based discovery of a novel, possibly synthetic, acid-tolerance mechanism in Clostridium acetobutylicum involving non-coding RNAs and ribosomal RNA processing. *Metabolic Engineering* **12:** 268-281

Breaker RR (2008) Gene Regulation by Riboswitches. *Faseb J* **22**

Bringer S, Finn R, Sahm H (1984) Effect of oxygen on the metabolism of Zymomonas mobilis. *Arch Microbiol* **139:** 376-381

Brynildsen MP, Liao JC (2009) An integrated network approach identifies the isobutanol response network of Escherichia coli. *Mol Syst Biol* **5:** 277

Burgess-Brown NA, Sharma S, Sobott F, Loenarz C, Oppermann U, Gileadi O (2008) Codon optimization can improve expression of human genes in Escherichia coli: A multi-gene study. *Protein expression and purification* **59:** 94-102

Caldelari I, Chao Y, Romby P, Vogel J (2013) RNA-mediated regulation in pathogenic bacteria. *Cold Spring Harbor perspectives in medicine* **3:** a010298

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. *BMC bioinformatics* **10:** 421

Carey VC, Ingram LO (1983) Lipid composition of Zymomonas mobilis: effects of ethanol and glucose. *J Bacteriol* **154:** 1291-1300

Castellana M, Wilson MZ, Xu Y, Joshi P, Cristea IM, Rabinowitz JD, Gitai Z, Wingreen NS (2014) Enzyme clustering accelerates processing of intermediates through metabolic channeling. *Nat Biotechnol* **32:** 1011-1018

Chao Y, Papenfort K, Reinhardt R, Sharma CM, Vogel J (2012) An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J* **31:** 4005-4019

Chappell J, Takahashi MK, Meyer S, Loughrey D, Watters KE, Lucks J (2013) The centrality of RNA for engineering gene expression. *Biotechnology journal* **8:** 1379-1395

Chen Y, Indurthi DC, Jones SW, Papoutsakis ET (2011) Small RNAs in the genus Clostridium. *MBio* **2:** e00340-00310

Cho SH, Haning K, Contreras LM (2015) Strain engineering via regulatory noncoding RNAs: not a one-blueprint-fits-all. *Current Opinion in Chemical Engineering* **10:** 25-34

Cho SH, Ju SH, Contreras LM (2016) Synthetic chimeras with orthogonal ribosomal proteins increase translation yields by recruiting mRNA for translation as measured by profiling active ribosomes. *Biotechnology progress*

Cho SH, Lei R, Henninger TD, Contreras LM (2014) Discovery of Ethanol-Responsive Small RNAs in Zymomonas mobilis. *Appl Environ Microb* **80:** 4189-4198

Chong H, Huang L, Yeow J, Wang I, Zhang H, Song H, Jiang R (2013) Improving ethanol tolerance of Escherichia coli by rewiring its global regulator cAMP receptor protein (CRP). *PloS one* **8:** e57628

Chong HQ, Geng HF, Zhang HF, Song H, Huang L, Jiang RR (2014) Enhancing E. coli Isobutanol Tolerance Through Engineering Its Global Transcription Factor cAMP Receptor Protein (CRP). *Biotechnology and bioengineering* **111:** 700-708

Cloonan N, Forrest AR, Kolle G, Gardiner BB, Faulkner GJ, Brown MK, Taylor DF, Steptoe AL, Wani S, Bethel G, Robertson AJ, Perkins AC, Bruce SJ, Lee CC, Ranade SS, Peckham HE, Manning JM, McKernan KJ, Grimmond SM (2008) Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods* **5:** 613-619

Conrado RJ, Wu GC, Boock JT, Xu H, Chen SY, Lebar T, Turnsek J, Tomsic N, Avbelj M, Gaber R, Koprivnjak T, Mori J, Glavnik V, Vovk I, Bencina M, Hodnik V, Anderluh G, Dueber JE, Jerala R, DeLisa MP (2012) DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. *Nucleic Acids Res* **40:** 1879-1889

Contreras-Martinez LM, Boock JT, Kostecki JS, DeLisa MP (2012) The ribosomal exit tunnel as a target for optimizing protein expression in Escherichia coli. *Biotechnology journal* **7:** 354-360

Contreras-Martinez LM, DeLisa MP (2007) Intracellular ribosome display via SecM translation arrest as a selection for antibodies with enhanced cytosolic stability. *Journal of molecular biology* **372:** 513-524

Corbino KA, Barrick JE, Lim J, Welz R, Tucker BJ, Puskarz I, Mandal M, Rudnick ND, Breaker RR (2005) Evidence for a second class of S-adenosylmethionine riboswitches and other regulatory RNA motifs in alpha-proteobacteria. *Genome biology* **6**

Costa S, Almeida A, Castro A, Domingues L (2014) Fusion tags for protein solubility, purification and immunogenicity in Escherichia coli: the novel Fh8 system. *Frontiers in microbiology* **5:** 63

De Graaf AA, Striegel K, Wittig RM, Laufer B, Schmitz G, Wiechert W, Sprenger GA, Sahm H (1999) Metabolic state of Zymomonas mobilis in glucose-, fructose-, and xylose-fed continuous cultures as analysed by 13C- and 31P-NMR spectroscopy. *Arch Microbiol* **171:** 371-385

DiChiara JM, Contreras-Martinez LM, Livny J, Smith D, McDonough KA, Belfort M (2010) Multiple small RNAs identified in Mycobacterium bovis BCG are also expressed in Mycobacterium tuberculosis and Mycobacterium smegmatis. *Nucleic Acids Res* **38:** 4067-4078

Doelle HW, Kirk L, Crittenden R, Toh H, Doelle MB (1993) Zymomonas mobilis--science and industrial application. *Critical reviews in biotechnology* **13:** 57-98

Doran-Peterson J, Cook DM, Brandon SK (2008) Microbial conversion of sugars from plant biomass to lactic acid or ethanol. *The Plant journal : for cell and molecular biology* **54:** 582-592

Douka E, Christogianni A, Koukkou AI, Afendra AS, Drainas C (2001) Use of a green fluorescent protein gene as a reporter in Zymomonas mobilis and Halomonas elongata. *Fems Microbiol Lett* **201:** 221-227

Dueber JE, Wu GC, Malmirchegini GR, Moon TS, Petzold CJ, Ullal AV, Prather KL, Keasling JD (2009) Synthetic protein scaffolds provide modular control over metabolic flux. *Nat Biotechnol* **27:** 753-759

Dunlop MJ (2011) Engineering microbes for tolerance to next-generation biofuels. *Biotechnol Biofuels* **4:** 32

Dunlop MJ, Dossani ZY, Szmidt HL, Chu HC, Lee TS, Keasling JD, Hadi MZ, Mukhopadhyay A (2011) Engineering microbial biofuel tolerance and export using efflux pumps. *Molecular Systems Biology* **7**

Eddy SR (2001) Non-coding RNA genes and the modern RNA world. *Nat Rev Genet* **2:** 919-929

Ellis T, Wang X, Collins JJ (2009) Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat Biotechnol* **27:** 465-471

Engler C, Kandzia R, Marillonnet S (2008) A one pot, one step, precision cloning method with high throughput capability. *PloS one* **3:** e3647

Filipovska A, Rackham O (2013) Specialization from synthesis: how ribosome diversity can customize protein function. *FEBS letters* **587:** 1189-1197

Fiocco D, Capozzi V, Goffin P, Hols P, Spano G (2007) Improved adaptation to heat, cold, and solvent tolerance in Lactobacillus plantarum. *Applied microbiology and biotechnology* **77:** 909-915

Fortman JL, Chhabra S, Mukhopadhyay A, Chou H, Lee TS, Steen E, Keasling JD (2008) Biofuel alternatives to ethanol: pumping the microbial well. *Trends Biotechnol* **26:** 375-381

Gaida SM, Al-Hinai MA, Indurthi DC, Nicolaou SA, Papoutsakis ET (2013) Synthetic tolerance: three noncoding small RNAs, DsrA, ArcZ and RprA, acting supra-additively against acid stress. *Nucleic Acids Res* **41:** 8726-8737

Gao Q, Zhang M, McMillan JD, Kompala DS (2002) Characterization of heterologous and native enzyme activity profiles in metabolically engineered Zymomonas mobilis strains during batch fermentation of glucose and xylose mixtures. *Applied biochemistry and biotechnology* **98-100:** 341-355

Gelderman G, Contreras LM Discovery of posttranscriptional regulatory RNAs using next generation sequencing technologies.

Gelderman G, Sivakumar A, Lipp S, Contreras L (2015) Adaptation of Tri-molecular fluorescence complementation allows assaying of regulatory Csr RNA-protein interactions in bacteria. *Biotechnology and bioengineering* **112:** 365-375

Georg J, Hess WR (2011a) cis-antisense RNA, another level of gene regulation in bacteria. *Microbiol Mol Biol Rev* **75:** 286-300

Georg J, Hess WR (2011b) Regulatory RNAs in cyanobacteria: developmental decisions, stress responses and a plethora of chromosomally encoded cis-antisense RNAs. *Biol Chem* **392:** 291-297

Gottesman S (2005) Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends in genetics : TIG* **21:** 399-404

Gottesman S, McCullen CA, Guillier M, Vanderpool CK, Majdalani N, Benhammou J, Thompson KM, FitzGerald PC, Sowa NA, FitzGerald DJ (2006) Small RNA regulators and the bacterial response to stress. *Cold Spring Harb Symp Quant Biol* **71:** 1-11

Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR (2003) Rfam: an RNA family database. *Nucleic Acids Res* **31:** 439-441

Gripenland J, Netterling S, Loh E, Tiensuu T, Toledo-Arana A, Johansson J (2010) RNAs: regulators of bacterial virulence. *Nat Rev Microbiol* **8:** 857-866

Gruel N, Chapiro J, Fridman WH, Teillaud JL (2001) Purification of soluble recombinant human FcgammaRII (CD32). *Preparative biochemistry & biotechnology* **31:** 341-354

Gudapaty S, Suzuki K, Wang X, Babitzke P, Romeo T (2001) Regulatory interactions of Csr components: the RNA binding protein CsrA activates csrB transcription in Escherichia coli. *J Bacteriol* **183:** 6017-6027

Guisbert E, Rhodius VA, Ahuja N, Witkin E, Gross CA (2007) Hfq modulates the sigmaE-mediated envelope stress response and the sigma32-mediated cytoplasmic stress response in Escherichia coli. *J Bacteriol* **189:** 1963-1973

Gumbart J, Schreiner E, Wilson DN, Beckmann R, Schulten K (2012) Mechanisms of SecM-mediated stalling in the ribosome. *Biophysical journal* **103:** 331-341

Harris LM, Blank L, Desai RP, Welker NE, Papoutsakis ET (2001) Fermentation characterization and flux analysis of recombinant strains of Clostridium acetobutylicum with an inactivated solR gene. *J Ind Microbiol Biotechnol* **27:** 322-328

He MX, Wu B, Shui ZX, Hu QC, Wang WG, Tan FR, Tang XY, Zhu QL, Pan K, Li Q, Su XH (2012) Transcriptome profiling of Zymomonas mobilis under ethanol stress. *Biotechnol Biofuels* **5:** 75

Ingolia N (2014) Ribosome Profiling Analysis of In Vivo Translation. In *Regulatory Nascent Polypeptides*, Ito K (ed), 7, pp 119-133. Springer Japan

Ingram LO, Conway T, Clark DP, Sewell GW, Preston JF (1987) Genetic engineering of ethanol production in Escherichia coli. *Appl Environ Microbiol* **53:** 2420-2425

Ivan E, Colovai AI (2006) Human Fc receptors: critical targets in the treatment of autoimmune diseases and transplant rejections. *Human immunology* **67:** 479-491

Jang YS, Park JM, Choi S, Choi YJ, Seung do Y, Cho JH, Lee SY (2012) Engineering of microorganisms for the production of biofuels and perspectives based on systems metabolic engineering approaches. *Biotechnology advances* **30:** 989-1000

Jarboe LR, Grabar TB, Yomano LP, Shanmugan KT, Ingram LO (2007) Development of ethanologenic bacteria. *Advances in biochemical engineering/biotechnology* **108:** 237-261

Jeffries TW (2005) Ethanol fermentation on the move. *Nat Biotechnol* **23:** 40-41

Jewett MC, Fritz BR, Timmerman LE, Church GM (2013) In vitro integration of ribosomal RNA synthesis, ribosome assembly, and translation. *Mol Syst Biol* **9:** 678

Jiang Y, Xu C, Dong F, Yang Y, Jiang W, Yang S (2009) Disruption of the acetoacetate decarboxylase gene in solvent-producing Clostridium acetobutylicum increases the butanol ratio. *Metab Eng* **11:** 284-291

Jin C, Yao MF, Liu HF, Lee CFF, Ji J (2011) Progress in the production and application of n-butanol as a biofuel. *Renew Sust Energ Rev* **15:** 4080-4106

Jung ST, Kang TH, Georgiou G (2010) Efficient expression and purification of human aglycosylated Fcgamma receptors in Escherichia coli. *Biotechnology and bioengineering* **107:** 21-30

Junker F, Ramos JL (1999) Involvement of the cis/trans isomerase Cti in solvent resistance of Pseudomonas putida DOT-T1E. *J Bacteriol* **181:** 5693-5700

Kalscheuer R, Stolting T, Steinbuchel A (2006) Microdiesel: Escherichia coli engineered for fuel production. *Microbiology* **152:** 2529-2536

Kang A, Chang MW (2012) Identification and reconstitution of genetic regulatory networks for improved microbial tolerance to isooctane. *Molecular bioSystems* **8:** 1350-1358

Kang Z, Zhang C, Zhang J, Jin P, Zhang J, Du G, Chen J (2014) Small RNA regulators in bacteria: powerful tools for metabolic engineering and synthetic biology. *Applied microbiology and biotechnology* **98:** 3413-3424

Kast P (1994) pKSS--a second-generation general purpose cloning vector for efficient positive selection of recombinant clones. *Gene* **138:** 109-114

Kawano M, Reynolds AA, Miranda-Rios J, Storz G (2005) Detection of 5'- and 3'-UTR-derived small RNAs and cis-encoded antisense RNAs in Escherichia coli. *Nucleic Acids Res* **33:** 1040-1050

Kim Y, Ingram LO, Shanmugam KT (2007) Construction of an Escherichia coli K-12 mutant for homoethanologenic fermentation of glucose or xylose without foreign genes. *Appl Environ Microbiol* **73:** 1766-1771

Kiran MD, Prakash JS, Annapoorni S, Dube S, Kusano T, Okuyama H, Murata N, Shivaji S (2004) Psychrophilic Pseudomonas syringae requires trans-monounsaturated fatty acid for growth at higher temperature. *Extremophiles : life under extreme conditions* **8:** 401-410

Ko JH, Altman S (2007) OLE RNA, an RNA motif that is highly conserved in several extremophilic bacteria, is a substrate for and can be regulated by RNase P RNA. *Proceedings of the National Academy of Sciences of the United States of America* **104:** 7815-7820

Koharyova M, Kolarova M (2008) Oxidative stress and thioredoxin system. *General physiology and biophysics* **27:** 71-84

Koo JT, Alleyne TM, Schiano CA, Jafari N, Lathem WW (2011) Global discovery of small RNAs in Yersinia pseudotuberculosis identifies Yersinia-specific small, noncoding RNAs required for virulence. *Proceedings of the National Academy of Sciences of the United States of America* **108:** E709-717

Kramer G, Rauch T, Rist W, Vorderwulbecke S, Patzelt H, Schulze-Specking A, Ban N, Deuerling E, Bukau B (2002) L23 protein functions as a chaperone docking site on the ribosome. *Nature* **419:** 171-174

Kristensen O, Gajhede M (2003) Chaperone binding at the ribosomal exit tunnel. *Structure* **11:** 1547-1556

Lamichhane G, Arnvig KB, McDonough KA (2013) Definition and annotation of (myco)bacterial non-coding RNA. *Tuberculosis* **93:** 26-29

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357-359

Lee JY, Jang YS, Lee J, Papoutsakis ET, Lee SY (2009) Metabolic engineering of Clostridium acetobutylicum M5 for highly selective butanol production. *Biotechnology journal* **4:** 1432-1440

Lee JY, Yang KS, Jang SA, Sung BH, Kim SC (2011) Engineering butanol-tolerance in escherichia coli with artificial transcription factor libraries. *Biotechnology and bioengineering* **108:** 742-749

Lee KY, Park JM, Kim TY, Yun H, Lee SY (2010) The genome-scale metabolic network analysis of Zymomonas mobilis ZM4 explains physiological features and suggests ethanol and succinic acid production strategies. *Microbial cell factories* **9:** 94

Lee RC, Feinbaum RL, Ambros V (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75:** 843-854

Lee SK, Chou H, Ham TS, Lee TS, Keasling JD (2008a) Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Curr Opin Biotechnol* **19:** 556-563

Lee SY, Park JH, Jang SH, Nielsen LK, Kim J, Jung KS (2008b) Fermentative butanol production by Clostridia. *Biotechnology and bioengineering* **101:** 209-228

Lewandowski B, De Bo G, Ward JW, Papmeyer M, Kuschel S, Aldegunde MJ, Gramlich PM, Heckmann D, Goldup SM, D'Souza DM, Fernandes AE, Leigh DA (2013) Sequence-specific peptide synthesis by an artificial small-molecule machine. *Science* **339:** 189-193

Lewis JA, Elkon IM, McGee MA, Higbee AJ, Gasch AP (2010) Exploiting natural variation in Saccharomyces cerevisiae to identify genes for increased ethanol resistance. *Genetics* **186:** 1197-1205

Li F, Wang Y, Gong K, Wang Q, Liang Q, Qi Q (2014) Constitutive expression of RyhB regulates the heme biosynthesis pathway and increases the 5-aminolevulinic acid accumulation in Escherichia coli. *FEMS Microbiol Lett* **350:** 209-215

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25:** 1754-1760

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26:** 589-595

Livny J (2012) Bioinformatic discovery of bacterial regulatory RNAs using SIPHT. *Methods in molecular biology* **905:** 3-14

Livny J, Waldor MK (2007) Identification of small RNAs in diverse bacterial species. *Current opinion in microbiology* **10:** 96-101

Loh E, Dussurget O, Gripenland J, Vaitkevicius K, Tiensuu T, Mandin P, Repoila F, Buchrieser C, Cossart P, Johansson J (2009) A trans-Acting Riboswitch Controls Expression of the Virulence Regulator PrfA in Listeria monocytogenes. *Cell* **139:** 770-779

Lynch SA, Desai SK, Sajja HK, Gallivan JP (2007) A high-throughput screen for synthetic riboswitches reveals mechanistic insights into their function. *Chemistry & biology* **14:** 173-184

Maenaka K, van der Merwe PA, Stuart DI, Jones EY, Sondermann P (2001) The human low affinity Fcgamma receptors IIa, IIb, and III bind IgG with fast kinetics and distinct thermodynamic properties. *J Biol Chem* **276:** 44898-44904

Maier T, Guell M, Serrano L (2009) Correlation of mRNA and protein in complex biological samples. *FEBS letters* **583:** 3966-3973

Mars RA, Nicolas P, Ciccolini M, Reilman E, Reder A, Schaffer M, Mader U, Volker U, van Dijl JM, Denham EL (2015) Small regulatory RNA-induced growth rate heterogeneity of Bacillus subtilis. *PLoS Genet* **11:** e1005046

Martin VJ, Pitera DJ, Withers ST, Newman JD, Keasling JD (2003) Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nat Biotechnol* **21:** 796-802

Masse E, Salvail H, Desnoyers G, Arguin M (2007) Small RNAs controlling iron metabolism. *Curr Opin Microbiol* **10:** 140-145

Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proceedings of the National Academy of Sciences of the United States of America* **101:** 7287-7292

McCoy J, La Ville E (2001) Expression and purification of thioredoxin fusion proteins. *Current protocols in protein science / editorial board, John E Coligan   [et al]* **Chapter 6:** Unit 6 7

Meyer MM, Hammond MC, Salinas Y, Roth A, Sudarsan N, Breaker RR (2011) Challenges of ligand identification for riboswitch candidates. *RNA biology* **8:** 5-10

Miyazaki K (2015) Molecular engineering of a PheS counterselection marker for improved operating efficiency in Escherichia coli. *BioTechniques* **58:** 86-88

Modi SR, Camacho DM, Kohanski MA, Walker GC, Collins JJ (2011) Functional characterization of bacterial sRNAs using a network biology approach. *Proc Natl Acad Sci U S A* **108:** 15522-15527

Mohagheghi A, Linger J, Smith H, Yang S, Dowe N, Pienkos PT (2014) Improving xylose utilization by recombinant Zymomonas mobilis strain 8b through adaptation using 2-deoxyglucose. *Biotechnol Biofuels* **7:** 19

Moreau RA, Powell MJ, Fett WF, Whitaker BD (1997) The effect of ethanol and oxygen on the growth of Zymomonas mobilis and the levels of hopanoids and other membrane lipids. *Current microbiology* **35:** 124-128

Morris KV, Mattick JS (2014) The rise of regulatory RNA. *Nat Rev Genet* **15:** 423-437

Nahvi A, Barrick JE, Breaker RR (2004) Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Res* **32:** 143-150

Nahvi A, Sudarsan N, Ebert MS, Zou X, Brown KL, Breaker RR (2002) Genetic control by a metabolite binding mRNA. *Chemistry & biology* **9:** 1043

Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, Floden EW, Gardner PP, Jones TA, Tate J, Finn RD (2015) Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res* **43:** D130-137

Nechooshtan G, Elgrably-Weiss M, Sheaffer A, Westhof E, Altuvia S (2009) A pH-responsive riboregulator. *Genes Dev* **23:** 2650-2662

Nettleship JE, Assenberg R, Diprose JM, Rahman-Huq N, Owens RJ (2010) Recent advances in the production of proteins in insect and mammalian cells for structural biology. *Journal of structural biology* **172:** 55-65

Ohta K, Beall DS, Mejia JP, Shanmugam KT, Ingram LO (1991) Genetic improvement of Escherichia coli for ethanol production: chromosomal integration of Zymomonas mobilis genes encoding pyruvate decarboxylase and alcohol dehydrogenase II. *Appl Environ Microbiol* **57:** 893-900

Oliva G, Sahr T, Buchrieser C (2015) Small RNAs, 5' UTR elements and RNA-binding proteins in intracellular bacteria: impact on metabolism and virulence. *Fems Microbiol Rev* **39:** 331-349

Olivarius S, Plessy C, Carninci P (2009) High-throughput verification of transcriptional starting sites by Deep-RACE. *BioTechniques* **46:** 130-132

Osman YA, Ingram LO (1985) Mechanism of ethanol inhibition of fermentation in Zymomonas mobilis CP4. *J Bacteriol* **164:** 173-180

Papenfort K, Pfeiffer V, Lucchini S, Sonawane A, Hinton JC, Vogel J (2008) Systematic deletion of Salmonella small RNA genes identifies CyaR, a conserved CRP-dependent riboregulator of OmpX synthesis. *Mol Microbiol* **68:** 890-906

Paredes CJ, Alsaker KV, Papoutsakis ET (2005) A comparative genomic view of clostridial sporulation and physiology. *Nat Rev Microbiol* **3:** 969-978

Picard F, Milhem H, Loubiere P, Laurent B, Cocaign-Bousquet M, Girbal L (2012) Bacterial translational regulations: high diversity between all mRNAs and major role in gene expression. *BMC genomics* **13:** 528

Qi LS, Arkin AP (2014) A versatile framework for microbial engineering using synthetic non-coding RNAs. *Nat Rev Microbiol* **12:** 341-354

Rabhi M, Espeli O, Schwartz A, Cayrol B, Rahmouni AR, Arluison V, Boudvillain M (2011) The Sm-like RNA chaperone Hfq mediates transcription antitermination at Rho-dependent terminators. *EMBO J* **30:** 2805-2816

Repoila F, Majdalani N, Gottesman S (2003) Small non-coding RNAs, co-ordinators of adaptation processes in Escherichia coli: the RpoS paradigm. *Mol Microbiol* **48:** 855-861

Reyes LH, Almario MP, Kao KC (2011) Genomic library screens for genes involved in n-butanol tolerance in Escherichia coli. *PloS one* **6:** e17678

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. *Nat Biotechnol* **29:** 24-26

Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS (2002) Comparative genomics of thiamin biosynthesis in procaryotes - New genes and regulatory mechanisms. *J Biol Chem* **277:** 48949-48959

Rogers PL, Jeon YJ, Lee KJ, Lawford HG (2007) Zymomonas mobilis for fuel ethanol and higher value products. *Advances in biochemical engineering/biotechnology* **108:** 263-288

Romby P, Charpentier E (2010) An overview of RNAs with regulatory functions in gram-positive bacteria. *Cellular and molecular life sciences : CMLS* **67:** 217-237

Rosano GL, Ceccarelli EA (2014) Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in microbiology* **5:** 172

Roth A, Winkler WC, Regulski EE, Lee BW, Lim J, Jona I, Barrick JE, Ritwik A, Kim JN, Welz R, Iwata-Reuyl D, Breaker RR (2007) A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. *Nature structural & molecular biology* **14:** 308-317

Rude MA, Schirmer A (2009) New microbial fuels: a biotech perspective. *Current opinion in microbiology* **12:** 274-281

Said N, Rieder R, Hurwitz R, Deckert J, Urlaub H, Vogel J (2009) In vivo expression and purification of aptamer-tagged small RNA regulators. *Nucleic Acids Res* **37:** e133

Saida F (2007) Overview on the expression of toxic gene products in Escherichia coli. *Current protocols in protein science / editorial board, John E Coligan   [et al]* **Chapter 5:** Unit 5 19

Salis HM, Mirsky EA, Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* **27:** 946-950

Seo JS, Chong H, Park HS, Yoon KO, Jung C, Kim JJ, Hong JH, Kim H, Kim JH, Kil JI, Park CJ, Oh HM, Lee JS, Jin SJ, Um HW, Lee HJ, Oh SJ, Kim JY, Kang HL, Lee SY, Lee KJ, Kang HS (2005) The genome sequence of the ethanologenic bacterium Zymomonas mobilis ZM4. *Nat Biotechnol* **23:** 63-68

Seo SW, Yang JS, Kim I, Yang J, Min BE, Kim S, Jung GY (2013) Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency. *Metab Eng* **15:** 67-74

Shapiro RS, Cowen LE (2012) Thermal control of microbial development and virulence: molecular mechanisms of microbial temperature sensing. *MBio* **3**

Shui ZX, Qin H, Wu B, Ruan ZY, Wang LS, Tan FR, Wang JL, Tang XY, Dai LC, Hu GQ, He MX (2015) Adaptive laboratory evolution of ethanologenic Zymomonas mobilis strain tolerant to furfural and acetic acid inhibitors. *Applied microbiology and biotechnology* **99:** 5739-5748

Sittka A, Lucchini S, Papenfort K, Sharma CM, Rolle K, Binnewies TT, Hinton JC, Vogel J (2008) Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS genetics* **4:** e1000163

Slobodin B, Gerst JE (2010) A novel mRNA affinity purification technique for the identification of interacting proteins and transcripts in ribonucleoprotein complexes. *Rna* **16:** 2277-2290

Sorensen HP, Kristensen JE, Sperling-Petersen HU, Mortensen KK (2004) Soluble expression of aggregating proteins by covalent coupling to the ribosome. *Biochemical And Biophysical Research Communications* **319:** 715-719

Sorensen HP, Mortensen KK (2005) Advanced genetic strategies for recombinant protein expression in Escherichia coli. *Journal of biotechnology* **115:** 113-128

Soukup JK, Soukup GA (2004) Riboswitches exert genetic control through metabolite-induced conformational change. *Curr Opin Struc Biol* **14:** 344-349

Sridhar J, Gunasekaran P (2013) Computational small RNA prediction in bacteria. *Bioinformatics and biology insights* **7:** 83-95

Stanley D, Bandara A, Fraser S, Chambers PJ, Stanley GA (2010) The ethanol stress response and ethanol tolerance of Saccharomyces cerevisiae. *J Appl Microbiol* **109:** 13-24

Stephanopoulos G (2007) Challenges in engineering microbes for biofuels production. *Science* **315:** 801-804

Stockbridge RB, Lim HH, Otten R, Williams C, Shane T, Weinberg Z, Miller C (2012) Fluoride resistance and transport by riboswitch-controlled CLC antiporters. *Proceedings of the National Academy of Sciences of the United States of America* **109:** 15289-15294

Storz G, Vogel J, Wassarman KM (2011) Regulation by small RNAs in bacteria: expanding frontiers. *Molecular cell* **43:** 880-891

Suess B, Fink B, Berens C, Stentz R, Hillen W (2004) A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. *Nucleic Acids Res* **32:** 1610-1614

Swings J, De Ley J (1977) The biology of Zymomonas. *Bacteriological reviews* **41:** 1-46

Thanonkeo P, Sootsuwan K, Leelavacharamas V, Yamada M (2007) Cloning and transcriptional analysis of groES and groEL in ethanol-producing bacterium Zymomonas mobilis TISTR 548. *Pak J Biol Sci* **10:** 13-22

Thiebaut F, Grativol C, Carnavale-Bottino M, Rojas CA, Tanurdzic M, Farinelli L, Martienssen RA, Hemerly AS, Ferreira PC (2012) Computational identification and analysis of novel sugarcane microRNAs. *BMC genomics* **13:** 290

Toledo-Arana A, Dussurget O, Nikitas G, Sesto N, Guet-Revillet H, Balestrino D, Loh E, Gripenland J, Tiensuu T, Vaitkevicius K, Barthelemy M, Vergassola M, Nahori MA, Soubigou G, Regnault B, Coppee JY, Lecuit M, Johansson J, Cossart P (2009) The Listeria transcriptional landscape from saprophytism to virulence. *Nature* **459:** 950-956

Tomas CA, Beamish J, Papoutsakis ET (2004) Transcriptional analysis of butanol stress and tolerance in Clostridium acetobutylicum. *J Bacteriol* **186:** 2006-2018

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

Tomes DT, Lakshmanan P, Songstad D (2011) *Biofuels : global impact on renewable energy, production agriculture and technological advancements*,   New York: Springer.

Topp S, Reynoso CM, Seeliger JC, Goldlust IS, Desai SK, Murat D, Shen A, Puri AW, Komeili A, Bertozzi CR, Scott JR, Gallivan JP (2010) Synthetic riboswitches that induce gene expression in diverse bacterial species. *Appl Environ Microbiol* **76:** 7881-7884

Torres-Quesada O, Reinkensmeier J, Schluter JP, Robledo M, Peregrina A, Giegerich R, Toro N, Becker A, Jimenez-Zurdo JI (2014) Genome-wide profiling of Hfq-binding RNAs uncovers extensive post-transcriptional rewiring of major stress response and symbiotic regulons in Sinorhizobium meliloti. *RNA Biol* **11:** 563-579

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols* **7:** 562-578

Tsai CH, Baranowski C, Livny J, McDonough KA, Wade JT, Contreras LM (2013) Identification of Novel sRNAs in Mycobacterial Species. *PloS one* **8:** e79411

Tsai CH, Liao R, Chou B, Palumbo M, Contreras LM (2015) Genome-wide analyses in bacteria show small-RNA enrichment for long and conserved intergenic regions. *J Bacteriol* **197:** 40-50

Varallyay E, Burgyan J, Havelda Z (2008) MicroRNA detection by northern blotting using locked nucleic acid probes. *Nature protocols* **3:** 190-196

Vazquez-Anderson J, Contreras LM (2013a) Regulatory RNAs Charming gene management styles for synthetic biology applications. *RNA biology* **10:** 1778-1797

Vazquez-Anderson J, Contreras LM (2013b) Regulatory RNAs: charming gene management styles for synthetic biology applications. *RNA biology* **10:** 1778-1797

Venkataramanan KP, Jones SW, McCormick KP, Kunjeti SG, Ralston MT, Meyers BC, Papoutsakis ET (2013) The Clostridium small RNome that responds to stress: the paradigm and importance of toxic metabolite stress in C. acetobutylicum. *BMC genomics* **14:** 849

Villaverde A, Carrio MM (2003) Protein aggregation in recombinant bacteria: biological role of inclusion bodies. *Biotechnology letters* **25:** 1385-1395

Vogel J, Bartels V, Tang TH, Churakov G, Slagter-Jager JG, Huttenhofer A, Wagner EG (2003) RNomics in Escherichia coli detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res* **31:** 6435-6443

Vogel J, Luisi BF (2011) Hfq and its constellation of RNA. *Nat Rev Microbiol* **9:** 578-589

Vogel J, Sharma CM (2005) How to find small non-coding RNAs in bacteria. *Biol Chem* **386:** 1219-1238

Waldo GS, Standish BM, Berendzen J, Terwilliger TC (1999) Rapid protein-folding assay using green fluorescent protein. *Nat Biotechnol* **17:** 691-695

Wallace JG, Zhou Z, Breaker RR (2012) OLE RNA protects extremophilic bacteria from alcohol toxicity. *Nucleic Acids Res* **40:** 6898-6907

Wassarman KM (2002) Small RNAs in bacteria: diverse regulators of gene expression in response to environmental changes. *Cell* **109:** 141-144

Waters LS, Storz G (2009) Regulatory RNAs in bacteria. *Cell* **136:** 615-628

Wegrzyn RD, Hofmann D, Merz F, Nikolay R, Rauch T, Graf C, Deuerling E (2006) A conserved motif is prerequisite for the interaction of NAC with ribosomal protein L23 and nascent chains. *J Biol Chem* **281:** 2847-2857

Weinberg Z, Barrick JE, Yao Z, Roth A, Kim JN, Gore J, Wang JX, Lee ER, Block KF, Sudarsan N, Neph S, Tompa M, Ruzzo WL, Breaker RR (2007) Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res* **35:** 4809-4819

Widiastuti H, Kim JY, Selvarasu S, Karimi IA, Kim H, Seo JS, Lee DY (2011) Genome-scale modeling and in silico analysis of ethanologenic bacteria Zymomonas mobilis. *Biotechnology and bioengineering* **108:** 655-665

Will S, Joshi T, Hofacker IL, Stadler PF, Backofen R (2012) LocARNA-P: accurate boundary prediction and improved detection of structural RNAs. *RNA* **18:** 900-914

Wilson DN, Gupta R, Mikolajka A, Nierhaus KH (2001) Ribosomal Proteins: Role in Ribosomal Functions. In *eLS*. John Wiley & Sons, Ltd

Winkler W, Nahvi A, Breaker RR (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419:** 952-956

Winkler WC, Breaker RR (2003) Genetic control by metabolite-binding riboswitches. *Chembiochem : a European journal of chemical biology* **4:** 1024-1032

Wright PR, Georg J, Mann M, Sorescu DA, Richter AS, Lott S, Kleinkauf R, Hess WR, Backofen R (2014) CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. *Nucleic Acids Res* **42:** W119-W123

Wright PR, Richter AS, Papenfort K, Mann M, Vogel J, Hess WR, Backofen R, Georg J (2013) Comparative genomics boosts target prediction for bacterial small RNAs. *Proceedings of the National Academy of Sciences of the United States of America* **110:** E3487-E3496

Xu M, Zhao J, Yu L, Tang IC, Xue C, Yang ST (2015) Engineering Clostridium acetobutylicum with a histidine kinase knockout for enhanced n-butanol tolerance and production. *Applied microbiology and biotechnology* **99:** 1011-1022

Xue C, Zhao XQ, Liu CG, Chen LJ, Bai FW (2013) Prospective and development of butanol as an advanced biofuel. *Biotechnology advances* **31:** 1575-1584

Yang S, Franden MA, Brown SD, Chou YC, Pienkos PT, Zhang M (2014a) Insights into acetate toxicity in Zymomonas mobilis 8b using different substrates. *Biotechnol Biofuels* **7:** 140

Yang S, Land ML, Klingeman DM, Pelletier DA, Lu TY, Martin SL, Guo HB, Smith JC, Brown SD (2010a) Paradigm for industrial strain improvement identifies sodium acetate tolerance loci in Zymomonas mobilis and Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America* **107:** 10395-10400

Yang S, Pan C, Hurst GB, Dice L, Davison BH, Brown SD (2014b) Elucidation of Zymomonas mobilis physiology and stress responses by quantitative proteomics and transcriptomics. *Frontiers in microbiology* **5:** 246

Yang S, Pan C, Tschaplinski TJ, Hurst GB, Engle NL, Zhou W, Dam P, Xu Y, Rodriguez M, Jr., Dice L, Johnson CM, Davison BH, Brown SD (2013) Systems biology analysis of Zymomonas mobilis ZM4 ethanol stress responses. *PloS one* **8:** e68886

Yang S, Pappas KM, Hauser LJ, Land ML, Chen GL, Hurst GB, Pan C, Kouvelis VN, Typas MA, Pelletier DA, Klingeman DM, Chang YJ, Samatova NF, Brown SD (2009a) Improved genome annotation for Zymomonas mobilis. *Nat Biotechnol* **27:** 893-894

Yang S, Pelletier DA, Lu TY, Brown SD (2010b) The Zymomonas mobilis regulator hfq contributes to tolerance against multiple lignocellulosic pretreatment inhibitors. *BMC Microbiol* **10:** 135

Yang S, Tschaplinski TJ, Engle NL, Carroll SL, Martin SL, Davison BH, Palumbo AV, Rodriguez M, Jr., Brown SD (2009b) Transcriptomic and metabolomic profiling of Zymomonas mobilis during aerobic and anaerobic fermentations. *BMC genomics* **10:** 34

Yao Z, Weinberg Z, Ruzzo WL (2006) CMfinder--a covariance model based RNA motif finding algorithm. *Bioinformatics* **22:** 445-452

Yomano LP, York SW, Ingram LO (1998) Isolation and characterization of ethanol-tolerant mutants of Escherichia coli KO11 for fuel ethanol production. *J Ind Microbiol Biotechnol* **20:** 132-138

Yomano LP, York SW, Zhou S, Shanmugam KT, Ingram LO (2008) Re-engineering Escherichia coli for ethanol production. *Biotechnology letters* **30:** 2097-2103

Yoshikawa K, Tanaka T, Furusawa C, Nagahisa K, Hirasawa T, Shimizu H (2009) Comprehensive phenotypic analysis for identification of genes affecting growth under ethanol stress in Saccharomyces cerevisiae. *FEMS Yeast Res* **9:** 32-44

Zhang A, Wassarman KM, Rosenow C, Tjaden BC, Storz G, Gottesman S (2003) Global analysis of small RNA and mRNA targets of Hfq. *Molecular microbiology* **50:** 1111-1124

Zhang H, Chong H, Ching CB, Song H, Jiang R (2012) Engineering global transcription factor cyclic AMP receptor protein of Escherichia coli for improved 1-butanol tolerance. *Applied microbiology and biotechnology* **94:** 1107-1117

Zhang M, Chou YC, Howe W, Eddy C, Evans K, A M. (2007) Zymomonas pentose-sugar fermenting strains and uses thereof., US patent, Vol. US7223575.

Zhang M, Eddy C, Deanda K, Finkelstein M, Picataggio S (1995) Metabolic Engineering of a Pentose Metabolism Pathway in Ethanologenic Zymomonas mobilis. *Science* **267:** 240-243

Zheng M, Doan B, Schneider TD, Storz G (1999) OxyR and SoxRS regulation of fur. *J Bacteriol* **181:** 4639-4643

Zheng Y, Chen X, Shen Y (2008) Commodity chemicals derived from glycerol, an important biorefinery feedstock. *Chem Rev* **108:** 5253-5277

Zuroff TR, Xiques SB, Curtis WR (2013) Consortia-mediated bioprocessing of cellulose to ethanol with a symbiotic Clostridium phytofermentans/yeast co-culture. *Biotechnol Biofuels* **6:** 59

VITA

Seung Hee Cho was born in Seoul, South Korea. Seung Hee Cho attended Sehwa girl's high school, Seoul, Korea. In 1999, she entered the Department of Biotechnology, College of Engineering, Yonsei University, Seoul, Korea. Upon receiving her Bachelor of Science from Yonsei University in August 2003, she entered graduate school in Yonsei University and received her master's degree of science in biotechnology on February 2006. In 2006, she worked as a researcher in Dong-A Pharmaceutical R&D center, Biotech team II. In September 2008, she entered Graduate School of the University of Texas at Austin to pursue a doctoral degree.

.

Permanent Contact: seungheecho0325@gmail.com

This manuscript was typed by Seung Hee Cho.