The Dissertation Committee for Chen-Hsun Tsai Certifies that this is the approved version of the following dissertation:

Identification and characterization of novel regulatory small RNAs in *Deinococcus radiodurans*

Committee:

Lydia M. Contreras, Supervisor

Hal Alper

Jeffrey Barrick

Bryan Davies

George Georgiou

Identification and characterization of novel regulatory small RNAs in *Deinococcus radiodurans*

by

Chen-Hsun Tsai

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

December 2017

Identification and characterization of novel regulatory small RNAs in *Deinococcus radiodurans*

Chen-Hsun Tsai, PhD

The University of Texas at Austin, 2017

Supervisor: Lydia M. Contreras

Gene expression regulation mechanisms during stress recovery are ubiquitous across all organisms. Some organisms have evolved more robust and distinct regulatory systems to survive exposures of extreme conditions. For instance, *Deinococcus radiodurans* is one of the most radioresistive organisms that have been identified. It can withstand 15,000 Gray of gamma radiation, which is 5,000 times of human and 250 times of E. coli. Many studies have been done to decipher the radioresistance mechanisms of *D. radiodurans*. However, most analysis with the genome and proteome are not conclusive. Although a recent study suggested that the small molecules that protect the proteome of *D. radiodurans* under irradiation are a key to its radioresistance, the mechanisms of gene regulation under irradiation are still unclear. In this study, we have introduced small non-coding RNAs (sRNA) as potential regulators in the radioresistance mechanisms. In bacteria, sRNAs have been identified in multiple species and have been shown to play essential roles in responsive mechanisms to environmental stresses. A key property of sRNAs is their ability to up or down regulate global metabolic pathways in response to lethal environmental signals. The versatility and specificity of sRNA regulation make it highly relevant for developing engineering applications.

Since very little about sRNAs is known or discovered in *D. radiodurans*, we have used bioinformatics algorithms to analyze genomics patterns of sRNAs coding regions in 13 bacterial species to develop a pipeline and have identified novel transcripts in bacteria

based on the conservation level and size of the intergenic regions. For example, alone with computation predictions, we found 40 novel transcripts in *Mycobacteria smegmatis* and *Mycobacteria bovis*. The combination of this bioinformatics approach and experimental verification has led us to identify 41 sRNAs in *D. radiodurans*, and many of them show differential expression during recovery from exposures to various environmental stresses. Among them, one transcript Dsr2 showed a unique expression pattern under ionizing radiation and the deletion mutant has a reduced survival rate under irradiation. We designed multiple experiments to verify the regulatory networks of Dsr2, including HITS-CLIP analysis, transcriptome analysis and electrophoresis mobility shift assay. We have found 14 mRNAs that have direct interaction with Dsr2 sRNA and many of them have important functions related stress response. For example, PprM protein has been verified as a modulator of PprA protein, which activates DNA end-repair under stresses. This pathway is also regulated by PprI but the regulatory mechanism of PprM was previously unclear. We found Dsr2 is able to bind with *pprM* and the transcriptome analysis suggested its expression level is correlated with Dsr2 expression. Dsr2 was also shown to have interactions with many ribosomal proteins and translation initiation factors. In summary, our work provides insight and evidence of how non-coding transcripts regulate radioresistance mechanisms in *D. radiodurans*, and how bioinformatics analysis can lead us to the discovery of more regulatory transcripts in bacteria of interest.

Table of Contents

List of Tables x

List of Figures xii

Chapter One 1

Introduction and background 1

1.1 Regulatory ncRNAs, challenges and important findings 1

1.2 Regulatory role of sRNAs in bacterial stress responses and its lack of discovery in *D. radiodurans* 5

1.3 Current bioinformatics and experimental methods of sRNAs identification 8

1.4 *D. radiodurans* and current understandings of its radioresistance mechanisms 11

1.5 Summary of research objectives and accomplishments 14

Chapter Two 17

*Develop bioinformatics pipeline with genomic comparative analysis to refine experimental identification of novel sRNAs in bacteria 17

2. 1 Introduction 17

2.2 Results 20

2.2.1 Analysis of intergenic regions in different species shows conservation of a large number of intergenic regions 20

2.2.2 Conserved intergenic regions are enriched for small RNAs 24

2.2.3 Refined conservation analysis based on phylogenetic distance strengthens observations of sRNA enrichment in conserved intergenic 26

2.2.4 Conservation of sRNAs relative to conservation of flanking coding regions 30

2.2.5 Isolated genomic regions are enriched in sRNAs 32

2.2.6 Prediction of sRNAs in silico using SIPHT 33

2.2.7 17 Novel sRNAs Identified in Mycobacterium Smegmatis 34

2.2.8 23 Novel sRNAs Identified in Mycobacterium Bovis BCG 36

2.2.9 Deep-RACE Identifies sRNA 5´ and 3´ Ends 38
2.3 Discussion 41
2.4 Material and Methods 48
2.4.1 Targeted bacterial species 48
2.4.2 Genome-wide extraction of intergenic and extended intergenic region sequences 48
2.4.3 Conservation analysis of different genomes by BLAST 49
2.4.4 Collection of experimentally observed sRNAs from published works 51
2.4.5 Phylogenetic distance calculation 51
2.4.6 Comparisons of all intergenic regions with experimentally observed sRNAs 51
2.4.7 Strains and Plasmids 52
2.4.8 Phylogenetic Selection of Computationally Predicted sRNA Candidate 52
2.4.9 RNA Isolation and Northern Blot Analysis 52
2.4.10 ChIP-seq Analysis 53
2.4.11 Deep 5´ and 3´ RACE 53

Chapter Three 54

*Identify novel ncRNAs in *D. radiodurans* with bioinformatics and transcriptome analysis 54
3.1 Introduction 54
3.2 Results 58
3.2.1 Deep sequencing reveals hundreds of potential transcripts from non-coding regions 58
3.2.2 sRNA transcripts verified by Northern blotting analysis from deep sequencing result 59
3.2.3 RT-PCR and co-transcription experiment identified more sRNA candidates 61
3.2.4 5' ends of the sRNA candidates were mapped by Deep-RACE .63
3.2.5 Confirmation of additional sRNAs from computational predictions 63

3.2.6 Differential expression of sRNAs during genome recovery after ionizing irradiation....66
3.2.7 Identification of conserved sRNAs in D. geothermalis 69
3.4 Discussion and conclusions 72
3.4 Methods....76
3.4.1 Bacterial strains and growth conditions....76
3.4.2 Preparation of protein lysate and Western blotting analysis....76
3.4.3 Total RNA extraction....76
3.4.4 Whole transcriptome deep sequencing and data analysis 77
3.4.5 Selection of computational predicted candidates....77
3.4.6 Ionizing irradiation....78
3.4.7 Northern blotting analysis....78
3.4.8 RT-PCR and co-transcription 79
3.4.9 Deep 5' RACE 79
3.4.10 ELISA analysis 80
Chapter Four 81
Idnetification and characterization of regulatory networks of small RNAs in *D. radiodurans* that contribute to radioresistance 81
4.1 Introduction....81
4.2 Results....85
4.2.1 Selection of Dsr2 D. radiodurans as a pioneering model to understand regulatory role of sRNAs in radioresistance 85
4.2.2 Defective phenotypes of deletion strains confirm Dsr2 as an important contributor to acute and chronic radiation resistance.87
4.4.3 Genome-wide transcriptional and translational effect of Dsr2 ...89
4.2.4 HITS-CLIPS analysis identifies direct potential Dsr2 targets 90
4.2.5 EMSA Analysis shows that Dsr2 directly binds 14 mRNA targets involved in PprM and translation regulation *in-vitro* 93
4.2.6 Go-term analysis of Dsr2 targets reveals additional layers to radiation resistance regulation in D. radiodurans 97
4.3 Discussion....99

4.4. Methods.....103
4.4.1 Bacterial growth conditions and stresses induction. 103
4.4.2 RNA extraction and Northern blotting 103
4.4.3 Construction of gene overexpression and deletion strains in *D. radiodurans*.....104
4.4.4 Survival rate measurements 105
4.4.5 HITS-CLIP analysis (MS2 pull-down analysis).....105
4.4.6 Transcriptome analysis 107
4.4.7 Electrophoretic mobility shift assay.....107
4.4.8 Proteomic Identification of Dsr2 Targets 108
Chapter Five.....111
Conclusions and perspectives 111
Appendices.....116
Appendix A: Supplementary data for Chapter Two 116
Supplementary figures for Chapter Two.....116
Supplementary Tables for Chapter Two 130
Appendix B: Supplementary data for Chapter Three 147
Supplementary figures for Chapter Three.....147
Supplementary Tables for Chapter Three 152
Appendix C: Supplementary data for Chapter Four 161
Supplementary figures for Chapter Four 161
Supplementary figures for Chapter Four 165
References.....180

List of Tables

Table 2.1. Novel sRNAs confirmed by Northern blotting analysis in *M. smegmatis* 37

Table 2.2. Novel sRNAs confirmed by Northern blotting analysis in *M. bovis*. 39

Table A.1. The distribution of the size of intergenic regions in analyzed species 130

Table A.2. BLAST target species 130

Table A.3. Intergenic analysis for all 13 species 138

Table A.4. Conservation level comparison of sRCRs and RIGRs 138

Table A.5. *M. Smegmatis* MC155 8596 ALL SIPHT Predictions 141

Table A.6. *M. Bovis* BCG_008769 ALL SIPHT Predictions 141

Table A.7. Probes sequences used in Northern blotting analysis 141

Table A.8. Transcription factor ChIP-seq peaks located within 100 bp upstream and 20 bp downstream of sRNA 5' ends. 144

Table A.9. List of sRNAs from database and literature 144

Table A.10. Minimum length of long intergenic regions with different thresholds in each species 145

Table A.11. The oligonucleotide sequence of all primers used for Deep-RACE PCR 145

Table B.1. Potential sRNAs identified with deep sequencing and computational prediction 152

Table B.2. sRNAs predicted by Rfam 152

Table B.3. Probes for Northern Blotting Analysis 152

Table B.4. RT-PCR Primers 155

Table B.5. Predicted sRNA by QRNA and SIPHT 158

Table B.6. Target RNA prediction 158

Table B.7. Probes for 5' RACE.....159

Table C.1. Differential expressed sRNA under acute irradiation from previous study 165

Table C.2. Differential expressed genes in Dsr2KD and Dsr2OE 165

Table C.3. Potential mRNA candidates of Dsr2 verified with MS2 pull-down (HITS-CLIP) 165

Table C.4. Proteomics analysis of the 47 MS2-pull down mRNAs......167

Table C.5. Predicted targets of Dsr2 by CopraRNA 168

Table C.6. Oligonucleotides used in this study.....171

Table C.7. Sequences of mRNA used for EMSA analysis.....174

List of Figures

Figure 2.1. Extended intergenic region. 21

Figure 2.2. Conservation patterns of intergenic regions in selected species...23

Figure 2.3. Enrichment of sRNAs in outside-genus conserved intergenic regions. 25

Figure 2.4. sRNA-coding-region enrichments in intergenic regions conserved in species with different phylogenetic distances and sRNAs enrichments with a refined species set. 27

Figure 2.5. Illustration of the sRNA-coding region in an intergenic region and comparison of conservation levels. 29

Figure 2.6. Enrichment of sRNAs in long and conserved intergenic regions (IGR). 31

Figure 2.7. Schematic for sRNA identification......33

Figure 2.8. Northern blotting confirmation of sRNA candidates in *M. smegmatis*. 35

Figure 2.9. Northern blotting confirmation of sRNA candidates in *M. bovis* BCG. 38

Figure 2.10. Identification of sRNA 5′ and 3′ ends by Deep RACE.....40

Figure 2.11. ChIP-seq peaks associated with predicted sRNA homologues in *M. tuberculosis*.....47

Figure 3.1. Novel small RNA candidates in D. radiodurans confirmed by Northern blotting and/or RT-PCR deep sequencing......60

Figure 3.2. Images of Northern blotting analysis for confirmed sRNAs candidates from deep sequencing analysis. 62

Figure 3.3. Novel small RNA candidates confirmed with Northern blotting analysis in D. radiodurans from computational predicted candidates 65

Figure 3.4. Images from Northern blotting analysis for confirmed sRNAs candidates from computational prediction.....66

Figure 3.5. Scheme of experimental procedure.....67

Figure 3.6. Differential expression of selected sRNAs.....69

Figure 3.7. Novel small RNA candidates confirmed with Northern blotting analysis in *D. geothermalis* and homologous counterparts in *D. radiodurans* 70

Figure 3.8. Images of Northern blotting analysis for sRNAs candidates with homology to *D. geothermalis* 72

Figure 4.1. Experiments show potential regulatory sRNA Dsr2 has unique response to ionizing irradiation. 86

Figure 4.2. Dsr2KD showed a significant reduced survival level under acute and chronic irradiation.....88

Figure 4.3. GO-term analysis of Dsr2 mutants under irradiation. 91

Figure 4.4. HITS-CLIP analysis and differential expression of 47 pull-down mRNA......92

Figure 4.5. *In-vitro* interaction of Dsr2-mRNAs were verified by EMSA for 14 potential targets.....94

Figure 4.6. Go-term analysis and prediction binding sites of the 14 targets ..96

Figure 4.7. Dsr2 as a global regulator in radioresistance pathways 98

Figure A.1. Conservation distribution and patterns in other ten species.....116

Figure A.2. sRNAs coding region enrichments in intergenic region with different conservation level for outside genus species 119
Figure A.3. The distribution of the size of intergenic regions in analyzed species 120
Figure A.4. Northern blotting analysis for *M. smegmatis* sRNAs 122
Figure A.5. Northern blotting analysis confirmation of sRNA candidates in *M. tuberculosis* with *M. smegmatis* and *M. bovis* probes. 123
Figure A.6. Northern blotting analysis confirmation of sRNA candidates in *M. bovis* BCG. 124
Figure A.7. Deep-RACE mapped reads of all sRNAs and adjacent gene annotations 125
Figure B.1. Deep sequencing read distribution 147
Figure B.2. Expression of the Y RNA was confirmed with the deep-seq data 147
Figure B.3. RT-PCR for sRNAs identified in *D. radiodurans* 148
Figure B.4. RT-PCR for sRNA co-transcription verification 149
Figure B.5. Selected examples of transcription starting sites identified by 5' Deep-RACE. 149
Figure B.6. Venn diagram of computational predicted sRNA candidates and deep-sequencing identified (transcriptome) sRNA candidates in D. radiodurans. 150
Figure B.7. Cells survival rate after 15 kGy irradiation 150
Figure B.8. Selected sRNAs of non-differential expression under irradiation 151
Figure B.9. Expression of DR_A0234 is confirmed with the deep sequencing analysis 151

Figure C.1. Northern blotting analysis for sRNAs of interest under different environmental stresses.....161
Figure C.2. Quantitative RTPCR of Dsr2 and DR_0141 under 10kGy acute irradiation.....162
Figure C.3. (A) Venn diagram of differential expressed gene in Dsr2KD and Dsr2OE under sham condition overlapping with the MS2 pull-down candidates (B) Expression profile of genes in Dsr2KD and Dsr2OE.
.....162
Figure C.4. Quantitative RTPCR of Dsr2 and EMSA verified in-vitro targets under acute irradiation.....163
Figure C.5. Predictions of Dsr2-mRNA binding sites by IntaRNA.....164
Figure C.6. Northern blotting analysis confirmation of MS2BD-Dsr2 transcript
.....164

Chapter One

Introduction and background

1.1 Regulatory ncRNAs, challenges and important findings

RNA has been at the center of novel research of genomic regulation since its functions other than the transmission of the Central Dogma was discovered [1]. The versatility of the regulatory functions of RNAs has created a new field of molecular biology in gene regulation. RNA interacts with other molecules and exerts regulation though binding its targets (via sequence homology) and exerting structural changes, enabled by its unique intrinsic abilities of binding and folding [2]. Regulatory RNAs have been identified in various organisms ranging from human to bacteria, and their function also varies. In bacteria, many regulatory RNAs are known to relate with the bioprocesses of pathogens or stress response mechanisms [3]. These regulatory RNAs are expressed under specific conditions and help the cells to survive. Understanding of these mechanisms will improve drug development and the discovery of novel pathways in extremophiles which is of great interest in bioengineering applications [4], [5]. Therefore, identification and characterization of regulatory RNAs have been an emerging topic in RNA research. This dissertation focuses on the identification of regulatory sRNAs and their potential mechanisms. The known sRNAs in bacteria were analyzed with bioinformatics approaches to develop a pipeline to refine experimental confirmation of novel ncRNAs in bacteria of interest, i.e. *Mycobacteria smegmatis*, *Mycobacteria bovis* and *D. radiodurans*. We further characterized the potential regulatory sRNAs in *D. radiodurans* that may contribute to radioresistance and verified direct interactions between sRNA and mRNAs. The study in this dissertation also provides insights of possible pathways and mechanisms of sRNA regulation in *D. radiodurans*. The first

regulatory ncRNA was found in 1980s. Small nuclear RNA, also known as snRNA, was identified to play a part in the intron splicing process [6]. At 1900s microRNA (miRNA) was discovered in *C. elegans*, which has 22 nucleotides long and control larval development by suppressing *lin-14* expression [7]. This encouraged further studies in regulatory RNAs. Today, regulatory RNAs have been identified in all domains of life, from human to bacteria, and new technologies such as deep sequencing and computational analysis have been facilitating the discovery effectively [8]–[11]. Various novel functions have been also discovered and overthrown many rules of RNA biology. For example, many RNAs are now known to be enzymatic, RNA processing is known to take place beyond the 3'end, the human genome can encode long ncRNAs and bacteria can use regulatory RNAs to repress or promote gene expression [3], [12]. One of the most exciting discoveries is that now scientists can use a regulatory system CRISPRs, which is originally used in bacteria against virus invasion, to edit designated sites in other organisms' genome including human and mouse cells [13]. All these studies indicate that understanding of ncRNA regulatory mechanisms can impact bioengineering and also drug discovery, medical treatment development, and environmental remediation.

In bacteria, three unique types of regulatory ncRNAs have been identified to participate in different cellular mechanisms: (i) riboswitches (ii) T box leaders and (iii) small RNAs. First, riboswitches are regulators that are induced mostly by small-molecules and regulate gene expression *in cis* [14]. Upon binding with small-molecule, riboswitches change their structures and free or block sites for translation initiation or termination [5]. Riboswitches are also found in eukaryotes and can also be induced by temperature changes or by cooperative binding with other regulatory biomolecules, [15], [16]. Secondly, T boxes (tRNA-binding leader element) are riboswitches that bind to uncharged tRNAs and prevent the formation of termination loops of certain genes [[17]].

T boxes are widely used and conserved in Gram-positive bacteria to regulate aminoacyl-tRNA synthetase [6]. Last, small RNAs (sRNAs) are antisense or structured RNAs that bind to an mRNA and up or down regulate its expression [18]. It can also induce mRNA degradation or prevent degradation [5]. sRNAs are the focus of this work. They represent the most versatile regulatory transcripts in bacteria and have great potential for regulating multiple mechanisms and pathways under different conditions of the cell. Thus, characterization and identification of sRNAs are key to understand bacteria cellular regulation. However, sRNA identification could be highly challenging due to the fact that they are typically short transcripts that display complex structure.

To identify sRNAs in bacteria, early studies rely on serendipity and biochemical evidence such as protein or RNA co-purification [ref]. In 1980s, *micF* sRNA was found in *E. coli* to repress OmpF expression under environmental stresses such as oxidative compounds [19]. *micF* one of the first discovered *trans*-encoded sRNA in bacteria. Since then, many sRNAs have been found in *E. coli* serendipitously [20]. In 2001, several studies of genome-wide searches of sRNAs in *E. coli* discovered more than 50 of sRNAs [21], and many were also found in other organisms [22]. However, experimental confirmation of sRNAs can be challenging. For example, many labeling or staining protocols require the transcript to be relatively abundant, and the results can be ambiguous [20]. Co-purification requires sRNAs to be tightly associated with their binding partners, and microarray detection can be expensive and inconsistent [20]. Thus, a combinatory of bioinformatics and experimental methods has been applied to reduce the time and increase the efficiency of sRNAs discovery. For example, 88 sRNAs were found in *Streptococcus pneumoniae* by deep sequencing and comparative analysis of the genome [23]. Many algorithms have also been created to predict the loci of potential sRNAs, such as QRNA, SIPHT and sRNAPredict [24]. Most of these tools utilize

sequence and structure conservation to score loci for sRNAs prediction. However, most of them still have low accuracy and require experimental confirmation.

Although sRNAs have been discovered in many bacterial species, their existence and mechanisms still remain unclear in most organisms, especially the ones without well-annotated genomes. sRNAs have been known to participate in stress regulation [3], and this could be an interesting missing piece for mechanistic studies of stress response in many extremophiles, non-model organisms that can survive in extreme environmental conditions such as high temperature, desiccation or radiation. One of the most interesting extremophile is *Deinococcus radiodurans*. *D. radiodurans* is a Gram-positive bacterium that can survive acute ionizing radiation up to 5000 times of human and 200 times of *E. coli* [25]. *D radiodurans* is known for its exceptional ability to repair its genome after irradiation which introduces numerous DNA double-strand-breaks that is fatal to most organisms [26]. *D. radiodurans* was first discovered at 1956 when researchers tried to sterilize canned foods with gamma radiation (destined to kill all known forms of life [27]). Since then, many studies have been done to decipher the outstanding radioresistance of *D. radiodurans*. Early studies showed that the genomic sequence of *D. radiodurans* is not unique, and that its DNA repairing proteins are similar to other radiation sensitive organisms [28]. In was not until recently that scientists found that the proteome of *D. radiodurans* was under well protection by a Mn^{2+}/orthophosphate complexes [29]. These complexes not only just actively remove superoxides but also prevent protein oxidation [29]. However, gene regulation under irradiation in *D. radiodurans* is still vague. For example, PprM is a modulator of the PprA protein, a protein that promotes DNA end-repairing in the non-homologous end joining pathway [30]. PprM is known as a cold shock protein homolog but previous studies showed it was unclear how PprM was induced [ref]. This is one of many evidences that suggest that

more studies are required to understand the regulatory mechanisms of radioresistance in *D. radiodurans*.

To contribute to this field, In chapter 3 I present how we identified novel regulatory transcripts in *D. radioduans* with bioinformatics pipeline and transcriptome analysis. In chapter 4 I show how we characterized sRNAs of interest and identified potential regulatory pathways that may under the control of sRNAs. The following sections introduce the three main parts of this study: (A) Regulatory mechanisms of sRNAs in bacteria (B) Current bioinformatics and experimental methods of sRNAs identification (C) *D. radiodurans* and current understanding of its radioresistance mechanisms.

1.2 REGULATORY ROLE OF SRNAS IN BACTERIAL STRESS RESPONSES AND ITS LACK OF DISCOVERY IN *D. RADIODURANS*

Small RNAs (sRNAs) represent an emerging class of regulatory transcripts in bacterial species. sRNAs can range from 40 to 400 nucleotides long and mostly do not encode functional proteins, although recent evidence has been found to contradict this claim [31]. sRNAs have been known as cellular regulators of transcription and translation with versatile functions [5]. Compared to other cellular regulators, sRNAs are *trans*-encoded and have imperfect base-pairing with their targets [32]. Many of them can bind the ribosomal binding sites (RBS) of their mRNA targets, thus blocking translation. Others exert more complicated mechanisms such as inhibiting mRNA secondary structures via base-pairing. They can also (up or down) regulate the stability of a target mRNA via base-pairing [3]. While most of the *trans*-encoded sRNAs down-regulate the activity of their targets, sRNAs can lead to activation [5]. In addition, many sRNAs are able to bind multiple targets due to their intrinsic nature of imperfect base-pairing [18].

The binding regions of sRNA-mRNA are typically short (~10-25 nt) and only a few nucleotides play critical roles in binding. For example, the SgrS sRNA has a 23 base pairs binding region with its target *ptsG*, but only four single mutations affect the interaction [5]. In Gram-negative bacteria, the RNA chaperone Hfq is required to facilitate RNA-RNA interactions [33]. Hfq increases the annealing rate of sRNA-mRNA by stabilizing the duplexes or by promoting structure remodeling [34]. It has been shown that sRNAs loose its binding ability in Hfq mutants of *E. coli* [33]. However, in some Gram-negative and many Gram-positive bacteria, sRNAs can function without the help of Hfq. For example, in *V. cholerae* the *VrrA* RNA can repress the expression of OmpA without Hfq existence [35]. In *D. radiodurans*, the homolog of Hfq has not been identified, but the Rsr protein in *D. radiodurans* shows a similar behavior by binding to a Y RNA, which is discussed in chapter 3.

In addition to interacting with mRNAs, sRNAs can also modulate protein activity. The CsrB and CsrC sRNA can modulate CsrA protein activity in *E. coli* via sequestration of CsrA from its mRNA targets [36]. CsrA is a Carbon storage regulator that can inhibit translation by binding to the Shine-Dalgarno sequence of mRNAs [36]. 6S RNA can mimic the structure of an open promoter and request sigma-70 factor, thus repress the sigma-70 promoted genes [37], [38]. It can also serve as a template to transcribe short product of RNAs, which is known as pRNA and can regulate 6S and affect cell viability [39]. The GlmY and GlmZ RNAs promote the synthesis of GlmS glucosamine-6-phosphate synthase, and they also regulate the activity of each other [40]. Importantly, GlmY does not direct activate GlmS translation but stabilizes GlmZ by binding to the YhbJ protein [40]. These examples demonstrate the versatile functions and great potential of sRNAs in bioprocesses.

Among all different sRNAs regulatory mechanisms, the stress response regulation is one of the most interesting. Many sRNAs are known to regulate certain pathways under specific environmental stresses and help the cell to survive. For example, RyhB RNA can down-regulate iron-storage and iron-using proteins under iron starvation [41], [42]. It is repressed by Fur protein and Fe^{2+} complex [43]. Studies showed that RyhB can regulate up to 18 operons encoding 56 proteins, demonstrating the potential of multi-pathway regulation of sRNAs [43]. As a second example, two sRNAs OxyR and OxyS are induced under oxidation stress and regulate gene expression of alkyl hydroperoxide reductases AhpF and superoxide dismutases Sod to protect cells from chemically induced mutagenesis [44], [45]. As a third example, GadY RNA has been reported in *E. coli* to bind to the operon of *gadX-gadW* in response to acid stress, stabilizing *gadX* and up-regulating acid resistance genes [46]. As a last third, FNR protein can sense oxygen deficiency, and its activation can repress aerobic function genes and active anaerobic pathways [47]. It can also activate a sRNA FnrS to repress enzymes that are dispensable under anaerobic stress [47]. Lastly, the CRISPR system is a bacterial defense mechanism for virus infection. It contains genetic records of potential threats such as phages, virus or plasmids and provides acquired immunity [48]. CRISPR has been modified to enable gene editions in many species as the CRISPR/Cas9 system. The Cas9 nuclease complex can cut designated loci of genome with the help of a synthetic guide RNA (gRNA) and remove existed genes or add new ones [13].

In summary, sRNAs have been discovered in many stress response mechanisms and are critical for cells to survive. However, while sRNAs regulatory mechanisms under stresses are common in many bacteria, very little about sRNAs in *D. radiodurans* is known. It is a missing piece in *D. radiodurans* and sRNAs regulatory study of radioresistance will impact the bioengineering community and drug or medical research.

In our study, we focus on the search and characterization of sRNAs that can facilitate ionizing radiation resistance in *D. radiodurans* and investigate radioresistance regulatory mechanisms in chapter 4.

1.3 CURRENT BIOINFORMATICS AND EXPERIMENTAL METHODS OF SRNAS IDENTIFICATION

Since the first sRNA was identified in *E. coli*, scientists have been developing novel methods to identify regulatory RNAs in a more effective and accurate manner. While the first few sRNAs were found serendipitously, scientists have been encouraged to develop systematic genomic-wide methods to high-throughput verify sRNAs after many sRNAs showed to be important in many bioprocess regulation. Many experimental techniques have been adopted to identify sRNAs, such as RNA labeling, functional genetic screen, co-purification and shotgun cloning [20], [49]. RNA labeling can easily detect abundant transcripts in cells, but the result could be interfered by other abundant RNAs such as tRNA or rRNA [20]. Functional screens track sRNAs based on its potential functionality (mostly via genetic deletions) and verifies its physiological function by expressing a multiple-copies of sRNAs libraries in cells [20]. However, it has many technical limitations and toxic or stress related sRNAs may not be able to be identified with this method [50]. RNA co-purification can pinpoint RNA-protein or RNA-RNA interactions readily and many sRNAs have been found in *E. coli* by Hfq co-purifications [50]. This method is limited to a certain type of sRNA that has a relatively strong association with other molecules, and specific antibodies are required for protein purifications. Shotgun cloning, also known as RNomics, is a technique that first isolates a specific size range of transcripts for making cDNAs libraries and then proceeds for sequencing [51]. This technique in theory can detect all RNAs within the size range that

are expressed under certain conditions. It does not require previous knowledge of sRNA coding regions and can identify non-conical sRNAs [51]. However, highly structured sRNAs may be biased in cDNA synthesis. For example, the level of 6S RNA has been shown to be underrepresented due to its structure [52].

To support and facilitate the experimental sRNA identification, many computational tools or bioinformatics algorithms have been developed. Most of them focus on sRNA predictions, sRNAs target predictions and on innovating new methods to analyze transcriptome data that can be used for these purposes. The sRNA prediction algorithms screen the genomic sequences of organisms of interest and generate a list of potential ncRNAs based on different features. Most of them analyze known sRNAs to find rules based on sequence or structure conservation and use them to screen the genome for novel sRNA predictions [53]. A very common algorithm is to predict sRNA loci by genome comparative analysis. The assumption is that sRNAs should display conservation of at least a certain sequence length or of secondary structure among a group of related species. For example, 24 putative sRNAs were found by applying BLAST program on *E. coli* against Salmonella bacteria [54]. In another study, 50 out of 55 known sRNAs were identified by using conservation analysis between *E. coli* and *Shigella flexneri* [55]. Programs such as QRNA, SIPHT and sRNAPredict have been developed with comparative analysis. In addition, SIPHT and sRNAPredict also use the feature of Rho-dependent terminators, while QRNA also consider secondary structure for predictions [56]. However, most sRNAs do not share a unified common feature. For example, unlike tRNAs (which have very distinctive structures), the secondary structure of sRNAs can vary. The size of sRNAs also varies from 40 to 400 base pairs and it is not clear that homology of sRNA sequences are conserved across bacteria. Thus, the accuracy of most computational prediction tools is not impressive and the results from different algorithms

do not overlap to a large extent [57]. Still, the use of computational algorithm can be a useful start to the identification of sRNAs given the technical challenges of conducting purely experimental searches.

Besides sRNA predictions, using transcriptome data from next generation sequencing (also known as deep-seq) is another emerging method to identify sRNAs in various organisms. In this case, cDNA libraries can be generated from total RNA or RNA that is prepared from cells under certain stresses, and process for high-throughput sequencing [58]. The sequenced reads generated are then mapped to designated genomes with computational alignment tools such as Bowtie or BWA [59], [60]. With this technique, massive amount of putative sRNAs have been identified in various organisms [61]–[63]. As described in Chapter 4, we have taken advantage of these methods to sequence the transcriptome of *D. radiodurans* under recovery from acute radiation and have identified 41 novel transcripts that have been experimentally confirmed using Northern blotting analysis. As documented in the literature, in our work we have also found that analysis of transcriptome data could be challenging, especially in species without well-annotated genomes. Interpreting real sRNA from processed mRNA fragment could also be challenging [64].

In addition to sRNA prediction tools, many bioinformatics tools have also also been developed to verify sRNA targets. Most of these tools incorporate common sRNA interaction features such as Hfq binding sites and structures, flanking regions around translation starting sites and conservation of sRNAs and their targets [57]. For example, TargetRNA uses algorithms to score hybridizations based on local sequence alignment and RNA folding models [65]. IntaRNA calculate minimum extended hybridization energy and search for optimal interactions between sRNAs and mRNAs [66]. RNApredator uses similar algorithms, but also take target accessibility into consideration

[67], where accessibility is defined as the energy required to make transcripts single-stranded. Nevertheless, more samples are required to sharpen these bioinformatics tools, and more RNA-protein interaction prediction program is also required. In this work, we have taken advantage of these tools (with limited success). We have particularly used IntaRNA and TargetRNA, to predict targets for one of our most interesting novel sRNAs Dsr2 (Table C.5).

1.4 *D. RADIODURANS* AND CURRENT UNDERSTANDINGS OF ITS RADIORESISTANCE MECHANISMS

Deinococcus radiodurans is one of the most radioresistant organisms. It is a Gram-positive, aerobic and nonpathogenic bacterium. The cells have pink color and can form dyads or tetrads [68]. It can survive up to 15-20 kGy of ionizing radiation and not loose viability up to 10 kGy. It can also survive high dosage of oxidative stress such as hydrogen peroxide, desiccation and heat shock [ref]. The ionizing radiation can introduce fatal DNA double-strand-breaks (DSB) (~ 200 DSB at D_{37}) and reactive oxygen species (ROS) that cause oxidative damage to the DNA [68]. The extraordinary radioresistance of *D. radiodurans* has been scrutinized by many researchers but not until very recently we have started to understand its unique mechanisms of protection. The early studies assumed *D. radiodurans* exhibited special tools or mechanisms to protect and repair its genome from radiation damage. However, genomic studies showed that its DNA repairing systems is not remarkably different than radiation sensitive species [28]. Many DNA repairing components are actually missing in *D. radiodurans*. For example, DNA dioxygenase and photolyases, that typically repairs DNA Alkylation damage and thymine leisons, have not been found in the genome of *D. radiodurans* [69]. Still, many genes showed to be critical for the radioresistance mechanisms and distinct features have also

been identified in *D. radiodurans*. For example, *ddrA*, *ddrB* and *pprI* are unique genes found in *D. radiodurans* that exhibit differential expression under ionizing radiations, and the deletion of *pprI* decrease the survival rate of the cell [69]. Irradiated cells were enriched with RecA, DdrA, DdrB as well as PprA [69], [70]. PprA protein binds to the end of double–strand DNA and is essential for irradiation recovery [70]. It was reported to be regulated by PprI protein, a global regulator which activates at least 31 proteins under irradiation [71]. Expression of PprI in *E. coli* also increases its resistance of osmotic stress [72]. PprA is also regulated by PprM protein, which is a cold shock protein homolog [73]. However, the detail of how PprM is regulated is unclear. The other conserved genes related with DNA repairing such as RecA, RecF and RecR still play critical roles, since the deletion strains of these genes are all more susceptible to irradiation [74]. On the other hand, the antioxidative enzymes such as SodA and SodC in *D. radiodurans* do not have more copy numbers than radiation sensitive strains but show a higher activity compared to the counterpart in *E. coli*, and the deletion of these genes leads to lower survival of cells under irradiation [69].

Although the number of DNA double stranded breaks that accumulate in *D. radiodurans* (~200 DSB at D_{37}) is even higher then radiation sensitive organisms such as *E. coli* (~10 DSB at D_{37}) [69], the induced protein damages under these radiation conditions are lower in *D. radiodurans* [75]. In addition, protein-free extracts from *D. radiodurans* increase the survival rate of *E. coli* under X-ray, suggesting small molecules in *D. radiodurans* could be critical to radioresistance [69]. Studies have also shown that *D. radiodurans* has a higher ratio of Mn/Fe compared to radiation sensitive species, and this trend was also observed in other radioresistant organisms [76]. It was later found that Mn^{2+} forms an organic complex that protects the proteome by removing superoxide and prevents protein oxidation [29]. It was also demonstrated that the radioresistance of

bacteria depends on how well the proteome is protected, especially how susceptible they are to oxidation [75]. Despite progress in understanding role f protective species in radioresistance, what is not clear is how these mechanisms are regulated within the larger context of cellular activity (i.e. translation, DNA repair etc). Therefore, understanding how protein damage and proteome protection are regulated in transcription and translation level is very important to help us to understand the regulatory mechanisms of *D. radiodurans*. So far it has been proposed in the literature that proteome damage might lead to the regulation of cellular bioprocess for DNA repair [75]. Early studies of the chronicle of DNA repair physiology also suggest ionizing radiation induced damage is highly regulated. For instance, studies showed the metabolic activities after irradiation in *D. radiodurans* are highly ordered. DNA replication and protein synthesis are halted right after irradiation, and chromosomal digestion is activated then damaged nucleotides are released from the damaged cells before DNA replication resumes [77], [78].

In this study, we use bioinformatics pipelines and transcriptome data combined experimental techniques (e.g. Northern blotting analysis) to identify many novel transcripts in *D. radiodurans*. Many of these ncRNAs shows distinct expression patterns under different stresses. Our study also identifies potential regulatory role of sRNAs during recovery from IR, and provides evidence of the regulatory networks of sRNAs in *D. radiodurans*. This in the first study that shows sRNA regulation can contribute to the radioresistance in *D. radiodurans.*

1.5 SUMMARY OF RESEARCH OBJECTIVES AND ACCOMPLISHMENTS

The following chapters discuss the research that I performed at the University of Texas at Austin, collected into three main works that have been already published or are under preparation.

Chapter 2 is a description of the development of our bioinformatics analysis to accelerate experimental methods of sRNAs discovery. In this study, we analyzed the enrichment patterns of sRNAs in 13 well-annotated bacterial species using existing transcriptome and experimental data. All intergenic regions were analyzed by WU-BLAST to examine conservation levels relative to species within or outside their genus. In total, more than 900 validated bacterial sRNAs and 23,000 intergenic regions were analyzed. The results indicate that sRNAs are enriched in intergenic regions, which are longer and more conserved than the average intergenic regions in the corresponding bacterial genome. We also found that sRNA-coding regions have different conservation levels relative to their flanking regions. This work provides a way to analyze how noncoding RNAs are distributed in bacterial genomes and also shows conserved features of intergenic regions that encode sRNAs. These results also provide insight into the functions of regions surrounding sRNAs and into optimization of RNA search algorithms. We also applied this approach to predict sRNA candidates in the mycobacterial species *M. smegmatis* and *M. bovis* BCG and confirmed the expression of many sRNAs using Northern blotting. We have also applied a high-throughput technique (Deep-RACE) to map the 5' and 3' ends of many of these sRNAs and identified potential regulators of sRNAs by analysis of existing ChIP-seq datasets.

Chapter 3 presents the study of how we used similar approach and identify novel transcripts in *D. radiodurans*. In this study, we merged computational and experimental techniques to identify novel potential sRNAs in *D. radiodurans*. We used computational

tools to find hundreds of loci that were predicted to be sRNA candidates and applied a previously developed criterion to further filter most plausible sRNA candidates [ref]. We also used deep sequencing techniques with total RNA from *D. radiodurans* and discovered 199 sRNA candidates in intergenic regions (IGRs). Upon confirmation by Northern blotting analysis, we uncovered the expression of 41 novel sRNAs in *D. radiodurans*, 8 of which showed differential expression following recovery from ionizing radiation. We also found and experimentally validated the presence of homologous sRNA candidates in a closely related radioresistant species, *Deinococcus geothermalis*. Our studies also identify other well-characterized noncoding RNAs that have not been previously annotated in current versions of the *D. radiodurans* genome (NCBI GenBank accession numbers NC_001263.1 and NC_001264.1).

Chapter 4 presents the regulatory sRNAs and its proposed targets we found in *D. radiodurans.* We investigated and characterized our previously found novel ncRNAs in *D. radiodurans* and some candidates appears to show distinctive expression patterns under ionizing irradiation. We created overexpression and deletion mutants of sRNAs of interest and found the one sRNA deletion strain Dsr2KD showed a reduced survival rate under ionizing irradiation compared to wild type. We then designed sRNA target identification experiments such as HITS-CLIP, transcriptome analysis and EMSA analysis to verify potential binding target and regulatory pathways of Dsr2 sRNA. We created mutants that express Dsr2 sRNA conjoined with MS2 protein binding sites for pull-down assay and generate potential binding targets with deep-seq, and used EMSA to verify direct bindings of sRNA-mRNA. We also use transcriptome data and qPCR to investigate gene differential expressions in sRNA mutants. Our results indicate that Dsr2 could play a role in PprA regulation by modulating the expression of PprM protein alone

with other genes that in the PprI pathway, and have a potential impact on translation regulation under ionizing irradiation.

Chapter Two

*Develop bioinformatics pipeline with genomic comparative analysis to refine experimental identification of novel ncRNAs in bacteria

2. 1 INTRODUCTION

Recently, small non-coding RNAs (sRNAs) have been under higher scrutiny as mediators and regulators of gene expression [18], [81]–[84]. This class of RNAs has been found to play a variety of roles in important cell functions [85] [5]. Typically composed of 50 to 500 nucleotides in length, sRNAs are known to control plasmid replication, bacterial virulence, and various stress responses [86]–[89].

An interesting aspect of sRNAs is the wide diversity of their functional mechanisms. sRNAs can repress or stimulate gene-expression post-transcriptionally by pairing their targets through base complementarity; a target can be, but is not limited to, an mRNA or a protein.. sRNAs that regulate other RNAs can be *cis*-encoded or *trans*-encoded. A *cis*-encoded sRNA is typically encoded adjacent to its regulatory target on the same strand as a riboswitch, or on the opposite strand as an antisense sRNA. In most cases, they will base pair to their targets or change the secondary structure to inhibit ribosome binding [90]–[92]. On the contrary, a *trans*-encoded sRNA is encoded away from its target, has a lower base complementarity with its target, and can potentially bind multiple targets [93].

With advances in high-throughput sequencing technologies [94], it is now possible to sequence gigabases of nucleotides in a matter of hours [95]. Aided by sRNA

*This work was published in:

[79] C. H.Tsai, R.Liao, B.Chou, M.Palumbo, andL. M.Contreras, “Genome-wide analyses in bacteria show small-RNA enrichment for long and conserved intergenic regions,” *J. Bacteriol.*, vol. 197, no. 1, pp. 40–50, 2015. (Author contribution to the work: data analysis, research concepts discussion and manuscript writing)

[80] C. H.Tsai, C.Baranowski, J.Livny, K. aMcDonough, J. T.Wade, andL. M.Contreras, “Identification of Novel sRNAs in Mycobacterial Species.,” *PLoS One*, vol. 8, no. 11, p. e79411, Jan.2013. (Author contribution to the work: data analysis, experiment performing and manuscript writing)

prediction algorithms, these large data sets are paving the way for continual sRNA discovery [90], [96], [97]. However, sRNA validation as well as determination of mechanistic function remains elusive. This is mainly due to the complexity of sRNA regulatory mechanisms. As a result, a plethora of computational approaches for sRNA prediction have gained popularity [98], [99]. Some of the most widely-used methods include eQRNA [24], RNAz [100], sRNAPredict3/SIPHT [12], and Nucleic Acid Phylogenetic Profiling (NAPP) [101]. These methods rely on the search of a variety of patterns: compensatory mutations consistent with base-paired secondary structure, thermodynamic stability and structural conservation, regions of primary sequence conservation followed by transcriptional termination signals, and noncoding sequence clusters based on cross-genome conservation profiles. While different computational methods of sRNA identification include a multitude of criteria, even the most popularly applied methods tend to have low precision and sensitivity. Indeed, a previous study has reported a mean precision between 4% and 12% for eQRNA, RNAz, sRNAPredict3, and NAPP across 10 data sets [98]. Thus, a significant challenge stems from the fact that computational approaches tend to generate a large bank of potential sRNA sequences that result in only a handful of accurate hits.

Various approaches are routinely used to complement computational sRNA identification; these include cloning, high-throughput sequencing, Northern Blotting, and micro-arraying. While microarrays have been the most common method for transcriptome analysis [49], [102], [103], this method is limited by indirect recording of expression levels and by typically not encompassing the entire transcriptome. Most recently, RNA sequencing (RNA-seq) has become a powerful technique [11], [104], [105]. However, RNA-seq also has drawbacks; one of the major limitations being that certain sRNAs expressed during a particular cellular condition may not be present during

cellular harvesting for RNA preparation. For the most part, Northern Blotting analysis has become an accepted method for verification of potential sRNA candidates that stem from prediction techniques and RNA-seq data. Even so, a significant amount of RNA is required for detection by Northern Blotting analysis, and sRNAs with fewer copy numbers can be difficult to detect.

In this study, we performed a genome-wide analysis of conservation and length distribution patterns for all the intergenic regions in thirteen selected species that have well-annotated genomes and experimental RNA-seq analysis with significant genome coverage (all greater than 50%). Using highly stringent criteria, we compared the query genomes to species both inside and outside their genera and determined the conservation level of all intergenic regions. Previous studies that have focused on the analysis of only the sRNA-coding regions do not indicate a consistent trend in sRNA conservation levels [12], [106]–[109]. In this study, we take a different approach by considering the entire intergenic region where an sRNA is housed. We also analyzed the lengths of the intergenic regions where experimentally observed sRNAs were found in their native genomes. This large-scale study encompasses thirteen different species for analysis of a total of more than 900 validated bacterial sRNAs and of more than 23000 total intergenic regions. Our genome-wide analysis has yielded trends that provide clues to various questions regarding (1) how distant and/or independently trans-acting sRNAs have evolved from coding regions, (2) how large intergenic regions that encode sRNAs are relative to the average size of intergenic regions in their native genomes, and (3) how conserved sRNAs are relative to the intergenic regions where they are found.

This study takes advantage of detailed transcriptomic work that has now been completed in a diverse set of bacterial species with sequenced genomes. As such, this analysis contributes to our understanding of conservation patterns in sRNA-encoding

intergenic regions and of sRNA evolution among bacterial species of varying phylogenetic distances. This contributes new insights to possible refinement strategies that can improve current identification of transcribed intergenic sRNA sequences.

2.2 RESULTS

2.2.1 Analysis of intergenic regions in different species shows conservation of a large number of intergenic regions

To confirm the orthology of analyzed intergenic regions, we compared the conservation level of intergenic regions and extended intergenic regions. This eliminates the possibility that the intergenic region is co-conserved with the adjacent protein-coding region, or potential untranslated regions (UTR) that are not annotated.

WU-BLAST was used to analyze the conservation level of all intergenic regions in the 13 selected bacterial species in this study (see Table A.1 in the supplemental material). Our selection of the 13 species used for this study ensures that identification of sRNAs has been exhaustive since these species have all been well characterized at the transcriptome level. A list of "extended intergenic regions" for each species was created as a control for conservation analysis; Figure 2.1 illustrates how extended intergenic regions were determined. The extended region includes parts of the upstream and downstream coding regions that are the same size as the original intergenic region along with the intergenic sequence, so that the combined region is three times the size of the original intergenic region. The parameters and criteria for conservation analysis are described in Materials and Methods. In brief, an intergenic or extended intergenic region is considered conserved in a species if a hit is returned by WU-BLAST and satisfies the filter criteria. Conservation levels of within-genus and outside-genus regions were calculated separately. The within-genus criterion refers to all species that are in the same

genus as the analyzed species, and the outside-genus criterion refers to a group that includes 38 species from different genera (see Table A.2 in the supplemental material for the full list). Since some genera have a limited number of species, such as Escherichia, this analysis can broaden the diversity of species for WU-BLAST and provide insight into how phylogenetic distance can affect the results of conservation analysis. Only intergenic regions with a conservation level equal to or higher than that of the extended intergenic regions are defined as "conserved intergenic regions."

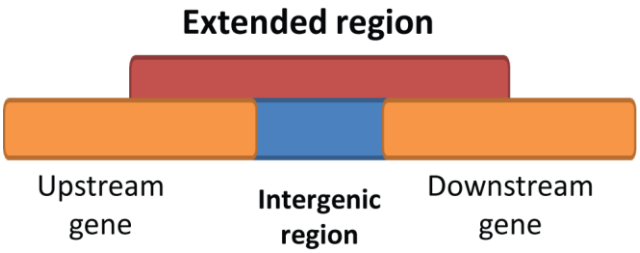


Figure 2.1. Extended intergenic region

The orange bars indicate the upstream and downstream protein-coding regions, the blue bar indicates the intergenic region, and the red bar indicates the extended intergenic region, which includes the intergenic region along with a part of the upstream and downstream region that equals the length of the intergenic region. The intergenic regions and extended regions were analyzed with WU-BLAST and compared with each other for conservation-level analysis.

Figure 2.2 shows representative within-genus and outside-genus conservation patterns of three selected species (others are shown in Figure A.1 in the supplemental material). All intergenic regions were grouped into non-conserved, equally conserved, and more highly conserved (where conservation levels are lower than, equal to, or higher than those of extended intergenic regions, respectively). Results show that while most

intergenic regions (68 to 82%) are not conserved within-genus or outside-genus among the species studied, a large enough fraction of all intergenic regions are either equally conserved or highly conserved relative to the surrounding (gene-carrying) regions; the latter cases were of the most interest, as we aimed to analyzed enrichment patterns of sRNAs in exceptionally conserved intergenic regions. Raw data from WU-BLAST can be found in Table A.3 in the supplemental material. To fully understand how heterogeneously distributed conserved intergenic regions were among species surveyed, we tabulated the distribution of all conserved intergenic regions among all with within-genus species. According to unique conservation patterns that were observed, we classified the target species into two categories: group 1 includes the majority of all species analyzed, where the main characteristic is that most intergenic regions are conserved in a way that is not specific to a single species. In contrast, group 2 is characterized by having most of its intergenic regions (>50%) conserved in a limited set of species. The two species that fall into this category are *M. bovis*, which has a 63.8% of its conserved intergenic regions conserved only in *Mycobacterium tuberculosis*, and *C. trachomatis*, which has 83.4% of its conserved intergenic regions conserved only in *C. muridarum*.

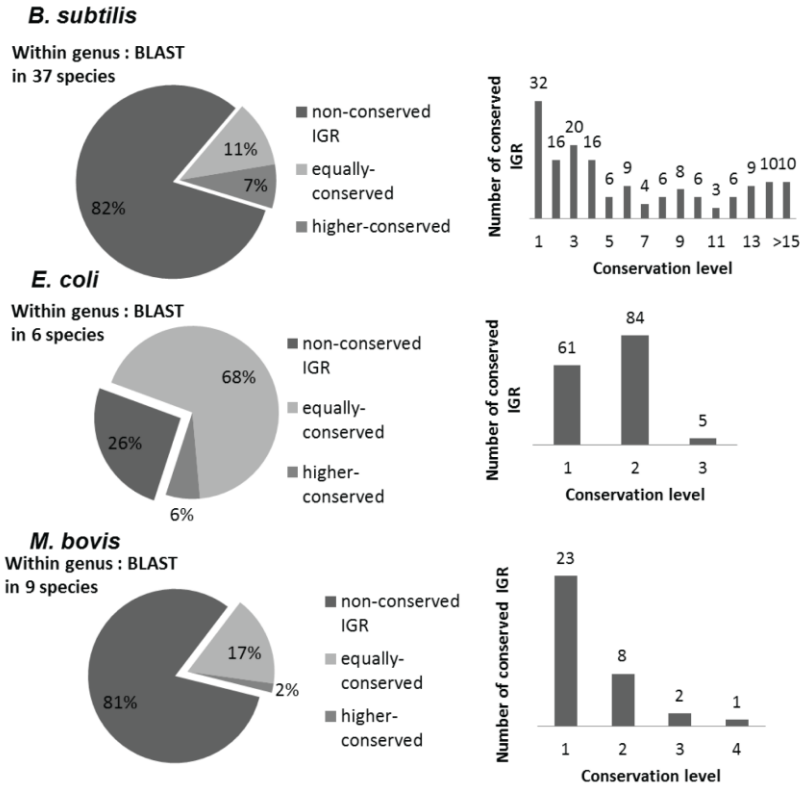


Figure 2.2. Conservation patterns of intergenic regions in selected species

The figure shows the conservation level distribution of intergenic regions in three selected. The conservation level is the number of within-genus or outside-genus organisms found to have homology of the intergenic region. The intergenic region would be marked as “non-conserved” if its conservation level is less than that of the extended intergenic region or as “equally conserved” or “more highly conserved” if the conservation level is equal to or higher than that of the extended intergenic region. The pie charts show how conservation levels are distributed in the more highly conserved intergenic regions, and the total numbers of within-genus organisms are shown above them.

2.2.2 Conserved intergenic regions are enriched for small RNAs

To understand if sRNAs were more likely encoded by conserved intergenic regions, we first cross-referenced the reported sRNA coordinates to all the intergenic regions for each species. The pools of experimentally observed sRNAs that were identified via transcriptome or Northern blotting for each species were collected from published works and online databases. References for all sRNAs collected are listed in Table A.1 [10], [21], [23], [33], [34], [51], [61], [62], [80], [110]–[143]. Experimentally observed sRNAs were mapped to their corresponding genomes to identify sRNA encoding regions. The antisense sRNAs are beyond the scope of this study and are excluded from our analysis. An important general observation that stems from this analysis (see Table A.1 in the supplemental material) is that a range of ∼2% to 12% of all intergenic regions encodes sRNAs. This is close to computational and experimental estimations of sRNAs in bacteria [144].

After mapping all experimentally observed sRNAs to their corresponding intergenic regions, we determined the percentage of within-genus or outside-genus non-conserved, conserved, and higher-conserved intergenic regions that encoded sRNAs. The percentages of sRNA-coding intergenic regions for all 13 species are showed in Figure 2.3. For the within-genus analysis, most of the species have a greater percentage of sRNA-coding regions in conserved intergenic regions than in non-conserved intergenic regions. We used Fisher's exact test to test the statistical significance ($P < 0.05$), and all but one species (*M. bovis*) show significant sRNA enrichment. Importantly, the more highly conserved intergenic regions in most species are even more enriched for sRNAs than the non-conserved intergenic regions, indicating that sRNAs are more likely to be encoded within highly conserved intergenic regions. Our general findings of sRNA enrichment in conserved intergenic regions in outside-genus species compared to non-

conserved intergenic regions (see Figure A.2 in the supplemental material) further support these results

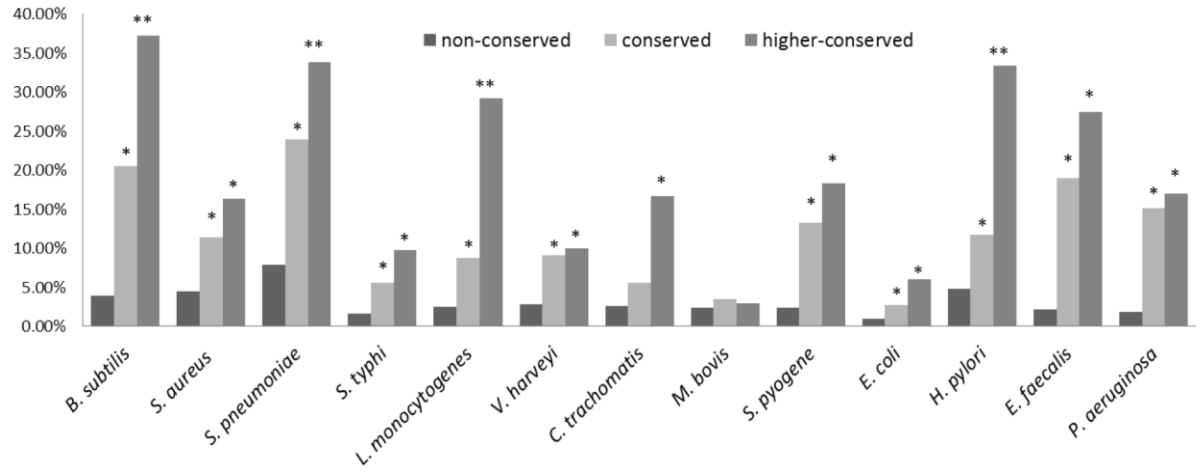


Figure 2.3. Enrichment of sRNAs in outside-genus conserved intergenic regions

The percentage is defined as the number of sRNA-coding intergenic regions relative to non-conserved, equally conserved, or more highly conserved (outside-genus) intergenic regions. A conserved intergenic region refers to any intergenic region that has a conservation level equal to or higher than that of the extended intergenic region. A single asterisk denotes statistically significant enrichment of sRNA compared to non-conserved regions by Fisher's exact test ($P < 0.05$), and double asterisks denote statistically significant enrichment of sRNA compared to the conserved intergenic region.

It is interesting that the enrichment of sRNA-coding regions is not as significant as the within-genus analysis across all species (as determined by Fisher's exact test). This is particularly the case with species that exhibit that fall into the second conservation pattern group, where conservation is observed among only a very limited set of species (i.e., *C. trachomatis* and *M. bovis*). It is also possible that many sRNAs remain unidentified in these species. We suspect that this might be the case in species such as *C.*

trachomatis and *E. faecalis,* where the percentage of all intergenic regions that has been identified as encoding sRNAs remains lower than 3% and the number of reported sRNAs remains low.

2.2.3 Refined conservation analysis based on phylogenetic distance strengthens observations of sRNA enrichment in conserved intergenic

To investigate how phylogenetic distance affects the enrichment of sRNAs in conserved intergenic regions, we selected two species, *B. subtilis* and *S. pneumoniae*, and analyzed how intergenic regions were conserved across differently phylogenetically distant sets of species. These species were selected due to the larger number of identified sRNAs and the larger set of within-genus species that has been sequenced and can be used as a basis for conservation analysis. For this analysis, we used MEGA5 to compute the phylogenetic distances between the specific species of interest (e.g., *B. subtilis*) and the respective within-genus species (e.g., other Bacillus species, listed in Table A.2 in the supplemental material). As shown in Figure 2.4, a wide variation in evolutionary spread was observed among the species we tested. For instance, the distances between all the within-genus Bacillus species and *B. subtilis* range from 0.019 (*Bacillus amyloliquefaciens*) to 0.123 (*Bacillus pseudofirmus*). In contrast, the phylogenetic spread is lower in *S. pneumoniae* (0.004 to 0.077) than in other Streptococcus species. As such, the latter genus clusters more closely in terms of phylogenetic distance than the Bacillus species. As a reference, the phylogenetic distance from *E. coli* (a bacterium from a different genus) was evaluated for the three selected organisms to gain a sense of how these evolutionary measurements could be interpreted. The distance for *B. subtilis* is 0.231, and that for *S. pneumoniae* is 0.229.

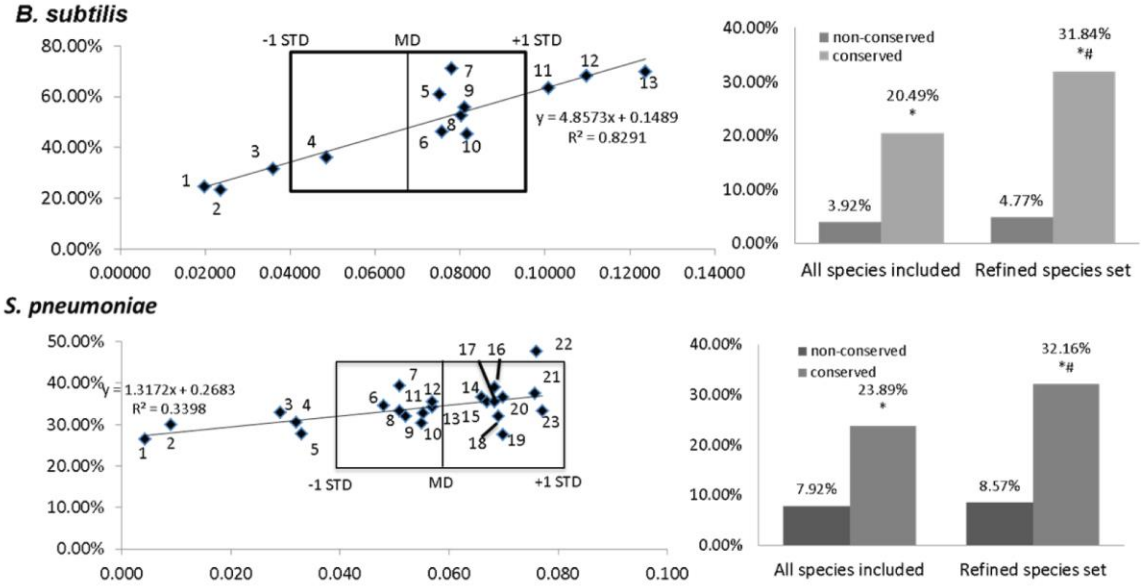


Figure 2.4. sRNA-coding-region enrichments in intergenic regions conserved in species with different phylogenetic distances and sRNAs enrichments with a refined species set

The phylogenetic distances between the within-genus species were calculated for *B. subtilis* and *S. pneumoniae* with MEGA5. The percentages of all sRNA-encoding intergenic regions that are conserved in a certain species were also calculated. For instance, in B. subtilis, 9.3% of the intergenic regions that are conserved in *B. amyloliquefaciens* (dot 1) were found to encode sRNAs. The mean distance (MD) and standard deviation (STD) to the within-genus species were calculated for *B. subtilis* and *S. pneumoniae* and marked in each graph. The following species were included in the plot. (Top) *1, B. amyloliquefaciens; 2, B. atrophaeus; 3, B. licheniformis; 4, B. pumilus; 5, B. anthracis; 6, B. cereus; 7, B. halodurans; 8, B. megaterium; 9, B. weihenstephanensis; 10, B. thuringiensis; 11, B. clausii; 12, B. selenitireducens; 13, B. pseudofirmus. (Bottom) 1, S. mitis; 2, S. oralis; 3, S. sanguinis; 4, S. gordonii; 5, S. parasanguinis; 6, S. salivarius; 7, S. constellatus; 8, S. pasteurianus; 9, S. intermedius; 10,S. lutetiensis; 11, S. gallolyticus; 12, S. macedonicus; 13, S. infantarius; 14, S. iniae; 15, S. dysgalactiae; 16, S. agalactiae; 17, S. mutans; 18, S anginosus; 19, S. suis; 20, S. equi; 21, S. uberis; 22, S. parauberis; 23, S. pyogenes*. A refined set of

organisms was selected by phylogenetic distance. Any species within one standard deviation (within the boxed area) of the mean distance was included in the analysis. The percentages of sRNA-coding intergenic regions in conserved and non-conserved intergenic regions were calculated and compared. The asterisk denotes enrichment of sRNA compared to non-conserved regions as determined by Fisher's exact test ($P < 0.05$), and the pound sign denotes a statistically significant difference compared to conserved intergenic regions of all included species.

The level of enrichment for intergenic regions that encode sRNAs was analyzed relative to the phylogenetic distance of within-genus species for Bacillus, and Streptococcus (Figure 2.4). The sRNA-coding-region percentages were calculated for intergenic regions that were conserved in different species and plotted against the phylogenetic distance for each within-genus species. A positive correlation of sRNA enrichment and phylogenetic distance was observed, indicating that intergenic regions, which are conserved in more distant species, are more likely to encode sRNA. This trend was consistent up to a certain threshold distance, where the species were too distant to have significant homology. Representative data are shown for *B. subtilis* and *S. pneumoniae* in Figure 2.4.

A pattern observed in some species from our phylogenetic-distance analysis was that certain within-genus species that were analyzed appeared significantly closer (in evolutionary distance) to the species under analysis than the rest of the within-genus species. One example was S. pneumoniae, for which two outlier species (*Streptococcus mitis* and *Streptococcus oralis*) were more than two standard deviations farther than the mean distance of all other within-genus species (Figure 2.4). The conservation analysis also showed that a large number (71.4%) of all the intergenic regions in *S. pneumoniae* are also conserved in *S. mitis*. In this case, we rationalized that intergenic region

conservation likely stems from general genome-wide conservation and not from any functional sRNA feature that poses an evolutionary advantage to these species. We reasoned that this could weaken our ability to observe true sRNA enrichment in conserved intergenic regions, since conservation of intergenic regions among organisms that are evolutionarily very close may not possess biological importance.

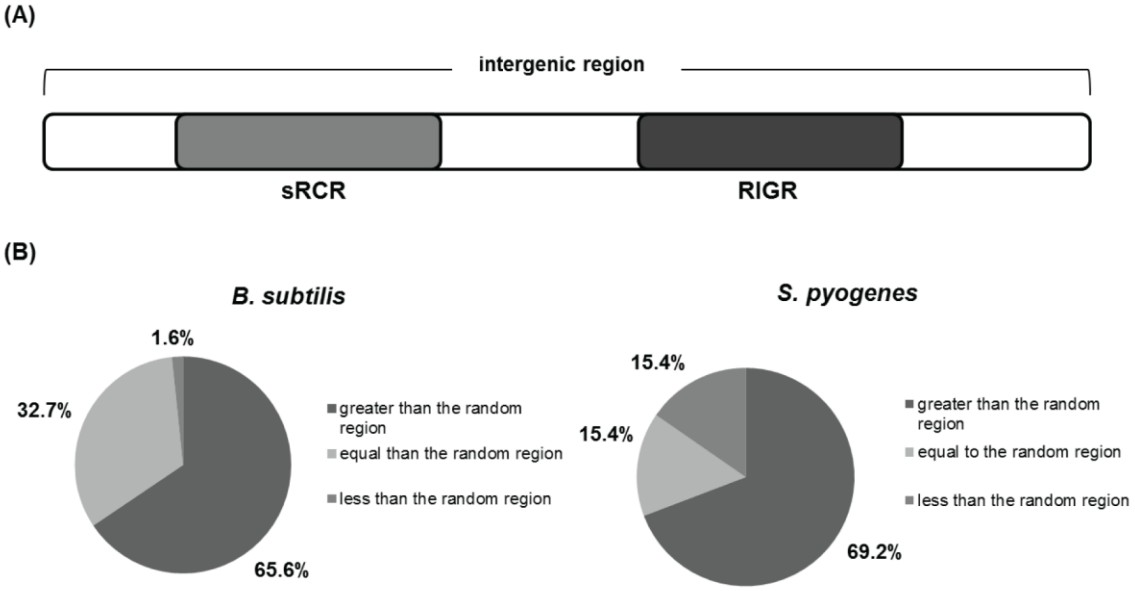


Figure 2.5. Illustration of the sRNA-coding region in an intergenic region and comparison of conservation levels

(A) Sketch of an sRNA-coding intergenic region (sRCR) and a randomly selected, non-overlapping region of the same length as the sRCR in the same intergenic region (RIGR) used for comparison of conservation levels. (B) Comparisons of the conservation levels of sRCRs and corresponding RIGRs. The percentages show how many sRCRs have a conservation level that is greater than, equal to, or lower than that of the respective RIGRs. For instance, 69.2% of the sRCRs have a greater conservation level than the RIGRs, while 15.4% of the sRCRs have a lower conservation level than the respective RIGRs.

To evaluate the effect of phylogenetic distance on the sRNA enrichment observed in intergenic regions, we repeated the conservation analysis with a refined set of target organisms. The refined set includes only species with a phylogenetic distance within one standard deviation from the mean phylogenetic distance. As shown in Figure 2.5, in the context of *S. pneumoniae* and *B. subtilis*, a statistically significant enrichment of sRNAs in conserved intergenic regions was observed after the outliers were removed. These results were consistent with those for other species analyzed, where species that were originally too close or too far were used to determine conservation. These results demonstrated that the refinement of the set of organisms used in the conservation studies to a more appropriate phylogenetic distance results in even higher sRNA enrichment in conserved intergenic regions. Importantly, these data also highlight the importance of selecting an appropriate set of species to make valid conclusions regarding sRNA conservation.

2.2.4 Conservation of sRNAs relative to conservation of flanking coding regions

Since sRNA-encoding intergenic regions were observed to be more conserved than all other intergenic regions, we wanted to test whether conservation was specific to regions encoding sRNAs. This initially caught our attention due to the high number of sRNAs that we observed to be encoded from a small fraction of all intergenic regions, leaving a large fraction of all intergenic regions seemingly idle. For this analysis, we determined the relationship between the conservation levels of sRNA coding regions (sRCRs) and random intergenic regions (RIGRs) (Figure 2.5A). A random intergenic region was defined as a segment of the intergenic region of the same length as the sRNA-coding region with no overlapping sequences.

In order to collect statistically meaningful data, we first generated a list of suitable target genomes in which a high number of conserved intergenic regions were more than double the length of the encoding sRNA in that same region. We selected Bacillus subtilis 168 and *Streptococcus pyogenes* MGAS5005 for this analysis because each species included over 25 conserved intergenic regions that met the set criteria. These species were analyzed for conservation against two groups, the within-genus and outside-genus groups, as described above. Intergenic regions that contained sRNAs with lengths that were <40% of the entire intergenic region were considered for analysis. Figure 2.5B shows analysis of 61 conserved intergenic regions in B. subtilis and 28 in *S. pyogenes*, where 65% and 69% of sRNA-encoding regions, respectively, were more conserved than the respective RIGR control (see Table A.4 in the supplemental material). Importantly, this result indicates that fragments that encode sRNAs are significantly more conserved than a random region of the same size within the same conserved intergenic region. This interesting result supports our underlying hypothesis that conserved intergenic regions are enriched in sRNAs, as these represent biologically important regions that are beneficial to bacteria.

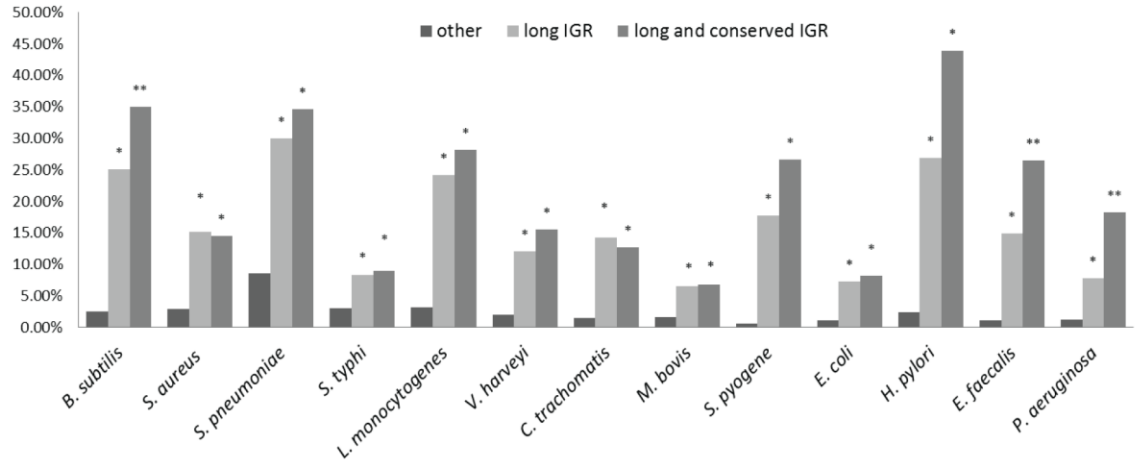


Figure 2.6. Enrichment of sRNAs in long and conserved intergenic regions (IGR)

The percentage is defined as the number of sRNA-coding intergenic regions relative to long intergenic regions (top 20% long), long and conserved (within-genus) intergenic regions, and other intergenic regions. A conserved intergenic region refers to any intergenic region that has a conservation level equal to or higher than that of the extended intergenic region. The asterisk denotes statistically significant enrichment of sRNA compared to other regions, as determined by Fisher's exact test ($P < 0.05$), and double asterisks denote values that are statistically significant relative to those for the long intergenic region.

2.2.5 Isolated genomic regions are enriched in sRNAs

One last interesting question that we explored concerned the presence of sRNAs in large intergenic regions that were previously thought to be noncoding. We suspect that these large intergenic regions isolated from protein coding regions potentially serve some purpose. After examination of all 13 genomes in this study, we found that the size distributions of all their intergenic regions are highly similar (see Figure A.3 in the supplemental material) despite pronounced differences in their genome sizes (ranging from 1 to 6.8 million nucleotides). We therefore speculated that in addition to conservation, the presence of isolated (long intergenic) regions in the genome could be another signature of the presence of sRNAs. Given the recent findings of a large number of noncoding RNAs in bacterial genomes, it is also informative to determine what percentage of the genome is indeed noncoding. Our analysis of long intergenic regions (as defined by the top 20% longest intergenic regions), showed significant sRNA enrichment for all species analyzed compared to that of all the intergenic regions (Figure 2.6).

Figure 2.6 also shows the combined enrichment effect we observed for long and conserved intergenic regions. For this analysis, we calculated the percentage of sRNAs

found in intergenic regions of the longest 20% of regions that are also conserved or highly conserved (within-genus and outside-genus). We consistently observed a significant level of sRNA enrichment in intergenic regions that were both conserved and long.

2.2.6 Prediction of sRNAs in silico using SIPHT

Using SIPHT, we identified 93 candidate sRNAs in *M. smegmatis* (refseq: NC_008596) (Table A.5) and 144 candidate sRNAs in *M. bovis* BCG (refseq: NC_008769) (Table A.6). Tables A.5 and A.6 include a detailed description of the predicted coordinates, orientations, sizes and neighboring upstream and downstream genes. Northern probes were designed according to SIPHT prediction. Figure 2.7 summarizes the overall approach that was employed in this work for sRNA identification and confirmation.

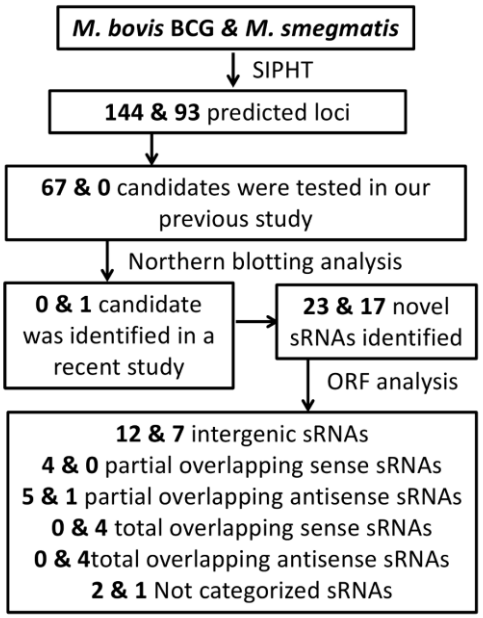


Figure 2.7. Schematic for sRNA identification

This schematic shows the combination of computational approaches and Northern blotting analysis used to identify the reported novel sRNAs in *M. bovis* BCG and *M. smegmatis.*

2.2.7 17 Novel sRNAs Identified in Mycobacterium Smegmatis

All 93 *M. smegmatis* sRNA candidates were tested by Northern blot using oligonucleotides in both orientations; expression was confirmed for 18 sRNA (listed in Table 2.1; see blot pictures in Figure 2.8 and Figure A.4). One of them (Sm32/33) was identified in recent work as IGR-1 with similar size, coordinates and same orientation [145]. Thus, 17 *M. smegmatis* sRNAs identified here have not been experimentally demonstrated in any previous studies. In our previous study [128], we reported homologs of 6 *M. smegmatis* sRNA candidates (Sm32/Sm33, Sm35, Sm46, Sm47, and Sm74) in *M. bovis* BCG (Mpr13/Mcr14, Mpr20, Mpr3, Mpr4, and Mpr5, respectively). These were confirmed directly in *M. smegmatis* by Northern blotting in current study and listed in the 17 novel confirmed sRNAs. A homologue of Sm76 was previously identified in *M. tuberculosis* by RNA-seq [82] and microarray analysis [146] but not otherwise experimentally confirmed. All of the validated sRNAs were in the same orientation to that predicted by SIPHT. This suggests that the sequence specificity of SIPHT for this prediction is higher than in our previous work, in which 9 out of 37 of the validated sRNAs were in the opposite orientation to the prediction [128]. All confirmed sRNAs were assigned gene names according to a recently-proposed nomenclature [147].

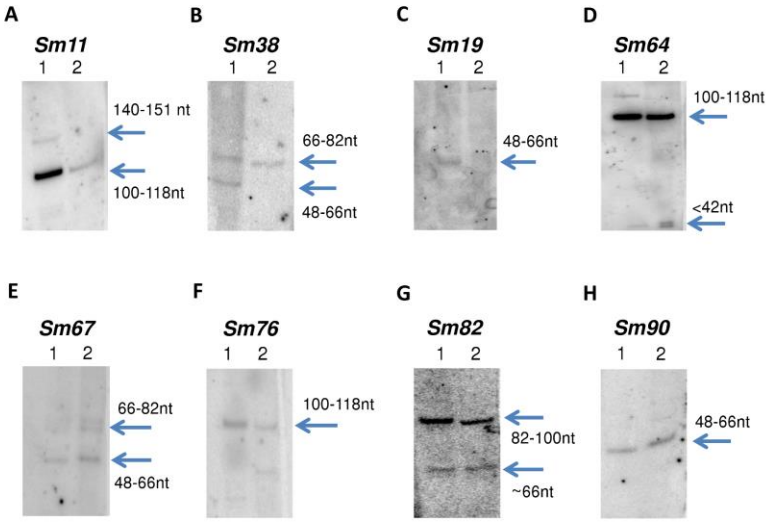


Figure 2.8. Northern blotting confirmation of sRNA candidates in *M. smegmatis*

Selected images of Northern blotting analysis for validated *M. smegmatis* sRNAs; the remaining images are included in Figure A.4. Lane 1 and 2 indicate total RNA samples extracted from *M. smegmatis* and *M. bovis* BCG, respectively. We used Phi-X174/Hae III Marker for the size prediction. The probes we used for this analysis are listed in Table A.7.

Given the practical convenience of testing RNA from both species simultaneously to search for novel sRNA candidates, we used the designed probes for sRNA detection in *M. smegmatis* to also probe expression of these candidates in *M. bovis* BCG and *M. tuberculosis*. Although our focus was to validate *M. smegmatis* predictions, we fortuitously discovered homologues of 9 candidates in *M. bovis* BCG and 4 candidates in *M. tuberculosis* (Table 2.1, Figure A.4). Since these probes were not specifically designed for the other two species, lack of detection could be due to either the absence of sRNA expression or to non-optimization of the probe sequence that was used for hybridization to the targeted region in the *M. bovis* BCG and *M. tuberculosis* genome.

Also, differences in culture medium might contribute to the low number of expressed homologous sRNAs of *M. smegmatis* in *M. tuberculosis* as expression of these sRNAs could be specific to different conditions in *M. tuberculosis*. Given our focus in sRNA identification, specific conditions that could lead to differences in sRNA expression will be explored in future work.

2.2.8 23 Novel sRNAs Identified in *Mycobacterium Bovis* BCG

Twenty-one of the sRNA candidates for *M. bovis* BCG (Bo12, Bo15, Bo41, Bo52, Bo58, Bo67, Bo68, Bo75, Bo80, Bo85, Bo99, Bo100, Bo111, Bo113, Bo115, Bo117, Bo122, Bo125, Bo126, Bo137, and Bo139) were previously identified, under the nomenclature Mpr 1–21, respectively [128]. Forty-six other candidates were also tested previously but showed no signal; therefore, only the remaining 77 candidates were tested using Northern blotting analysis in this study, and we confirmed expression of 23 new sRNA candidates (Figure 2.8 and Figure A.5). A homologue of Bo46 was previously identified in M. tuberculosis by RNA-seq [82] but not otherwise experimentally validated. All of the validated sRNAs were in the same orientation as that predicted by SIPHT. We also applied the probes to *M. smegmatis* and *M. tuberculosis* and identified 20 and 5 sRNA homologues, respectively (Table 2.1; Figure 2.8; Figure A.5). All the confirmed sRNAs in *M. bovis* BCG and *M. tuberculosis* are listed in Table 2.2, along with the new nomenclature for sRNAs.

Table 2.1. Novel sRNAs confirmed by Northern blotting analysis in *M. smegmatis*

			Homology confirmed by Northern analysis in:			
	5' end*	**3' end***	BCG	TB	**New nomenclature**	**Potential Mtb orthologs**
Intergenic sRNAs						
Sm19	5029661	5029530			ncMSMEG14931Ac	
Sm49	1086797	1087035			ncMSMEG11016A	
Sm64	2523008	**2522888#**	✓	✓	ncMSMEG12439Ac	ncRv12904A
Sm76	**3690377**	**3690280#**	✓	✓	ncMSMEG13628Ac	ncRv11846A
Sm82	4392939/ 4392970	**4393039**	✓		ncMSMEG14302A	
Non-intergenic sRNAs						
Sm11	2835860	**2835984/ 2835999#**	✓		ncMSMEG12771A	
Sm38	2237170	**2237220#/ 2237253**	✓		ncMSMEG2161A	
Sm41	3815700/ 3815647	**3815581**	✓		ncMSMEG3749Ac	
Sm42	4290417/ 4290487	**4290537**	✓	✓	ncMSMEG4206A	
Sm67	2600405/ 2600425	**2600485#**			ncMSMEG2514A	
Sm68	2600389	2600701			ncMSMEG2514B	
Sm90	6845964	6846035	✓	✓	ncMSMEG6799A	ncRv11847A
Sm93	**858482**	**858588**	✓		ncMSMEG0774A	

*: The coordinates in bold are verified by 5' or 3' Deep-RACE. Coordinates that not in bold are estimated according to the size from Northern blotting analysis. For those don't have any Deep-RACE verified ends (i.e. Sm19, Sm49, Sm68 and Sm90), the coordinates of the ends are predicted by SIPHT. #: Located in or close to an AT rich region.

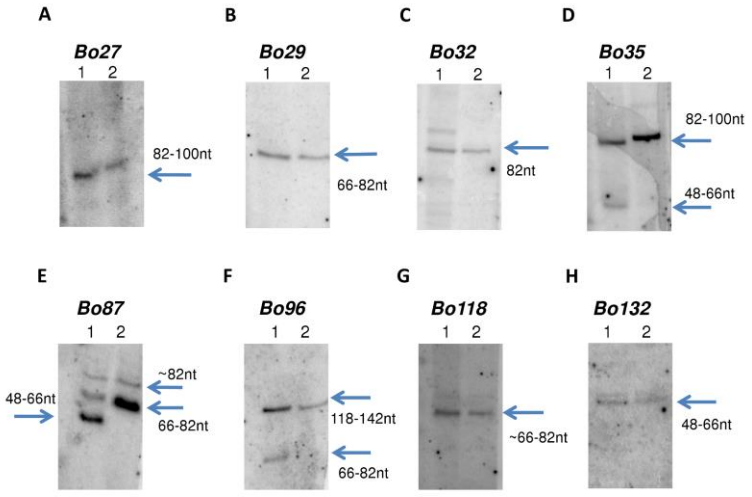


Figure 2.9. Northern blotting confirmation of sRNA candidates in *M. bovis* BCG

Selected images of Northern blotting analysis for validated *M. bovis* BCG sRNAs; the remaining images are included in Figure A.6. Lane 1 and 2 indicate total RNA samples extracted from *M. bovis* BCG and *M. smegmatis*, respectively. We used Phi-X174/Hae III Marker for the size prediction. The probes we used for this analysis are listed in Table A.7.

2.2.9 Deep-RACE Identifies sRNA 5′ and 3′ Ends

We used Deep-RACE, a previously described approach that combines conventional RACE and deep sequencing to identify 5′ and 3′ ends of selected RNAs [148], [149]. In total, we identified 5′ ends for 9 sRNAs and 3′ ends for 21 sRNAs. Examples are shown in Figure 2.9. For some sRNAs we identified multiple 5′/3′ ends. Multiple 5′ ends could be due to multiple transcription start sites or RNA processing. Multiple 3′ ends could be due to RNA processing or may indicate imprecise Rho-dependent termination of transcription.

Table 2.2. Novel sRNAs confirmed by Northern blotting analysis in *M. bovis*

			Homology confirmed by Northern analysis in:			
	5' end*	3' end*	Smeg.	TB	New nomenclature	Potential Mtb orthologs
Intergenic sRNAs						
Bo13	3126934/ 3126964	**3127039**			ncBCG12882A	
Bo27	**2157804**	2157704	✓		ncBCG11948Ac	
Bo29	1769910	**1769830**	✓		ncBCG11603Ac	
Bo32	**817571**	**817483**#	✓		ncBCG10734Ac	
Bo46	2603016	**2602916**	✓	✓	ncBCG12368Ac	ncRv12345Ac
Bo48	**3028955/ 3028965**	**3028876**	✓	✓	ncBCG12782Ac	ncRv12765Ac
Bo53	1044606	**1044706/ 1044720/ 1044727**	✓		ncBCG10960A	
Bo73	1647800	**1647871/ 1647880**	✓		ncBCG11504A	
Bo81	218700	**2187796**	✓		ncBCG10195A	
Bo82	2235286	**2235196**#	✓		ncBCG12024Ac	
Bo86	2325795	2325960	✓		ncBCG12107A	
Bo87	**2351000/ 2351046**	**2350915**# **/2350874**	✓	✓	ncBCG12128Ac	ncRv12111Ac
Bo96	**2686849**	**2687033**#	✓		ncBCG12441A	
Bo118	3765977	**3765917**	✓		ncBCG13438Ac	
Bo132	4260533	**4260610**	✓		ncBCG13885A	
Non-intergenic sRNAs						
Bo35	**576179**	**576067**#**/ 576104**	✓		ncBCG10493Ac	
Bo47	**2705925/ 2705838**	**2705735**			ncBCG12462Ac	
Bo60	1247638	**1247538**	✓		ncBCG11150Ac	
Bo71	1588810	**1588733/ 1588747**	✓	✓	ncBCG11448Ac	ncRv11387Ac
Bo78	**207337/ 207273**	**207179**	✓	✓	ncBCG10186Ac	ncRv10150Ac
Bo101	2919337	**2919277**	✓		ncBCG2654Ac	
Bo105	3073466	**3073546**			ncBCG12831A	
Bo130	413414/4 13464	**413324**	✓		ncBCG0352Ac	

*: The coordinates in bold are verified by 5' or 3' Deep-RACE. Coordinates that not in bold are estimated according to the size from Northern blotting analysis. For those don't have any Deep-RACE verified ends (i.e. Bo86), the coordinates of the ends are predicted by SIPHT. #: Located in or close to an AT rich region.

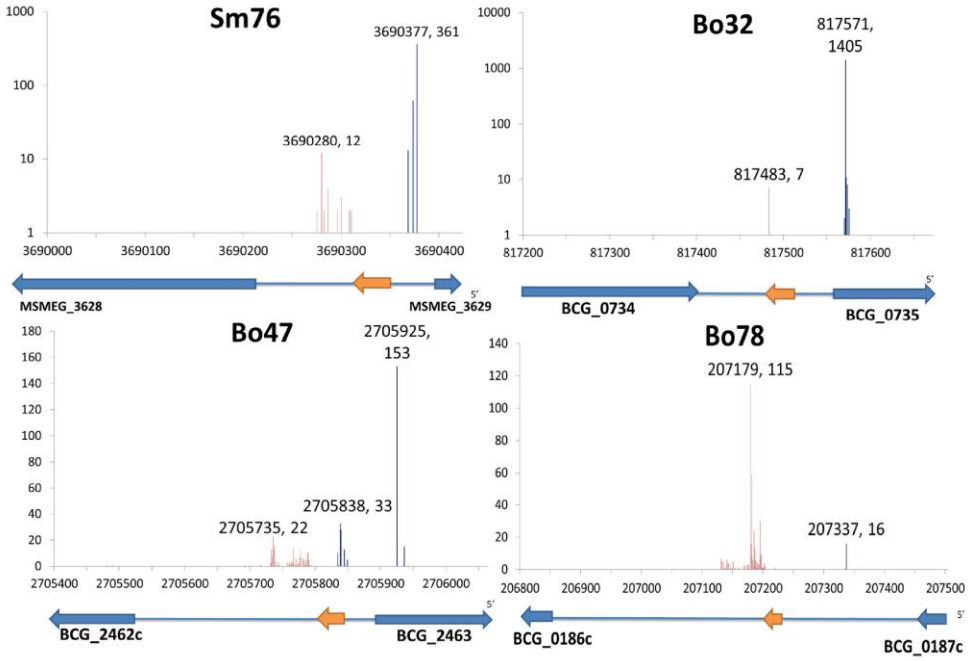


Figure 2.10. Identification of sRNA 5′ and 3′ ends by Deep RACE

Blue lines show the number of 5′ RACE reads mapped to respective genome, while red lines show the number of 3′ RACE reads. The coordinates with the highest number of mapped reads (the peak) indicate the likely 5′ and 3′ ends of sRNAs and are labeled in the figure. The orange arrow under the chart shows where the Northern probes base-paired and the blue arrows are the adjacent annotated coding regions. Results for other sRNAs can be found in Figure A.7.

As noted in our earlier study [128], the predicted size of the candidate sRNAs correlates only weakly with experimental observations. Only about 17% of the confirmed sRNAs were within 10% of their predicted sizes. Additionally, in many cases, multiple bands were detected by Northern analysis, suggesting the presence of multiple start sites, multiple termination sites, and/or sRNA processing. This is consistent with the Deep RACE data (Figure 2.10; Figure A.7). Deep RACE identified both 5′ and 3′ ends for

seven sRNAs. In these cases, the sizes determined by Deep RACE are similar to those confirmed by Northern blotting.

2.3 DISCUSSION

Advances in experimental and computational techniques have led to continual identification of a vast number of sRNAs in bacteria. We now understand that sRNA structures and sequences can be conserved between evolutionarily close organisms [150]. However, conservation patterns of functional sRNAs are more complex than those observed in coding regions. For example, some sRNAs are always co-conserved adjacent to coding regions, other sRNAs have similar sequences but perform different roles in different organisms, and, even in the same organism, some sRNAs can have multiple genomic copies that serve different regulatory functions [151]. Thus far, most of the conservation properties of bacterial sRNAs are not well understood. Given this, the evolution of bacterial sRNAs continues to be puzzling; this is particularly intriguing in the case of intergenic sRNAs that have evolved outside genomic coding regions.

For our analysis, we collected data for experimentally observed sRNAs in intergenic regions from 13 different bacterial species that have been widely studied and well annotated. Given the dependency of this analysis on selected species whose sRNAs we used and collected, we collected a vast amount of data to ensure statistical significance. Despite our selection of species that possess a well-annotated genome, have more comprehensive transcriptome data, and are more commonly used in sRNA studies and our use of only experimentally observed sRNAs, our data could be inherently biased based on our current selection of bacteria that have been sequenced and characterized extensively for medical or biotechnological purposes. Moreover, sRNAs that were

identified by different techniques could weigh differently, and some regulatory sRNAs may be expressed only under certain environmental conditions. While ideally this study can be done with sRNAs that all come from the same experimental technique (such as Northern blotting), this would yield only a relatively small number of sRNA candidates in some species that lack large-scale Northern blotting confirmation. Given the large numbers of sRNAs and the broad sample of organisms analyzed that validate the trends that we have observed, we believe that these patterns will hold for an even larger and more comprehensive data set. Furthermore, to assess the possible conservation bias from different techniques, we compared the conservation level of intergenic regions that encode sRNAs identified from Northern blotting to that of sRNAs identified with other techniques (microarray, RNA-seq, etc.) and found no significant difference in the conservation levels between these two groups of sRNAs (Table A.5 in the supplemental material shows the classification of sRNAs according to how they were experimentally identified).

A second key observation that results from our work is that intergenic regions that are conserved are enriched for sRNAs relative to non-conserved intergenic regions. Since some sRNA might be conserved along with adjacent coding regions, and to eliminate the possibility that high conservation levels of intergenic regions are due to 5′ or 3′ UTRs, we define as conserved only the intergenic regions that have a higher conservation level than flanking regions. Since most intergenic regions carry functional sequences, they are expected to be less conserved than protein-coding regions. This is a different approach from others that have been used in the literature to study conservation of intergenic regions [152], [153]. In most of the analyzed species, more than 20% of the intergenic regions have a conservation level equal to or higher than that of the extended region. As a result, it is possible that more functional sequences are yet to be identified in these highly

conserved intergenic regions. These results support the hypothesis that intergenic regions that are conserved across multiple species encode functional entities that are important for survival. This is further stressed by our findings that the actual sRNA-encoding regions are even more conserved than random regions within the same intergenic area.

The above results also depend on the technicalities of the WU-BLAST analysis. We used two different groups for WU-BLAST: the within-genus and outside-genus groups. The two groups yielded similar results, indicating that the number of species (outside-genus groups include more species than most genera) is not a critical parameter in this analysis. We hypothesized that by using an optimal phylogenetic distance to select species for WU-BLAST, we could eliminate species that are too close or too far from the interested species and yield more significant results. Our analysis of B. subtilis and S. pneumoniae supports this idea, while it was less significant for other species (data not shown), as these appeared to be evolutionarily clustered within a more optimal distance. Nevertheless, we believe that this approach can be further improved by systematically performing a cross-genus analysis to find the optimal phylogenetic distance applied for all species. The dependency on appropriate phylogenetic distance for conservation analysis is not surprising given that phylogenetic distances that are too close will obscure identification of intergenic regions that are truly conserved due to the potential importance of their encoded function. In contrast, organisms that are phylogenetically too far away will not show enough conservation among intergenic regions for meaningful analysis.

A third observation of our study is that the average sizes and distributions of intergenic-region lengths are very similar among the species analyzed, regardless of their genome size. Furthermore, intergenic areas that are significantly longer than the average are largely enriched in sRNAs. Indeed, this trend was observed to increase as intergenic

regions increased in length. This suggests that bacteria use their genome space highly efficiently, without the presence of large "unused regions" that do not encode functional transcripts. Interestingly, not many intergenic regions in our analysis were observed to encode more than one sRNA, and the few intergenic regions that did encode multiple sRNAs (no more than two) were not significantly longer. A more fundamental question is whether these long intergenic regions are long because they encode sRNAs or whether sRNAs are more likely to be encoded in long intergenic regions. Based on this study, we believe that most long intergenic regions could have encoded functional sequences. This is not limited to sRNAs but also applies other functional noncoding transcripts or sequences in other organisms [154], [155]. Long intergenic regions have more space to house noncoding RNAs, and it would be interesting to look for unknown sRNAs in long intergenic regions in which no functional transcripts have been found yet.

For the novel sRNAs identification in Mycobacteria, although we aimed to find intergenic sRNAs, half of the candidates we identified in this study overlap partially or entirely protein-coding genes in either the sense or antisense orientation (Table 2.1, Table 2.2). We categorized sRNAs into different classes according to their position relative to adjacent coding regions. Where possible, we used 5′/3′ end information from Deep-RACE data. For sRNAs that have only one end mapped by Deep-RACE, the other end was estimated according to the length confirmed by Northern blotting analysis (Figure 2.8). For sRNAs that have neither end mapped by Deep-RACE, the farthest possible ends were estimated according to Northern blotting analysis and the sRNAs would be categorized as "not determined" if multiple class options exist.

Nine sRNAs in *M. smegmatis* (Sm19, Sm32/33, Sm35, Sm46, Sm49, Sm64, Sm76, Sm82) and twelve sRNAs in *M. bovis* BCG (Bo35, Bo48, Bo53, Bo60, Bo71, Bo73, Bo78, Bo86, Bo101, B0105, Bo118, Bo132) were mapped completely to

intergenic regions. Four sRNAs in *M. smegmatis* (Sm38, Sm41, Sm90, Sm93) were mapped to the sense strand of annotated protein-coding genes, and four were mapped to the antisense strand (Sm42, Sm67, Sm68, Sm74). One sRNA in *M. smegmatis* (Sm11) and five in *M. bovis* BCG (Bo32, Bo47, Bo81, Bo96, Bo130) overlap partially with adjacent genes in the antisense orientation, and four sRNAs in *M. bovis* BCG (Bo27, Bo46, Bo82, Bo87) overlap partially with adjacent genes in the sense orientation. One sRNA in *M. smegmatis* (Sm47) and two in *M. bovis* BCG (Bo13, Bo29) were not classified.

The location of sRNAs relative to protein-coding genes also gives clues as to their function. Regulatory sRNAs that are completely intergenic typically function by base-pairing with distally-encoded mRNAs; however, some of the sRNAs are close to the 5′ end or 3′ end of adjacent genes, suggesting possible alternative regulatory roles. sRNAs antisense to ORFs or UTRs can regulate expression of the overlapping gene [29]. sRNAs located within UTRs or ORFs in the sense orientation may be degradation products or mRNAs or could be important cis-acting regulatory elements such as riboswitches.

sRNAs can be transcribed independently or generated by processing of mRNA UTRs. Several features of the sRNAs identified in this work are consistent with the sRNAs being independently transcribed from their own promoters. First, the Northern blots showed no evidence of larger bands that could correspond to pre-processed mRNAs. Second, 13 sRNAs (Sm35, Sm42, Sm67, Sm68, Sm74, Bo13, Bo32, Bo60, Bo71, Bo73, Bo81, Bo118, Bo130) are orientated away from the surrounding genes. Third, 5 sRNAs (Sm64, Sm82, Bo47, Bo105, Bo132) are located >200 bp from the nearest gene start/stop. Nineteen sRNAs are close to (<200 bp) upstream or downstream coding regions (Sm11, Sm19, Sm32/33, Sm46, Sm47, Sm49, Sm76, Bo27, Bo29, Bo35, Bo46, Bo48, Bo53, Bo78, Bo82, Bo86, Bo87, Bo96, Bo101) and four (Sm38, Sm41,

Sm90, Sm93) overlap coding regions in the sense orientation. It is formally possible that these sRNAs are generated by mRNA processing or premature termination, although the Northern blot analysis argues against this. Regardless, sRNAs processed from mRNAs could still have important regulatory functions [3], [30], [31]. Indeed, a recent study identified 3′ UTRs as an abundant source of regulatory sRNAs in Salmonella enterica [32]. Alternatively, sRNAs generated by processing of mRNAs could indicate cis-acting regulatory elements such as riboswitches.

The regulation of sRNAs can provide important clues as to their biological functions. However, very little is currently known about regulation of mycobacterial sRNAs. The genome-wide binding profiles of many *M. tuberculosis* transcription factors have recently been determined using ChIP-seq and these data are publicly available [25]. Although we identified sRNAs in *M. bovis* BCG, it is highly likely that these sRNAs are conserved in *M. tuberculosis* given the extremely high similarity of the *M. bovis* BCG and *M. tuberculosis* genomes [33]. Hence, we searched existing ChIP-seq datasets of *M. tuberculosis* for transcription factors that bind close to sRNA 5′ ends, including sRNAs identified in earlier studies [12]. We identified 10 ChIP-seq peaks (indicative of a transcription factor binding site) located between 100 bp upstream and 20 bp downstream of sRNA 5′ ends (Table A.8). Thus, we have identified likely examples of sRNA regulation. In some cases, the ChIP-seq peak is also close to the start of an annotated protein-coding gene. Hence, the transcription factor may regulate the protein-coding gene rather than the sRNA. Nevertheless, in four cases, the ChIP-seq peak is unambiguously associated with an sRNA 5′ end. The two examples with highest ChIP-seq signal are shown in Figure 2.11. For each of these examples, the transcription factor is otherwise uncharacterized.

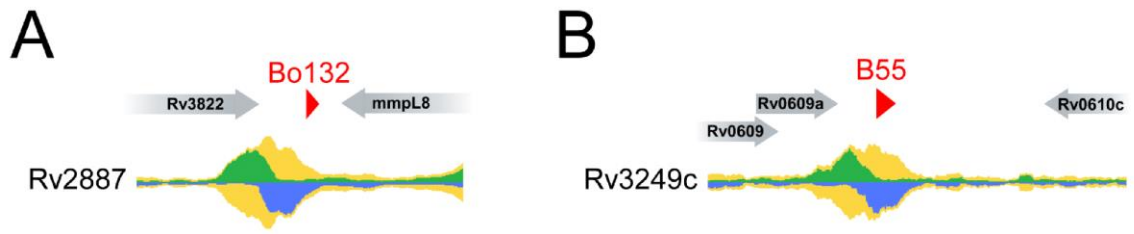


Figure 2.11. ChIP-seq peaks associated with predicted sRNA homologues in *M. tuberculosis*

ChIP-seq peaks that are unambiguously associated with sRNA 5′ ends. Raw ChIP-seq data from www.tbdb.org are shown for two transcription factors, (A) Rv2887, and (B) Rv3249c. Data are shown for genomic regions surrounding (A) sRNA Bo132 (this work), and (B) sRNA B55 [12]. The green and blue graphs indicate the relative number of sequence reads mapping to the plus and minus strands, respectively. The yellow graphs indicate the sum of plus and minus strand reads. Annotated genes are shown as gray arrows. sRNAs are shown as red triangles.

In summary, the evolution of sRNA in bacteria is an intriguing subject. A major challenge in this field is that some sRNAs with the same function could have different sequences in different organisms, or the same sRNA sequence could have different functions in different organisms. This study provides insight into some critical questions that remain unanswered about sRNA evolution in bacteria. A future approach could incorporate the use of structural homology prediction models in addition to sequence homology methods to better identify and understand sRNA conservation patterns in terms of function (89). An advantage of the strategy we have used is the ability to look at sRNAs in the context of the entire genetic region in which they are found. We have also identified 17 novel sRNAs in *M. smegmatis* and 23 novel sRNAs in *M. bovis* BCG, verified 5′ and 3′ ends, and list these sRNAs according to a recently-proposed annotation

nomenclature. Our analysis of sRNA position relative to protein-coding genes suggests various potential roles for these sRNAs in gene regulation. Although the specific biological function of these, and all other known mycobacterial sRNAs, is not understood, we speculate that some of these sRNAs contribute to the biology of pathogenic mycobacterial species.

2.4 Material and Methods

2.4.1 Targeted bacterial species

In this study, we selected 13 bacterial species: *Bacillus subtilis str.* 168, *Chlamydia trachomatis* L2b/UCH-1/proctitis, *Enterococcus faecalis* V583, *Escherichia coli str.* K-12 *substr.* MG1655, *Helicobacter pylori 26695, Listeria monocytogenes* EGD, *Mycobacterium bovis* BCG Pasteur, *Pseudomonas aeruginosa* PAO1, *Salmonella enterica subsp. enterica serovar Typhi* str. Ty2, *Staphylococcus aureus* N315, *Streptococcus pneumoniae* TIGR4, *Streptococcus pyogenes* MGAS5005, and *Vibrio cholerae* El Tor. These species were selected due to the availability of detailed transcriptome analysis data that have been reported for their genomes using high-throughput sequencing or other traditional methods. The list of species, along with the Gram stain results, pathogenicity, and reference to the corresponding published transcriptome study, is given in Table A.1 in the supplemental material.

2.4.2 Genome-wide extraction of intergenic and extended intergenic region sequences

Data for all the sample genomes were found in the J. Craig Venter Institute (JCVI) database or in the National Center for Biotechnology Information (NCBI) genome database (36). To prevent conservation bias due to the presence of protein-coding sequences, the analysis of sRNA candidates was limited to sequences that were

completely intergenic (as determined by the most recent genome annotations) and showed negligible overlap with nearby annotated open reading frames. Sequences that had up to a 10-nucleotide (nt) overlap upstream and/or downstream of the candidate sRNAs were included in the analysis, to accommodate for any potential annotation errors. The list of "extended intergenic regions" was generated by including a part of the upstream and downstream coding regions along with each intergenic region sequence. An intergenic region with a length of n nucleotides was extended for n nucleotides upstream and downstream. As a result, extended intergenic regions were three times the length of the original intergenic regions.

2.4.3 Conservation analysis of different genomes by BLAST

WU-BLAST (BLASTN 2.0MP-WashU [4 May 2006]) (W. Gish, personal communication) was used to perform the sequence conservation analysis of intergenic and extended intergenic regions. Intergenic sequences with a minimum length of 60 nt were used to avoid spurious hits. However, the conservative expectation value (E value) established for the WU-BLAST analysis rarely returned hits for short (<60-nt) sequences when used with genome-sized databases. WU-BLAST outputs were filtered with a PERL script to a stringent threshold of at least 50% query sequence coverage with 50% identity in the conserved regions. The filtering restricted the hits for the search of homologous sequences to ones with a "high-to-extreme similarity" regime. These parameters were selected according to search criteria that have been developed to analyze conservation levels of protein-encoding sequences, where the expected level of conservation is much higher [156].

Two measures of conservation were used: "within-genus" and "outside-genus." For the within-genus criterion, the homology of a specific sRNA candidate and/or

intergenic region was determined relative to a list of specific genomes of species within the genus. For instance, for all C. trachomatis intergenic region sequences, homology was analyzed relative to *C. trachomatis* intergenic region sequences; homology was analyzed relative to *C. psittaci, C. pneumonia, C. pecorum* and *C. muridarum* (members of the Chlamydia genus). The full list of genomes that apply to each species is included in Table A.2 in the supplemental material. A measure of within-genus homology was obtained by counting the number of organisms within the genus where homology was observed. The length of all the query sequences, the resulting hit score, and the E values were summarized in Table A.3 in the supplemental material.

For the outside-genus criterion, the homology of a specific sRNA candidate and/or intergenic region was determined relative to any species within a specific list of the following genera: *Agrobacterium, Bacillus, Bacteroides, Bordetella, Borrelia, Brucella, Burkholderia, Chlamydia, Clostridium, Deinococcus, Desulfo, Enterobacter, Enterococcus, Escherichia, Geobacter, Haemophilus, Helicobacter, Lactobacillus, Listeria Mycobacterium, Mycoplasma, Neisseria, Pseudomonas, Rhizobium, Rhodobacter, Rhodococcous, Rickettsia, Shigella, Salmonella, Streptococcus, Streptomyces, Staphylococcus, Synechococcus, Thermotoga, Vibrio, Xanthomonas, Yersinia,* and *Zymomonas*. This list was generated as a way to further control the searches conducted for all sample species in a way that broadly sampled across all bacterial species. The length of all query sequences, the resulting hit score, and the E values were recorded as for the outside-genus analysis. The BLAST data are presented in Table A.3 in the supplemental material.

The NCBI BLASTn discontiguous Megablast tool was used to determine sequence conservation of sRNA-coding regions (sRCR) and an adjacent random sequence in the same intergenic region (RIGR). Stringent conservation parameters were

used: an E value of <0.001, ≥50% query coverage, and ≥50% identity. Using discontiguous Megablast, each sRNA-coding region and a random selected region of the respective intergenic region were analyzed. The number of hits returned from species of the same genus (within-genus group) and the number of hits returned of genera that differed from the target species (outside-genus group) are summarized in Table A.4 in the supplemental material.

2.4.4 Collection of experimentally observed sRNAs from published works

For each species analyzed with WU-BLAST, coordinates of experimentally observed sRNAs were collected from online databases or published reports (all sources used are listed in Tables A.1 and A.9 in the supplemental material). All pooled sRNAs were identified either by experimental techniques, such as Northern blotting or cloning, or by transcriptome sequencing techniques, such as RNA-seq or microarray analysis.

2.4.5 Phylogenetic distance calculation

Phylogenetic distances were estimated by MEGA5 (Molecular Evolutionary Genetics Analysis), a tool for aligning sequences and computing nucleotide pairwise distances [157]. 16S RNA sequences were retrieved from the NCBI database and aligned by ClustalW (a MEGA5 built-in algorithm). The P-distance model was used to estimate the phylogenetic distance between each species.

2.4.6 Comparisons of all intergenic regions with experimentally observed sRNAs

The list of all intergenic regions generated from the JCVI or NCBI database was compared to all sRNAs that have been experimentally observed (see Table A.1 in the supplemental material). Any intergenic region within the genome that contained one or more experimentally observed sRNAs was identified as an sRNA-coding intergenic region. Further criteria were applied to the data to explore any possible correlations

between the likelihood of intergenic regions being sRNA-containing regions and the length or conservation level of those regions.

A survey of the longest intergenic regions is shown in Table A.10 in the supplemental material, where the top 20% longest regions of all intergenic regions within a species were defined as "long intergenic regions." Conservation data from the WU-BLAST analysis were also used to verify correlations between conserved intergenic regions and sRNA-coding intergenic regions. An intergenic region was considered conserved if the hit number returned by WU-BLAST was at least 1 and was higher than the hit number of the extended region.

2.4.7 Strains and Plasmids

M. bovis BCG (Pasteur strain, Trudeau Institute), and *M. tuberculosis* H37Rv were grown in mycomedium (as previously reported, [128]). *M. bovis* BCG and *M. tuberculosis* cultures were grown for 7 days, with shaking, to late-log phase. Cultures of *M. smegmatis* MC2155 were grown shaking at 37°C, in trypticase soy media supplemented with 0.05% Tween 80 for 18 hours with shaking (late-log phase).

2.4.8 Phylogenetic Selection of Computationally Predicted sRNA Candidate

Small RNA candidates of *M. smegmatis* were predicted using the SIPHT program with the same parameters as described previously [97], [158]. SIPHT identifies potential sRNA candidates based on the presence of intergenic sequence conservation upstream of putative Rho-independent terminators. SIPHT has been widely applied in sRNA studies [159]–[161], and its reliability has been tested and compared with other algorithms [98].

2.4.9 RNA Isolation and Northern Blot Analysis

RNA was isolated as previously reported [128]. Northern blot analysis was performed as previously reported [128]; probes were designed according to SIPHT

predicted sequences and tested in *M. bovis* BCG, *M. smegmatis* and *M. tuberculosis* [128].All the oligonucleotides that were used in this study are listed in Table A.1.

2.4.10 ChIP-seq Analysis

We analyzed existing ChIP-seq datasets for 55 *M. tuberculosis* transcription factors extracted from a previous study [162]. ChIP-seq peak positions were compared to the 5' end positions of *M. bovis* BCG and *M. tuberculosis* sRNAs from the current study and two previous studies [12, 15]. For *M. bovis* BCG sRNAs, we first identified the equivalent region of the *M. tuberculosis* H37Rv genome. Possible sRNA regulators were selected if the ChIP-seq peak was located within 100 bp upstream and 20 bp downstream of an sRNA 5′ end.

2.4.11 Deep 5′ and 3′ RACE

Deep 5′ RACE and Deep 3′ RACE were performed as previously described [149] with the following exceptions. Deep 5' RACE libraries and Deep 3' RACE libraries were pooled and sequenced together using an Ion Torrent 316 chip (Wadsworth Center Applied Genomic Technologies Core Facility). For Deep 5' RACE, sequence reads were identified by the presence of the expected adapter sequence at the read 5' end. Adapter sequences were removed and reads of >40 nt were mapped to the reference genomes using BWA [164]. For Deep 3' RACE, sequence reads were identified by the presence of the expected adapter sequence. Adapter sequences were removed. The oligo-dT stretch was removed by identifying the first consecutive pair of bases not including a "T" and removing all sequence upstream of this. Sequences of >40 nt were mapped to the reference genomes using BWA [164]. For both Deep 5′ RACE and Deep 3′ RACE, 5′ and 3′ ends were identified as the position with the most sequence reads, and with a minimum of 5 reads. Sequences of all primers used for Deep RACE are listed in Table A.11.

Chapter Three

*Identify novel ncRNAs in *D. radiodurans* with bioinformatics and transcriptome analysis

3.1 INTRODUCTION

Small RNAs (sRNAs), ranging from 21 to more than 400 nucleotides, are an intriguing class of RNAs that typically do not encode functional proteins but have demonstrated intrinsic roles as cellular regulators of transcription and translation [2], [13], [44], [165]. A key property of sRNAs is their ability to simultaneously turn on and off a variety of metabolic pathways in response to environmental signals, such as the change of temperature, pH, and other potentially lethal stressors [87], [165]–[167]. To exert their function, sRNAs can either base-pair with messenger RNAs (mRNAs) to prevent or promote protein translation, or sequester proteins into ribonucleic protein complexes to intervene protein activity [4]. Although a variety of mechanisms for sRNA function continue to be reported, it is well documented that sRNAs are highly dependent on their secondary structure and on their ability to undergo rapid conformational changes to exert their regulatory effects [87], [165]–[167]. Noncoding RNAs can be broadly categorized into two classes based on where they are encoded relative to their targets [163], [168] but this classification is continuously evolving. For instance, a newly tRNA-derived sRNA from E. coli and other organisms continues to challenge these classifications [169]. While most of the cis-encoded sRNAs control one specific target,

This work was published in:
[63] CH.Tsai, R.Liao, B.Chou, and L. M.Contreras, "Transcriptional Analysis of Deinococcus radiodurans Reveal Novel sRNAs that are Differentially Expressed under Ionizing Radiation," *Appl. Environ. Microbiol.*, vol. 81, no. 5, pp. 54–64, 2015. (Author contribution to the work: experiment designing and performing, research concept discussion, data analysis and manuscript writing.)

some trans-encoded sRNAs are capable of binding and regulating multiple targets [87], [165]–[167]. Recent studies have also discovered more potential mechanisms of sRNA function, such as a coupled action with riboswitch elements that are regulated by different ligands [170]. Moreover, the versatile role and the specificity of sRNA targeting are gaining increasing traction for engineering applications, particularly in the context of metabolic engineering [41], [42], [171], [172].

Currently, with the development of advanced prediction and sequencing techniques, an increased number of sRNAs have been identified throughout bacteria [5]. Hundreds of sRNA candidates have been computationally predicted with different algorithms, such as sRNApredict, QRNA or NAPP [98]. On the other hand, many sRNAs have been identified experimentally by deep sequencing techniques (e.g. RNA-seq) and other experimental techniques (e.g. Northern blotting analysis and microarray analysis) [10], [121], [146], [173].

Deinococcus species represents an interesting group of bacteria given their ability to survive extraordinarily high doses of ionizing radiation. *Deinococcus radiodurans* can survive acute doses up to 12-20 kGy, which cause massive DNA damage, and can grow under chronic irradiation at dose-rates as high as 60 Gy/h, without inducing mutations [28]. Moreover, *D. radiodurans* is amenable to genetic engineering and has been subjected to whole-genome sequencing and functional genomics [28], [92], [174]–[177]. By comparison, vertebrates and Escherichia coli cannot typically survive doses higher than 5 Gy and 1 kGy, respectively. This makes *D. radiodurans* a leading model for

studies of DNA repair and a top candidate for bioremediation of radioactive waste sites [25], [178], [179].

Various hypotheses have been tested to understand extreme radioresistance in *D. radiodurans* [76], [180]–[183]. This phenotype is complex, relying on a set of DNA repair proteins which operate far more efficiently than in naturally radiation-sensitive organisms [76], [180]–[183]. The molecular basis for the high efficiency of DNA repair proteins in *D. radiodurans* appears to include the accumulation of manganese antioxidants which prevent the inactivation of enzymes by reactive oxygen species (ROS)[26], [179]. Over the last 15 years, a diverse set of genes (including some involved in metabolism, DNA repair, and ROS-scavenging) have been shown to be differentially regulated following high-dose exposures (5-16 kGy) [180], [181]; however, most of the up-regulated genes were subsequently shown not to be essential to radioresistance [26]. Since then, the main strategy to delineating a minimal set of genes involved in extreme resistance has been to compare the whole-genome sequences of phylogenetically related but distinct Deinococcus species, whereby unique genes were ruled out but shared genes have been pooled as candidates for involvement in resistance. This bioinformatics approach eliminated almost all the novel genes first implicated in the extreme radiation resistance of *D. radiodurans* [28], and few unique genes in Deinococcus spp, such as recA, remain implicated in contributing to its remarkable DNA repair capacity [184], [185]. Indeed, the conserved set of radiation resistance determinants of *D. radiodurans* consists mainly of genes present in many other organisms [26]. For instance, a common

palindromic DNA motif of a dedicated transcriptional regulator (HucR) was predicted within the set of conserved genes [186].

The question of how radioresistance in *D. radiodurans* is regulated remains unresolved, and we have hypothesized that sRNAs in *D. radiodurans* may be important based not only on their regulatory roles in other bacteria, but also on their small size. The linear density of DNA damage caused by ionizing radiation in *D. radiodurans* and other organisms is very similar [25], [178]. Approximately 0.005 DSBs/Gy/Mb of DNA double strand breaks (DSBs) are introduced to the genome and this rate is approximately 10 times greater for single stranded breaks (SSBs). We hypothesize that the small size of sRNA genes would leave them largely undamaged at 15 kGy and could contribute to irradiation resistance. Previous work had hinted at the existence of non-coding RNAs of potential importance to Deinococcus spp. A transcript that resembles a Y RNA has been identified in *D. radiodurans* [187]. This particular non-coding RNA is able to bind Rsr, a Ro protein ortholog that contributes to radioresistance and is structurally similar to Hfq [188].

In this study, we merged computational and experimental techniques to identify novel potential sRNAs in *D. radiodurans*. We used computational tools to find hundreds of loci that were predicted to be sRNA candidates and applied a previously developed criteria to further filter most plausible sRNA candidates [24], [97], [189]. We also used deep sequencing techniques using total RNA from *D. radiodurans* and discovered 199 sRNA candidates in intergenic regions. Upon confirmation by Northern blotting and analysis, we uncovered expression of 41 novel sRNAs in *D. radiodurans*, 8 of which

showed differential expression following recovery from ionizing radiation. We also found and experimentally validated the presence of homologous sRNA candidates in a closely related radioresistant species *Deinococcus geothermalis*. Our studies also identify other well-characterized non-coding RNAs that have not been previously annotated in current versions of the *D. radiodurans* genome (NCBI reference sequence: NC_001263.1 & NC_001264.1). We suspect that validation of sRNA expression in *D. radiodurans* will contribute another dimension to ongoing studies of the mechanisms of radiation resistance.

3.2 RESULTS

3.2.1 Deep sequencing reveals hundreds of potential transcripts from non-coding regions

A range of 8,610,676 to 9,473,672 reads were generated per library in our Illumina RNA-seq analysis of total RNA extracted from wild-type *D. radiodurans* that was cultured to exponential phase (OD_{600}= 1); of these, more than 80% were mapped to the genome of *D. radiodurans* by Bowtie2 [190]. Reads that mapped to annotated rRNA and tRNA (around 30% of the total reads, Figure B.1) were excluded in our analysis. Intergenic regions that potentially encode for non-coding RNAs were mapped and visualized by Integrated Genome Viewer (IGV). As a result, the sRNA candidates were manually annotated by inspecting the intergenic regions. All intergenic loci mapped with over 100 reads per base and longer than 30 nt were annotated as potential sRNA candidates. As a point of reference, reads per base for annotated tRNAs ranged from 3000 to 20000 counts. Depending on the genome location, the candidate sRNA was categorized as overlapping coding region or intergenic region (IGR) transcript.

Given the continual evolution of genome annotation of *D. radiodurans*, we arbitrarily considered any sRNA candidates that overlapped less than 10 nt with the annotated coding region as intergenic. As a result, 199 sRNA candidates were identified. This includes a tmRNA which had been computationally identified [191] and the YRNA-like sRNAs that had been identified with deep-seq data (Figure B.2)[188]. The full list of 199 potential sRNA candidates is included in Table B.1. Among all the small RNA candidates, 46% of them are encoded entirely in intergenic regions and 54% overlap with the 5' or 3' end of their adjacent coding region. These overlapping candidates, if on the same strand of the coding region, could be potential functional untranslated regions (UTRs). A functional UTR can form secondary structure and interact with a coding region to regulate translation. We found 56 overlapping candidates that are longer than 100 nt, which could support functional structures. On the other hand, these overlapping candidates can also act as *cis*-encoded sRNAs if they are on the opposite strand. Upon further bioinformatics analysis with the Rfam database [109] we also found two other candidates to be homologous with the T-box leader sequence (Table B.2)[191]. The high number of reads observed for these candidates was consistent with the high level of expression expected from this class of RNAs.

3.2.2 sRNA transcripts verified by Northern blotting analysis from deep sequencing result

We selected 54 (Dsr1 to Dsr54) out of all 199 sRNA candidates identified from our deep sequencing analysis for further experimental verification. We hypothesized that these particular transcripts could be more easily detectable since they showed higher expression levels relative to the average intergenic region (e.g. mapped with more than 500 reads), these candidates also exhibited higher conservation levels in close species (the conservation level analysis is discussed in a later section). Unique probes were

designed for each candidate to target the most expressed loci (Table B.3). Northern blotting was performed using total RNA prepared from exponential phase cultures, and 27 sRNAs were confirmed to be expressed (Figures 3.1 and 3.2). To our knowledge, these sRNA candidates had not been identified previously. While most of the identified sRNA candidates are intergenic, eight candidates (Dsr1, Dsr7, Dsr12, Dsr17, Dsr35, Dsr40, Dsr46 and Dsr48) were categorized as 5' or 3' overlapping sRNAs. Expression of

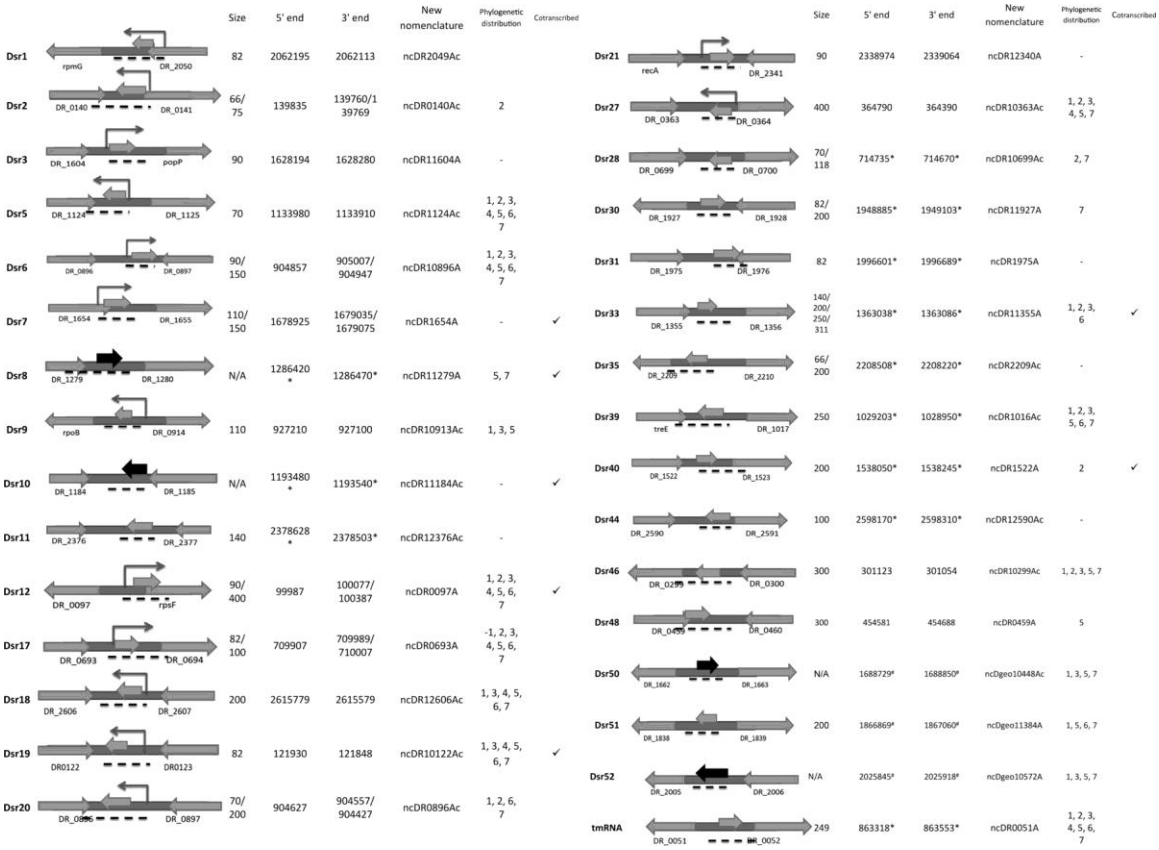


Figure 3.1. Novel small RNA candidates in D. radiodurans confirmed by Northern blotting and/or RT-PCR deep sequencing

The elbow arrows show the transcription starting site confirmed by 5' Deep-RACE. The middle arrows show where the Northern probes bind. The black middle arrows are sRNAs that identified with RT-PCR

but not with Northern blotting analysis. Thearrows on the sides are the annotated flanking protein coding regions. The black dash lines indicate the estimated sRNA coding loci. . The black arrows are sRNAs that identified with RT-PCR but not with Northern blotting analysis. The size of the RNAs was estimated by comparing to phiX174 ladder. The 5' ends were either identified by 5'-RACE or (with * mark) estimated using deep sequencing data. The 3' ends are estimated with 5' end coordinates and RNA size from Northern analysis or deep sequencing data. The phylogenetic distribution shows the species where homologous small RNAs were found: 1. *Deinococcus gobiensis* 2. *Deinococcus proteolyticus* 3. *Deinococcus deserti* 4. *Deinococcus peraridilitoris* 5. *Deinococcus geothermalis* 6. *Deinococcus maricopensi* 7. *Deinococcus swuensis.* The cotranscribed sRNA that identified with RT-PCR are marked in the last column.

an RNA homologous to tmRNA was also validated by Northern blotting analysis (Figures 3.1 and 3.2). All confirmed sRNAs are renamed with a recently proposed nomenclature which use *Mycobacterium* as an example but can be applied to all bacterial species [147].

While it is known that sRNAs can be independently transcribed from the genome or processed from an mRNA, most of the sRNAs identified in this study only showed one band by Northern blotting analysis. For the remaining candidates, multiple or larger bands could suggest post-transcriptional processing of the sRNA or potential riboswitches.

3.2.3 RT-PCR and co-transcription experiment identified more sRNA candidates

To further confirm expression of all sRNA candidates that were verified by Northern analysis, we conducted an RT-PCR analysis. These experiments were

particularly beneficial for candidates that were expressed at very low levels and that yielded ambiguous results from Northern analysis. All sRNA candidates that were verified by Northern blotting were also confirmed with RT-PCR, except Dsr5 (Figure B.3). We also tested 6 additional sRNA candidates that were originally not detected by Northern probing (presumably because of their lower expression levels, as seen in our transcriptome data) but were conserved in *D. geothermalis* or had predicted relevance to the radioresistance phenotype. This resulted in 4 more sRNAs candidates identified: Dsr8, Dsr10 Dsr50 and Dsr52 (Figure 3.1).

Since many identified sRNAs overlap or are in proximal distance to the upstream or downstream open reading frame, we hypothesized that some of these sRNAs could be co-transcribed with flanking genes. After testing for co-transcription, with primers that amplify the sRNA and corresponding flanking gene, we found Dsr8, Dsr10, Dsr12, Dsr19, Dsr33 and Dsr40 to be co-transcribed with their flanking coding regions (Figure B.4). These data also suggest that sRNAs could be post-transcriptionally processed from larger transcripts.

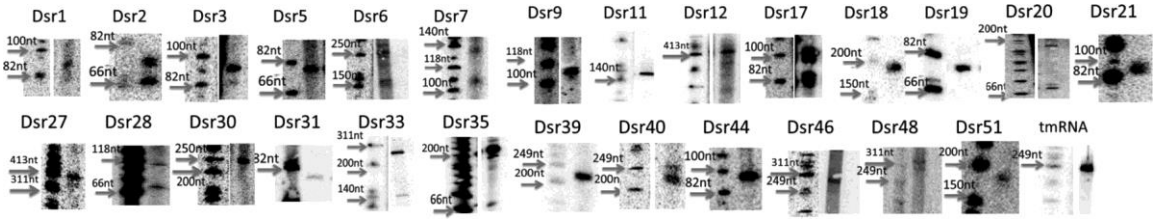


Figure 3.2. Images of Northern blotting analysis for confirmed sRNAs candidates from deep sequencing analysis.

Cells were cultured to exponential phase (OD600 = 1). Analysis was performed for 8 ug of total RNA sample. The expressions of predicted sRNAs were confirmed, and the sizes of the transcripts were

estimated relative to the phiX174/Hinfl Marker. The images of the ladder and the sRNA lanes might cut from the same gel in different parts with contrast adjusted to show clearer image.

3.2.4 5' ends of the sRNA candidates were mapped by Deep-RACE

To verify the exact coordinates of the transcription starting site (TSS) for each sRNA candidate, a 5' Deep-RACE analysis was conducted on the confirmed sRNA candidates [148]. After construction of a cDNA library for RNA samples prepared from exponential phase, reverse primers (same ones used as probes for Northern blotting analysis) were used to amplify the 5' end of each sRNA candidate (Table B.4). The amplicons were collected and sequenced. The sequenced reads were mapped to the genome of *D. radiodurans* as described in the Methods section. Annotations of the 5' ends were done by manual inspection with IGV (Figure B.5), and sequenced reads were successfully mapped to the 5' end of 14 sRNA candidates. Other sRNAs were not possible to map, potentially due to their lower expression levels. The identified coordinates of TSSs are shown in Figure 3.1.

3.2.5 Confirmation of additional sRNAs from computational predictions

A total of 391 computationally predicted sRNA candidates in *D. radiodurans* were collected (Table B.5); 256 were generated using QRNA and 126 were predicted by using SIPHT (previous known as sRNApredict3)[97][192]. It is worth noting that 17 candidates overlapped among these two sets of predictions (Figure B.6). Of the 17 sRNAs predicted by both QRNA and SIPHT, 7 of them were also identified with deep sequencing as potential sRNAs.

Based on a filter used successfully in previous works in our lab [189], we narrowed these predictions to include a smaller set of candidates that we rationalized

would be more likely to be true sRNAs. Briefly, sRNA candidates that were encoded in larger (top 20% long) and more isolated intergenic regions were selected for Northern blotting analysis. Based on those results, we tested 60 additional candidates (36 from published QRNA predictions and 24 candidates from our SIPHT predictions). Importantly, 10 more sRNA candidates were validated by Northern blotting analysis: 6 from QRNA predictions (Qpr1~Qpr6) and 4 from SIPHT analysis (Spr1~Spr4) (Figure 3.3, 3.4). The poor overlap between the two prediction methods or the high number of false positives is not surprising giving that these algorithms are based on two different criteria (sequence homology and secondary structure conservation) and not on any functional information; similar observations have been found in other bacterial studies [80]. Interestingly, only two confirmed sRNAs (Spr1, Qpr1) were detected by deep sequencing. This is not surprising, since some of the sRNAs might be degraded or not present in the cDNA pool due to biases in reverse-transcription. These results confirm our previous observation that better efficiency of sRNA experimental identification could be achieved by incorporating both computational and sequencing methods.

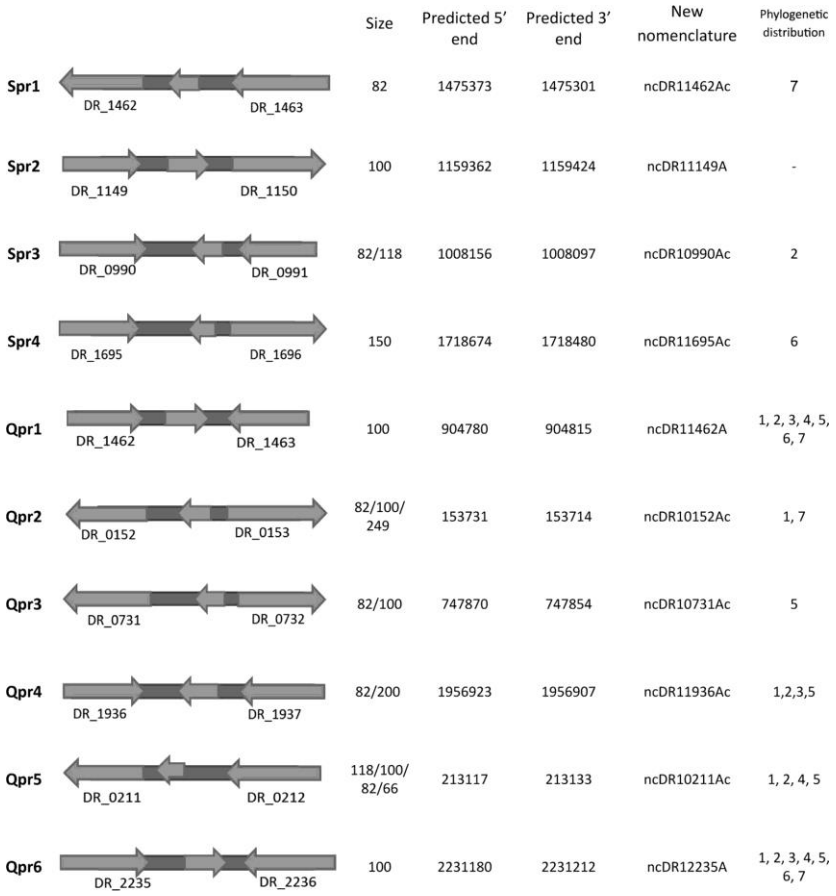


Figure 3.3. Novel small RNA candidates confirmed with Northern blotting analysis in *D. radiodurans* from computational predicted candidates

Spr1-4 are predicted by SIPHT, and Qpr1-6 are predicted by QRNA. The middle arrows show where the Northern probes bind. The arrows on the sides are the annotated flanking protein coding regions. The size of the RNAs was estimated by comparing to phiX174/Hinfl ladder. The 5' and 3' ends were predicted computationally (SIPHT or QRNA). The phylogenetic distribution shows the species that homologous small RNAs were found: 1. *Deinococcus gobiensis* 2. *Deinococcus proteolyticus* 3. *Deinococcus deserti* 4. *Deinococcus peraridilitoris* 5. *Deinococcus geothermalis* 6. *Deinococcus maricopensis* 7. *Deinococcus swuensis.*

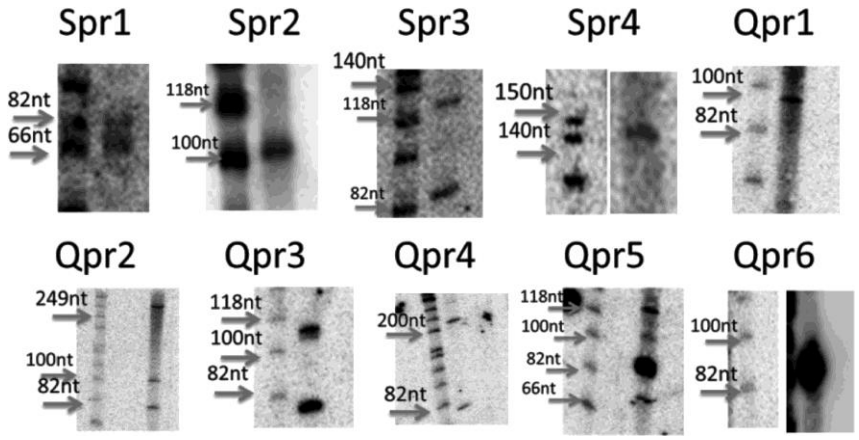


Figure 3.4. Images from Northern blotting analysis for confirmed sRNAs candidates from computational prediction

The total RNA samples were extracted from *D. radiodurans* cell culture at exponential phase (OD_{600} = 1). The expressions of predicted sRNAs were confirmed, and the sizes of the transcripts were estimated relative to the phiX174/Hinfl ladder. The images of the ladder and the sRNA lanes might cut from the same gel in different parts with contrast adjusted to show clearer image.

3.2.6 Differential expression of sRNAs during genome recovery after ionizing irradiation

Following discovery and validation of sRNAs in *D. radiodurans*, we investigated the possibility that these sRNAs were differentially expressed during recovery from high-dose irradiation, as an early indicator of their potential functional importance. For this analysis, we assayed differential expression of all confirmed novel sRNAs following ionizing radiation (15 kGy), relative to sham irradiation controls. A scheme of the ionizing radiation procedure used is shown in Figure 3.5A. To confirm consistency of overall biological trends observed after ionizing radiation [193], [194], we first verified that cells were viable after recovery post-irradiation by making growth curves and

counting colony forming units with plated cells (Figure B.7). We found the dose used in this study is equivalent to 15-18 kGy of gamma radiation [195]. It is worth noting that this initial high radiation exposure was designed to elicit a strong enough response that would allow detectable potential differential expression of sRNAs [76].

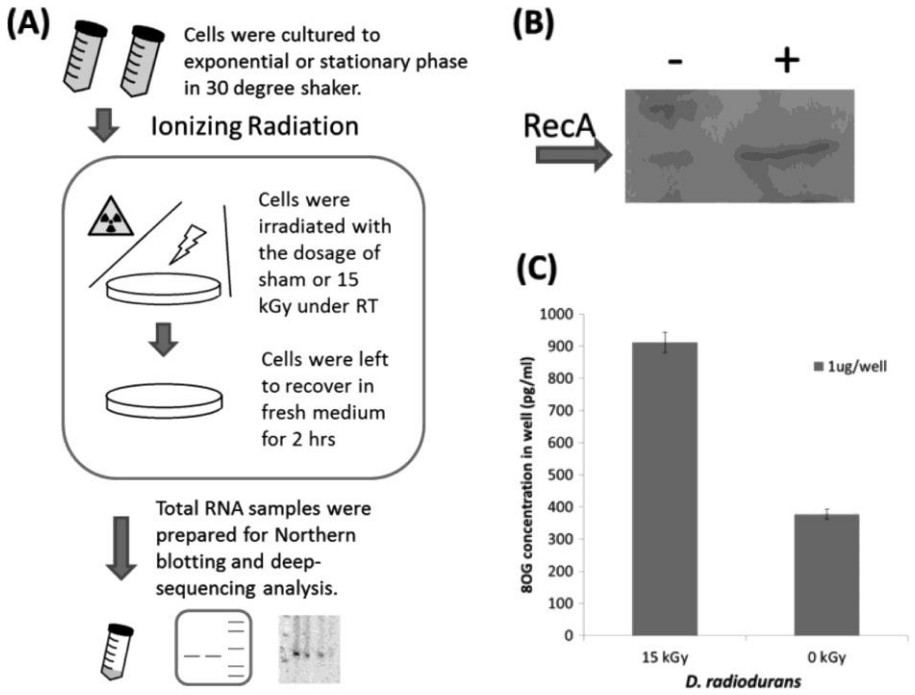


Figure 3.5. Scheme of experimental procedure

(A) *D. radiodurans* R1 (ATCC 13939) cells were cultured in TGY medium to exponential (OD_{600} = 1) or stationary phase (OD_{600} = 3). Cells were then irradiated while cold, and then recovered in fresh medium for 120 minutes. RNA and protein total lysates were prepared for analysis. (B) Western blotting analysis for RecA expression under 15 kGy ionizing radiation. (C) The ELISA test (1ug of RNA were used for each test) quantifies the oxidation damage of the DNA in irradiated samples and control.

Next, to quantitatively verify efficient irradiation, we also probed expression levels of RecA in irradiated samples (relative to non-irradiated samples) via Western blotting analysis (Figure 3.5B). Increase of the RecA protein expression level has been known as an indirect indication of irradiation stresses and this trend was confirmed in our study (Figure 3.5B). Lastly, expected higher nucleic acid damages were confirmed when using the enzyme-linked immunosorbent assay (ELISA) to detect 8-oxoguanine levels (Figure 3.5C). Higher levels of 8-oxoguanine have been previously reported to accumulate in nucleic acids under oxidative stress [181], [196]. Collectively, the above analysis confirmed radiation-induced damage while confirming viability and stress-response.

To test differential sRNA expression during irradiation recovery, we prepared total RNA from sham and 15 kGy irradiated samples from both exponential (OD_{600} = 1) and stationary (OD_{600} = 3) growth phases. These RNA samples were probed with radiolabeled oligonucleotides with complementarity to all identified novel sRNAs. The intensity changes were normalized by the average change of tRNA expression levels, as we assumed that the expression level changes of tRNAs was only due to loading or degradation effects. We found 8 sRNAs that showed differential expression following a 2 hour genome recovery from a 15 kGy irradiation (relative to sham irradiation). Figure 3.6 shows sRNAs that exhibited at least a twofold decrease or increase in band intensity after tRNA normalization (as quantified by GelQuant). Most of them show the same trend both in exponential and stationary phase, while some sRNAs exhibit these trends more obviously in one growth phase relative to the other. sRNA blots that do not suggest differential expression are included in Figure B.8.

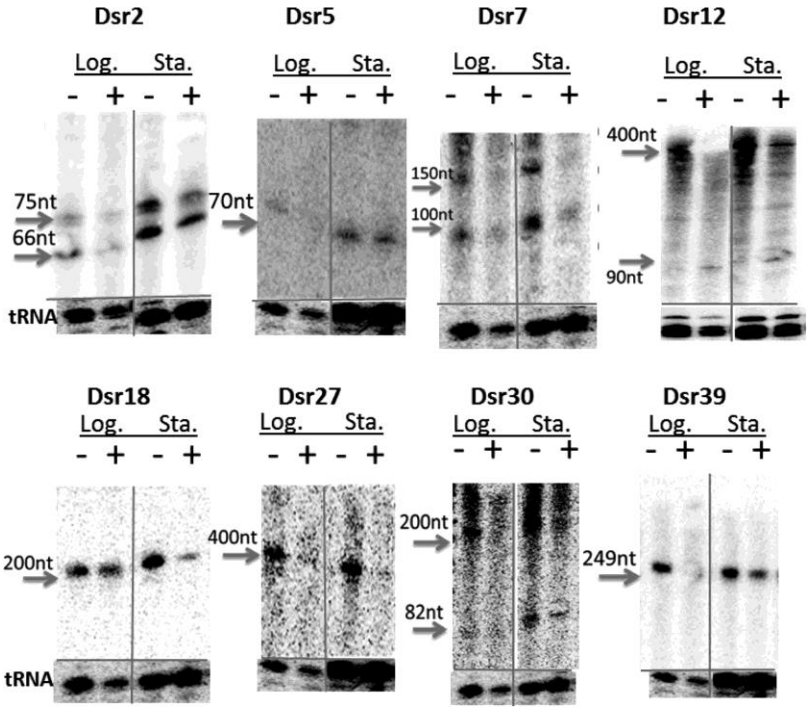


Figure 3.6. Differential expression of selected sRNAs

In each Northern blotting image, The left two lanes are RNA samples from exponential phase (OD_{600} = 1) cells, and the right two lanes are from stationary phase (OD_{600} = 3) cells. The first and third lanes are the control (sham irradiation) RNA samples, and the second and fourth lanes are the 15kGy irradiated RNA samples. The images of first two lanes and last two lanes are cut from the same gel in different parts. The band intensity change of each candidate is normalized to tRNA levels showed at the bottom of each blot. All the blots showed here either have a two-fold decrease in activity after irradiation (Dsr2, Dsr5, Dsr7, Dsr18, Dsr27, Dsr30, Dsr39) or two-fold increase in activity (Dsr12) after irradiation.

3.2.7 Identification of conserved sRNAs in D. geothermalis

To test the conservation level of the sRNA candidates in related radioresistant species, we used BLAST (Basic Logic Alignment Search Tool, NCBI) to

bioinformatically identify homologous sRNAs (E-value < 0.01) in a sample of 6 representative *Deinococcae species* (*D. gobiensis*, *D. proteolyticus*, *D. deserti*, *D. peraridilitoris*, *D. geothermalis*, and *D. maricopensis*). All these species have been shown to be highly resistant to ionizing and UV radiation [197]. We also blasted all sRNAs candidates to other bacterial species to test if these sRNAs (particularly those differentially expressed) were only conserved in *Deinococcus spp*. Only the sRNAs that are conserved in *Deinococcus spp* but not in other bacteria are included in Figure 3.1, 3.3 and 3.7. Most of these sRNAs did not show sequence conservation in other species beyond *Deinococcus spp*.; this does not however rule out the possibility that the function of the identified sRNAs is conserved in other bacteria.

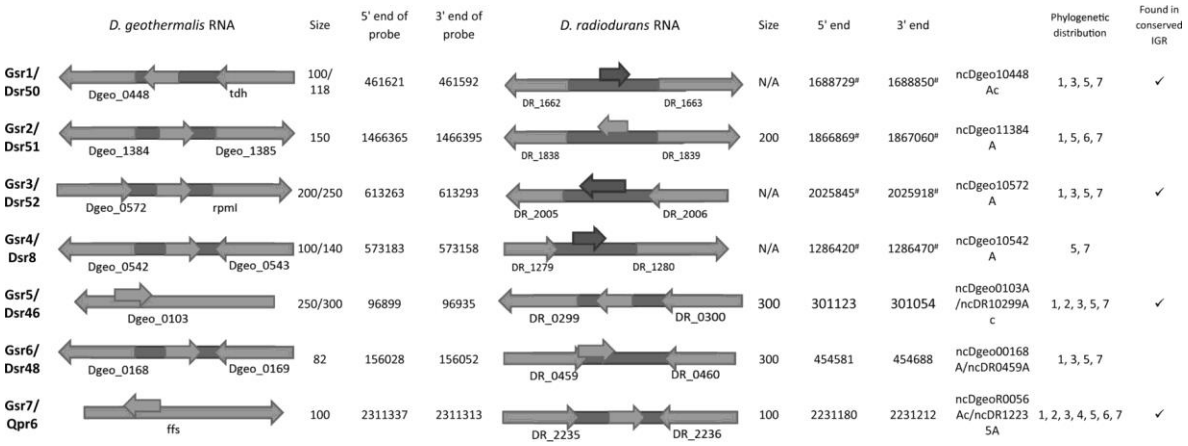


Figure 3.7. Novel small RNA candidates confirmed with Northern blotting analysis in *D. geothermalis* and homologous counterparts in *D. radiodurans*

The middle arrows show where the Northern probes bind. The gray middle arrows are sRNAs that identified with RT-PCR but not with Northern blotting analysis. The arrows on the sides are the annotated flanking protein coding regions. The size of the RNAs was estimated by comparing to phiX174 ladder. The 5' end and 3' end of *D. geothermalis* RNA are not applicable; instead, the coordinates of 5' end and 3' end

of the probe are shown. The phylogenetic distribution shows the species that homologous small RNAs were found: 1. *Deinococcus gobiensis* 2. *Deinococcus proteolyticus* 3. *Deinococcus deserti* 4. *Deinococcus peraridilitoris* 5. *Deinococcus geothermalis* 6. *Deinococcus maricopensis* 7. *Deinococcus swuensis.*

We selected 10 sRNA candidates from the deep sequencing data that were highly conserved in *Deinococcus spp* (in at least two species) from the above analysis, and used Northern blotting analysis to confirm sRNAs expressions in *D. geothermalis and D. radiodurans*. We used BLAST to locate the homologous loci in the genome of *D. geothermalis* and designed complimentary probes that were specific to *D. geothermalis* for Northern blotting analysis. Seven sRNAs were identified in *D. geothermalis* (Figures 7&8). Two of the *D. geothermalis* sRNAs were found in the antisense strand of a coding region (Gsr5, Gsr7), while others are intergenic. All of these candidates were also confirmed in *D. radiodurans* with Northern blotting or RT-PCR, and four of the sRNAs reside in intergenic regions conserved between *D. radiodurans* and *D. geothermalis* (Figure 3.8). We also found Gsr1/Dsr50 and Gsr3/Dsr52 to have downstream and upstream genes with similar functions. It is important to note that the observed difference in sizes of the conserved sRNAs between *D. geothermalis* and *D. radiodurans* is not surprising since it is likely that they are processed differently in the two organisms; these trends have been reported before between homologous sRNAs across species [128].

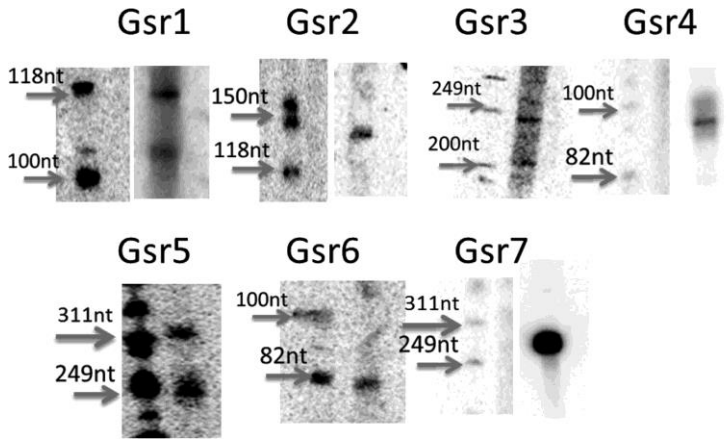


Figure 3.8. Images of Northern blotting analysis for sRNAs candidates with homology to *D. geothermalis*

The RNA samples were extracted from non-irradiated *D. geothermalis* exponential phase (OD_{600} = 1) cell culture and the size was estimated relative to the phiX174/Hinfl ladder.

3.4 DISCUSSION AND CONCLUSIONS

Genomic expression by whole-transcriptome analysis of *D. radiodurans* recovering from acute exposures to 15 kGy was previously reported [198][199]. During the early and mid-phases of recovery, *D. radiodurans* fails to grow, but within this interval hundreds of genes within diverse functional groups are differentially regulated [198]. After an exposure of 15 kGy, about 150 DSBs are inflicted randomly over *D. radioduran's* four genomic partitions [200], followed by extensive exonucleolytic DNA degradation [201]. In acutely irradiated cells, this causes a substantial lowering of the copy number of the more heavily damaged, larger genomic partitions compared to smaller ones in the first hours of recovery, with global levels of RNA expression post-irradiation shown to be inversely related to partition-size. Within the broader context of partition-specific expression, some of the genes were predicted to encode sRNAs. For

example, DRA0234 is 171 nt long, shows no similarity to any protein sequences, and has a transcript that was predicted to form a stable stem–loop structure [198]; previous studies shown DRA0234 was up-regulated very early in recovery, displaying a 12-fold increase in irradiated *D. radiodurans* within the first 1.5 hours of recovery [198]. The expression of DRA0234 was also identified in our deep sequencing analysis (Figure B.9). This gene, and perhaps other similar ones, might encode uncharacterized regulatory sRNAs.

In this study we found 199 potential sRNA transcripts using a combined approach that involved deep-sequencing analysis and computational predictions. We tested 125 candidates by Northern blotting analysis and RT-PCR, and confirmed expression of 41 sRNAs in *D. radiodurans* and 7 sRNAs in *D. geothermalis.* We confirmed that a variety of potential sRNAs could be encoded by this organism, as observed in data acquired by deep sequencing analysis. One interesting example is Dsr12. It overlaps with the 5' end of the *rpsF* gene, which encodes the 30S ribosomal protein S6. Under sham irradiation the size of the Dsr12 transcript approximately equals the length of the identified TSS to the 3' end of *rpsF*, indicating co-transcription of Dsr12 and *rpsF*. However, the signal of the larger transcript becomes weaker following acute irradiation, and a smaller transcript appears. One possibility is that Dsr12 can serve mechanistically as a functional UTR that can change structure under radiation stress and induce downstream RNA degradation or post-transcriptional processing.

Besides Dsr12, most of the sRNAs that show differential expression during recovery from 15 kGy are down-regulated following irradiation. For example, Dsr2 is down-regulated during irradiation recovery. Dsr2 is also predicted to bind the 5' UTR of the *recA* mRNA, which is critical for homologous recombination. Therefore, one potential mechanism of Dsr2 is that under no radiation stress Dsr2 could bind with *recA*

and block translation; on the contrary, under irradiation stress, the downregulation of Dsr2 leads to an increase expressions of *recA*. It is worth noting that, since most sRNAs do not exhibit expression change during irradiation recovery, we suspect that the observed effects of sRNA downregulation are specific to these transcripts and not just a general byproduct of irradiation.

In this work we have also found by BLAST analysis that many sRNA candidates in *D. radiodurans* are conserved in other *Deinococcus spp*. We have experimentally confirmed 7 of them to be expressed in *D. geothermalis* (Figures 3.8 and 3.9)[202], [203]. While generally conserved sRNAs could act as housekeeping regulatory components, exclusively conserved sRNAs in radioresistant *Deinococcus* spp (indicated in Figures 3.1, 3.3 and 3.7) could have unique functionality and play critical roles in radioresistant species. The lack of experimental validation of all sRNAs predicted to be conserved in both organisms could be attributed to the possibility that these are found at much lower levels in one of the two organisms (making it difficult to detect by Northern blotting analysis).

Upon using TargetRNA2 to predict potential sRNA binding targets in *D. radiodurans,* we found predicted mRNA targets that could be functionally related to radioresistance mechanisms (those with an E-value smaller than 0.05 are listed in Table B.6)[204]. Many predicted mRNA targets found in this study are also coding proteins that contribute to radiation survival, such as RecA, RuvA and RadA [181][185][205]. Other mRNAs do not encode for proteins directly associated with DNA repair, but are associated with stress response mechanisms or global gene regulation such as the TetR family transcriptional regulator DR_0074, CRP/FNR family transcriptional regulator DR_0097, and the MerR family transcriptional regulator [206]–[208].

The extraordinary survival of *D. radiodurans* cells exposed to ionizing radiation has been explicated with the hypothesis that the critical proteins for DNA repairing and replication are protected by a mechanism of small-molecule Mn^{2+} antioxidants during irradiation [29], [76], [209]. Proteins in *D. radiodurans* grown in TGY medium and exposed to massive acute doses of gamma-rays (>10 kGy) are shielded from oxidation, preserving the functions of cytoplasmic enzymes as well as the transcriptional and translational potential of the cell [26]. In contrast, naturally radiation-sensitive prokaryotes, such as *E. coli* and *Psudomonas putida*, exposed to doses greater than 1 kGy display extensive global protein oxidation, which inactivates their repair and replication systems, rendering even minor DNA damage lethal. Evidently, the survival of cells exposed to high doses of radiation rests on a functional proteome [76]. Within this context, the amount of genome damage caused per unit length is directly proportional to the dose of radiation. We argue that if regulation of DNA repair genes in *D. radiodurans* is mediated by sRNAs, the small size of sRNA genes (<400 bp) would leave them largely intact compared to protein-encoding genes (~1000-2000 bp) at the outer limits of *D. radiodurans* survival (15 kGy). This might yield functional sRNAs transcribed from genomic DSB fragments before the onset of DNA repair, acting on the pool of existing repair proteins (e.g., RecA) present in the cells prior to irradiation [183], [184]. Our future studies will focus on better understanding the targets of the newly discovered sRNA candidates in this study. Other environmental stresses that cause DNA damage could also be tested, such as desiccation and ultraviolet C light, yielding a more comprehensive view of sRNA regulatory pathways in oxidative stress responses.

3.4 METHODS

3.4.1 Bacterial strains and growth conditions

The strains *Deinococcus radiodurans* R1 (ATCC-13939) and *Deinococcus geothermalis DSM 11300* were cultured according to previous reports [202], [210], [211]. In brief, cells were grown overnight at 30° (*D. radiodurans*) or 37° C (*D. geothermalis*) in TGY broth (1% tryptone/0.1% glucose/0.5%yeast extract) to exponential phase (OD_{600} = 1) or stationary phase (OD_{600} = 3).

3.4.2 Preparation of protein lysate and Western blotting analysis

Cell pellets were harvested by centrifugation, washed with sterilized water and resuspended in a lysis buffer (1 mm Tris-HCl, pH 8.0, containing 1 mm PMSF). The cells were frozen with liquid nitrogen, thawed on ice and lysed by sonication. The cell free extracts were collected by centrifugation and the concentrations were measured with Direct Detect (EMD Millipore). Proteins were resolved by SDS-PAGE using 12% stacking and 5% resolution polyacrylamide gels and immunoblotted according to standard protocols [128]. Anti RecA *Escherichia coli* Rabbit (cosmo bio co., BAM-61-003-EX) and Goat Anti-Rabbit IgG (H+L) HRP Conjugate (Fisher, PR-W4011) were used as primary and secondary antibodies, respectively.

3.4.3 Total RNA extraction

Whole RNA was extracted as previously reported with minor changes [80]. Briefly, cells were pelleted and resuspended in TRIzol reagent (Invitrogen, 15596-026), and lysed using a bead-beater (Bio Spec Products Inc, 3110BX) with four 100s pulses. The top aqueous phase containing RNA was extracted with chloroform:isoamyl (24:1) alcohol and precipitated with isopropanol. The resulting pellet was dissolved in RNAse-free water. RNA concentration was measured by spectrophotometer and stored in

-20° C for short term use. The integrity and purity of total RNAs were verified with spectrometer (OD_{260}/OD_{280}) and RNA gel staining.

3.4.4 Whole transcriptome deep sequencing and data analysis

cDNA libraries were prepared from total RNAs that extracted from irradiated or non-irradiated cells using standard methods [105]. We used NEBNext® Small RNA Libray Prep Set for Illumina® (New England Biolabs Inc. E7330S) to prepare cDNA for all samples with the protocol provided by the manufacturer. The cDNA library was then sequenced with Illumina HiSeq 2000 on one run with 200 cycles. All sequenced reads were trimmed to remove the adapter sequence for mapping to *D. radiodurans R1* genome (NC_001263.1 and NC_001264.1) with Bowtie2 Aligner [190]. For alignment visualization, Integrated Genome Viewer (IGV) was used to identify expression in non-coding regions that could indicate potential sRNA candidates.

3.4.5 Selection of computational predicted candidates

Computationally predicted sRNA candidates were selected from both previous literature reports and from predictions made with SIPHT [97], [192]. In a previous study, 265 potential sRNAs were predicted by QRNA with a comparative algorithm and 127 sRNA candidates were predicted by SIPHT with default parameters [97], [192]. The genome coordinates of all computationally predicted candidates have been included in Table B.6. We used criteria from our previous study to narrow the list of highly potential sRNA candidates for confirmation by Northern blotting analysis[189]. In short, sRNA candidates found in longer and generally conserved intergenic regions (among all bacteria) were selected for Northern blotting analysis. In total, 35 candidates from QRNA prediction and 24 from SIPHT were selected and added to the list compiled from transcriptomics.

3.4.6 Ionizing irradiation

D. radiodurans cells were cultured to exponential (OD_{600} = 1) and stationary (OD_{600} = 3) phase, packed, frozen with dry ice and transported in sterilized plastic bags for irradiation. Exponential and stationary phase were determined by growth curve and tested via spectrophotometer (OD_{600}), as previously mentioned [182]. Samples were thawed at room temperature at the radiation facility before irradiation. These samples were kept cold on wet ice (0°C) while irradiated with a 10 MeV, 18 kW LINAC □ ray source at the National Center for Electron Beam Research, Texas A&M University. Cell samples were subjected to sham and 15 kGy (250Gy/s) exposures. This initial high radiation exposure was designed to elicit a strong enough response that would allow detectable differential expression of sRNAs by Northern blotting [26]. Cells were diluted 4-5 fold to OD_{600} =1 and recovered in fresh culture (TGY) medium for 2 hours at 30° C immediately following irradiation and processed for RNA extraction or stored at -80° C for future analysis. Cell survival rates were measured by plating recovered sham and irradiated cells on TGY plates for colony forming units (CFU) comparison.

3.4.7 Northern blotting analysis

Total RNA was prepared from exponential-phase and stationary-phase *D. radiodurans* cells exposed to 0 kGy and 15 kGy. Northern blotting analysis (performed as previously described)[128] was used to confirm sRNA expression and to evaluate differential expression as a result of irradiation. We used 10% polyacrylamide gel for total RNA electrophoresis (under denaturing conditions), and a total of 10 µg of RNA was loaded on each lane for sampling. Radioactive labeled phiX174 DNA/HinfI ladder (Promega, E3511) was used as a size marker. The separated RNAs were transferred to a positively charged membrane (Hybond N+, GE Life Sciences, RPN119B) and crosslinked with 254nm UV light. PerfectHyb™ Plus Hybridization Buffer (Sigma-

Aldrich, H7033) was used for probe hybridizations over 3 hour incubation at 42° C. Radioactivity was recorded by phosphor storage imaging (Typhoon, GE). The probes were designed to have a complementary sequence towards the target sRNA and radiolabeled with γ-^{32}P by T4 Polynucleotide Kinase (New England Biolabs, M0236S). A full list of the probes is included as Table B.3. Each candidate was tested with probes in the forward and reverse orientation. Each sRNA was experimentally verified only in one direction and at least two sets of probes were used to test each candidate.

3.4.8 RT-PCR and co-transcription

cDNAs were prepared with Super Script III kit (Invitrogen, USA) using the manufacturer's protocol. Forward and reverse primers with 18 to 22 nt long were designed for each sRNA candidate based on RNA-seq data. Primers were also designed to amplify the upstream or downstream ORF by itself or by including the candidate sRNA coding region to test the possibility of sRNA co-transcription with the flanking gene. The PCR reactions were ran with cDNA and a minus reverse transcriptase (RT) control. The annealing temperature was optimized depending on the melting temperature of each primer (Table B.3 for PCR primer sequence).

3.4.9 Deep 5' RACE

Deep 5' RACE was performed as previously described with minor modifications [80]. Deep 5' RACE libraries were pooled and sequenced using an Ion Torrent 316 chip (Wadsworth Center Applied Genomic Technologies Core Facility)[80]. For Deep 5' RACE, sequence reads were identified by the presence of the expected adapter sequence at the 5' end of the transcript. Adapter sequences were removed and reads of >20 nt were mapped to the reference genomes (NC_001263.1 and NC_001264.1) using Bowtie2 [60]. The 5' ends were identified as the farthest position with sequenced reads mapped to from

the 5’ end of the probe. Sequences of all primers used for Deep RACE are listed in Table B.7.

3.4.10 ELISA analysis

The total RNA samples from irradiated cells were analyzed with ELISA as another way to confirm the oxidative stress induced by irradiation. The test was done with OxiSelect™ Oxidative RNA Damage ELISA Kit (Cell Biolabs, Inc. STA-325) using the protocol provided by the manufacturer.

Chapter Four

Identification and characterization of regulatory networks of small RNAs in *D. radiodurans* that contribute to radioresistance

4.1 INTRODUCTION

Unraveling molecular components and mechanisms that are important to genome stability under mutagen stress is important to basic biology of aging, many cancers and neurological diseases. Ionizing radiation (IR) represents an extreme environmental stress that threatens cellular survival by introducing hundreds of DNA double strand breaks (DSBs), thousands of single strand breaks (SSBs), and thousands of mutations to each copy of the genome [212]. Indeed, there is significant interest in deciphering mechanisms that underlie radioresistance given their relevance to anticancer and antioxidant research [69]. *Deinococcus* species comprise a group of bacteria that can survive extreme conditions, such as desiccation, ionizing radiation and oxidative damage; as such, these extremophiles have been ideal models for mechanistic studies of radioresistance. It has been documented that *Deinococcus radiodurans* can survive acute doses of ionizing radiation of up to 12-20 kGy, and can grow under chronic irradiation at dose-rates as high as 60 Gy/h without inducing mutations [26], [28]. These doses are considerably higher than the documented doses of *E. coli*. In addition, *D. radiodurans* can also survive atypical levels of UV irradiation, desiccation and oxidative stresses [213]. The extreme survivability of *D. radiodurans* post ionizing radiation stress is truly remarkable given that ionizing radiation threatens cellular survival by introducing hundreds of DNA double strand breaks (DSBs), thousands of single strand breaks (SSBs), and thousands of mutations to each copy of the genome[212]. Indeed *D. radiodurans* is known for its exceptional ability of DNA double strand break repairing

Most research on *D. radiodurans* has focused on the characterization of DNA repair proteins and on establishing evolutionary relationships to other bacteria. Studies in *D. radiodurans* alone with other radioresistive organisms, such as *Thermococcus gammatolerans* and *Kineococcus radiotolerans*, have suggested that mechanisms of radioresistance in these extremophiles are similar to other stress responses such as heat shock or desiccation [25], [69]. Other studies have supported the absence of unique DNA repair systems in *D. radiodurans* by demonstrating the lack of complexity (and similarity) of DNA repair pathways in *D. radiodurans* as compared to radiation-sensitive species [69]. As the field has deviated from the search of unique genes and pathways, other important aspects of radioresistance have been uncovered.

Over the last 15 years, m of the genes shown to be differentially regulated following high IR-exposures (5-16 kGy), such as SodA or RecA, have been found not to be essential to radioresistance [26], [194], [198]. So far, the only essential aspect that has been established in radioresistant bacteria is that important Mn2+/orthophosphate complexes accumulate and act as scavengers of superoxide to protect the proteome from cellular reactive oxygen species [29], [76]. In this widely accepted model, protective mechanisms based on small-molecule Mn^{2+} antioxidants preserve the high efficiency of DNA repair and replication proteins as well as the transcriptional and translational potential of the cell during irradiation [75]. The notion of general proteome protection is consistent with classical models of radiation toxicity that assert that gamma rays indiscriminately damage cellular macromolecules, primarily indirectly by hydroxyl radicals. Yet, a major gap of the model is our understanding of how cells reconfigure essential metabolic pathways and regulate these responses during IR recovery.

Given the current chronicle of repair physiology, we suspect that response to IR-induced oxidative stress is highly regulated. For instance, it has been now well-

established that metabolic activities are highly ordered immediately after IR exposures: first, DNA replication and protein synthesis are halted; second, chromosomal digestion is activated; and last, damaged nucleotides and amino acids are exported, before replication resumes [77], [78]. Importantly, the halt in cellular activity raises the question of how cells coordinate these responses post-regulation. It is worth noting that mechanisms of energy conservation during stresses (i.e. phage infection, heat shock and oxidative damages) have been reported in other organisms [214]–[216]. For instance, in *E. coli* specific RNA-driven regulatory mechanisms (i.e. *MicF*/OmpF) have been found to reduce protein synthesis under stresses [217], [218]. Furthermore, in *D. radiodurans*, small molecules binding sites have been identified on L11 ribosomal proteins to control protein synthesis [219], but the regulatory mechanisms remained unclear.

Although very little has been done to understand how gene expression is regulated and coordinated post-stress, the contribution of global transcriptional regulators has recently been highlighted. For instance, three naturally up-regulated genes during ionizing radiation recovery are *pprI* (DR_0167, also referred to as IrrE), *ddrA* (DR_0423*)* and *ddrB* (DR_0070*)* [69]. *PprI* has been known to serve as a unique global regulator for DNA repairing pathways in *D. radiodurans* by inducing expression of proteins such as RecA (DR_2340, a key protein for DNA repair by homologous recombination) and PprA (DR_A0346, a DNA repair protein promote non-homologous end-joining pathway) by binding to their promoters during recovery from ionizing irradiation, [71]. Studies have also shown that a cold shock homolog protein PprM (DR_0309), that is likely under the regulation of PprI [73], can also modulate PprA expression. However, further details of this mechanism remain unclear.

Most recently, we have recently uncovered the presence of 24 novel small RNAs (sRNAs) in *D. radiodurans* in a RNA-seq study intended to identify global transcription

patterns (of non-coding RNAs), post IR-exposures [63]; 8 of the sRNAs discovered showed differential expression under acute dosage of ionizing radiation [63]. Small RNAs ranging from 21 to ~400 nt, are an unusual class of RNAs that typically do not encode functional proteins but have intrinsic roles as cellular regulators of transcription and translation [3], [32]. A key property of sRNAs is their ability to simultaneously turn on and off a variety of metabolic pathways in response to environmental signals [87], [165]–[167]. To exert their function, sRNAs can either base-pair with messenger RNAs (mRNAs) to prevent protein translation and promote degradation, or sequester proteins into ribonucleic—protein complexes to prevent their activity [4]. In recent studies, new mechanisms of sRNA functions have been discovered, such as a coupled action with riboswitch elements that are regulated by different ligands [170]. Although previous work in *D. radiodurans* did not lead to the functional characterization of the newly uncovered sRNAs, sRNAs have been demonstrated to be involved in various environmental response mechanisms in other bacteria; some of these include, oxidative stress (*oxyS*), pH stress (*gadY*), and/or anaerobic stress (*FnrS*) [3], [220]. As a result, we hypothesize that sRNAs could be important players in regulating stress responsive pathways induced by radiation.

In this study, we screened sRNAs of interest from last study for their potential roles in radioresistance and find one particular sRNA, Dsr2, which is naturally differentially expressed during recovery from acute ionizing radiation and contributes to the radioresistance of *D. radiodurans* under acute IR exposures and chronic IR exposures. We demonstrate that genetic deletion of Dsr2 reduces survival of *D. radiodurans* by up to 90% under 15 kGy acute radiations and up to 99% under 35 Gy/h chronic exposures. Lastly, we apply a previously published integrated FourD omics approach merged with biochemical experiments to map the gene networks regulated by

Dsr2 [221]. Upon confirmation by qRT-PCR and EMSA approaches, we find that Dsr2 regulates *pprM*, DR_1082 (a ribosome hibernation promotion factor) and many ribosomal protein. Importantly, this work proposes a novel model where sRNA regulation can play a part in a unique pathway in *D. radiodurans*, i.e. PprI regulatory pathway, and may contribute to the post-irradiation translation regulatory *in D. radiodurans*. A significant premise of this model is that sRNAs remain largely functional during irradiation due to their smaller sizes and ability to rapidly fluctuate expression since they are not translated. This represents an important step in beginning to reconcile the multi layers of regulatory mechanisms that are present in *D.radiorudans* during recovery from radiation, all ultimately enabled by the ability to protect the cellular transcription and translational machinery.

4.2 RESULTS

4.2.1 Selection of Dsr2 as a pioneering model to understand regulatory role of sRNAs in radioresistance

Among the 24 novel sRNA transcripts previously identified in *D. radiodurans* [9], Dsr2 represents one of 8 transcripts (all listed in Table C.1) that naturally exhibits differential expression during recovery from 15 kGy ionizing radiation during log and stationary growth phases. The decrease in expression (~up to 50% under 15kGy) of native levels of Dsr2 post- IR exposures is confirmed by Northern blotting and qRT-PCR analyses during recovery conditions post exposures to 0, 5, 10 and 15 kGy (Figure 4.1A/B). It is important to note that the Northern blotting analysis also showed that Dsr2 was intact (i.e. not degraded) after exposure to ionizing irradiation. Based on this observation, we hypothesized that Dsr2 could have an important role in radioresistance. We selected Dsr2 as a model system (from other differentially expressed sRNAs during

radiation recovery) given that this sRNA is relatively highly abundant in cells, easily detectable, and independently transcribed [9]. Moreover, changes in expression of Dsr2 were only observed during IR recovery and not during recovery from any other imposed oxidative stresses (e.g. 0-300 mM H_2O_2 and 5% desiccation, Figure 4.1C). This differential expression pattern was unique to Dsr2; other identified sRNAs (i. e. Dsr1, Dsr9, Dsr11 and Dsr20) did not exhibit differential expression during IR recovery and changed expression upon recovery from 100mM and 50mM hydrogen peroxide; others (i.e. Dsr5, Dsr12 and Dsr39) showed differential expression post exposure to any oxidative stress agent (i.e. IR, H2O2 and desiccation) (Figure C.1). Collectively, these results suggested that Dsr2 was uniquely relevant to native mechanisms of radioresistance in *D. radiodurans.*

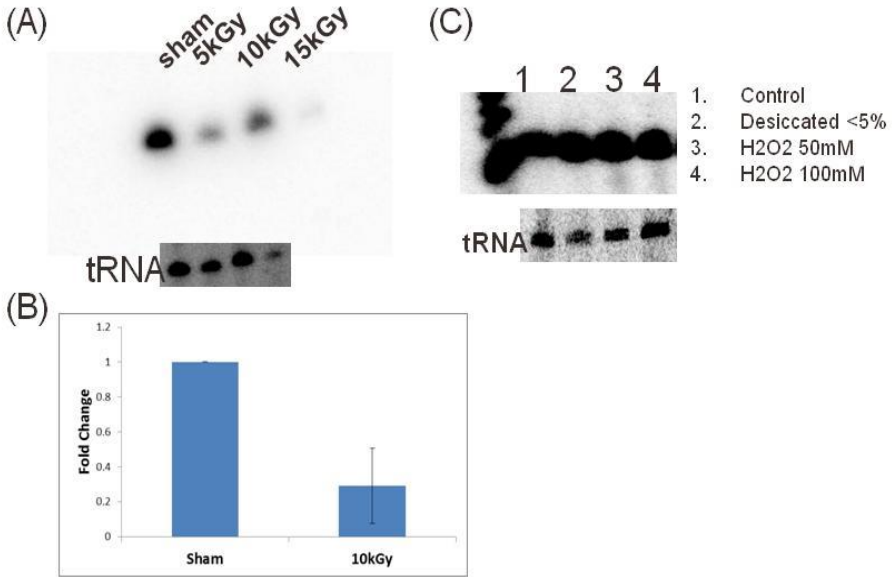


Figure 4.1. Experiments show potential regulatory sRNA Dsr2 has unique response to ionizing irradiation.

(A) Northern blotting analysis of Dsr2 expression level under different dosage of acute ionizing irradiation. After normalized with tRNA level, Dsr2 showed a 2 fold reduction under 15 kGy ionizing irradiation. (B) Quantitative RT-PCR analysis of Dsr2 level in R1 and Dsr2KD under 10kGy acute irradiation. The expression level in R1 is designated as 1. (C) Northern blotting analysis for Dsr2 under different dosage of H2O2 stress and desiccation. The dosages of H2O2 were 0, 100, 200 and 300mM. Cell cultures were dried on plate and incubated in driers (<5% room humidity) for two weeks.

4.2.2 Defective phenotypes of deletion strains confirm Dsr2 as an important contributor to acute and chronic radiation resistance

To investigate the potential direct contribution of Dsr2 to radioresistance, we constructed two isogenic strains of *D. radiodurans* R1: an overexpression strain, harboring plasmid (pRADGro) expressing Dsr2 and a genetic knockdown of Dsr2 (outlined in Figure 4.2A/B). The Dsr2 knockdown was confirmed via genomic PCR (Figure 4.2B). Furthermore, Northern blotting analysis and qPCR confirmed that Dsr2 expression was decreased 10 fold in the knockdown strain (Dsr2KD) (Figure 4.2A/B) and increased 2.5 fold in the overexpression strain (Dsr2OE). Although a complete Dsr2 deletion was attempted (via homologous recombination methods described in Materials and Methods), a homogenous deletion of Dsr2 could not be achieved. Importantly, phenotypic characterization of survival rates showed that Dsr2KD (but not the Dsr2OE strain) exhibited less than 10% survival compared to the wild type post exposure to 15 kGy during exponential phase (Figure 4.2C/D). This represents a stark contrast to other deletion mutants constructed as negative controls (Dsr1KO) which maintained the same survival level as the wild type post IR. Dsr1 was selected as a negative control since it does not show differential expression during recovery from IR and instead shows differential expression under recovery from other oxidative stress (confirmation of the

overexpression and deletion Dsr1 strains by PCR and Northern analysis is shown in Figure 4.2B). To eliminate the possibility that the survival decrease in the Dsr2KD strain was due to unintended disruption of the nearby gene (DR_0141), given the overlap of its 5'end with Dsr2, we conducted a qPCR analysis to evaluate DR_0141 expression in the Dsr2KD strain. As shown in Figure C.2, qPCR analysis confirmed the non-disrupted expression of the pseudogene DR_0141, demonstrating that it was not affected by the genomic insertion at the Dsr2 loci. Importantly, these results suggest a direct phenotypic effect of Dsr2 in *D. radiodurans* radioresistance.

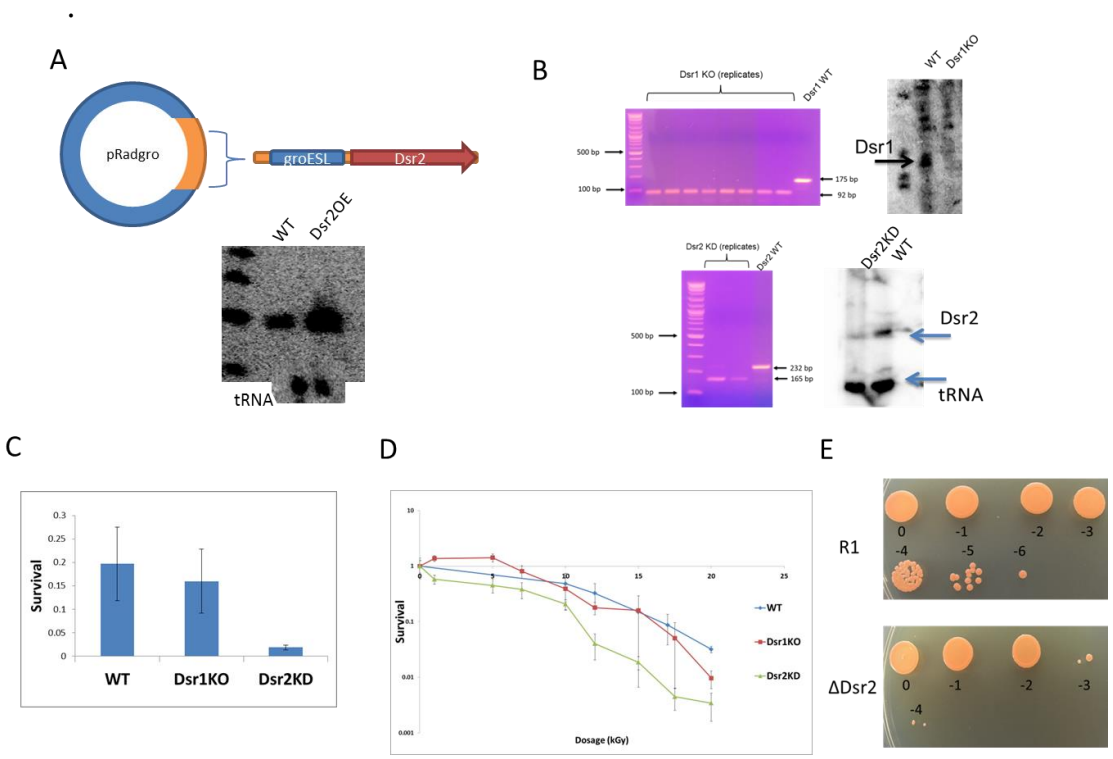


Figure 4.2. Dsr2KD showed a significant reduced survival level under acute and chronic irradiation

(A) Construction of Dsr2 overexpresion with pRADgro. Dsr2 coding sequence was amplified from the genomic DNA and cloned into pRADgro after the groESL promoter. The Northern blotting

analysis showed a increased (~2.5 fold) of Dsr2 expression in the transformed cell. (B) Dsr2 and Dsr1 were removed from the genomic DNA of *D. radiodurans* by homologous recombination. PCR confirmations of the strains showed Dsr1 is homogenously removed from the genome, but Dsr2 was not able to be fully removed. This was further confirmed with Northern blotting analysis. (C) Survival rate of the deletion mutants and wild type under 15kGy acute ionizing radiation (Survival level at sham = 1). (D) Survival curves of deletion mutants and wild type at different doses of ionizing radiation. (E) Survival rate of Dsr2KD and R1 under chronic irradiation of R1 and Dsr2KD at 35Gy/hour for 5 days. Numbers in the images indicate the dilution ratio of colonies.

4.4.3 Genome-wide transcriptional and translational effect of Dsr2

To understand global (direct and indirect) regulatory contributions of Dsr2 to the proteome and transcriptome, we conducted transcriptomics and proteomics analysis using Dsr2KD, Dsr2OE and the R1 strain. These experiments were conducted as biological triplicates under sham and 10kGy ionizing irradiation. A total of around 6,000,000 reads were mapped to the genome for each sample, and reads that mapped to rRNA or tRNA were then excluded in our analysis. All reads were counted and analyzed by DEseq2 algorithm to find statistically significant differential expressed transcripts. Importantly, in the Dsr2KD strain, 2142 genes showed at least a 2 fold increase or decrease in expression relative to the R1 strain (with padj value lower then 0.1) and 93 genes meet the same criteria too when we compared irradiated Dsr2KD to sham condition. Likewise, in the Dsr2OE 2035 genes meet the same criteria and qualified as differential expressed in relative to the R1 strain, and 27 genes meet the same criteria when we compared irradiated Dsr2OE to sham condition; these data is included in Table C.2. GO-term analysis was done for all the differential expressed genes in different samples with online

tool www.geneontology.org [222]. In the R1 under irradiation, we found the genes that were identified as differential expressed in previous literature are also identified in our data, such as RecA, PprA and Ssb (Table C.2) Although not much GO terms are enriched in the Dsr2KD and Dsr2OE at sham condition, we found many stress-response pathways were significantly enriched under irradiation in both Dsr2KD and Dsr2OE compared to sham condition (Figure 4.3A). These GO terms are also enriched in the R1 but only in less stringent criteria (without the Bonferroni correction). We found 28 GO terms enrich in DsrKD and 43 enriched in Dsr2OE. Interestingly, all 28 GO terms in Dsr2KO are all enriched in Dsr2OE. Under sham condition, no significant results were found in Dsr2KD when compared to R1, but protein-synthesis GO terms were enriched in DsrOE (Figure 4.3B). This may suggest the deletion of Dsr2 can be mitigated by other mechanisms.

4.2.4 HITS-CLIPS analysis identifies direct potential Dsr2 targets

To identify potential regulatory targets of Dsr2 and thereby obtain insight into Dsr2 regulatory mechanism, we conducted High-throughput sequencing of the RNAs that were isolated (in association with Dsr2) by crosslinking immunoprecipitation (HITS-CLIP sequencing). For the HITS-CLIP analysis, the MS2 binding sites (MS2BD) were fused to Dsr2 (at the 5'end), as shown in Figure 4.4. These construct was expressed in *D. radiodurans* with pRADgro plasmid, then purified and incubated with MS2 protein for affinity binding. The pull-downed transcripts were then sequenced and map to the genomes of *D. radiodurans* to find Dsr2-associating mRNA targets. As shown in Table C.3, 47 genes were enriched at least 2-fold in samples expressing Dsr2 ($p<0.05$), relative to a negative control that lacked Dsr2, just expressing the MS2BD. As shown in Table C.3, 10 of these genes were ribosomal proteins, 5 were tRNAs, 7 were associated with

stress response, 2 are associated with DNA binding or repairing, 8 are housekeeping proteins and others are uncharacterized proteins.

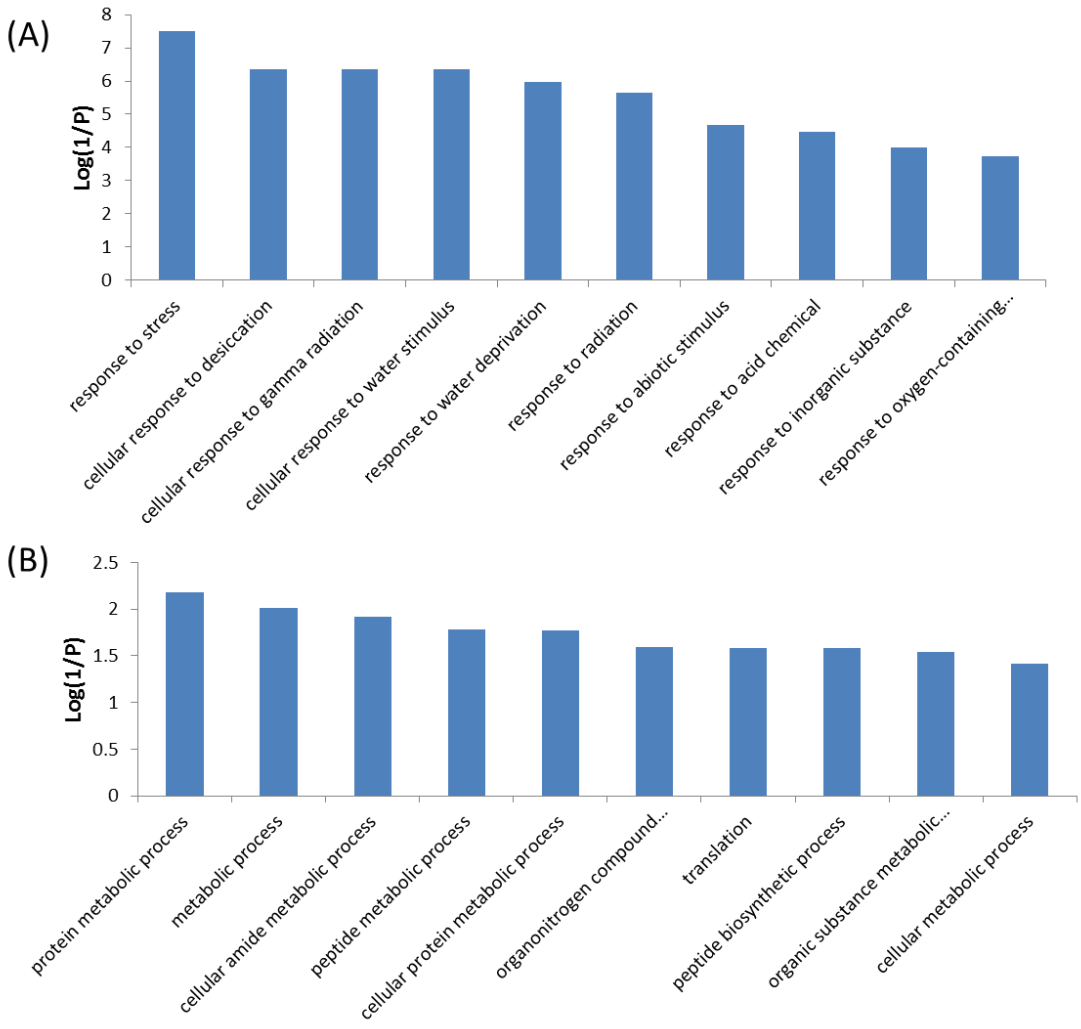


Figure 4.3. GO-term analysis of Dsr2 mutants under irradiation

(A) Significantly enriched GO terms in both Dsr2KD and Dsr2OE under irradiation. P-values were calculated with www.geneontology.org with Bonferroni correction to compare GO-term enrichment

in irradiated Dsr2KD and Dsr2OE compare to sham condition. (B) Significantly enriched GO terms in Dsr2OE compared to R1 under sham condition. P-values were calculated as mentioned above.

As shown by the heat map in Figure 4.4, 13 of the identified 47 potential mRNA targets were up-regulated in the Dsr2 overexpression strain and down-regulated in the knock-down mutant at the transcript level; this is an expected pattern for true direct sRNA targets. Interestingly, 15 mRNAs within this set of 47 exhibited up-regulation in the Dsr2 overexpression strain but a less drastic down-regulation ($p<0.05$ fold < 2) in the Dsr2 knock-down strain; we attributed this to an alternative mechanism that may compensate the deletion of Dsr2 at sham conditions. In general, since many of these genes are related to stresses response and have experimental identified interactions, this analysis suggests Dsr2 acts as a global up-regulator in different pathways at mRNA level.

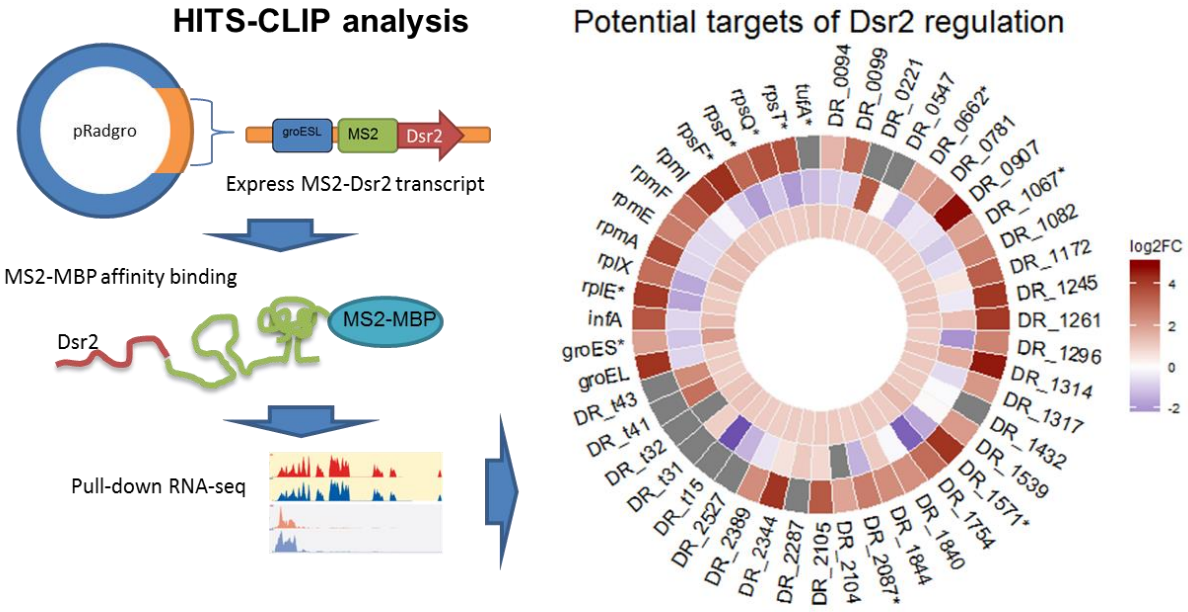


Figure 4.4. HITS-CLIP analysis and differential expression of 47 pull-down mRNA

MS2 binding sites (MS2BD) were conjoined with Dsr2 and cloned with pRADgro to express MS2BD-Dsr2 transcripts in vivo. The transcripts were pull-downed by MS2 protein and extracted for RNA high-throughput sequencing. The results were analyzed by DEseq2 and 47 potential targets were found. Inner circle: enriched folds of mRNAs in the MS2BD-Dsr2 pull down transcripts. Middle circle: fold changes of mRNA transcripts in Dsr2KD compared with wild type *D. radiodurans*. Outer circle: fold changes of mRNA transcripts in Dsr2OE compared with wild type *D. radiodurans*. *: Candidates that were up-regulated in Dsr2OE and down-regulated in Dsr2KD.

4.2.5 EMSA Analysis shows that Dsr2 directly binds 14 mRNA targets involved in PprM and translation regulation *in-vitro*

We then selected the top half of them alone with others that have functions of interest for electrophoretic mobility shift assay (EMSA). In total, we have tested 30 candidates for EMSA analysis. We found 14 of them are likely to have direct binding with Dsr2 sRNA (Fig 4.5). Among the 14 identified targets, three are ribosomal proteins (*rplE, rpml, rpsP*), three are translational initiation factors or elongation factor (*infA, DR_2087, DR_0309*), a stress response protein (*DR_0907*), an ABC transporter-binding protein (*DR_1571*), a ribosome hpf protein (*DR_1082*) and five uncharacterized proteins (*DR_2105, DR_1067, DR_2104, DR_0094, DR_1261*). Although they are not in the same metabolic pathway, most of them are associated with translation mechanism, and some of them are involved in the PprI regulatory mechanism. In addition, genes that were not verified binding by EMSA could still play a role in the regulatory pathway, since many of them have confirmed interactions with the 14 targets, such as *DR_0099* or *groES/groEL* [223], [224]. The EMSA analysis also suggested that among the 14 verified binding targets, some of them have a higher binding affinity to the Dsr2 sRNA, such as DR_1067, DR_1571 and DR_2104 (Figure 4.5).

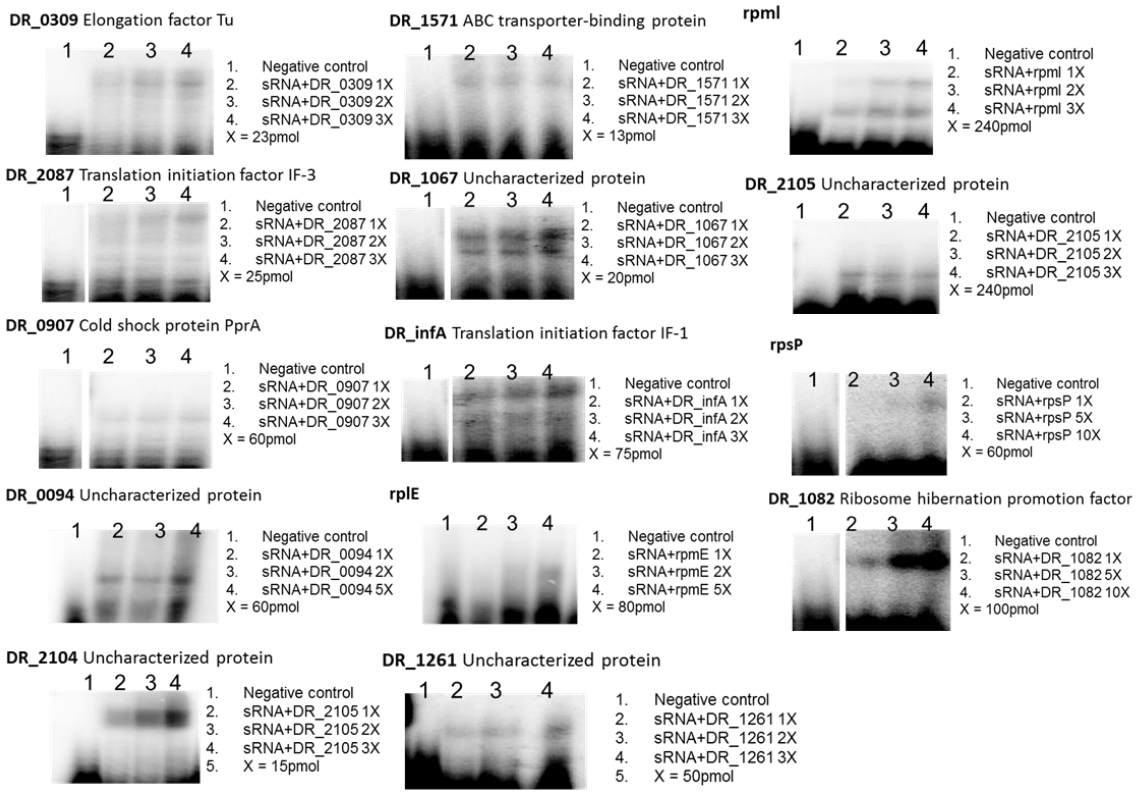


Figure 4.5. *In-vitro* interaction of Dsr2-mRNAs were verified by EMSA for 14 potential targets

5pmol of radiolabeled Dsr2 sRNA was incubated with mRNA binding fragments at 70 degree for 10 minutes and then 37 degree for an hour before EMSA. A sRNA only negative control was used to compare with each candidate. Concentration titrations were also used for each samples with 1, 2 and 3x, 1, 2, 5x or 1, 5 and 10X.

Among the 14 targets verified with EMSA, 6 of them are down-regulated in Dsr2KD and up-regulated in Dsr2OE. This pattern is statistically significant enriched (Fisher's exact test, $p<0.05$) compared to the overall genes that were differentially expressed, which have 166 out of 1474 showing this pattern (Figure C.3). While 6 out of 14 EMSA verifies targets showed a more significant fold change, most of the others were

also following the same trend but less drastic. The 6 genes that show at least a 2 fold reduced change in the Dsr2 knockdown and 2-fold up-regulated change in the overexpression mutants are *tufA*, *rpsP*, *rplE*, *DR_2087*, *DR_1067*, and *DR_1571*. In addition, 5 genes have less fold changes (ranges from a 1.1 to 1.96 reduced fold) in the knockdown strain are *DR_0094, DR_0907 (pprM), rpml, infA* and *DR_1082*. To see how mRNAs of interest response under irradiation, we used qPCR to compare the expression levels between sham and 10kGy irradiated samples of *D. radiodurans* R1. The results showed that most of the transcripts were down-regulated, while *infA* and DR_1571 is up-regulated (Figure C.4). So far the proteomic data does not provide much significant results. The proteomic data in previous studies does not overlap much with our 47 candidates. DR_1571 was identified in a previous proteomic study [225] but no significant changes were observed. Our proteomic analysis was able to identify 25 of the 47 candidates and 10 of the 14 targets. However, very little statistical significant changes were able to verify. Only 4 genes showed up-regulation and two showed down-regulation in the Dsr2OE with is not entirely consistent with transcriptome data (Table C.4).

A molecular analysis of Dsr2-mRNA binding was performed using IntaRNA. As shown in Figure 4.6 and Figure C.5. Most of the mRNA targets have predicted binding sites at their 5' UTR with Dsr2, while some of them are predicted to bind with Dsr2 downstream of their start codon. This could suggest two possible mechanism of Dsr2 regulation, but the transcriptome data does correlated with this prediction. Two mRNA target are predicted to have lower predicted energy (*rpmI* and *rplE*) and two are predicted to have higher binding energy than others (*rpsP* and *DR_2087*), and this is consistent with the EMSA gel results. It is worth mentioning that we initially performed target prediction analysis using TargetRNA2 algorithm and found *RecA* and *DR_0144* are predicted to bind with Dsr2 [9], but further evidence was not found in this study. The

target prediction was performed again with CopraRNA which uses sRNA homologs for target prediction [226]. This algorithm yielded 100 top target candidates and 2 of them are within our 47 pull-down target list (DR_0907 *pprM* and DR_0781) (Table C.5). It is interesting that *pprM* was identified since it can regulate PprA protein which induces DNA end repair mechanism. Other predicted stress response protein or DNA repair protein includes *dps1* and *mutS* but not identified experimentally in this study.

(A)

		FC in Dsr2KD	FC in Dsr2OE	binding site	energy
PprI pathway related					
DR_0907	Cold shock protein CSD family	0.66243125	29.80561693	130/-8	-6.23
DR_1082	Ribosome hibernation promotion factor	0.671873258	6.306792002	80	-6.7
Ribosomal protein or translation factors					
infA	Translation initiation factor IF-1	0.579617634	12.05759497	186/-5	-8.43
rplE	50S ribosomal protein L5	0.307350023	18.0110947	60	-3.5
rpmI	50S ribosomal protein L35	0.908993211	17.57081014	50	-3.2
rpsP	30S ribosomal protein S16	0.275315639	9.306632321	150/-25	-10.11
tufA	Elongation factor Tu	0.410524749	12.45959268	120/80	-6.39
Uncharacterized protein and other					
DR_2087	Translation initiation factor IF-3	0.363792272	6.793935068	-30	-11.7
DR_1571	ABC transporter-binding protein	0.324097764	19.63564641	35	-9.63
DR_0094	Uncharacterized protein	0.519350338	2.951565135	-30	-6.08
DR_1261	Uncharacterized protein	2.522846694	18.35268786	120	-5.62
DR_2104	Uncharacterized protein	0.868886051	3.764320342	160/25	-7.95
DR_2105	Uncharacterized protein	1.742728	11.37229147	-10	-6.02
DR_1067	Uncharacterized protein	0.496191307	3.769484823	100	-5.49

(B)

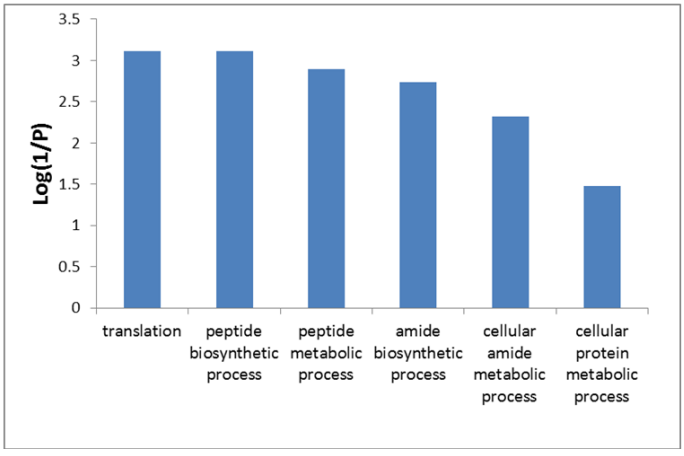


Figure 4.6. Go-term analysis and prediction binding sites of the 14 targets

The GO (gene ontology) terms of the 14 EMSA verified targets were analysis by online tool at www.geneontology.org to find enriched GO terms. 6 translation-related GO terms were found to be significantly enriched with Bonferroni correction. The binding sites (the numbers refer to basepairs from the start codon of mRNA) for mRNA-Dsr2 were predicted by IntaRNA (unit of energy: kcal/mol). FC in Dsr2KO: fold changes of genes in Dsr2KO compared to R1 at sham condition, number in red does not have significant p-value. FC in Dsr2OE: fold changes of genes in Dsr2OE compared to R1 at sham condition.

4.2.6 Go-term analysis of Dsr2 targets reveals additional layers to radiation resistance regulation in D. radiodurans

We also performed the gene-ontology term (GO term) analysis for the verified targets (Fig 4.6). All 14 genes were input into the online tool Panther (www.geneonbtology.org) [222] and two conditions were used to calculate the p-value of GO term enrichment. Without the Bonferroni correction, we have 49 GO terms enriched, and 5 GO terms were enriched with Bonferroni correction. Within the 49 GO terms, 33 are associated with translation or translation regulation, and other are associated with ribosomal assemble. The 5 stringently enriched GO terms are all associated with translation bioprocess, such as peptide biosynthetic process peptide metabolic process and cellular amide metabolic process (FIG 4.6). This alone with the transcriptome data suggest Dsr2 could have a critical role in the translation regulation that associated with radioresistance.

The above analysis suggests at least three different possible regulatory mechanisms of Dsr2 sRNA. First, *pprM* which encodes a cold shock protein homolog is verified as an in-vitro binding target of Dsr2 and the transcriptome data shows it is up-regulated by Dsr2. PprM was identified in previous studies as a PprA inhibitor, and its function is depend to PprI [227], [228]. PprI leads to a unique regulatory pathway in *D.*

radiodurans, and it would be interesting to show how Dsr2 contribute to the radioresistance of *D. radiodurans* by interacting or assisting this system. In addition, many translational proteins were also verified as *in-vitro* targets of Dsr2, including ribosomal proteins and translation initiation factors. This may suggest a mechanism that translation under irradiation can be regulated by sRNAs in *D. radiodurans*. The proteome of *D. radiodurans* are well-protected under irradiation, thus enable transcriptome machinery to function under stresses. Since our Northern blotting analysis showed Dsr2 was not degraded under irradiation (Figure 4.1), this supports that small regulatory elements such as sRNAs could be protected alone with this system or be rapidly produced under irradiation (Figure 4.7).

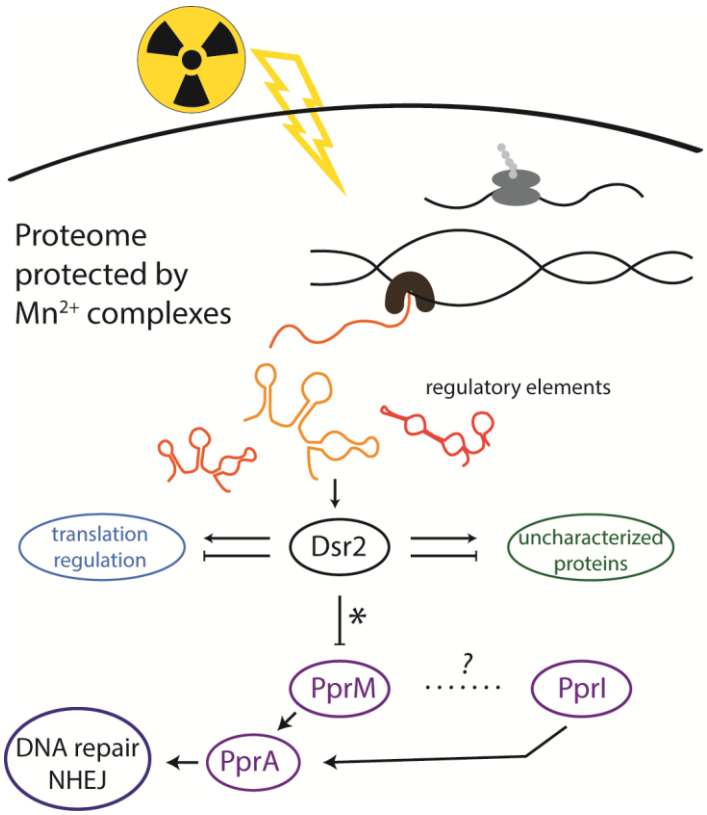


Figure 4.7 Dsr2 as a global regulator in radioresistance pathways

In this study we proposed Dsr2 as a stress response global regulator in *D. radiodurans* that at least participates in two pathways of regulation. DR_0907 mRNA (*pprM*) was confirmed as a direct binding target of Dsr2, and is up-regulated in Dsr2OE and down-regulated in Dsr2KD, which suggesting a possible novel regulatory mechanism. PprM was identified as a PprA inhibitor, which can induce DNA non-homologous end-joint repair (NHEJ). The GO-term analysis also showed that Dsr2 could modulate multiple targets in the translation mechanism and regulation post-stress translation. Our Northern blotting analysis of Dsr2 under irradiation suggests sRNAs under irradiation may be protected alone with proteome and trigger regulatory mechanisms. *: The mechanism of how Dsr2 modulate the PprM protein activity is unclear. ?: previous study of PprM showed its activity is depend on PprI, but the detail of regulation is unclear.

4.3 DISCUSSION

In this study, we examined the regulatory sRNA candidates from our previous study [9] to investigate the radioresistance mechanisms of *D. radiodurans*. We created mutants that can overexpress the sRNAs or have sRNAs deleted. We found the Dsr2 sRNA knock-down strain shows a lower survival rate and performed HITS-CLIP, transcriptome analysis and EMSA to identify the regulatory targets of Dsr2 sRNA. The EMSA analysis shows that 14 mRNA transcripts are able to interact with Dsr2 sRNA and could be important in the radioresistance mechanisms. We also found the 14 targets could be roughly grouped into three categories: translational factors or ribosomal proteins, PprM and associated proteins and uncharacterized proteins.

The most interesting group is the PprM and associated proteins. PprM is a modulator that can suppress PprA expression, which is also regulated by PprI protein [70], [227], [228]. The mutant *pprM* strain has become irradiation sensitive, suggesting the

critical role of PprM in radioresistance [228], [229]. PprM is also involved in many other regulatory mechanisms, such as the expression of *katE1* and desiccation and oxidative resistance [230]. PprM is a homolog of cold shock protein (Csp) but previous studies did not provide clean evidence of how PprM is regulated. Airo et al. showed that PprM is induced not by cold shock but heat shock [231], but this result were unable to be reproduced in another similar study [232]. In this study we found Dsr2 is able to interact with *pprM* and presumably modulate its expression level. The transcriptome data in this study showed that *pprM* is down-regulated in Dsr2KD and up-regulated in Dsr2OE, and two binding sites were predicted at the 130/-8 bp from the start codon of *pprM* (Figure 4.6). This suggests that Dsr2 could interact with the 5' UTR region of *pprM* to induce its expression. Our transciptome data also showed that the *pprA* mRNA is up-regulated in Dsr2KD, and this is consistent with the known regulatory pathway of *pprA*, which is suppressed by PprM. Recent studies also found a novel element DrRRA that cooperates with PprI in the radioresistance mechanism [233], and our transcriptome showed that *DrRRA* is up-regulated in Dsr2OE but no significant changes in Dsr2KD; on the other hand, pprI is down-regulated in Dsr2OE but no significant changes in Dsr2KD. These might suggest an indirect interaction between Dsr2 sRNA and *pprI* and *drRRA*. Interestingly, *Dr_1082*, a regulatory target of PprI, is also targeted by Dsr2 sRNA (Figure 4.6). *DR_1082* encodes a ribosome hibernation promotion factor and is regulated by PprI under irradiation, but it is the only one that does not predicted to be constitutively expressed among all the PprI targets. *DR_1082* is down-regulated in Dsr2KO and up-regulated in Dsr2OE, and has a predicted binding site of Dsr2 at 80bp from its start codon (Figure 4.6).

The second group is the ribosomal proteins and translation factors. Three transcripts that encode ribosomal protein (*rpsP*, *rplE*, *rpml*) were verified as Dsr2 targets

in this study, alone with two initiation factors (*infA, infC*) and a subunit of EF-Tu (*tufA*). Although ribosomal proteins are not common regulated targets in the radioresistance mechanisms of bacteria, some ribosomal proteins are involved in the stresses response pathways of eukaryote. For example, ribosomal S6 kinase can phosphorylate S6 and stimulate translation, and it is a target of the protein kinase mTOR, which is inhibited under ionizing irradiation in human cell [228]. In bacteria, evidences also show that ribosomal proteins or translation pathways can be regulated by small molecules or cellular regulators [219]. In our three ribosomal protein targets, *rpsP* has a much higher predicted binding affinity with Dsr2 sRNA, while the other two have the lowest of all 14 targets. It's interesting that RpsP is a small subunit (S16) and the other two are large subunit (L5 and L35), and this might imply that there are two different mechanism of Dsr2 on two different size of ribosomal proteins (Figure 4.6). The translation initiation factors InfA and InfC have been known to be regulated under cold shock [234]. In *E. coli*, *infA* and *infC* has two promoters and the one that is frequently used is induced by irradiation while the other is not [234]. In our study, *infA* has a strong predicted Dsr2 binding site in the upstream region, while *infC* has a strong predicted binding site but at the downstream of the start codon. Although their differential expression patterns are the same in the Dsr2 mutants (up-regulated in the Dsr2OE and down-regulated in Dsr2KD), further study is needed to verify the mechanisms of *infA* and *infC* with Dsr2 sRNA and how they are induced by cold shock and irradiation. *TufA* also has two predicted Dsr2 binding sites at upstream region, and has the same differential expression patterns as *infA* and *infC*. In conclude the GO-term analysis of the EMSA identified targets showed that translation related proteins are significantly enriched, and many of them can be involved in the pathway of radioresistance mechanisms. This provides evidence that energy

conservation and translation regulation could be important in *D. radiodurans* and is under the regulation of ncRNAs.

The third group is uncharacterized proteins. Although most of them do not have a known function, one of them (*DR_1261*) is predicted to play a role in the regulatory network of radioresistance in *D. radiodurans* with microarray data and bioinformatics algorithms [235]. Other interesting targets in this group are *DR_2105* and *DR_2104*. *DR_2105* is at the upstream of *DR_2104*, and one of the predicted binding sites of *DR_2104* is within the open reading frame of *DR_2105*. As a result, we also used the region around this prediction site for EMSA analysis. EMSA data verified that Dsr2 binds to two sites at the upstream of *DR_2104:* one is in the intergenic region of *DR_2104* and *DR_2105*, and the other is the one within the ORF of *DR_2105*. These two sites might play different roles under different conditions, but more evidence is needed to support they role in radioresistance mechanisms.

In summary, we have verified a novel regulatory transcript in *D. radiodurans* that contributes to the radioresistance mechanisms. The down-regulated of Dsr2 leads to a lower survival rate of *D. radioduran* under irradiation, and 14 transcripts are likely to have direct interaction with Dsr2. We propose Dsr2 as a global regulator that at least involves in two different pathways: PprI regulatory mechanism and translation regulation (Figure 4.7). The induction mechanism of Dsr2 and the detail of how Dsr2 interacts with its target are unclear and require more evidence to decipher the ncRNA regulatory network in *D. radiodurans*.

4.4. METHODS

4.4.1 Bacterial growth conditions and stresses induction.

Deinococcus radiodurans strain R1 (ATCC 13939) were grown according to methods reported previously [9]. In summary, *D. radiodurans* were in TGY broth (1% tryptone–0.1% glucose–0.5% yeast extract) at 33 °C to exponential phase (optical density at 600 nm [OD600] = 1). The cells were then packed and we followed the irradiation protocol that was reported previously [9]. In short, exponential phase cells (optical density at 600 nm [OD600] = 1) were packed and irradiated on ice with a 10-MeV, 18-kW linear accelerator (LINAC) β-ray source at the National Center for Electron Beam Research, Texas A&M University. Cells were subjected to dosages range from 0 to 15 kGy (250-Gy/s) exposures. Irradiated samples were recovered in fresh culture (TGY) medium for 2 h at 33°C immediately following irradiation and used for RNA preparation or plating for survival rate measurement which is discussed in flowing sections. The cells were also irradiated chronically at 35Gy/hour for 5 days and then were plated to compare survival levels. For hydrogen peroxide induction, cells were grown to exponential phase (optical density at 600 nm [OD600] = 1) as above and induced with 0-300mM of hydrogen peroxide at 4°C for an hour. The cell pellets were then collected and used for RNA extraction. For cell desiccation, exponential phase cell cultures (~10ml) were dried on plates and incubated in a drier (<5% room humidity) for two weeks. Cells were then washed from the plate for RNA extraction.

4.4.2 RNA extraction and Northern blotting

Total RNA extraction and Northern blotting analysis were performed as previously described [9]. The cells were collected and suspended with 1ml TRIzol reagent (catalog number 15596-026; Invitrogen) and lysed with beads-beating (catalof

numbe r3110BX, Bio Spec Products Inc.). Total RNA was then extracted with 300ul chloroform-isoamyl alcohol (24:1), precipitated with 300ul isopropanol overnight under -20°C and dissolved with 30ul nuclease-free water. RNA samples were loaded into a 10% polyacrylamide gel for total RNA electrophoresis under denaturing conditions and then were transferred onto a positively charged membrane (Hybond N+, catalog number RPN119B; GE Life Sciences) in TBE buffer for 18 hours. The Dsr2 probe (5' TTTGCGGTCTCCTGGGAGTGT 3') was designed and radiolabeled with γ-^{32}P for hybridization. All the oligonucleotides that were used in this study are listed in Table C.6.

4.4.3 Construction of gene overexpression and deletion strains in *D. radiodurans*.

We developed two types of mutants (overexpressions and deletions) of *D. radiodurans* to manipulate transcript expression levels. For overexpression strains, Dsr2 was amplified with genomic PCR (see Table C.6 for all the oligonucleotides that were used for strain construction), purified with PCR purification kit (illustra GFX PCR DNA and Gel Band Purification Kits, GE Healthcare 28-9034-70) and cloned into the Xbal site of pRADgro plasmid [236] then transformed into *D. radiodurans* using a previously reported protocol [237]. *D. radiodurans* R1 were grown to late log phase (optical density at 600 nm [OD600] = 1) and mixed with 30 mM $CaCl_2$ (J. T. Baker) and 10% glycerol (Sigma-Aldrich) to gain competence. Cells were there incubated on ice with plasmid DNA for 30 mins and then at 30°C for an hour. Fresh TGY medium was then added to transformed cells for overnight incubation and plated on chloramphenicol plates (3.5ug/ml) for selection. The sRNA overexpression strains were confirmed with Northern blotting analysis (Figure 4.1A). An empty vector of pRADgro (no sRNA cloning) was also transformed to make the pRADgroEV strain. To delete sRNA coding regions from the genomes of *D. radiodurans*, we use a suicide plasmid to introduce the desired

interruption by homologous recombination that was previously reported [237]. The deletions were confirmed with both genomic PCR and Northern blotting analysis (Figure 4.1B). Two deletion strains, Dsr2KD and Dsr1KO, were constructed in this study. The coordinates of the 5' end and 3' end deleted sRNA sequences were listed in Table C.1. The homogeneous deletion of Dsr2 was attempted but unable to accomplish. Thus, a knockdown (Dsr2KD) strain of Dsr2 was made instead of knockout. On the other hand, Dsr1 was removed homogeneously from *D. radiodurans* to construct Dsr1KO.

4.4.4 Survival rate measurements

After irradiation and recovery, 1ml of cell cultures of sham, 5, 7, 10, 12 and 15kGy irradiated samples were diluted to 10^{-3}, 10^{-4}, 10^{-5} and 10^{-6} folds with serial dilution. 100ul of biological triplicates of each sample with at least 3 different dilutions were then plated on TGY plates with proper antibiotics respectively. For R1, Dsr1KO and Dsr2KD strain, no antibiotics was used. For Dsr2OE and pRADgroEV, 3.5 ug/ml of chloramphenicol was added for selection.. For low dosage samples (<10kGy), dilution samples of 10^{-4}, 10^{-5} and 10^{-6} were used for plating, and dilution samples of 10^{-3}, 10^{-4}, 10^{-5} were used for plating for high dosage samples (>=10kGy). Multiple dilutions were used to reduce deviation and technique errors of plating. Plates were incubated for at least two days and then CFU were counted.

4.4.5 HITS-CLIP analysis (MS2 pull-down analysis)

Pull-down analysis was performed according to a protocol published previously [238]. Briefly, the sequences of MS2 protein binding sites (MS2BD) were added to the 5' end of the *Dsr2* sequence and cloned into *D. radiodurans* with pRADgro plasmid [237]. Full construct in the context of pRADgro plasmid is included in Table C.6. The expression of *Msr2BD-Dsr2* transcript was confirmed with Northern blotting analysis

(Figure C.6). Mutants were cultured to exponential phase (optical density at 600 nm [OD600] = 1) and collected for total RNA extraction. Total RNA was extracted as mentioned above with one difference: in addition to 500ul isopropanol, 10ul GlycoBlue™ was also used to precipitate RNAs. A fusion of MS2-MBP protein [238] that used as affinity tag was expressed in *E. coli* for purification. For purification, 100ml of the cells were culture to OD_{600} = 0.5 and induced with 1mM IPTG for 4 hours. Up to 100ml cells were then collected and resuspended in 10ml column buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, 10 mM β–mercaptoethanol pH7.4). After sonication in a 10ml conical tube on ice (Branson Sonifier 250), Cells were treated with DNAse for 1hr at 4°C and supernatants was collected after centrifuging. To extract MS2-MBP from lysates, 100ul amylose magnetic beads (New England Biolabs, E8035S) were suspended in 500ul lysates and incubated for 2-3 hrs at 4 °C. A magnet (Thermofisher scientific, Magjack rack) was then applied for 3 minutes and beads were washed with 1ml wash buffer (column buffer + 0.1mM maltose) after supernatants were removed. MS2-MBP protein was eluted with 50ul elution buffer (column buffer + 10mM maltose). 2ug of MBP protein were incubated with 100ul of total RNAs (~1ug/ul?), containing MS2BD-Dsr2 transcripts, for 1 hr at 4°C. This mixture was then incubated with amylose magnetic beads for another 2 hrs and supernatants were removed after magnets were applied. Beads were washed three times with wash buffer and the *MS2BD-Dsr2* MBP protein complexes were eluted. Total RNA was precipitated with equal volume of isopropanol and10 ul GlycoBlue™ overnight at -20°C, washed with 1ml cold 75% ethanol and resuspended in 20ul nuclease-free water.

4.4.6 Transcriptome analysis

The cDNA libraries were prepared from total RNAs using NEBNext® RNA First Strand Synthesis Module (NEB E7525L) and NEBNext® DNA Library Prep Master Mix Set for Illumina® (NEB E6040L). cDNA libraries were then analyzed with Illumina NextSeq 500 single-end platform. All sequenced reads were trimmed to remove the adapter sequence for mapping to the *D. radiodurans R1 genome* (GenBank accession numbers NC_001263.1 and NC_001264.1) with Bowtie2 Aligner. Differential expression of genes was normalized and calculated by DEseq2 algorithm [239]. Genes with expression level changes more than 2 folds increased or reduced with padj (adjusted p-value) < 0.1 would be considered as differentially expressed.

4.4.7 Electrophoretic mobility shift assay

Electrophoretic mobility shift assay (EMSA) was done to verify sRNA-mRNA binding. *Dsr2* sRNA was synthesized by IDT (Integrated DNA Technology, USA), and the coding sequencing of mRNAs were also synthesized; this is detailed in Table C.7, where all sequences of the mRNAs are included. Dsr2 was treated with alkaline phosphatase at 37°C for 1 hour and then radiolabeled with γ-^{32}P using T4 Polynucleotide Kinase (New England Biolabs, M0201) (1ul of T4 Polynucleotide Kinase was incubated with 100ng of Dsr2 and buffer at 37°C for 1 hour). A T7 promoter was added at the 5'end of the selected mRNA fragments and synthesized by IDT, and then were transcribed by MEGAscript T7 Transcription Kit (ThermoFisher Scientific, AM1333), following the manufacturers protocol. Instead of full coding sequence of mRNAs, only the upstream 200 and downstream 100 base-pair of the start codon were selected for T7 *in vitro* expression. The sRNA and target mRNAs were mixed (5-10 pmol of sRNA and at least 15 pmol of mRNA) and incubated at 70°C for 15 minutes and then at 37°C for 45 minutes. 20ul of each of these mixtures were then loaded on a non-denaturing 5%

polyacrylamide gel with running buffer TBE. Gels then were dried (Biorad, Gel Drier 583) and imaged (GE, Typhoon).

4.4.8 Proteomic Identification of Dsr2 Targets

D. radiodurans R1 (wild type), *D. radiodurans* R1 with a knock-down version of Dsr2 (10 folds of Dsr2 reduction is observed) (Dsr2KD) and *D. radiodurans* R1 harboring an overexpression plasmid of Dsr2, where 2.5 folds of Dsr2 increase in expression is observed (Dsr2OE) cultures were grown to an OD 600 of ~1.5 (50 mL cultures) and incubated at 4˚C for one hour before collecting the cell pellet; these conditions mimicked the conditions of irradiated *D. radiodurans* samples. The cell pellet was stored at -80˚C until lysis step. To obtain protein lysate, the pellet was briefly washed with 1xPBS then resuspended in 500 µL of 1x PBS and lysed using a probe sonicator (XL-2000 Microson Ultrasonic Liquid Processor, QSonica) voltage output of approximately 10V for 1 min for three bursts with 10 min rest on ice in between each burst to prevent overheating and denaturing of proteins.

Following sonication, the sample was centrifuged at 15,000 rpm to pellet the cellular debris and insoluble protein. The soluble protein lysate was collected by centrifugation to pellet the cellular debris and insoluble protein, followed by acetone precipitation of the soluble fraction (4x volume of acetone) for 16 hrs at -20˚C. Samples were centrifuged at 4,000 rpm for 10min to pellet the protein, before resuspension in 200µL of SDS-PAGE sample buffer (0.5M Tris-HCl, pH 6.8, 25% glycerol, 0.5% SDS, 0.5% (w/v) Bromophenol blue, 0.05% β-mercaptoethanol), and 50 µL was run 2 mm into the stacking layer of a 12% SDS-PAGE gel. The resulting coomassie stained gel band were cut and in-gel trypsin digested based on previously published protocols (Shevchenko, 2007; JO's PAPER). Briefly, cut gel bands were dehydrated with 100%

acetonitrile (Sigma-Aldrich) then reduced with 10mM DTT in 50 mM ammonium bicarbonate (for 30 min at room temperature and alkylated with 50 mM idoacetamide in 50 mM ammonium bicarbonate in the dark at room temperature for 30 min. Following a wash with 100 mM ammonium bicarbonate solution the gel was dehydrated with 100% acetonitrile. Trypsin digest of the gel utilized 20 ng/µL trypsin overnight at 37˚C. Protein was extracted from the gel using 5% formic acid and 1:2 (v/v) 5% formic acid: acetonitrile, then dried using a SpeedVac. Protein samples were resuspended in 0.1% formic acid and run through a Zip Tip with C18 resin prepared according to manufacturer's instructions. After a wash with 0.5% trifluoroacetic acid, the peptides were eluted from the resin using Elution Buffer (67% ACN; 32.8% Water; 0.2% TFA). The eluted protein samples were then dried and resuspended in 7µL of 0.1% formic acid. Samples were then injected into a Thermo Orbitrap Fusion hybrid linear ion trap FT-MS with Dionex 3000 nanospray UPLC and run for 2 hr per sample. Resulting protein spectral counts were analyzed using the Scaffold program NSAF (normalized spectral abundance factor) [or perSPECtives and normalized spectral count value]. This factor normalizes the number of spectral counts by the length of the protein followed by the overall PSMs. A one-way F-test (ANOVA) was performed and set at a significance boundary of alpha = 0.05 and a Benjamini-Hochberg correction was applied to control for overall type I error rate. Additionally, the false discovery rate (FDR) was set at 1% for peptides and 5% for protein identification.

4.4.9 Quantification RT-PCR analysis

Primers were designed to amplified 100-200 base pairs of the mRNA of interest and synthesized by IDT. cDNAs were prepared with Super Script III kit (Invitrogen, USA) using the manufacturer's protocol. The PCR reactions were ran with cDNA and a

minus reverse transcriptase (RT) control. The annealing temperature was optimized depending on the melting temperature of each primer (Table C.6 for PCR primer sequence). The real-time PCR were performed with ViiA 7 system (Applied Biosystems) with SYBR Select Master mix (Thermo Fisher scientific, 4472903) using the protocol provided by the manufacturer.

Chapter Five

Conclusions and perspectives

In this dissertation I have reviewed important characteristics and features of ncRNA, or sRNAs, in bacteria. I have also discussed important methods of sRNAs identifications both with bioinformatics tools and experimental techniques. This dissertation also discusses the important findings and theories of the regulatory mechanisms in *D. radiodurans*. By applying bioinformatics tools and novel experiment techniques, we were able to discover novel transcripts in *D. radiodurans* and verify potential radioresistance mechanisms. Collectively, this dissertation presents how we identified novel regulatory pathways and elements in bacteria display distinct phenotypes, such as extremophiles or pathogens.

In the work described in Chapter 2, I presented a series of bioinformatics analysis of multiple bacterial genomes and sRNAs and to identify potential sRNAs enriched regions with genomic intergenic region (IGR) conservation analysis. The conservation levels of intergenic regions from 13 bacterial species were analyzed within or outside their genus, and many sRNAs are encoded in highly conserved IGRs. This analysis also indicates that sRNAs encoded regions are more conserved compared to a random region, and using phylogenetic distances to normalize the calculation can further refine the results. The long or isolated IGRs are also shown to be more likely to encode sRNAs. With these observations we analyzed the computational predicted sRNAs in *M. smegmatis* and *M. bovis* BCG, and identified 40 novel transcripts. The 5' and 3' ends of these transcripts were verified with Deep-RACE analysis and proposed potential mechanisms of these sRNAs based on genetic context. This study demonstrates

bioinformatics pipelines can be powerful tools to refine computational predictions and expedite the discovery of novel transcripts in bacteria of interest.

In the third chapter, I discussed how to applied the previous mentioned concept to identify novel transcripts in an extremely radioresistive organism *D. radiodurans*. In this study, we used both computational prediction and transcriptome analysis to generate preliminary list of putative sRNAs and refined the list with bioinformatics filters for experimental confirmation. Northern blotting analysis and RT-PCR confirmed the expression of a total of 41 sRNAs, including 7 in *D. geothermalis*. The 5' and 3' ends of these novel trnascripts were also verified with Deep-RACE analysis, and the irradiation analysis showed that the expression level of many potential sRNAs were up or down-regulated under ionizing irradiation. The conservation analysis and computational targets prediction provides insight of the functions or regulatory mechanisms or these sRNAs, and some of them were predicted to interact with critical genes for radioresistance regulations. Overall, this study is a step stone to reveal potential sRNAs-regulated mechanisms that contributes to radioresistance.

Finally in the last chapter, I introduced how we identified a novel regulatory transcript in *D. radiodurans* that could play a part in the radioresistance mechanisms. One sRNAs from previous study stood out and has interesting expression pattern under ionizing irradiations. Mutants were constructed to overexpress or delete the transcripts of interest, and Dsr2 deletion mutant was found to have a phenotypic effect under irradiation. With various experimental techniques, we were able to verify mRNAs that can have a direct interaction with Dsr2 sRNA, suggesting potential novel mechanisms. The transcriptome data also providing clues and showed the expression levels of mRNAs of interest were correlated with Dsr2 sRNA levels. One of the most intriguing mechanisms that can be regulated by Dsr2 is the PprM protein. PprM as a cold shock

protein homolog does not have a clear induction pathway identified in previous studies. Our study provides a new aspect to investigate this unique regulatory pathway in *D. radiodurans*. In addition, Dsr2 was also recognized as a multi-target regulator, suggesting a more complex target network than we have identified.

The pioneering work detailed in this dissertation sets the grounds for broad applications that span from the high throughput identification of sRNA coding regions to the characterization of complex networks that control bacterial radioresistance in extremophiles. Besides sRNAs, more regulatory elements of *D. radiodurans* were under investigated in Dr. Contraras' lab. The Deinococcus subgroup has been devoted to study the novel regulatory mechanisms of radioresistance, and exploit the regulatory components for bioengineering applications. For example, we have been analyzing the transcriptome data to identify potential stress-response regulatory 5' untranslated regions and promoters. These regulatory elements will provide great assistance toward the bioengineering of *D. radiodurans*, such as novel molecular edition system expression. More mutants are also under constructions to examine their phenotypic effects under radiation stress. In addition, we are also studying how *D. radiodurans* produces nanoparticals under a various stresses. All these researches could lead to not only interesting fundamental molecular biology studies but also powerful bioengineering applications.

An important perspective that remains a challenge in the near future is to further verify the complete pathways of more regulatory sRNAs in *D. radiodurans*. In the last chapter we verified 14 targets of Dsr2 sRNA and many of them do not have an identified function. Thus, it is difficult to propose possible mechanisms without knowing the actual functions of target genes. These uncharacterized proteins were identified in our proteomic analysis and some of them showed differential expressions under the deletion or

overexpression of Dsr2, suggesting they may process functions that related with Dsr2 regulatory mechanisms. Our data also provide little information of how gene expression under irradiation could be regulated by protein damage. Since proteome protection is a primary mechanism in *D. radiodurans*, it would be crucial to understand how cellular regulation is correlated and regulated with proteome damage and protection. Dsr2 is shown to interact with many translational initiation factors and ribosomal proteins, but more evidence is require to find the missing piece that reveal the big picture of regulation of radiresistance in *D. radiodurans*.

Since we only revealed a small part of sRNAs regulatory mechanisms in *D. radiodurans*, there are much more unknown to discover for future researches. For example, we focused on Dsr2 in chapter 4 which is one of many potential regulatory sNRAs from the results of chapter 3, and I believe there are at least two or more that are interesting, i.e. Dsr7 and Dsr12. These two candidates also showed differential expressions under ionizing irradiation and even more drastic than Dsr2. The reason we did not include them in this study is we found they are overlapping and likely to be co-transcribed with flanking genes. Still, this does not rule out the possibility that they are mRNA fragments that are processed under stress-regulated mechanisms. Dsr18 is also worth for further investigation, although its differential expression pattern under irradiation is depends on growth phase, it may suggest Dsr18 is associated with cell growth regulatory mechanisms. Thus, it is interesting to study how they and their flanking genes affect the radioresistance of *D. radiodurans* by manipulating their expression level.

Although our study suggested Dsr2 to be an important factor in the PprM related regulatory mechanisms, it is unclear how Dsr2 interact with the other related proteins such as PprI and PprA. Our transcriptome indicates Dsr2 may modulate their expression

level, but more evidence is required to propose a more detailed pathway. One prospective study is to delete Dsr2 alone with these proteins and see how the phenotype changes in the deletion strains. On the other hand, we also showed that Dsr2 may regulate the translation of *D. radiodurans* under irradiation. Since previous studies have proposed the protein damage level is correlated with survival and the efficiency of DNA repair, it would be interesting to know how protein damage level is correlated with Dsr2 sRNA level or how it affect Dsr2 binding and regulation.

In summary, I expect that my PhD research becomes a pioneer work of regulatory elements in *D. radiodurans* of not only Dr. Contreras' group and also the Deinococcus research community. We have demonstrated a rational design of experiments and novel algorithms of data analysis to discover the territory of unknown mechanisms, and this work should be the step stone for future prospective studies of regulatory sRNAs in other organisms of interest.

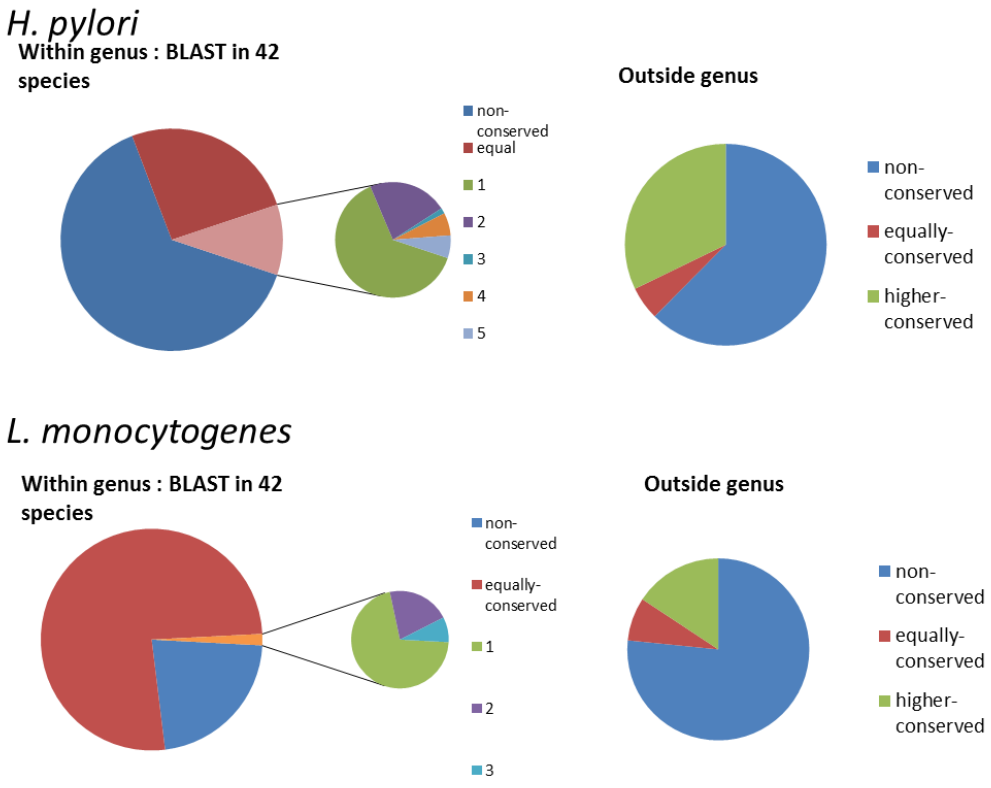
Appendices

APPENDIX A: SUPPLEMENTARY DATA FOR CHAPTER TWO

Supplementary figures for Chapter Two

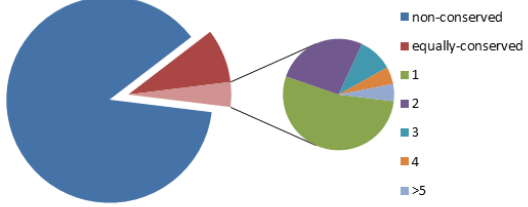
Figure A.1. Conservation distribution and patterns in other ten species

The conservation level is the number of within-genus or outside genus organisms found to have homology of the intergenic region. The intergenic region would be marked as "non-conserved" if its conservation level is less than the extended intergenic region, or "equally" and "higher" if the conservation level is equal or higher than the extended intergenic region. The smaller pie charts of "within-genus" show how conservation levels are distributed in the "higher-conserved" intergenic regions and the total numbers of "within-genus" organisms are showed above.

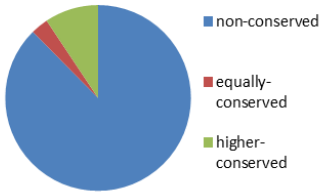


S. aureus

Within genus : BLAST in 20 species

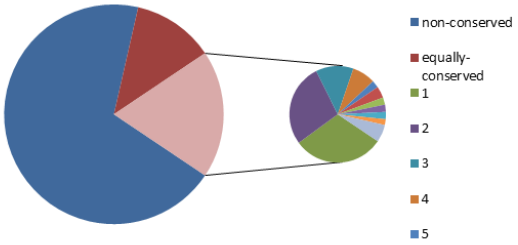


Outside genus

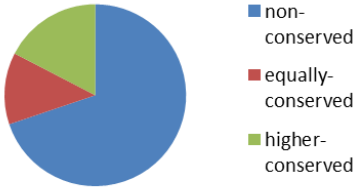


S. pneumoniae

Within genus : BLAST in 41 species

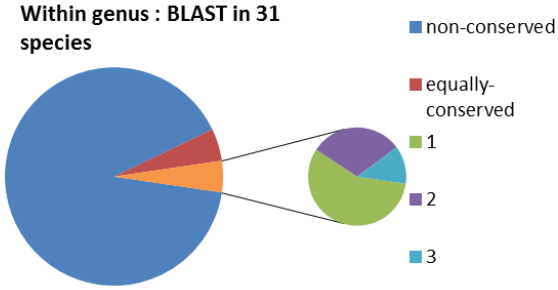


Outside genus

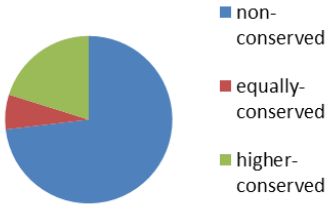


E. faecalis

Within genus : BLAST in 31 species

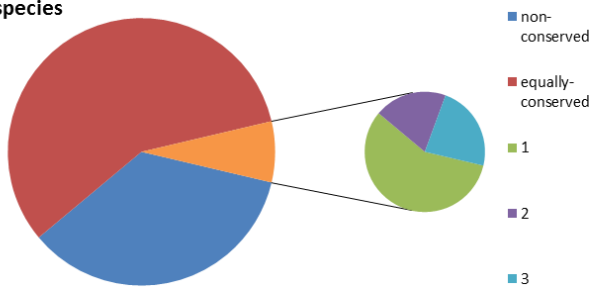


Outside genus

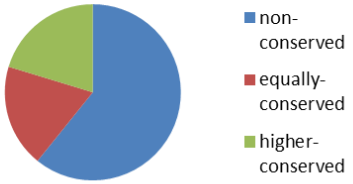


S. typhi

Within genus : BLAST in 7 species

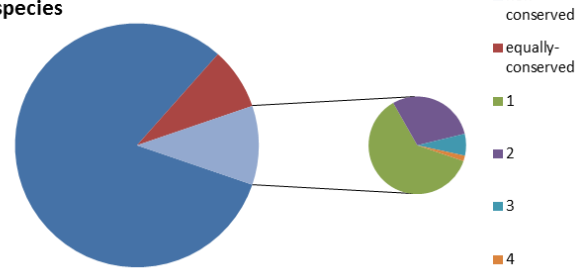


Outside genus

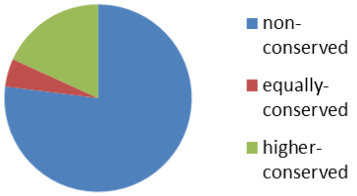


V. harveyi

Within genus : BLAST in 11 species

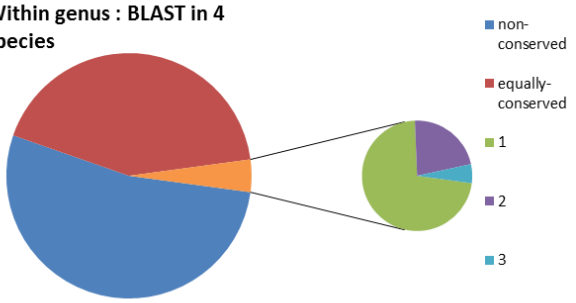


Outside genus

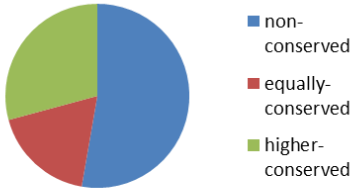


C. trachomatis

Within genus : BLAST in 4 species

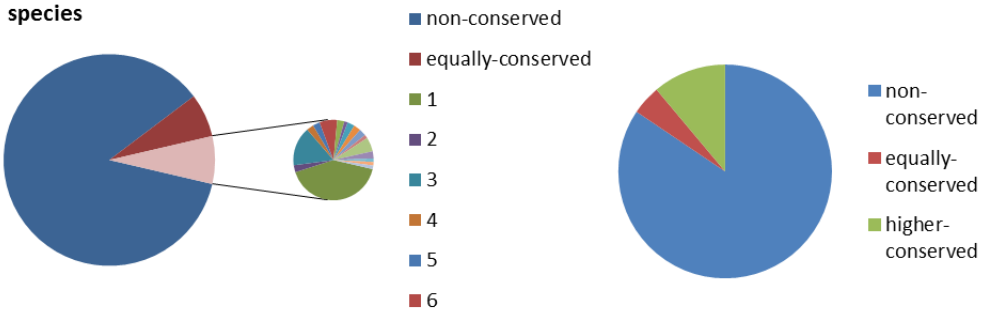


Outside genus



S. pyogene

Within genus : BLAST in 41 species



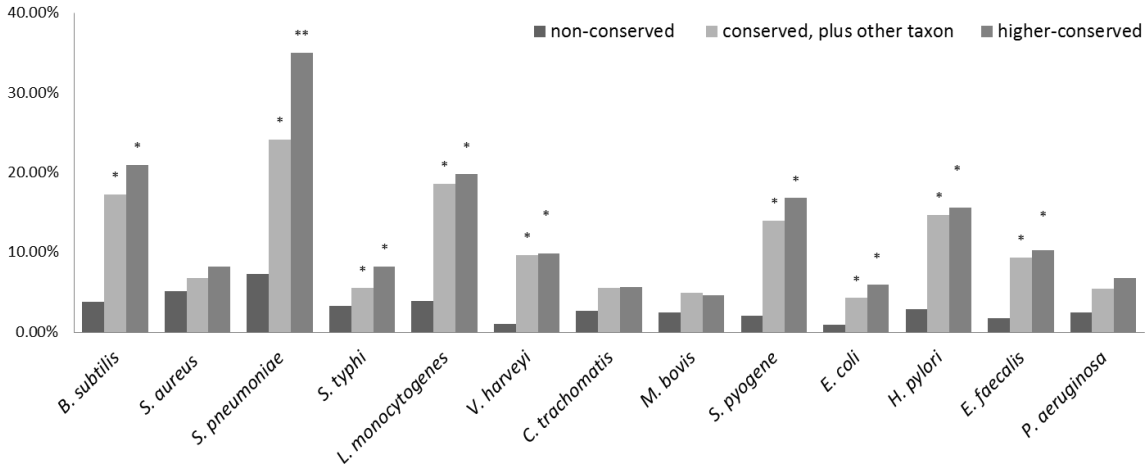
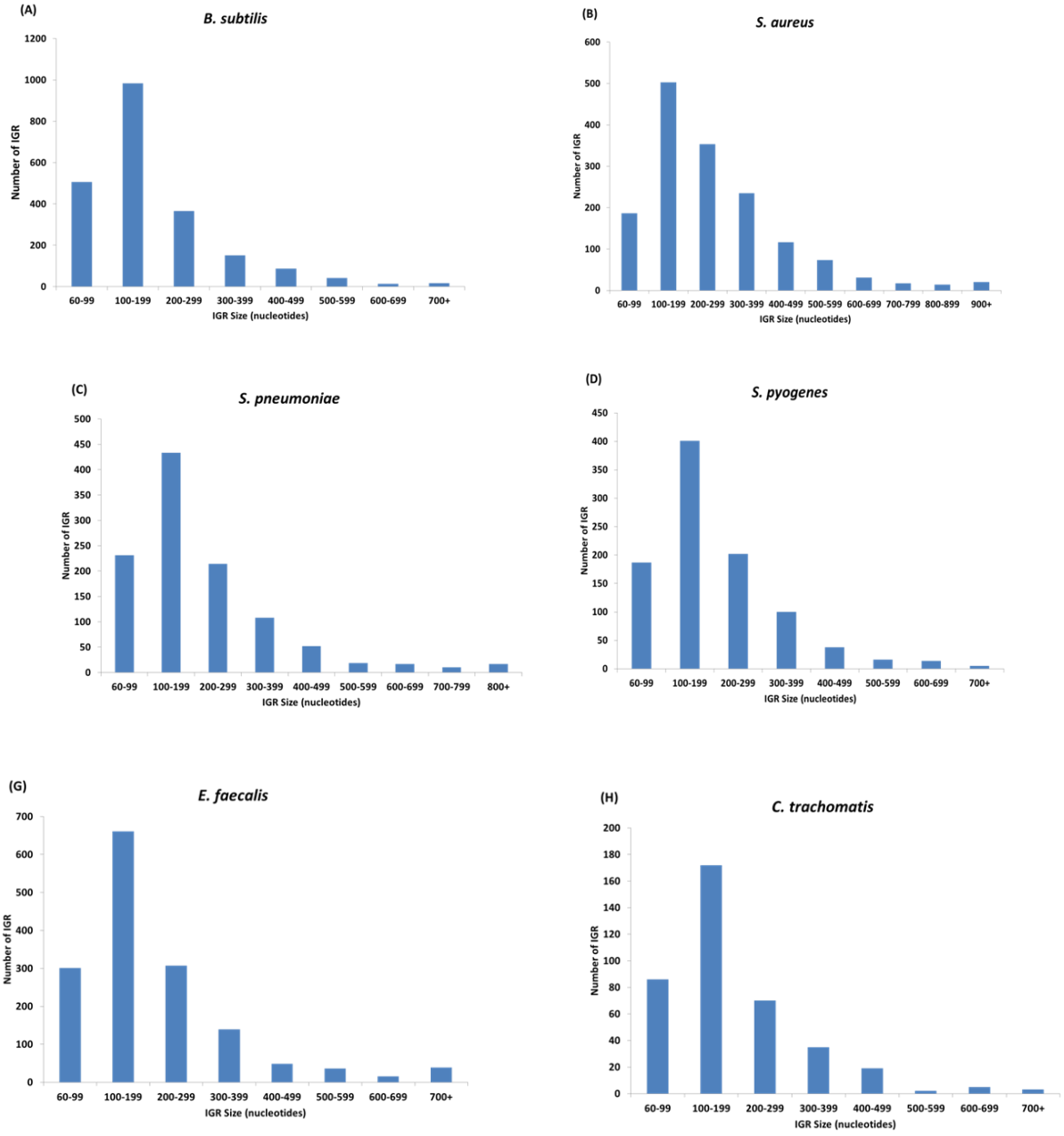


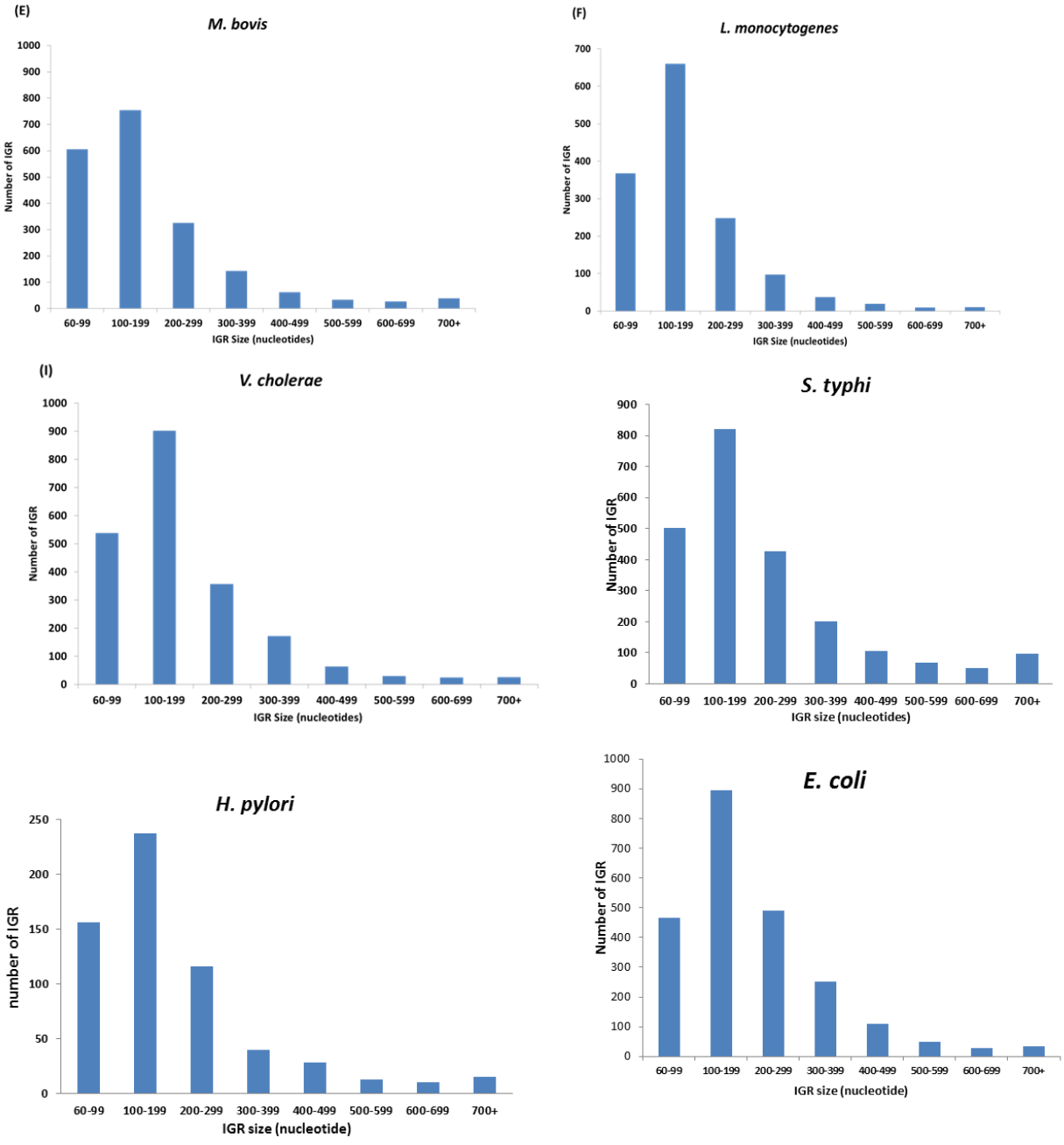
Figure A.2. sRNAs coding region enrichments in intergenic region with different conservation level for outside genus species

The chart above presents the enrichment of sRNAs in intergenic regions conserved "outside genus." The percentage is defined as the number of sRNA coding intergenic regions relative to all selected intergenic regions. For this figure, a conserved intergenic region refers to any intergenic region that has a conservation level greater than 0 for "outside genus." The enrichment of sRNAs in intergenic regions conserved "outside genus" was compared to the enrichment in all intergenic regions. The star (*) denotes: statistically significant enrichment of sRNA compared to non-conserved regions by Fisher Exact Test.

Figure A.3. The distribution of the size of intergenic regions in analyzed species

Intergenic regions with length shorter than 60 nucleotides were not included in this figure.




(E)
M. bovis
Number of IGR
IGR Size (nucleotides)
60-99
100-199
200-299
300-399
400-499
500-599
600-699
700+
(F)
L. monocytogenes
Number of IGR
IGR Size (nucleotides)
(I)
V. cholerae
Number of IGR
IGR Size (nucleotides)
S. typhi
Number of IGR
IGR size (nucleotides)
H. pylori
number of IGR
IGR size (nucleotide)
E. coli
Number of IGR
IGR size (nucleotide)

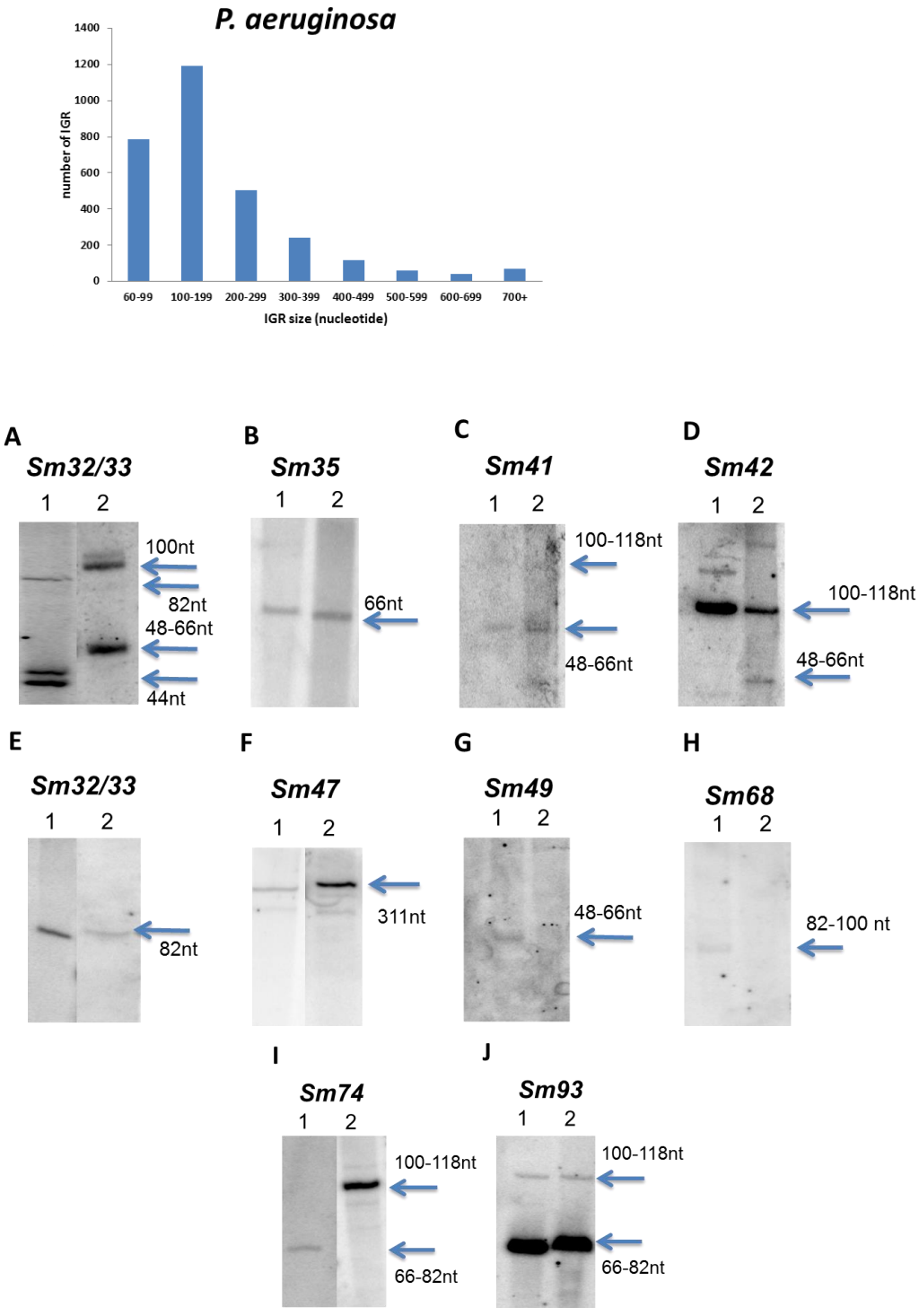


Figure A.4. Northern blotting analysis for *M. smegmatis* sRNAs

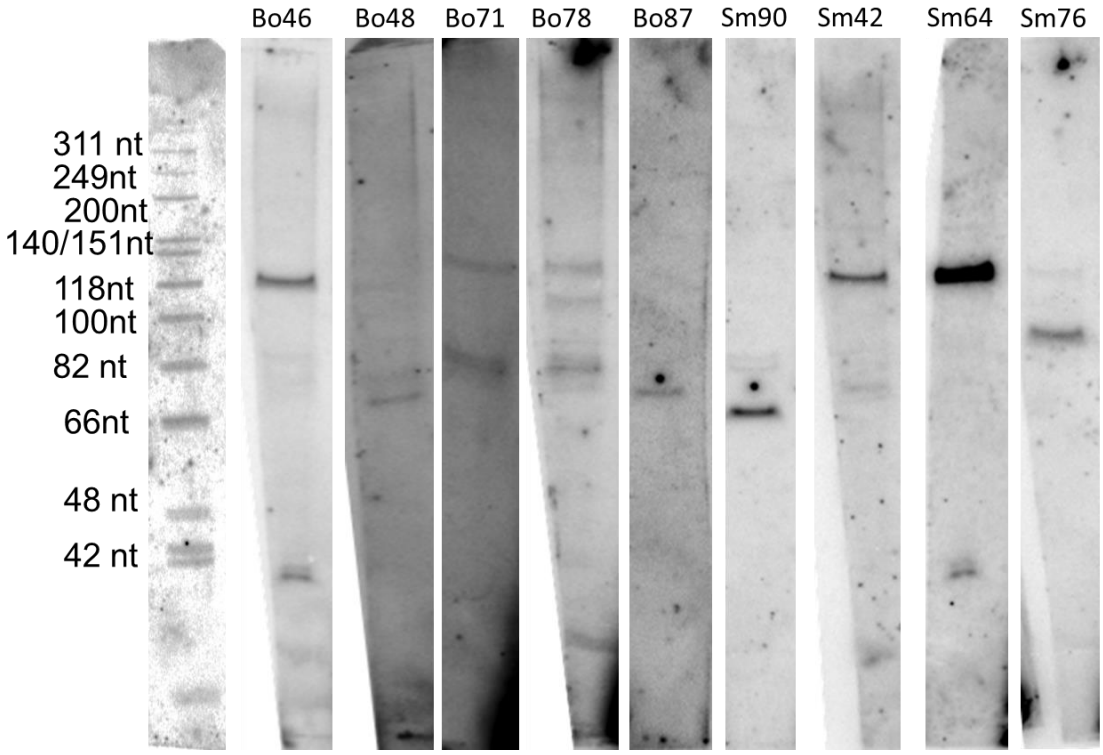


Figure A.5. Northern blotting analysis confirmation of sRNA candidates in *M. tuberculosis* with *M. smegmatis* and *M. bovis* probes.

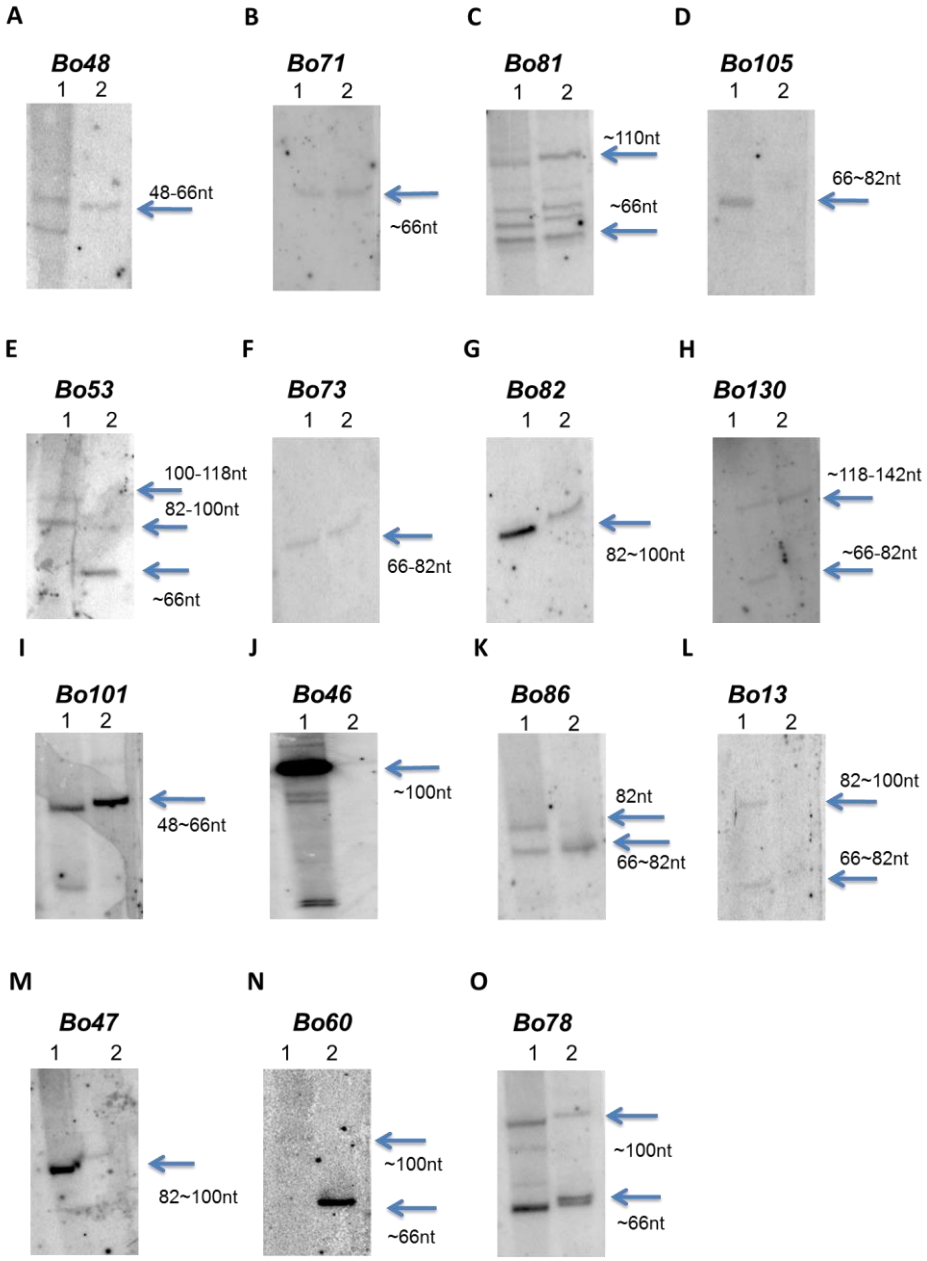
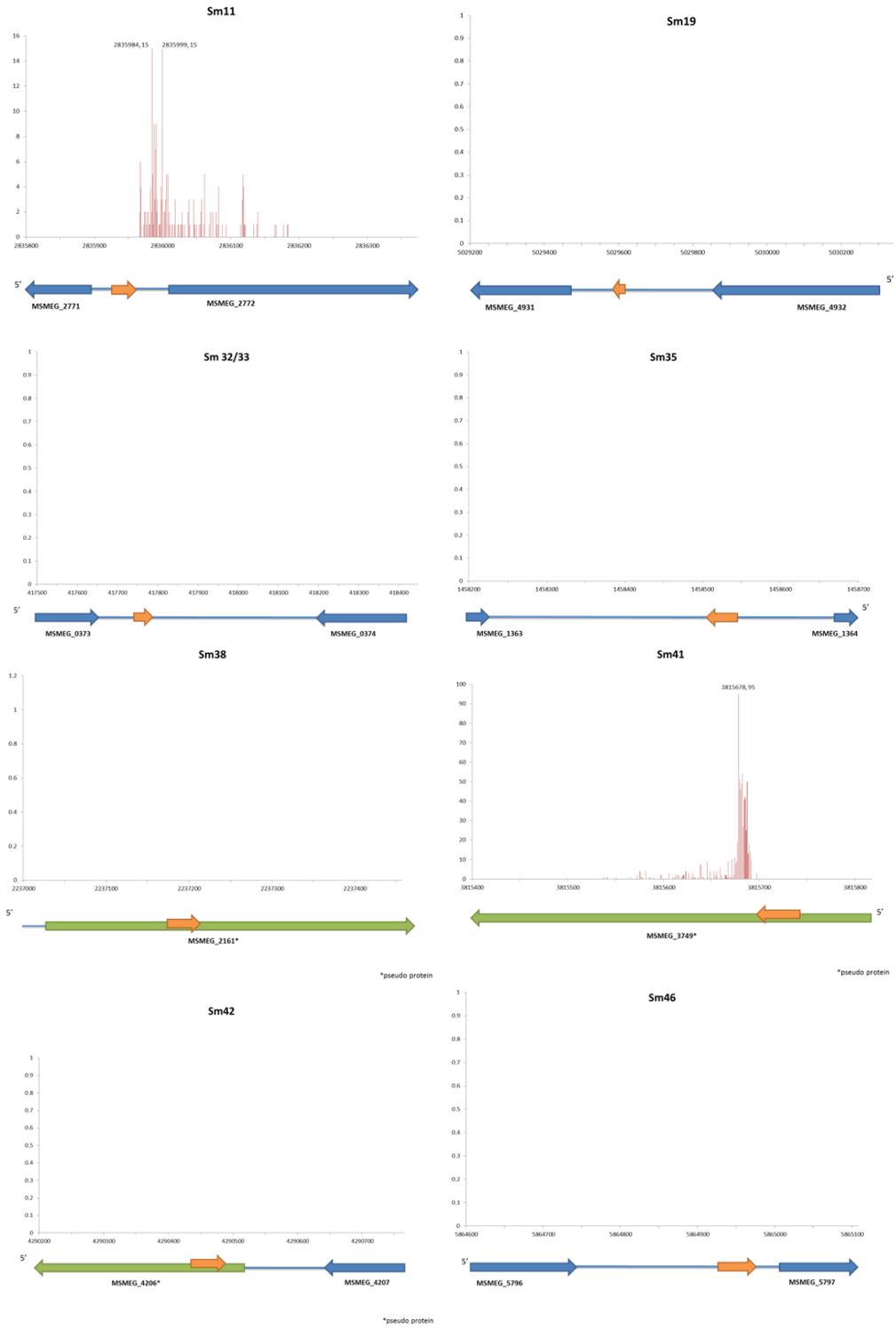
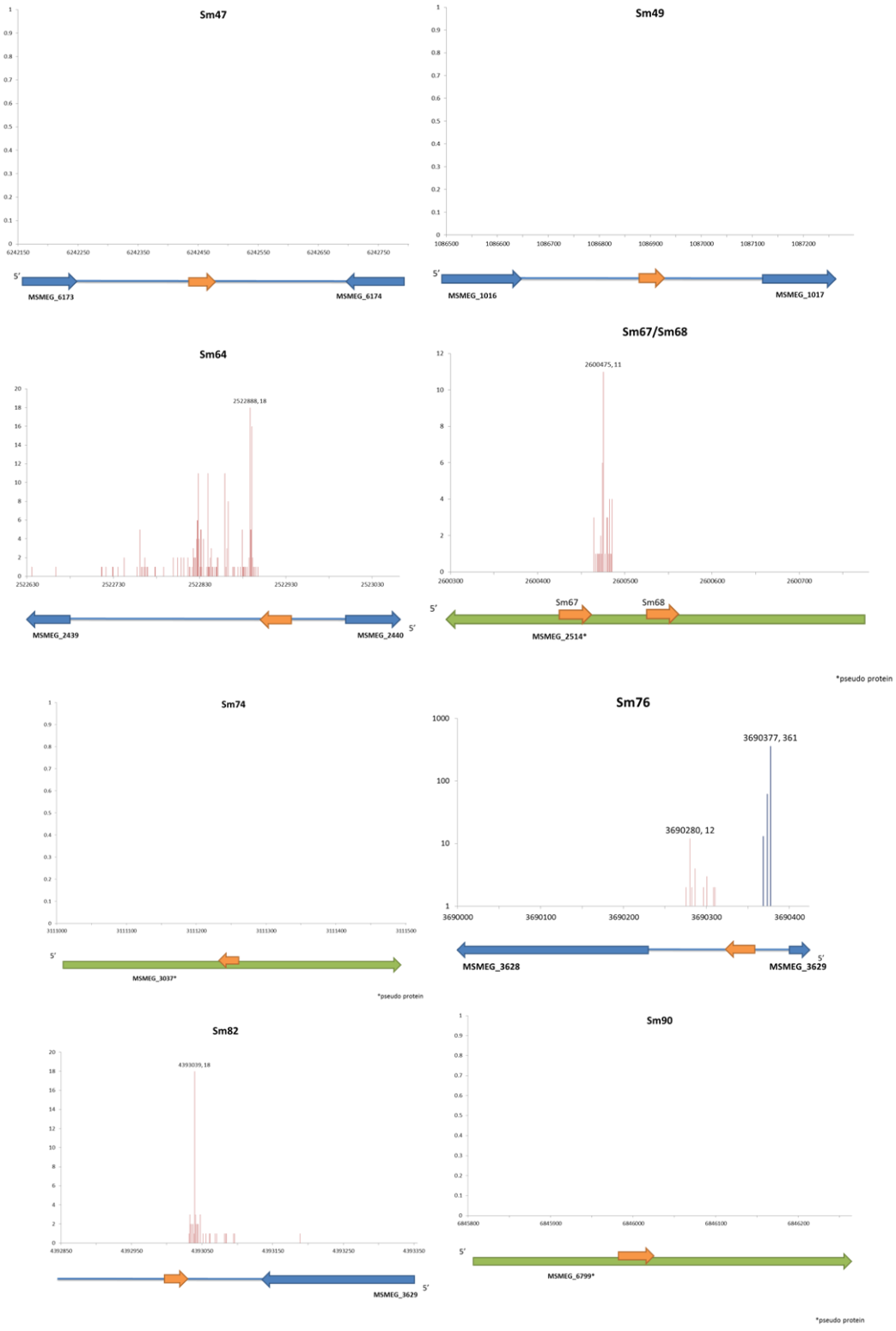
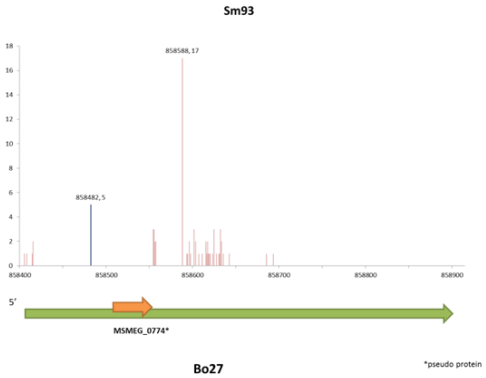


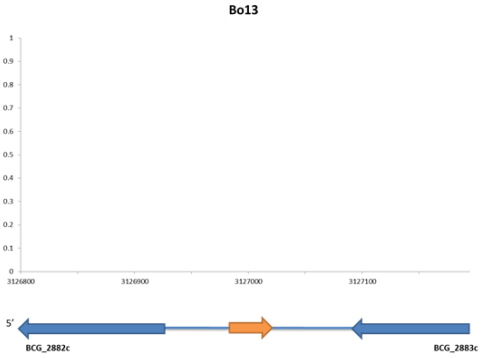
Figure A.6. Northern blotting analysis confirmation of sRNA candidates in *M. bovis* BCG.

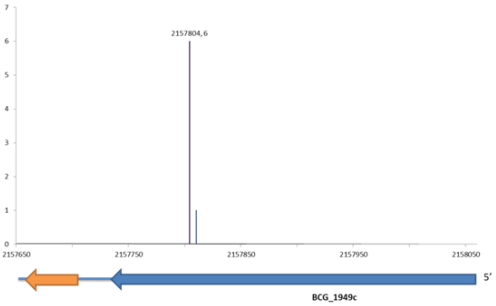
Figure A.7. Deep-RACE mapped reads of all sRNAs and adjacent gene annotations

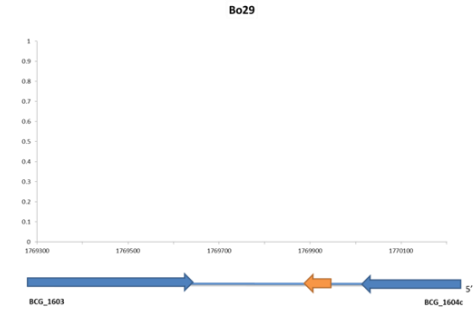


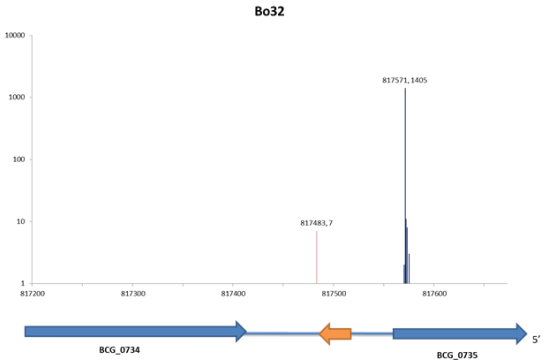

Sm47
6242150
6242250
6242350
6242450
6242550
6242650
6242750
5′
MSMEG_6173
MSMEG_6174
Sm49
1086500
1086600
1086700
1086800
1086900
1087000
1087100
1087200
5′
MSMEG_1016
MSMEG_1017
Sm64
2522888, 18
2522630
2522730
2522830
2522930
2523030
MSMEG_2439
MSMEG_2440
5′
Sm67/Sm68
2600475, 11
2600300
2600400
2600500
2600600
2600700
5′
Sm67
Sm68
MSMEG_2514*
*pseudo protein
Sm74
3111000
3111100
3111200
3111300
3111400
3111500
5′
MSMEG_3037*
*pseudo protein
Sm76
3690377, 361
3690280, 12
3690000
3690100
3690200
3690300
3690400
MSMEG_3628
5′
MSMEG_3629
Sm82
4393039, 18
4392850
4392950
4393050
4393150
4393250
4393350
MSMEG_3629
5′
Sm90
6845800
6845900
6846000
6846100
6846200
5′
MSMEG_6799*
*pseudo protein

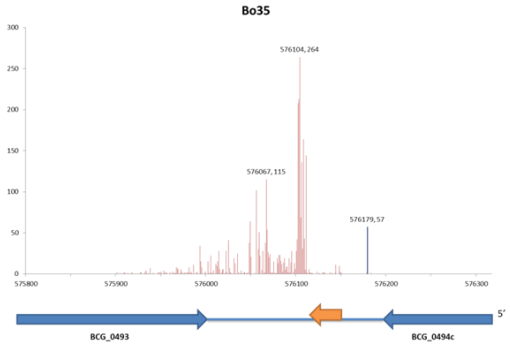

Sm93
858588, 17
858482, 5
5′
MSMEG_0774*
*pseudo protein

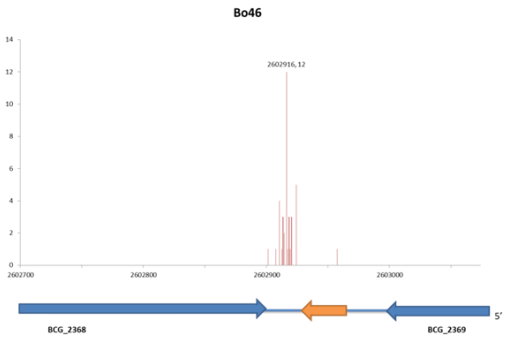

Bo13
5′
BCG_2882c
BCG_2883c

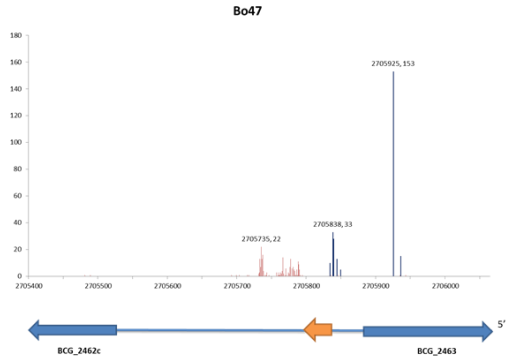

Bo27
2157804, 6
5′
BCG_1949c

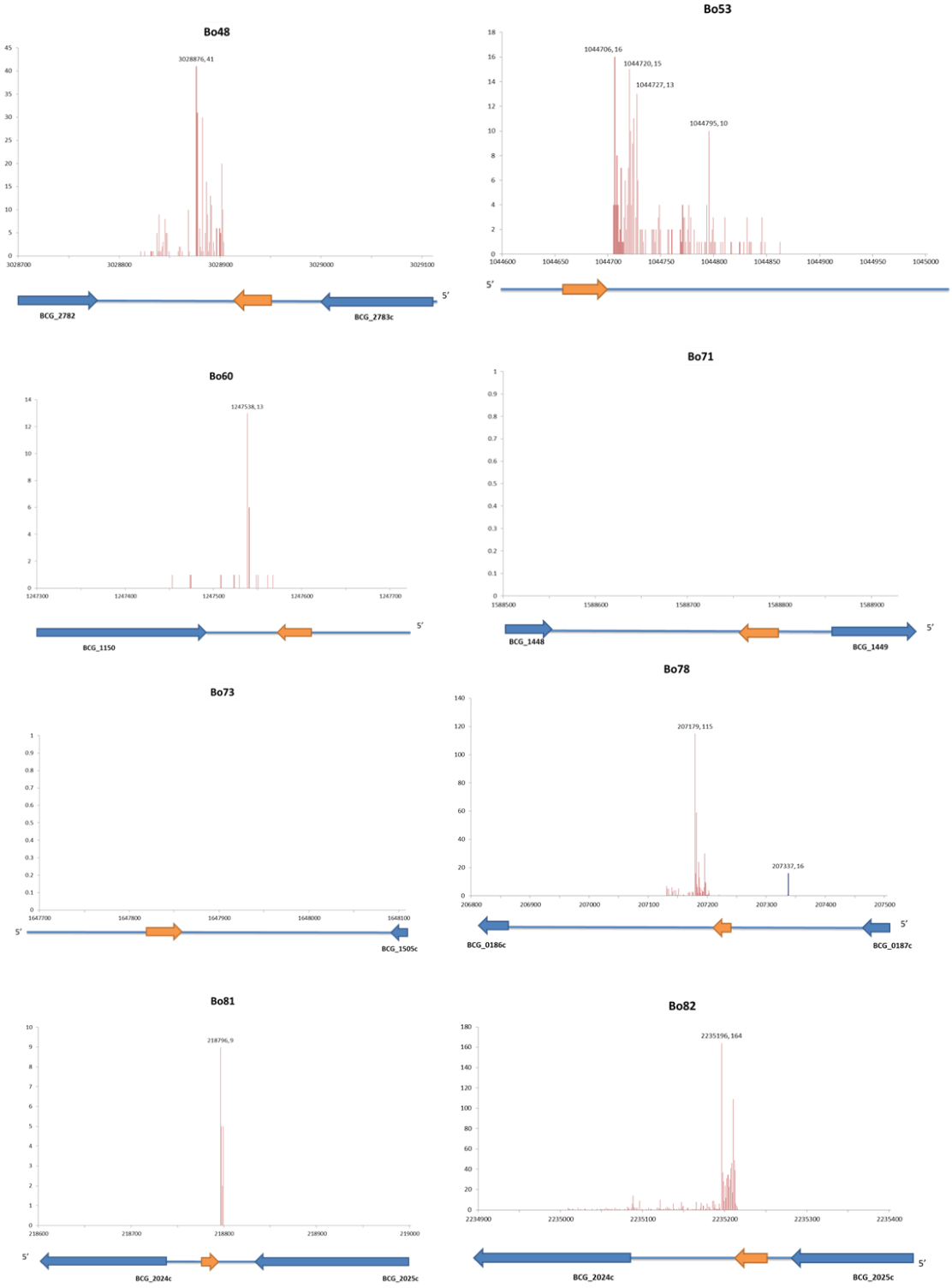

Bo29
5′
BCG_1603
BCG_1604c

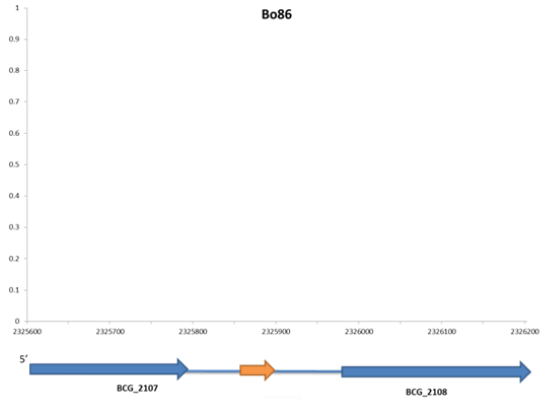

Bo32
817571, 1405
817483, 7
BCG_0734
5′
BCG_0735

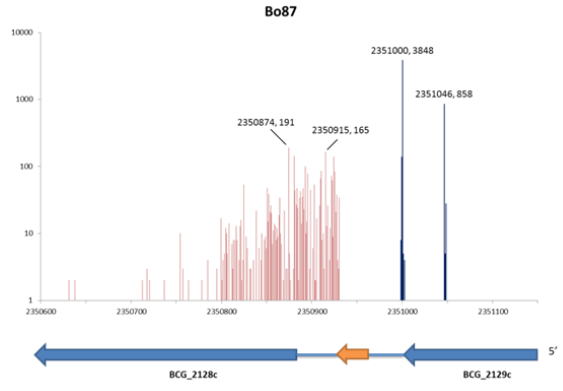

Bo35
576104, 264
576067, 115
576179, 57
BCG_0493
5′
BCG_0494c

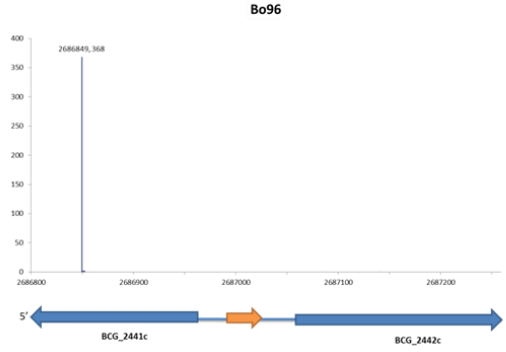

Bo46
2602916, 12
BCG_2368
5′
BCG_2369

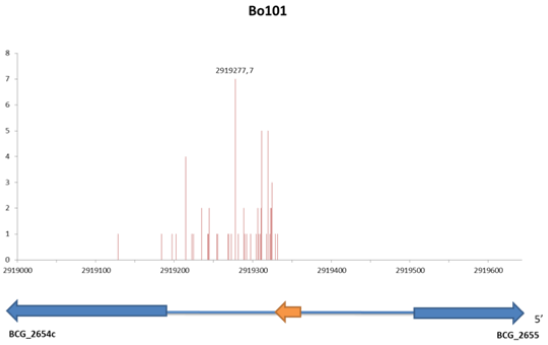

Bo47
2705925, 153
2705838, 33
2705735, 22
BCG_2462c
5′
BCG_2463

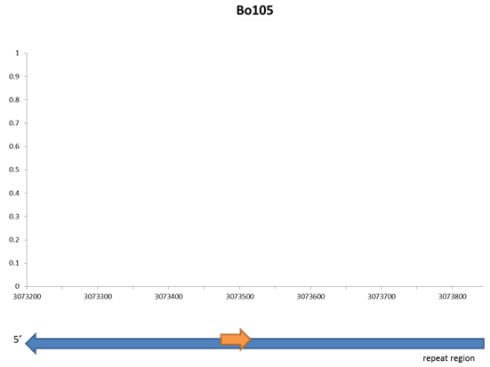

Bo48
3028876, 41
3028700
3028800
3028900
3029000
3029100
5′
BCG_2782
BCG_2783c
Bo53
1044706, 16
1044720, 15
1044727, 13
1044795, 10
1044600
1044650
1044700
1044750
1044800
1044850
1044900
1044950
1045000
5′
Bo60
1247538, 13
1247300
1247400
1247500
1247600
1247700
5′
BCG_1150
Bo71
1588500
1588600
1588700
1588800
1588900
BCG_1448
5′
BCG_1449
Bo73
1647700
1647800
1647900
1648000
1648100
5′
BCG_1505c
Bo78
207179, 115
207337, 16
206800
206900
207000
207100
207200
207300
207400
207500
BCG_0186c
5′
BCG_0187c
Bo81
218796, 9
218600
218700
218800
218900
219000
5′
BCG_2024c
BCG_2025c
Bo82
2235196, 164
2234900
2235000
2235100
2235200
2235300
2235400
BCG_2024c
BCG_2025c
5′

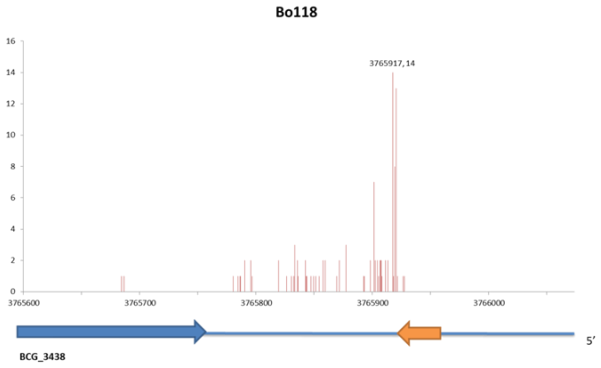

Bo86
5′
BCG_2107
BCG_2108

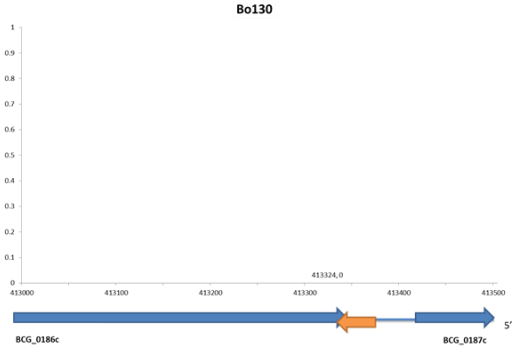

Bo87
2351000, 3848
2351046, 858
2350874, 191
2350915, 165
BCG_2128c
BCG_2129c
5′

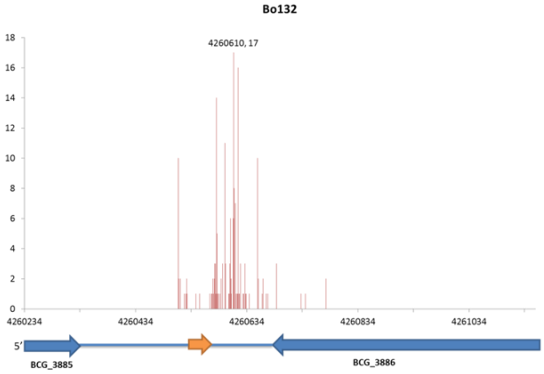

Bo96
2686849, 368
5′
BCG_2441c
BCG_2442c


Bo101
2919277, 7
BCG_2654c
BCG_2655
5′


Bo105
5′
repeat region


Bo118
3765917, 14
BCG_3438
5′


Bo130
413324, 0
BCG_0186c
BCG_0187c
5′


Bo132
4260610, 17
5′
BCG_3885
BCG_3886

Supplementary Tables for Chapter Two

Table A.1. The distribution of the size of intergenic regions in analyzed species

Species	Gram stain	Pathogenicity	References	NCBI accession code
V. cholerae	Negative	cholera	(39–42)	NC_002505.1
L. monocytogenes	Positive	listeriosis	(43–45)	NC_003210.1
C. trachomatis	Negative	chlamydia	(46)	NC_010280.2
S. aureus	Positive	boil or food poisoning	(47–51)	NC_002745.2
B. subtilis	Positive	None	(52, 53, 81)	NC_000964.3
S. pneumoniae	Positive	pneumonia	(54–56)	NC_003028.3
S. pyogenes	Positive	pharyngitis	(57)	NC_007297.1
E. faecalis	Positive	endocarditis or Bacteremia	(58, 59)	NC_004668.1
M. bovis	Positive*	bovine tuberculosis	(60, 61)	NC_008769.1
E. coli	Negative	None	(62–72)	NC_000913.3
H. pylori	Negative	chronic gastritis and gastric ulcers	(73)	NC_000915.1
S. typhi	Negative	Salmonellosis	(74,75,80)	NC_004631.1
P. aeruginosa	Negative	Pneumonia	(76–79)	NC_002516.2

***: acid-fast Gram-positive**

Table A.2. BLAST target species

(1) *Bacillus*	
Within genus	Outside genus
1. *Bacillus cereus* 2. *Bacillus thuringiensis* 3. *Bacillus anthracis* 4. *Bacillus mycoides* 5. *Bacillus weihenstephanensis* 6. *Bacillus cytotoxicus* 7. *Bacillus atrophaeus* 8. *Bacillus amyloliquefaciens* 9. *Bacillus licheniformis* 10. *Bacillus vallismortis* 11. *Bacillus mojavensis* 12. *Bacillus megaterium* 13. *Bacillus coagulans* 14. *Bacillus coahuilensis* 15. *Bacillus pumilus*	*Agrobacterium, Bacteroides, Bordetella, Borrelia, Brucella, Burkholderia, Chlamydia,Clostridium, Deinococcus, Desulfo, Enterobacter, Enterococcus, Escherichia, Geobacter, Haemophilus, Helicobacter, Lactobacillus, Listeria Mycobacterium, Mycoplasma, Neisseria, Pseudomonas, Rhizobium, Rhodobacter, Rhodococcous, Rickettsia, Salmonella, Shigella, Streptococcus, Streptomyces,*

16. *Bacillus methanolicus* 17. *Bacillus marmarensis* 18. *Bacillus safensis* 19. *Bacillus altitudinis* 20. *Bacillus aerophilus* 21. *Bacillus siamensis* 22. *Bacillus azotoformans* 23. *Bacillus vireti* 24. *Bacillus bataviensis* 25. *Bacillus horikoshii* 26. *Bacillus aquimaris* 27. *Bacillus simplex* 28. *Bacillus lentus* 29. *Bacillus smithii* 30. *Bacillus cellulosilyticus* 31. *Bacillus pseudomycoides* 32. *Bacillus pseudofirmus* 33. *Bacillus selenitireducens* 34. *Bacillus clausii* 35. *Bacillus halodurans* 36. *Bacillus psychrosaccharolyticus* 37. *Bacillus alcalophilus*	*Staphylococcus, Synechococcus, Thermotoga, Vibrio, Xanthomonas, Yersinia, and Zymomonas.*
(2) *Chlamydia*	
Within genus	Outside genus
1. *Chlamydia trachomatis* 2. *Chlamydia psittaci* 3. *Chlamydia pneumoniae* 4. *Chlamydia pecorum* 5. *Chlamydia muridarum*	*Agrobacterium, Bacillus, Bacteroides, Bordetella, Borrelia, Brucella, Burkholderia, Clostridium, Deinococcus, Desulfo, Enterobacter, Enterococcus, Escherichia, Geobacter, Haemophilus, Helicobacter, Lactobacillus, Listeria Mycobacterium, Mycoplasma, Neisseria, Pseudomonas, Rhizobium, Rhodobacter, Rhodococcous, Rickettsia, Salmonella, Shigella, Streptococcus, Streptomyces, Staphylococcus, Synechococcus, Thermotoga, Vibrio, Xanthomonas, Yersinia, and Zymomonas.*
(3) *Enterococcus*	

Within genus	Outside genus
1. *Enterococcus gallinarum* 2. *Enterococcus gilvus* 3. *Enterococcus haemoperoxidus* 4. *Enterococcus hawaiiensis* 5. *Enterococcus hermanniensis* 6. *Enterococcus hirae* 7. *Enterococcus inusitatus* 8. *Enterococcus italicus* 9. *Enterococcus lactis* 10. *Enterococcus malodoratus* 11. *Enterococcus moraviensis* 12. *Enterococcus mundtii* 13. *Enterococcus pallens* 14. *Enterococcus pernyi* 15. *Enterococcus phoeniculicola* 16. *Enterococcus plantarum* 17. *Enterococcus aquimarinus* 18. *Enterococcus asini* 19. *Enterococcus avium* 20. *Enterococcus azikeevi* 21. *Enterococcus caccae* 22. *Enterococcus camelliae* 23. *Enterococcus canintestini* 24. *Enterococcus canis* 25. *Enterococcus casseliflavus* 26. *Enterococcus cecorum* 27. *Enterococcus columbae* 28. *Enterococcus devriesei* 29. *Enterococcus dispar* 30. *Enterococcus durans* 31. *Enterococcus ENA07*	*Agrobacterium, Bacillus, Bacteroides, Bordetella, Borrelia, Brucella, Burkholderia, Chlamydia, Clostridium, Deinococcus, Desulfo, Escherichia, Geobacter, Haemophilus, Helicobacter, Lactobacillus, Listeria Mycobacterium, Mycoplasma, Neisseria, Pseudomonas, Rhizobium, Rhodobacter, Rhodococcous, Rickettsia, Salmonella, Shigella, Streptococcus, Streptomyces, Staphylococcus, Synechococcus, Thermotoga, Vibrio, Xanthomonas, Yersinia, and Zymomonas.*
(4) *Escherichia*	
Within genus	Outside genus
1. *Escherichia albertii* 2. *Escherichia fergusonii* 3. *Escherichia faecalis* 4. *Escherichia senegalensis* 5. *Escherichia vulneris* 6. *Escherichia hermannii*	*Agrobacterium, Bacillus, Bacteroides, Bordetella, Borrelia, Brucella, Burkholderia, Chlamydia, Clostridium, Deinococcus, Desulfo, Enterococcus, Geobacter, Haemophilus, Helicobacter, Lactobacillus, Listeria Mycobacterium, Mycoplasma, Neisseria, Pseudomonas, Rhizobium, Rhodobacter, Rhodococcous,*

	Rickettsia, Salmonella, Shigella, Streptococcus, Streptomyces, Staphylococcus, Synechococcus, Thermotoga, Vibrio, Xanthomonas, Yersinia, and Zymomonas.
(5) Helicobacter	
Within genus	Outside genus
1. *Helicobacter acinonychis* 2. *Helicobacter anseris* 3. *Helicobacter apodemus* 4. *Helicobacter aurati* 5. *Helicobacter baculiformis* 6. *Helicobacter bilis* 7. *Helicobacter bizzozeronii* 8. *Helicobacter brantae* 9. *Helicobacter callitrichis* 10. *Helicobacter canadensis* 11. *Helicobacter canis* 12. *Helicobacter cetorum* 13. *Helicobacter cholecystus* 14. *Helicobacter cinaedi* 15. *Helicobacter cynogastricus* 16. *Helicobacter equorum* 17. *Helicobacter felis* 18. *Helicobacter fennelliae* 19. *Helicobacter ganmani* 20. *Helicobacter heilmannii* 21. *Helicobacter hepaticus* 22. *Helicobacter macacae* 23. *Helicobacter magdeburgensis* 24. *Helicobacter marmotae* 25. *Helicobacter mastomyrinus* 26. *Helicobacter mesocricetorum* 27. *Helicobacter muricola* 28. *Helicobacter muridarum* 29. *Helicobacter mustelae* 30. *Helicobacter pametensis* 31. *Helicobacter peregrinus* 32. *Helicobacter pullorum* 33. *Helicobacter rappini* 34. *Helicobacter rodentium* 35. *Helicobacter salomonis* 36. *Helicobacter suis* 37. *Helicobacter suncus*	*Agrobacterium, Bacillus, Bacteroides, Bordetella, Borrelia, Brucella, Burkholderia, Chlamydia, Clostridium, Deinococcus, Desulfo, Enterococcus, Escherichia, Geobacter, Haemophilus, Lactobacillus, Listeria Mycobacterium, Mycoplasma, Neisseria, Pseudomonas, Rhizobium, Rhodobacter, Rhodococcous, Rickettsia, Salmonella, Shigella, Streptococcus, Streptomyces, Staphylococcus, Synechococcus, Thermotoga, Vibrio, Xanthomonas, Yersinia, and Zymomonas.*

38. Helicobacter trogontum *39. Helicobacter tursiopsae* *40. Helicobacter typhlonius* *41. Helicobacter vulpecula* *42. Helicobacter winghamensis*	
(6) *Listeria*	
Within genus	Outside genus
1. Listeria seeligeri *2. Listeria innocua* *3. Listeria welshimeri* *4. Listeria grayi*	*Agrobacterium, Bacillus, Bacteroides, Bordetella, Borrelia, Brucella, Burkholderia, Chlamydia, Clostridium, Deinococcus, Desulfo, Enterobacter, Escherichia, Geobacter, Haemophilus, Helicobacter, Lactobacillus, Mycobacterium, Mycoplasma, Neisseria, Pseudomonas, Rhizobium, Rhodobacter, Rhodococcous, Rickettsia, Salmonella, Shigella, Streptomyces, Synechococcus, Thermotoga, Vibrio, Xanthomonas, Yersinia, and Zymomonas.*
(7) *Mycobacterium*	
Within genus	Outside genus
1. Mycobacteria MCS *2. Mycobacteria JLS* *3. Mycobacteria KMS* *4. Mycobacteria abscessus* *5. Mycobacteria avium* *6. Mycobacteria ulcerans* *7. Mycobacteria leprae* *8. Mycobacteria tuberculosis* *9. Mycobacteria smegmatis*	*Agrobacterium, Bacillus, Bacteroides, Bordetella, Borrelia, Brucella, Burkholderia, Chlamydia, Clostridium, Deinococcus, Desulfo, Enterobacter, Escherichia, Geobacter, Haemophilus, Helicobacter, Lactobacillus, Listeria, Mycoplasma, Neisseria, Pseudomonas, Rhizobium, Rhodobacter, Rhodococcous, Rickettsia, Salmonella, Shigella, Streptomyces, Synechococcus, Thermotoga, Vibrio, Xanthomonas, Yersinia, and Zymomonas.*
(8) *Pseudomonas*	
Within genus	Outside genus

1. *Pseudomonas syringae* 2. *Pseudomonas savastanoi* 3. *Pseudomonas amygdali* 4. *Pseudomonas syringae group genomosp. 3* 5. *Pseudomonas fuscovaginae* 6. *Pseudomonas avellanae* 7. *Pseudomonas viridiflava* 8. *Pseudomonas coronafaciens* 9. *Pseudomonas pseudoalcaligenes* 10. *Pseudomonas mendocina* 11. *Pseudomonas oleovorans* 12. *Pseudomonas fluorescens* 13. *Pseudomonas tolaasii* 14. *Pseudomonas mandelii* 15. *Pseudomonas synxantha* 16. *Pseudomonas putida* 17. *Pseudomonas monteilii* 18. *Pseudomonas fulva* 19. *Pseudomonas stutzeri* 20. *Pseudomonas chlororaphis* 21. *Pseudomonas fragi* 22. *Pseudomonas brassicacearum* 23. *Pseudomonas alcaliphila* 24. *Pseudomonas kilonensis* 25. *Pseudomonas extremaustralis* 26. *Pseudomonas psychrotolerans* 27. *Pseudomonas agarici* 28. *Pseudomonas gingeri* 29. *Pseudomonas entomophila*	*Agrobacterium, Bacillus, Bacteroides, Bordetella, Borrelia, Brucella, Burkholderia, Chlamydia, Clostridium, Deinococcus, Desulfo, Enterobacter, Escherichia, Geobacter, Haemophilus, Helicobacter, Lactobacillus, Listeria Mycobacterium, Mycoplasma, Neisseria, Pseudomonas, Rhizobium, Rhodobacter, Rhodococcous, Rickettsia, Salmonella, Shigella, Streptomyces, Synechococcus, Thermotoga, Vibrio, Xanthomonas, Yersinia, and Zymomonas.*
(9) *Salmonella*	
Within genus	Outside genus
1. *Citrobacter farmeri* 2. *Citrobacter freundii complex* 3. *Citrobacter intermedius* 4. *Citrobacter koseri* 5. *Citrobacter rodentium* 6. *Salmonella bongori* 7. *Salmonella subterranea*	*Agrobacterium, Bacillus, Bacteroides, Bordetella, Borrelia, Brucella, Burkholderia, Chlamydia, Clostridium, Deinococcus, Desulfo, Enterobacter, Enterococcus, Escherichia, Geobacter, Haemophilus, Helicobacter, Lactobacillus, Listeria Mycobacterium, Mycoplasma, Neisseria, Pseudomonas, Rhizobium, Rhodobacter, Rhodococcous,*

	Rickettsia, Shigella, Streptococcus, Streptomyces, Synechococcus, Thermotoga, Vibrio, Xanthomonas, Yersinia, and Zymomonas.
(10) *Staphylococcus*	
Within genus	Outside genus
1. *Staphylococcus epidermidis* 2. *Staphylococcus lugdunensis* 3. *Staphylococcus warneri* 4. *Staphylococcus hominis* 5. *Staphylococcus saprophyticus* 6. *Staphylococcus pseudintermedius* 7. *Staphylococcus simulans* 8. *Staphylococcus capitis* 9. *Staphylococcus haemolyticus* 10. *Staphylococcus massiliensis* 11. *Staphylococcus lentus* 12. *Staphylococcus vitulinus* 13. *Staphylococcus equorum* 14. *Staphylococcus intermedius* 15. *Staphylococcus delphini* 16. *Staphylococcus simiae* 17. *Staphylococcus caprae* 18. *Staphylococcus carnosus* 19. *Staphylococcus hyicus* 20. *Staphylococcus xylosus*	*Agrobacterium, Bacillus, Bacteroides, Bordetella, Borrelia, Brucella, Burkholderia, Chlamydia, Clostridium, Deinococcus, Desulfo, Enterobacter, Enterococcus, Escherichia, Geobacter, Haemophilus, Helicobacter, Lactobacillus, Listeria Mycobacterium, Mycoplasma, Neisseria, Pseudomonas, Rhizobium, Rhodobacter, Rhodococcous, Rickettsia, Salmonella, Shigella, Streptococcus, Streptomyces, Synechococcus, Thermotoga, Vibrio, Xanthomonas, Yersinia, and Zymomonas.*
(11) *Streptococcus*	
Within genus	Outside genus
1. *Streptococcus agalactiae* 2. *Streptococcus pneumoniae* 3. *Streptococcus mutans* 4. *Streptococcus pyogenes* 5. *Streptococcus sobrinus* 6. *Streptococcus sanguinis* 7. *Streptococcus suis*	*Agrobacterium, Bacillus, Bacteroides, Bordetella, Borrelia, Brucella, Burkholderia, Chlamydia, Clostridium, Deinococcus, Desulfo, Enterobacter, Enterococcus, Escherichia, Geobacter, Haemophilus, Helicobacter,*

8. *Streptococcus mitis* 9. *Streptococcus thermophiles* 10. *Streptococcus oralis* 11. *Streptococcus dysgalactiae* 12. *Streptococcus equi* 13. *Streptococcus anginosus* 14. *Streptococcus intermedius* 15. *Streptococcus constellatus* 16. *Streptococcus parasanguinis* 17. *Streptococcus salivarius* 18. *Streptococcus infantis* 19. *Streptococcus gallolyticus* 20. *Streptococcus pseudopneumoniae* 21. *Streptococcus vestibularis* 22. *Streptococcus parauberis* 23. *Streptococcus pseudoporcinus* 24. *Streptococcus urinalis* 25. *Streptococcus downei* 26. *Streptococcus equinus* 27. *Streptococcus infantarius* 28. *Streptococcus macedonicus* 29. *Streptococcus lutetiensis* 30. *Streptococcus ratti* 31. *Streptococcus canis* 32. *Streptococcus pasteurianus* 33. *Streptococcus cristatus* 34. *Streptococcus australis* 35. *Streptococcus peroris* 36. *Streptococcus criceti* 37. *Streptococcus porcinus* 38. *Streptococcus ictaluri* 39. *Streptococcus macacae* 40. *Streptococcus iniae* 41. *Streptococcus uberis* 42. *Streptococcus gordonii*	*Lactobacillus, Listeria Mycobacterium, Mycoplasma, Neisseria, Pseudomonas, Rhizobium, Rhodobacter, Rhodococcous, Rickettsia, Salmonella, Shigella, Streptomyces, Synechococcus, Thermotoga, Vibrio, Xanthomonas, Yersinia, and Zymomonas.*
(12) *Vibrio*	
Within genus	Outside genus
1. *Vibrio alginolyticus* 2. *Vibrio caribbenthicus*	*Agrobacterium, Bacillus, Bacteroides, Bordetella, Borrelia,*

3. *Vibrio corallilyticus* 4. *Vibrio furnissi* 5. *Vibrio mimicus* 6. *Vibrio orientalis* 7. *Vibrio shilonii* 8. *Vibrio harveyi* 9. *Vibrio fischeri* 10. *Vibrio splendidus*	*Brucella, Burkholderia, Chlamydia, Clostridium, Deinococcus, Desulfo, Enterobacter, Escherichia, Geobacter, Haemophilus, Helicobacter, Lactobacillus, Listeria Mycobacterium, Mycoplasma, Neisseria, Pseudomonas, Rhizobium, Rhodobacter, Rhodococcous, Rickettsia, Salmonella, Shigella, Streptomyces, Synechococcus, Thermotoga, Xanthomonas, Yersinia, and Zymomonas.*

Table A.3. Intergenic analysis for all 13 species

This oversize Table was uploaded to Texas Digital Library.

Table A.4. Conservation level comparison of sRCRs and RIGRs

(A) *B. subtilis*	Conservation level within *Bacillus*	
Intergenic Region Coordinates	sRNA-coding section of intergenic region (sRCR)	Random section of intergenic region excluding sRNA-coding section (RIGR)
1018468_1018998	3	3
1056216_1056702	3	1
1076952_1077440	3	3
1150427_1150850	1	1
1219105_1219849 (length 100)	3	3
1219105_1219849 (length 241)	8	1
1233300_1233614	8	4
1435245_1435628	3	1
1446568_1447251	2	1
1451135_1451371	2	1
1467411_1467805	2	1
1483471_1484117	5	1
1527902_1528326	5	3

1596300_1596474	5	2
1780220_1780618	1	1
1868374_1868617	3	1
1900514_1901117	1	1
1901737_1902219	1	1
1917097_1917639	1	1
1925421_1925655	3	2
204890_205409	5	1
2053929_2054599	3	1
2069029_2069262	1	1
2069561_2070244	4	3
2078626_2079214	5	1
2099790_2100147	3	1
2208528_2208855	2	1
2225615_2227297	3	3
2282485_2283136	1	1
2283651_2283858	2	1
2316110_2316446	2	1
2540824_2541051	3	1
2692851_2692933	1	1
2734143_2734953	2	1
2751726_2752167	2	1
275561_275838	2	1
2913338_2913661	2	2
3072110_3072401	1	1
3105043_3105470	4	3
3145958_3146238	2	1
3302763_3303042	1	1
3572889_3573207	5	3
3625507_3625741	2	1
3631572_3631763	1	3
3738217_3738343	1	1
3851893_3852186	3	1
3856172_3857017	6	2
3988763_3989232	4	1
3999097_3999350	11	3
4035607_4035990	2	1
4122849_4123193	2	1
4171789_4172259	1	1
4187174_4187681	4	1
474225_474731	7	3

532552_532922	2	2
554475_554669	1	1
559464_560151	12	6
663027_663601	1	1
678951_679390	2	1
694281_694662	4	2
795867_796314	2	1

(B) *S. pyogenes*	Conservation level within *Streptococcus*	
Intergenic Region Coordinates	sRNA-coding section of intergenic region (sRCR)	Random section of intergenic region excluding sRNA-coding section (RIGR)
125779_125973	1	1
195504_196059	6	3
214258_214798	6	2
257186_257517	1	1
270844_271455	1	1
277177_277431	1	1
335986_336478	1	1
358476_358935	1	1
558841_559706	1	1
638386_638734	1	1
914250_914572	6	1
961799_962205	2	3
968871_969241	1	1
1016199_1016649	1	1
1018348_1018705	2	1
1173257_1173399	3	3
1175357_1176511	1	1
1251734_1252159	8	2
1354931_1355292	1	1
1532755_1533053	1	1
1603784_1604445	23	1
1678751_1679185	3	3
1719722_1720207	1	1
1719722_1720208	3	1
1719722_1720209	3	2

1719722_1720210	1	2

sRNA candidate	Probe Sequence

1719722_1720211	7	1
1719722_1720212	1	1

Table A.5. *M. Smegmatis* MC155 8596 ALL SIPHT Predictions

This oversize Table was uploaded to Texas Digital Library.

Table A.6. *M. Bovis* BCG_008769 ALL SIPHT Predictions

This oversize Table was uploaded to Texas Digital Library.

Table A.7. Probes sequences used in Northern blotting analysis

Sm11	2442 GCTGTAGCGTTCCGGGTGCACGGGTTGCGATAGCGTCG
Sm46	SEE REF. [110]
Sm76	2547 GGACCGGGGGTCTCTGCAGCCCTCAGGACTCCGGC TC
Sm64	2536 GCGCAGGACCGGGCTGAGTAGTGCCTGCCTGCTGCG
Sm49	2474 GGTGGCCTGTCGGTCTCTCAGACACTACACCTAGTGGCCCC
Sm82	2553 GGCCGGGCGGCCGCGGCATCAGCCTGATGTCGAGG
Sm32-33	SEE REF. [110]
Sm90	2561 CGCGACGTTCGCGTCTGCCGGTTGCGGGGTGTCCCCGGG
Sm38	2465 CCGTCGGCCGCAGCGGCTCCCAGGGTCGCGATCGCCTCGGCG
Sm42	2469 CGCATGCTCGTTCTGCGGTGTCGGGTGCGGGATCGAGGTGG
Sm47	SEE REF. [110]
Sm35	SEE REF. [110]
Sm74	SEE REF. [110]
Sm19	2450 GGAAAAAGAGGCGGACAAAAAACAACAAACAAAAACCACC
Sm41	2468 GTGGCGTCGGCCCTGGCCCCGCCCAGCAGGTGCAGGCCCG
Sm93	2564 GGTGTGTGTCCGAGCTGCACTGCGGCAGTGCCGACGCG
Sm67	2539 GCAACCCCGCCGACGAGGCCCCGTGTGTCTCCGGC
Sm68	2540 CGGCCCCCAACAACCCGACGGTTCCTACCACCTGCGCGGC
Sm75	2546 GCACCCCGGCGCGCGCCCCCGATCGGTCGCCGAC GG
Bo52	SEE REF. [110]
Bo80	SEE REF. [110]
Bo99	SEE REF. [110]
Bo100	SEE REF. [110]
Bo125	SEE REF. [110]
Bo15	SEE REF. [110]
Bo137	SEE REF. [110]
Bo58	SEE REF. [110]
Bo75	SEE REF. [110]
Bo41	SEE REF. [110]
Bo67	SEE REF. [110]
Bo12	SEE REF. [110]

Bo85	SEE REF. [110]
Bo117	SEE REF. [110]
Bo139	SEE REF. [110]
Bo13	2702 GCTCCGGCGGTTCGCGGTGCCCGCGACAGCCAGCATGTGGG
Bo32	2715 GCGTCTGCTCGCGAAAATGCCCAGCGTGCGGCGCTACGC
Bo35	2716 GGGCTGTCCCCCGAATGGTGGACAACATTTCCGGGTTCGTTG
Bo46	2720 GGGGATACCCGTACGCTGGCGCGTGTGGCCGTCGACCTAGGC
Bo47	2721 CGGGTGGTGACGTCATCCGGGTTGGACCGCTGATGGCTGCGGC
Bo48	2722 CGATGATGATTCAGCCGACGCCGGTCCGCGGTGCGCCCCG
Bo53	2725 GCAAGACCAGCCCTACCGAAGCCATCAATGGCCGCCTGG
Bo60	2728 CGCACACGCTTGCTTGAACATCGGGTGGAGCCGGTGG
Bo71	2732 CCGCGAGTGATCCCCGGCACTGCGAGTTGCGACGCCACC
Bo73	2734 GAAAAGTCAGCGGCCCTGACAGAGCAGCTGCGCGG
Bo78	2738 CGGGCTGCCCCTGGCCGGTCGCACCAAGACGCCGCATACG
Bo81	2739 GCTTCCCGGCGGGCGCGCTCTAGGCTCTAAGGGCCC
Bo82	2740 GGTTCACCCGACCGCCAGCGGGATTCACGCTCCCCCAGGC
Bo86	2742 GCCAACTCACCAGTTCCAGGTGATCGCGTGCACCCG
Bo87	2743 GCGCGTCACACGCCCGCTGTCTTTCTCTACCCTACCGG
Bo94	2756 GGACCGCCGTAATGGAGTTCGCGCCCGGCGCCGTCG
Bo96	2758 GGCGCTGGTGCGCCCGCTTATCACGCGTTGTTGGCCCACGGC
Bo101	2760 GGTGCCGCAGCCCGGCCAGCACGCCGTCAGAGTTTCACGGGG
Bo105	2761 CGGGGAGCCGATCAGCGACCACCGCACCCTGTCAGTCGTC
Bo118	2768 CGCGTCCAGCTACCACCACCGTCAGCGGTGACACCTTCACCGG
Bo130	2771 GGCTCTGGGTGAGCCGCGTTCCCCGGAGCTGGCCCCGTCGGTG
Bo132	2773 CCAAAAGGAAGACCTCGGCGTGTCTGCCCGAGGTCC

Bo135	2775 CGGCGACACGTATCGCCGAGTGTGAATCCCGCGACGCCGCA CCGG
Bo27	2711 GGCTAGCGTGACAGGCGTCTGCTAGGACCCGA TCGCCCCG
Bo29	2713 CGGCTGCCGCGAAATCCGGCTTTC TACCACGACGGATCC
Bo56	2726 CCACAGCGATCGCGGCCAGTGGCAATGCGAACCTCACCG

Table A.8. Transcription factor ChIP-seq peaks located within 100 bp upstream and 20 bp downstream of sRNA 5' ends.

sRNA	New nomenclature	Transcription Factor	5' end - peak distance
B55[a]	ncRv10609A	Rv3249c	-29[e]
B11[a]	ncRv13660Ac	Rv0081[d]	-56
B11[a]	ncRv13660Ac	Rv0081[d]	-45
B11[a]	ncRv13660Ac	Rv3249c	-25
B11[a]	ncRv13660Ac	Rv2034	-20.5
Mpr5[b]	ncBCG11109A/ncRv11051A	Rv0324	-27
Mcr11[b]	ncBCG11323Ac/ncRv11264A	Rv3574	-52
Mcr8/Mpr4[b]	ncBCG13719A/ncRv13661A	Rv0821c (PhoY2)	-38[e]
Bo29[c]	ncBCG11603Ac	Rv2021c	-19
Bo132[c]	ncBCG13885A	Rv2887	-44[e]

[a] Arnvig and Young, 2009
[b] DiChiara *et al*., 2010
[c] this work
[d] replicate ChIP-seq datasets are available for Rv0081 and we required that sRNA-proximal ChIP-seq peaks be present in both replicates for inclusion in this table
[e] unambiguously associated with an sRNA 5' end

Table A.9. List of sRNAs from database and literature

This oversize Table was uploaded to Texas Digital Library.

Table A.10. Minimum length of long intergenic regions with different thresholds in each species

Species	Size of genome ($\times 10^6$ bp)	# of total intergenic region	Average length*	length of intergenic regions within top 20% long
B. subtilis	4.2	2167	188 bp	≥252 bp
C. trachomatis	1	393	195 bp	≥280 bp
E. coli	4.6	2327	215 bp	≥302 bp
E. faecalis	3.2	1547	229 bp	≥282 bp
H. pylori	1.6	615	205 bp	≥282 bp
L. monocytogenes	2.9	1447	162 bp	≥242 bp
M. bovis	4.3	1991	185 bp	≥260 bp
P. aeruginosa	6.2	3016	206 bp	≥282 bp
S. aureus	2.8	1548	270 bp	≥381 bp
S. pneumoniae	2.2	1101	235 bp	≥300 bp
S. pyogenes	1.8	963	204 bp	≥287 bp
S. typhi	4.7	2271	235 bp	≥325 bp
V. cholerae	5	2112	190 bp	≥265 bp

*: The average length of all intergenic regions longer than 60 base pairs.

Table A.11. The oligonucleotide sequence of all primers used for Deep-RACE PCR

Sm11	CCTCTCTATGGGCAGTCGGTGATTGTAGCGTTCCGGGTGC
Sm76	CCTCTCTATGGGCAGTCGGTGATGGGTCTCTGCAGCCCTC
Sm64	CCTCTCTATGGGCAGTCGGTGATGCAGGACCGGGCTGAGTA
Sm49	CCTCTCTATGGGCAGTCGGTGATGGCCTGTCGGTCTCTCAGAC
Sm82	CCTCTCTATGGGCAGTCGGTGATGCGGCATCAGCCTGATGTC
Sm90	CCTCTCTATGGGCAGTCGGTGATCGACGTTCGCGTCTGCC
Sm38	CCTCTCTATGGGCAGTCGGTGATTCGCGATCGCCTCGGC
Sm42	CCTCTCTATGGGCAGTCGGTGATATGCTCGTTCTGCGGTGTC
Sm19	CCTCTCTATGGGCAGTCGGTGATGGAAAAAGAGGCGGACAAAAAACAAC
Sm41	CCTCTCTATGGGCAGTCGGTGATCCAGCAGGTGCAGGCC
Sm93	CCTCTCTATGGGCAGTCGGTGATTGTGTGTCCGAGCTGCAC
Sm67	CCTCTCTATGGGCAGTCGGTGATCCGCCGACGAGGC
Sm68	CCTCTCTATGGGCAGTCGGTGATCCAACAACCCGACGGTTC
Bo13	CCTCTCTATGGGCAGTCGGTGATCGGTTCGCGGTGCCC
Bo32	CCTCTCTATGGGCAGTCGGTGATGCGTCTGCTCGCGAAAATGC
Bo35	CCTCTCTATGGGCAGTCGGTGATTGTCCCCCGAATGGTGGAC
Bo46	CCTCTCTATGGGCAGTCGGTGATATACCCGTACGCTGGCGC

Bo47	CCTCTCTATGGGCAGTCGGTGATCGGGTGGTGACGTCATCC
Bo48	CCTCTCTATGGGCAGTCGGTGATCGATGATGATTCAGCCGACGC
Bo53	CCTCTCTATGGGCAGTCGGTGATAAGACCAGCCCTACCGAAGC
Bo60	CCTCTCTATGGGCAGTCGGTGATCGCACACGCTTGCTTGAACATC
Bo71	CCTCTCTATGGGCAGTCGGTGATCGAGTGATCCCCGGCAC
Bo73	CCTCTCTATGGGCAGTCGGTGATGAAAAGTCAGCGGCCCTGAC
Bo78	CCTCTCTATGGGCAGTCGGTGATGCACCAAGACGCCGCATAC
Bo81	CCTCTCTATGGGCAGTCGGTGATGGGCGCGCTCTAGGC
Bo82	CCTCTCTATGGGCAGTCGGTGATGGTTCACCCGACCGCC
Bo86	CCTCTCTATGGGCAGTCGGTGATCCAACTCACCAGTTCCAGGTGATC
Bo87	CCTCTCTATGGGCAGTCGGTGATTCACACGCCCGCTGTCTTTC
Bo96	CCTCTCTATGGGCAGTCGGTGATCGCTTATCACGCGTTGTTGGC
Bo101	CCTCTCTATGGGCAGTCGGTGATCAGCACGCCGTCAGAGTTTC
Bo105	CCTCTCTATGGGCAGTCGGTGATGAGCCGATCAGCGACCACC
Bo118	CCTCTCTATGGGCAGTCGGTGATGCGTCCAGCTACCACCAC
Bo130	CCTCTCTATGGGCAGTCGGTGATTCTGGGTGAGCCGCGTTC
Bo132	CCTCTCTATGGGCAGTCGGTGATAAGGAAGACCTCGGCGTGTC
Bo27	CCTCTCTATGGGCAGTCGGTGATGCTAGCGTGACAGGCGTC
Bo29	CCTCTCTATGGGCAGTCGGTGATCTGCCGCGAAATCCGGC
Universal 5’ primer	CCATCTCATCCCTGCGTGTCTCCGACTCAGCACTGCGTTTGCTGGCTTTGATG

APPENDIX B: SUPPLEMENTARY DATA FOR CHAPTER THREE

Supplementary figures for Chapter Three

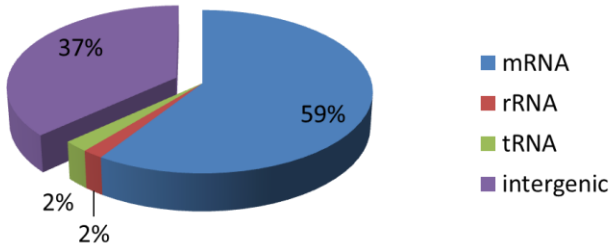


Figure B.1. Deep sequencing read distribution

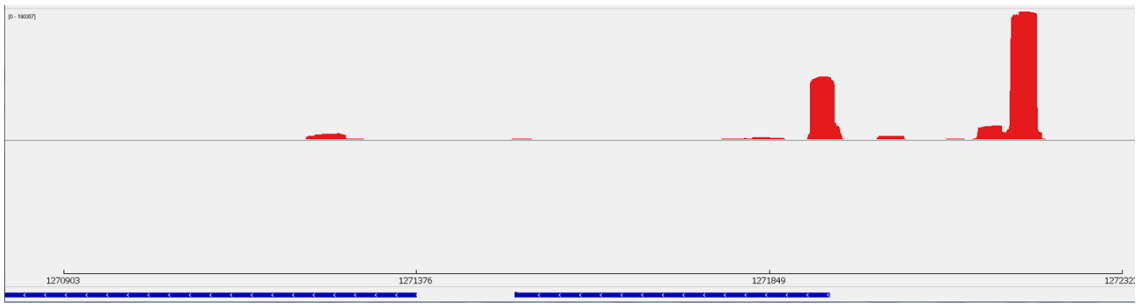


Figure B.2. Expression of the Y RNA was confirmed with the deep-seq data

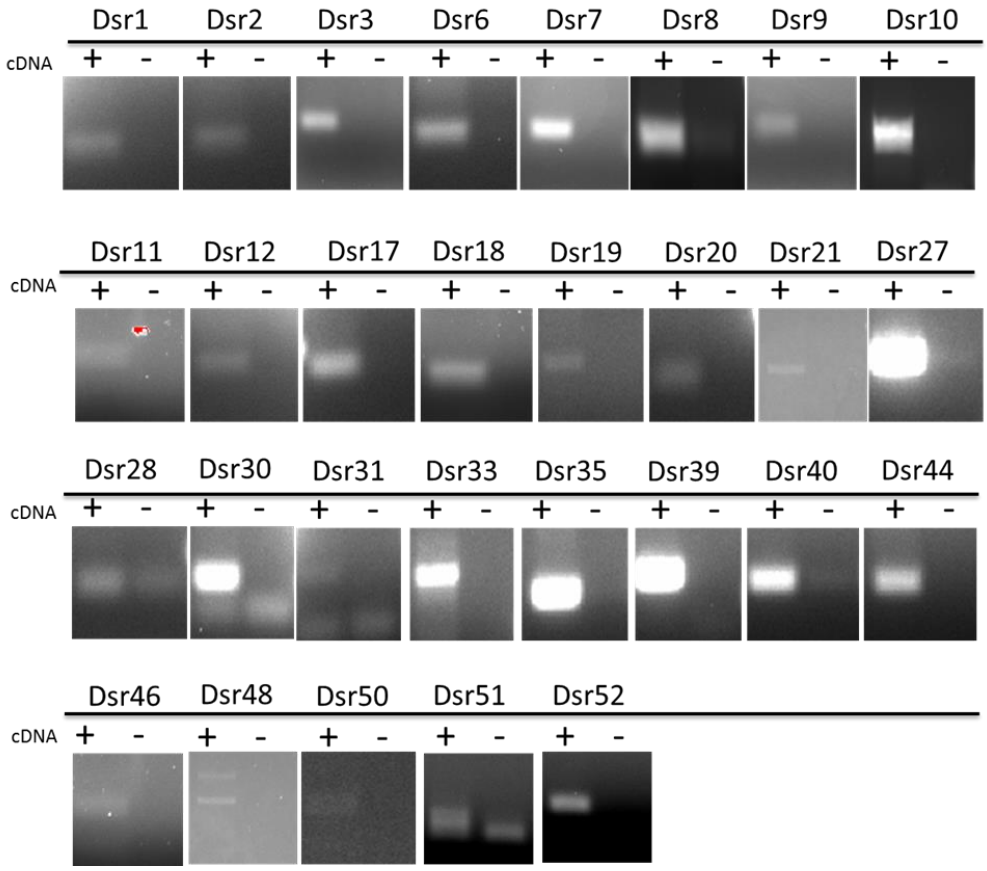


Figure B.3. RT-PCR for sRNAs identified in *D. radiodurans*

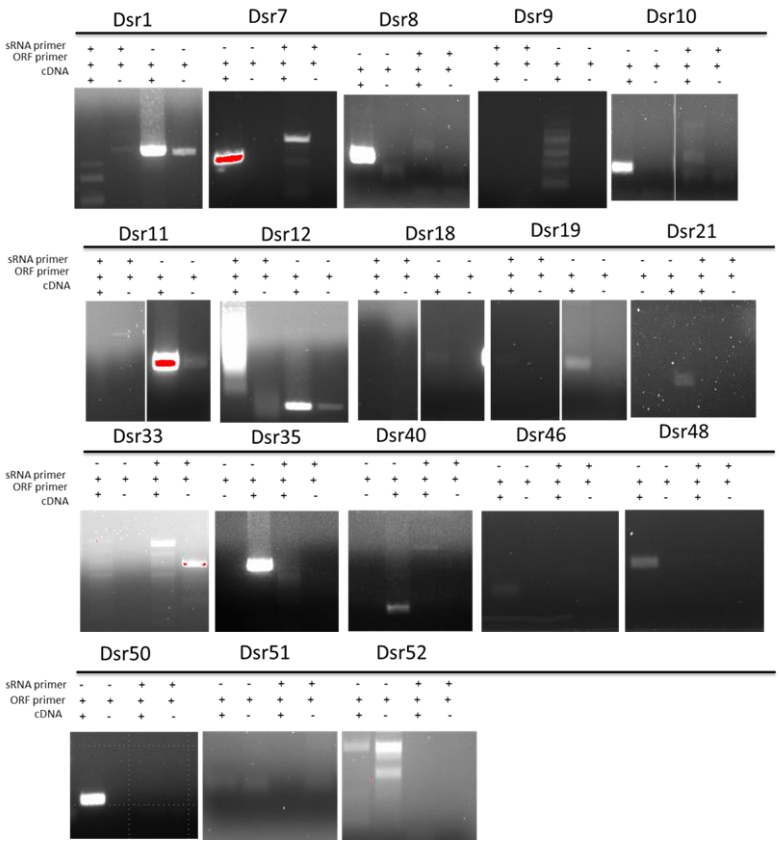


Figure B.4. RT-PCR for sRNA co-transcription verification

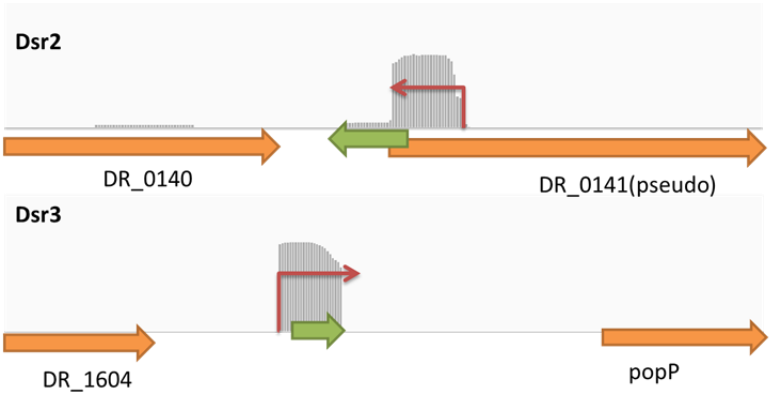


Figure B.5. Selected examples of transcription starting sites identified by 5' Deep-RACE.

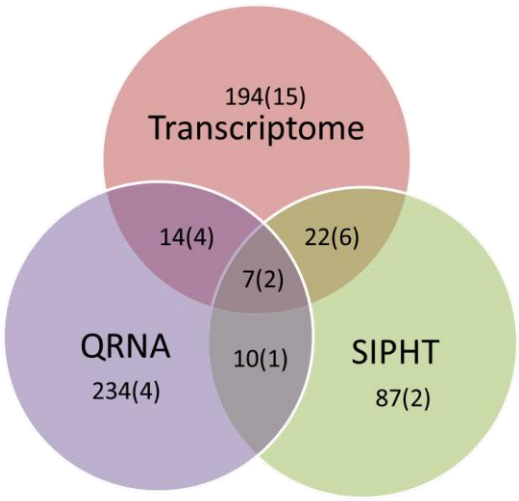
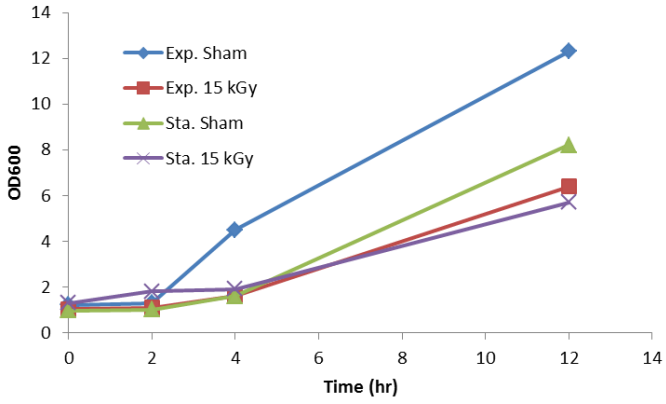


Figure B.6. Venn diagram of computational predicted sRNA candidates and deep-sequencing identified (transcriptome) sRNA candidates in D. radiodurans.

The numbers indicate the sRNA candidates that were predicted using each method, with the numbers of sRNA confirmed by Northern blotting analysis in parentheses.



Survival rate (CFU)	15 kGy after 2hrs
exponential cells	8.13E-02
stationary cells	1.38E-02

Figure B.7. Cells survival rate after 15 kGy irradiation

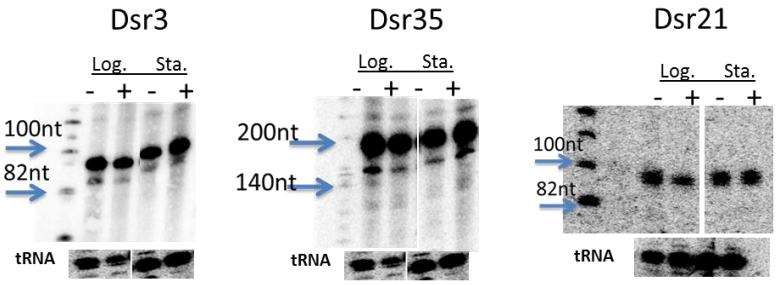


Figure B.8. Selected sRNAs of non-differential expression under irradiation

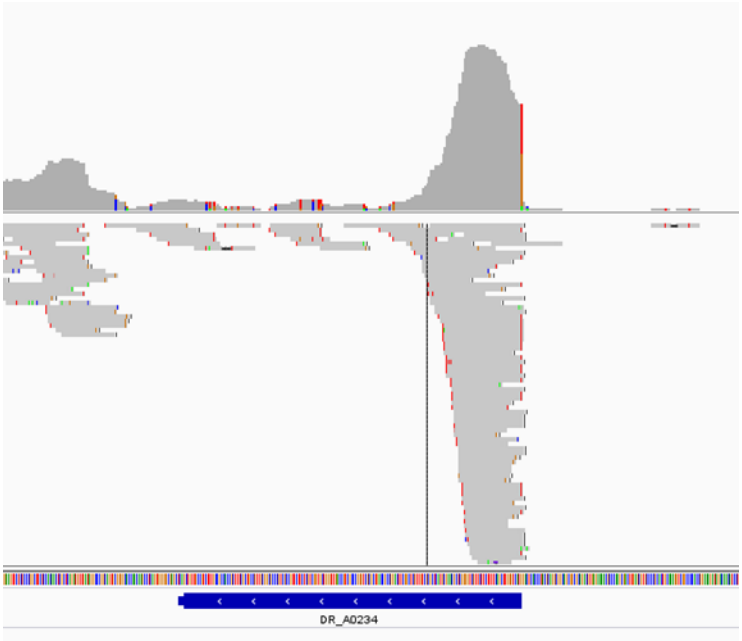


Figure B.9. Expression of DR_A0234 is confirmed with the deep sequencing analysis

Supplementary Tables for Chapter Three

Table B.1. Potential sRNAs identified with deep sequencing and computational prediction

This oversize Table was uploaded to Texas Digital Library.

Table B.2. sRNAs predicted by Rfam

Name	Left end Coordinate	Right end Coordinate	Size	Rfam Predictio	Upstream Gene		Direct ion	Downstre am Gene		Direct ion
gi|15805042|ref|NC_001263.1|	150539	150734	195	Tbox	DR_0149	MutT/nudix family protein	<<<	DR_0150	hypothetical protein	<<<
gi|15805042|ref|NC_001263.1|	2073216	2073494	278	T-Box	DR_2058	hypothetical protein	>>>	DR_2059	glycyl-tRNA synthetase	>>>
gi|15805042|ref|NC_001263.1|	863318	863553	235	tmRNA	DR_0851	4-hydroxybenzoate octaprenyltransferase	<<<	DR_00852	hypothetical protein	>>>

Table B.3. Probes for Northern Blotting Analysis

	Reverse	Forward
Dsr1	TTT CCT TCA GAG TAT TAC TCC A	T GGA GTA ATA CTC TGA AGG AAA
Dsr2	CTC CTG GGA GTG TGA AAT CAA CA	TG TTG ATT TCA CAC TCC CAG GAG
Dsr3	CCC AAA GAC AAG AGA CGG AAA TT	AAT TTC CGT CTC TTG TCT TTG GG
Dsr4	CCC CTT CCA GAC TTC TTT TCG T	A CGA AAA GAA GTC TGG AAG GGG
Dsr5	CCG CCG AGC TTT GAG ACC CCC T	A GGG GGT CTC AAA GCT CGG CGG
Dsr6	TGG CAA CAG GCG TCC CAG ACA	GCC TGA TCA TCT GTC TGG GAC
Dsr7	CTG CGC CAG CTA AAG ACA CCG A	TCG GTG TCT TTA GCT GGC GCA G
Dsr8	GCG AAG TAA GAT TTT CAC TGA CAA G	CTT GTC AGT GAA AAT CTT ACT TCG C
Dsr9	GTG ATT CAC GTT CCG GCA AAG AA	TT CTT TGC CGG AAC GTG AAT CAC
Dsr10	CAA AAA TCC CCT CCC ACG ACG GGG AGG G	CAA AAC CCC CTC CCC GTC GTG GGA GGG GAT TTT TG
Dsr11	GAA CAG CCG TTG CCT GAT CCG ACT	AGT CGG ATC AGG CAA CGG CTG TTC
Dsr12	CAT GAA CCA GTA CGA TCT CAA	TTG AGA TCG TAC TGG TTC ATG
Dsr13	ACT CCA CGA GTA TCA GGG CAA	GCA CTT CCT TGC CCT GAT ACT C
Dsr14	CCG AAG TTC AGG ACA TTG AAG TGA	CCT CAC TTC AAT GTC CTG AAC

Dsr15	GCA GTC AGC ATC AGT GTC TTT	CCT AAG ATG TGA CTC GCT ATG AA
Dsr16	CGC ATG GCC TCT TTT CCT GCC	GGC AGG AAA AGA GGC CAT GCG
Dsr17	CAT CAC AAG GTC GGG CCA TCA	TGA TGG CCC GAC CTT GTG ATG
Dsr18	ATT CAG TCT TTG TCA AGG CTA TGC	GCA TAG CCT TGA CAA AGA CTG AAT
Dsr19	TTA TCC ACA GTG AAA CGT AGA CCT	AGG TCT ACG TTT CAC TGT GGA TAA
Dsr20	GCG GGG AAA CTC CTG GTC GCG	CGC GAC CAG GAG TTT CCC CGC
Dsr21	CTG TGT TCA CGC TTC GGG CC	GGC CCG AAG CGT GAA CAC AG
Dsr22	CGG GAA TGT TCG CAA AGT CTT CGT C	GAC GAA GAC TTT GCG AAC ATT CCC G
Dsr23	CTC AGA AAC CTC GGG TTC AGT	TGC CGA CAC TGA ACC CGA GGT
Dsr24	CCTGACCTTTAGTGTCTCGCGCTTGGTAGGG	TAC CAA GCG CGA GAC ACT AAA GGT C
Dsr25	GGT GTA GAC TTC AGG CAC CCA AAC ATC TCT	AGA GAT GTT TGG GTG CCT GAA GTC TAC A
Dsr26	GGATACGAACTGCCCGCCTAGTGACAGTGA TCGT	ACGATCACTGTCACTAGGCGGGCAGTTCGTA TCC
Dsr27	CTT GAG GCT TCT ATG ATC AAT TTC ATC GTC AGA CGA	TCG TCT GAC GAT GAA ATT GAT CAT AGA AGC CTC AAG
Dsr28	GAC CTT TAG ACC CTC TTC CCT TAG AAC TT	AAG TTC TAA GGG AAG AGG GTC TAA AGG TC
Dsr29	GTACAAGTTAGCGTGTGAGGCTCAGATTTTC CACC	GGTGGAAAATCTGAGCCTCACACGCTAACTT GTAC
Dsr30	AGA AAA CAG TTC TCG AAT GGT CAG TGG CGC AAG CTG CT	GGA CGA GGG CTG TCA GGT GCA CAC
Dsr31	AAC AGT TCT CGA ATG GTC AGT GGC GCA A	GGT TGA CGC TGA CCG GGA CAT AAA
Dsr32	ATG AGG GCA GTC GAG ATG TAT TCC ATT GGG GT	CCC AAT GGA ATA CAT CTC GAC TGC CCT CAT
Dsr33	TTT ATC CAG AGT CGG CGC AGG GAC C	GGT CCC TGC GCC GAC TCT GGA TAA A
Dsr34	CGA CCC TCA ACC ACG GCC TGA CCA GCA T	ATG CTG GTC AGG CCG TGG TTG AGG GTC G

Dsr35	GGT TGT TCT TGT TGA TTT CGG TCA TC	GA TGA CCG AAA TCA ACA AGA ACA ACC
Dsr36	TAA GAC CTG CTT AGA AGC TT	CTT CTA AGC AGG TCT TAG CG
Dsr37	GCC GCG ATT CCA GAG TTC GGA AGA T	GCC CAT CTT CCG AAC TCT GGA ATC G
Dsr38	CTG AGC GCA ATG CCT GAA CGT GT	ACA CGT TCA GGC ATT GCG CTC AG
Dsr39	TTG GTG GTG AGG TCG AGG TCA GAC GAA TAC A	T GTA TTC GTC TGA CCT CGA CCT CAC CAC CAA
Dsr40	AAG TGT GGC AAC CTA ATC GGC TTA CC	GGT AAG CCG ATT AGG TTG CCA CAC TT
Dsr41	CAC CTC ACA CAA GCG AGA CGA CGA CAT GAA	TTC ATG TCG TCG TCT CGC TTG TGT GAG GTG
Dsr42	CGCCTCGGACTCGAACCGAGAACCC	GGGTTCTCGGTTCGAGTCCGAGGCG
Dsr43	GCGAGTGAGAACAGAATTTGATGTGTTG	TGG GTT GAG GGT TTC GCT GGA GGT
Dsr44	CAC AGC TTT CGC CTG AAT GCC CGT CAC TC	CGT TTG CTG GTC ACC GGG GAA CTT CTC C
Dsr45	CAA TCC ACT CTG ACA GGC CCC TGA CAG	CTG TCA GGG GCC TGT CAG AGT GGA TTG
Dsr46	AAA GGA GTT ACG CAC GCT CTG GCC GTC CA	TGG ACG GCC AGA GCG TGC GTA ACT CCT TT
Dsr47	CGC TCT GAG TCA AAG ACT CCG GCA GGC AGA A	TTC TGC CTG CCG GAG TCT TTG ACT CAG AGC G
Dsr48	TGA ATC TGG CGA GCT TCC AGT AAC CGA CAG GCC A	TGG CCT GTC GGT TAC TGG AAG CTC GCC AGA TTC A
Dsr49	GGG TCT TTC TAT GGG CTA CAG CTC AG	CTG AGC TGT AGC CCA TAG AAA GAC CC
Dsr50	CACAGTCGGCTACGTCGCCATGAAGGC	GCCTTCATGGCGACGTAGCCGACTGTG
Dsr51	GCA GCC ACT CCA GAA ATT CAC TCA GAC CTT TCA	TGA AAG GTC TGA GTG AAT TTC TGG AGT GGC TGC
Dsr52	CTT GTT GAG AAA GTT CTC AAA GAG CCG TTG GTC	GAC CAA CGG CTC TTT GAG AAC TTT CTC AAC AAG
Dsr53	CGA ATA CTG CAC CTG AAC TTG TTC ACC ACA GAG GAG GTC	GAC CTC CTC TGT GGT GAA CAA GTT CAG GTG CAG TAT TCG
Dsr54	GCGAGTCTGACCGGAGACTGTCC	GGACAGTCTCCGGTCAGACTCGC

Gsr1	TGT TCT TGC TGG TGG GCT TCG TGG CGG TGA	TCA CCG CCA CGA AGC CCA CCA GCA AGA ACA
Gsr2	CCA TGA AGG GCC TGC GCG AGT TGA TTG ATT G	CAA TCA ATC AAC TCG CGC AGG CCC TTC ATG G
Gsr3	TCG CCC TGC CCC ACT TGA CCA AGC GCA GAT T	AAT CTG CGC TTG GTC AAG TGG GGC AGG GCG A
Gsr4	AA AAA AGC CCC AGC CGG AGC CAG GGC	GCC CTG GCT CCG GCT GGG GCT TTT TT
Gsr5	G CGG CGA CCA GCC GAA TCA CGT CGA AGG AAT TGC GAA	TTC GCA ATT CCT TCG ACG TGA TTC GGC TGG TCG CCG C
Gsr6	A GCG GAG GGC CAG TGT TGG CGC CTC	GAG GCG CCA ACA CTG GCC CTC CGC T
Gsr7	AAG CAG CGG CAC CCG GAA ATC ATC C	GGA TGA TTT CCG GGT GCC GCT GCT T
Gsr8	CTC CTG CCC AAC CCA TCA TGG TTG GCC ATA G	CTATGGCCAACCATGATGGGTTGGGCAGGA G
Gsr9	AGT TAT CCA CAG TGA AAC GCG TGC CTG TGG ATA ACT	AGTTATCCACAGGCACGCGTTTCACTGTGGA TAACT
Gsr10	TAG TGT CTC GTG CTT GGT AGG GCC GAG GCG A	TCGCCTCGGCCCTACCAAGCACGAGACACTA

Table B.4. RT-PCR Primers

Dsr1_Left	GTAGGCAGTGTCCACCCTTG
Dsr1_Right	GTTTGGTGCCCCTTCTTTTT
Dsr1_Gene_left	ACGCCGACTACGTCAAGAAC
Dsr1_Gene_Right	CGTCGTGCGGAAGTAGAACT
Dsr2_Left	GCAGCGGATTCTGTTGATTT
Dsr2_Right	GGGTAACTGTTTGCGGTCTC
Dsr3_Left	TATCAACGCGACAGGAAAAA
Dsr3_Right	GACGGAAATTCATCGACAGG
Dsr5_Left	GCTCGGCGGAGCTTAGAA
Dsr5_Right	CCGGCCTGAAGTTCTTCTTC
Dsr6_Left	CCTCAAGCGACCATCCTG
Dsr6_Right	GTGCAAGACCCGACAGTG
Dsr11_Left	CTGACAAAACAGTGGCTTCC
Dsr11_Right	TGATCCGACTGACGACTGAG
Dsr11_Gene_left	GGCTACCTCGCCTACAAGG
Dsr11_Gene_right	GTCGCCTTGAACTTCTGGAG
Dsr12_Left	ATGTTCGGGTTCAGGATCAG
Dsr12_Right	GCACCGACAGCAACACTATG
Dsr12_Gene_left	GGCCTTCTTGGTCTTCCACT
Dsr12_Gene_Right	ATCCTGAACCCGAACATCAG
Dsr17_Left	CTCCAGACGAAAGCCTCATC

Dsr17_Right	CCATCAAGATTCACCGCTCT
Dsr18_Left	GGCTTTTTCGTTTTGGGTTT
Dsr18_Right	GACGTCAGGCCGATTATTCA
Dsr18_Gene_left	CGAGCACCTCGAACAAGAG
Dsr18_Gene_Right	GATCGGAAGCTCGATTTTGA
Dsr19_Left	TTTGAAGCGGATTTTCAAGC
Dsr19_Right	AAAACGCTGAACAGTCATGC
Dsr19_Gene_left	TTTTCATGGGGCAGTCACTC
Dsr19_Gene_right	GGCCTGTGGATTCTTGAGG
Dsr20_Left	ACCTTCCCTGCCGAAAAAG
Dsr20_Right	GCGGGGAAACTCCTGGTC
Dsr27_Left	CATCTGCTGTTGGTCCCTTT
Dsr27_Right	GGCGTGAGCTACACCTTCTT
Dsr28_Left	TGATGCTCCTTCGGAAAAGT
Dsr28_Right	CCTTTAGACCCTCTTCCCTTAGA
Dsr28_Gene_left	TCGGCTTCGTCGTAGAACTT
Dsr28_Gene_Right	GTGCAGAAGAAGCTCGATCC
Dsr30_Left	CTGGTCCTAAGCCATGCACT
Dsr30_Right	GACCGTATGACTCCCAGACC
Dsr31_Left	GCCACTGACCATTCGAGAAC
Dsr31_Right	CATAAGCCCCCAGAAAACAG
Dsr33_Left	CTGCGCCGACTCTGGATA
Dsr33_Right	ACGTGCAGGGTCACCTTTAT
Dsr33_Gene_left	GTCAGGTAGAGGCCGAACAC
Dsr33_Gene_right	ACGATTCTGGAGTGGTACGC
Dsr35_Left	CGAACCGTTTTCGTCTGAAT
Dsr35_Right	GCGTGCTTCAGTGTTTTGTC
Dsr35_Gene_left	TGAGGAGACGGATGACATTG
Dsr35_Gene_right	ATTGGGGAAAGTCGATGATG
Dsr39_Left	TGACCTCATGGGAAGCTCAT
Dsr39_Right	GTCAGGACGTGCAGGAAGAT
Dsr40_Left	CAAAATGCTCAGCAATGGAA
Dsr40_Right	TGGCAACCTAATCGGCTTAC
Dsr40_Gene_left	ATCGCGTGTTTCTCTGCTTT
Dsr40_Gene_Right	GCTTGAAGCGACCCTCTCT
Dsr44_Left	TGGCAAGTCCTGAGAGTCAA
Dsr44_Right	TGGAGTAGTTTCGGTGCTGTT
Dsr46_Left	CAGAGCGTGCGTAACTCCTT
Dsr46_Right	AGCCGGATCACGTCAAAG
Dsr46_Gene_left	GAGATCAACACCGTGGAGGT
Dsr46_Gene_Right	CGGTCTTTTCGAGGTCGTC
Dsr48_Left	CGCCTCTGCTCTCTCTGTTT
Dsr48_Right	ATCTGGCGAGCTTCCAGTAA
Dsr48_Gene_left	ATTCAGCGCCACGTTGTAG
Dsr48_Gene_right	CCCGAGAGTGTCAAGGTGTT
Dsr8_left	GTAGCGCTTCGAGACTTCGT
Dsr8_right	TTCTGGAACGTGGTGAACTG

Dsr8_gene_left GCGTCTTCGAACTTCTGCTT
Dsr8_gene_right CGTCGAACAACTCATTCAGC
Dsr10_left1 TGC AGG GTC ACC TTT ATC CA
Dsr10_left2 ATAAAGGAAGGCGGGTCCT
Dsr10_right TGGACGAAAAGAAGTCTGGAA
Dsr10_dgene_left GGTCGTCCAGATTGACGTG
Dsr10_dgene_right GACCTGACGCTGCTTATCGT
Dsr10_ugene_left GTAATGCACGCCGAGGTAAT
Dsr10_ugene_right GAGCCAGGATTTTCCTTTCC
Dsr13_left, GGGAATGACCCCCCGGTGGGTGTG
Dsr13_right, TTT ATC GGG ACA AAT GAA CAT CT
Dsr13_dgene_left1, GGTCGCTGTTCTTTTCGTTC
Dsr13_dgene_left2, ATGCCACCGAAGATGTTGAT
Dsr13_dgene_right1,ATTGCGCTGGATACCAAGTT
Dsr13_dgene_right2, CTACGAAGCGCGTGAAGTC
Dsr37_left, GGTATCCCGCCCATCTTC
Dsr37_right, GCCGCGATTCCAGAGTTC
Dsr37_Dgene_left1, CGGCACCACTGAACTCATTA
Dsr37_Dgene_right1, GCCACGAGGTCATTTTCAG
Dsr37_Dgene_left2, ATCGAACCGCAGGAGAAGTA
Dsr37_Dgene_right2, CACTCCTCGTCCACCCACT
Dsr37_Ugene_left1, ACCTGACCTTCGACCTGATG
Dsr37_Ugene_right1, ACAAACTCGCCCTTCTTGAC
Dsr37_Ugene_left2, CGAGCTGCACCTGAGTATCC
Dsr37_Ugene_right2, TGGGTCATGCTCAGAAACTG
Dsr50_left1, GTGATCTCGGCGTCTTTCTC
Dsr50_left2, GTGATCTCGGCGTCTTTCTC
Dsr50_right1, TCA AGT TGA GGG TGA TCT C
Dsr50_right2, GCATGATCAAGTACCGTAAGCA
Dsr50_Dgene_left1, GGAAGTTGACGTTCACGTTG
Dsr50_Dgene_left2, CCACGTTGATCCAGTGACTC
Dsr50_Dgene_right1, GACGACCTGACCCTTCAGAG
Dsr50_Dgene_right2, CCACGTTGATCCAGTGACTC
Dsr51_left1, TGCCTTCTTTTCTGACAATCC
Dsr51_left2, ACGCCGTTAATTGCCTTCTT
Dsr51_right1, AAGACTTCAGGGGTGGGAAG
Dsr51_right2, GGTGGGAAGGATTGTCAGAA
Dsr51_Ugene_left1, TGCTCGGTGTTTTGTACTCG
Dsr51_Ugene_left2, AGAAGACCAGCTCGCTCAAG
Dsr51_Ugene_right1, CCAGACTCAGGAAGGTCAGG
Dsr51_Ugene_left2, AAGATGGCACAGCTTTTTGG
Dsr52_left1, GACCAACGGCTCTTTGAGAA
Dsr52_left2, CAGGGCGACCAACGGCTCTTTGA
Dsr52_right1, GCCGCTCTCTTGTTGAGAAA
Dsr52_right2, GAA GCC TCC CCA GGG AGC
Dsr52_Ugene_left1, AGGTCATGGCGTTCAAGAGT
Dsr52_Ugene_left2, ACCAGAACACCGGCAAGA

Dsr52_Ugene_right1, CGGGAGCATGAGTTTCATC
Dsr52_Ugene_right2, GAC GAA ATT CGC GGC AAG G

Table B.5. Predicted sRNA by QRNA and SIPHT

This oversize Table was uploaded to Texas Digital Library.

Table B.6. Target RNA prediction

Gene	Target protein	Binding energy (kcal/mol)	Binding sRNA
DR_0095	ABC transporter ATP-binding protein	-12.51	Dsr10
DR_0095	ABC transporter ATP-binding protein	-14.02	Dsr13
ruvA	Holliday junction DNA helicase RuvA	-10.88	Dsr13
DR_2574	transcriptional regulator	-10.15	Dsr14rc
DR_0601	DNA primase	-10.11	Dsr14rc
DR_0335	ATP-dependent RNA helicase	-13.03	Dsr25
DR_1696	DNA mismatch repair protein MutL	-15.63	Dsr26
DR_1477	DNA repair protein	-16.08	Dsr26
recA	recombinase A	-11.5	Dsr26
DR_0601	DNA primase	-12.05	Dsr32rc
DR_1648	ABC transporter ATP-binding protein	-19.16	Dsr37
DR_1105	DNA repair protein RadA	-10.57	Dsr37rc
DR_0406	ABC transporter ATP-binding protein	-10.49	Dsr8
DR_1581	ABC transporter ATP-binding protein	-11.78	Dsr8
DR_0881	cation transport system protein	-15.25	Dsr8
DR_1105	DNA repair protein RadA	-11.27	Dsr8rc
DR_1581	ABC transporter ATP-binding protein	-10	NB-Dsr1
DR_1572	helicase-related protein	-10.62	NB-Dsr1
DR_0074	TetR family transcriptional regulator	-11.4	NB-Dsr1
DR_0475	ABC transporter ATP-binding protein	-11.99	NB-Dsr11
DR_0997	CRP/FNR family transcriptional regulator	-12.48	NB-Dsr12
DR_0416	ppGpp-regulated growth inhibitor suppressor ChpR/MazE	-12.57	NB-Dsr12
DR_2145	ABC transporter ATP-binding protein	-10.01	NB-Dsr17
DR_0601	DNA primase	-12.32	NB-Dsr17
DR_2418	DNA-binding response regulator	-13.72	NB-Dsr17
ruvA	Holliday junction DNA helicase RuvA	-12.32	NB-Dsr17

DR_0743	response regulator	-10.57	NB-Dsr18
recA	recombinase A	-13.29	NB-Dsr2
DR_0144	transposase	-28.76	NB-Dsr2
DR_0065	ATP-dependent helicase	-11.89	NB-Dsr3
DR_2519	MerR family transcriptional regulator	-11.29	NB-Dsr3
DR_1927	transposase	-20.9	NB-Dsr30
DR_0440	Holliday junction resolvase	-10.06	NB-Dsr31
DR_1379		-15.29	NB-Dsr33
DR_1581	ABC transporter ATP-binding protein	-18.17	NB-Dsr44
DR_1126	single-stranded-DNA-specific exonuclease	-10.69	NB-Dsr44
DR_1379	TetR family transcriptional regulator	-10.79	NB-Dsr44

Table B.7. Probes for 5' RACE

5'RACE_C1	CAAGCAGAAGACGGCATACGACTC CAG AAC TTT GGA AAC GAC GC
5'RACE_C2	CAAGCAGAAGACGGCATACGAGTC TCC TGG GAG TGT GAA ATC AAC A
5'RACE_C3	CAAGCAGAAGACGGCATACGACCG TCT CTT GTC TTT GGG CAG T
5'RACE_C4	CAAGCAGAAGACGGCATACGAGTC CCC TTC CAG ACT TCT TTT CGT
5'RACE_C5	CAAGCAGAAGACGGCATACGACCG AGC TTT GAG ACC CCC T
5'RACE_C6	CAAGCAGAAGACGGCATACGACAG CCT GAT CAT CTG TCT GGG AC
5'RACE_C7	CAAGCAGAAGACGGCATACGAGGT GTC TTT AGC TGG CGC A
5'RACE_C9	CAAGCAGAAGACGGCATACGAGTG ATT CAC GTT CCG GCA AAG AA
5'RACE_C11	CAAGCAGAAGACGGCATACGAAAC AGC CGT TGC CTG ATC C
5'RACE_C12	CAAGCAGAAGACGGCATACGAGGA TCA GGT TGA GAT CGT ACT GGT T
5'RACE_C14	CAAGCAGAAGACGGCATACGAGGG CCT CAC TTC AAT GTC CTG AA
5'RACE_C16	CAAGCAGAAGACGGCATACGACGC ATG GCC TCT TTT CCT GC
5'RACE_C17	CAAGCAGAAGACGGCATACGATGA TGG CCC GAC CTT GTG ATG
5'RACE_C18	CAAGCAGAAGACGGCATACGAGGG GTA TTC AGT CTT TGT CAA GGC T
5'RACE_C19	CAAGCAGAAGACGGCATACGACAC AGT GAA ACG TAG ACC TGT GGA
5'RACE_C20	CAAGCAGAAGACGGCATACGAAAA CTC CTG GTC GCG CCT
5'RACE_C21	CAAGCAGAAGACGGCATACGAGCC CGA AGC GTG AAC ACA
5'RACE_C26	CAAGCAGAAGACGGCATACGAAAC TGC CCG CCT AGT GAC A
5'RACE_C27	CAAGCAGAAGACGGCATACGATCA TCG TCA GAC GAC TGA TCC AGA T
5'RACE_C29	CAAGCAGAAGACGGCATACGATAC AAG TTA GCG TGT GAG GCT CAG A
5'RACE_C35	CAAGCAGAAGACGGCATACGACGG TTC GTG AAT AGG TTG TTC TTG T
5'RACE_C39	CAAGCAGAAGACGGCATACGAGGT GGT GAG GTC GAG GTC A
5'RACE_C42	CAAGCAGAAGACGGCATACGACCT CGG ACT CGA ACC GAG AA
5'RACE_C1B	CAAGCAGAAGACGGCATACGA GGT GCC CCT TTT CCT TCA GAG TAT T
5'RACE_C2A	CAAGCAGAAGACGGCATACGA AAA TCA ACA GAA TCC GCT GCC CA
5'RACE_C4B	CAAGCAGAAGACGGCATACGA GGT TCC AAC GGA CGG CTC A

5'RACE_C4A	CAAGCAGAAGACGGCATACGA GAA GGG GAC GCC CGT TTT
5'RACE_C9A	CAAGCAGAAGACGGCATACGA TCC AAC CCA TCC TGG TTG GC
5'RACE_C16B	CAAGCAGAAGACGGCATACGA GCC GTA TAA TAC CCG ATT CCG GAC
5'RACE_C19B	CAAGCAGAAGACGGCATACGA GAA CAG TCA TGC TTG AAA ATC CGC T
5'RACE_C20A	CAAGCAGAAGACGGCATACGA GGC AGG GAA GGT GAG GAA AGT
5'RACE_C26A	CAAGCAGAAGACGGCATACGA GAT CGT CCG GTC TGG GAG AA
5'RACE_C7M	CAAGCAGAAGACGGCATACGA CAG CTA AAG ACA CCG ACG CTG AA
5'RACE_C17M	CAAGCAGAAGACGGCATACGA CAA TCA TCA CAA GGT CGG GCC AT
5'RACE_C21M	CAAGCAGAAGACGGCATACGA CTG TGT TCA CGC TTC GGG C
5'RACE_C39B	CAAGCAGAAGACGGCATACGA GTT CGC GCA TGA GGG CTT T
5'RACE_Univ	AATGATACGGCGACCACCGCACTGCGTTTGCTGGCTTTGATG

APPENDIX C: SUPPLEMENTARY DATA FOR CHAPTER FOUR

Supplementary figures for Chapter Four

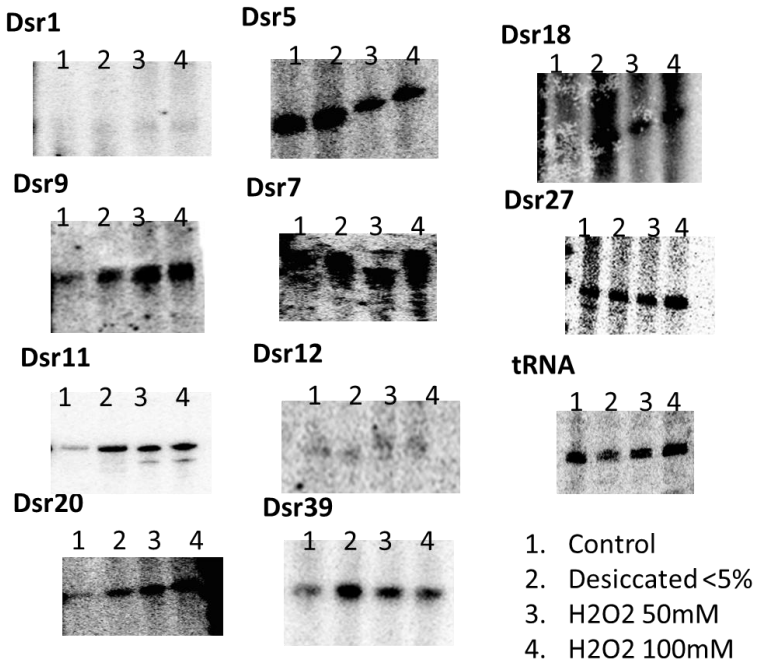


Figure C.1. Northern blotting analysis for sRNAs of interest under different environmental stresses

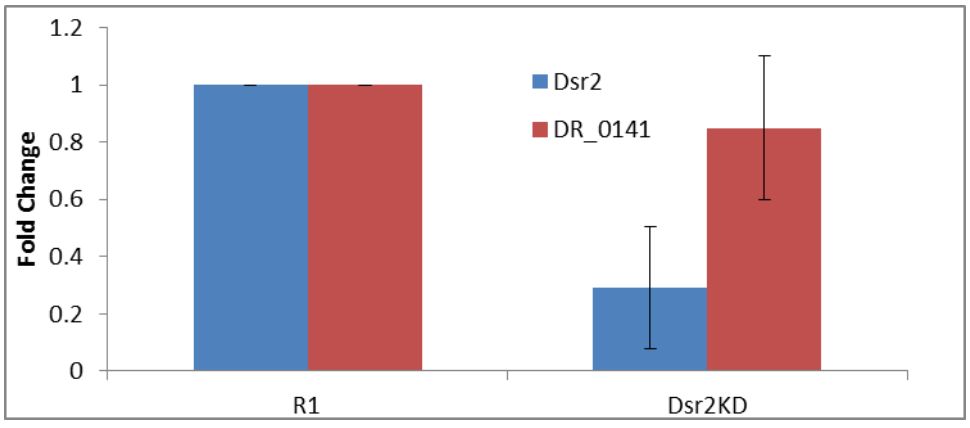


Figure C.2. Quantitative RTPCR of Dsr2 and DR_0141 under 10kGy acute irradiation

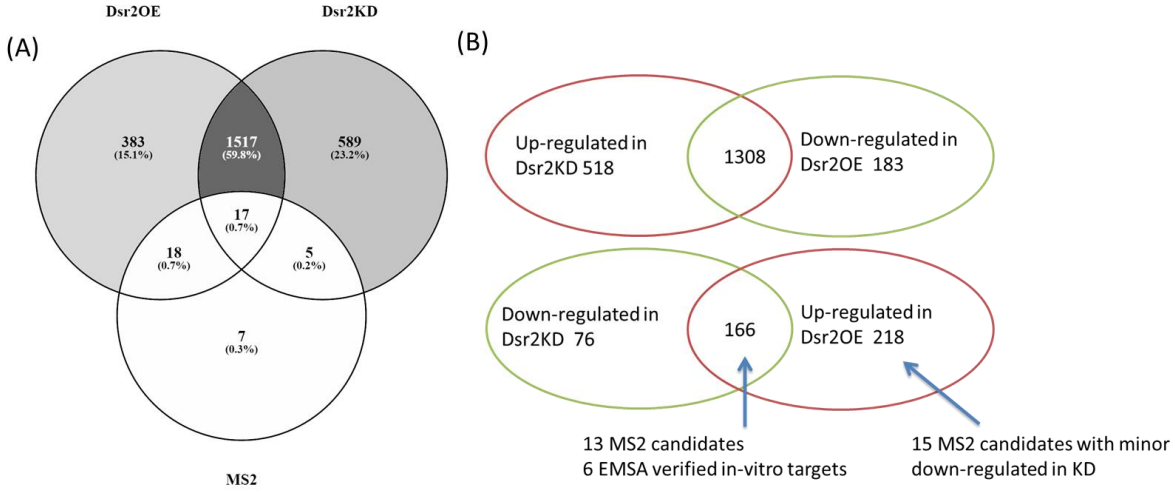


Figure C.3. (A) Venn diagram of differential expressed gene in Dsr2KD and Dsr2OE under sham condition overlapping with the MS2 pull-down candidates (B) Expression profile of genes in Dsr2KD and Dsr2OE.

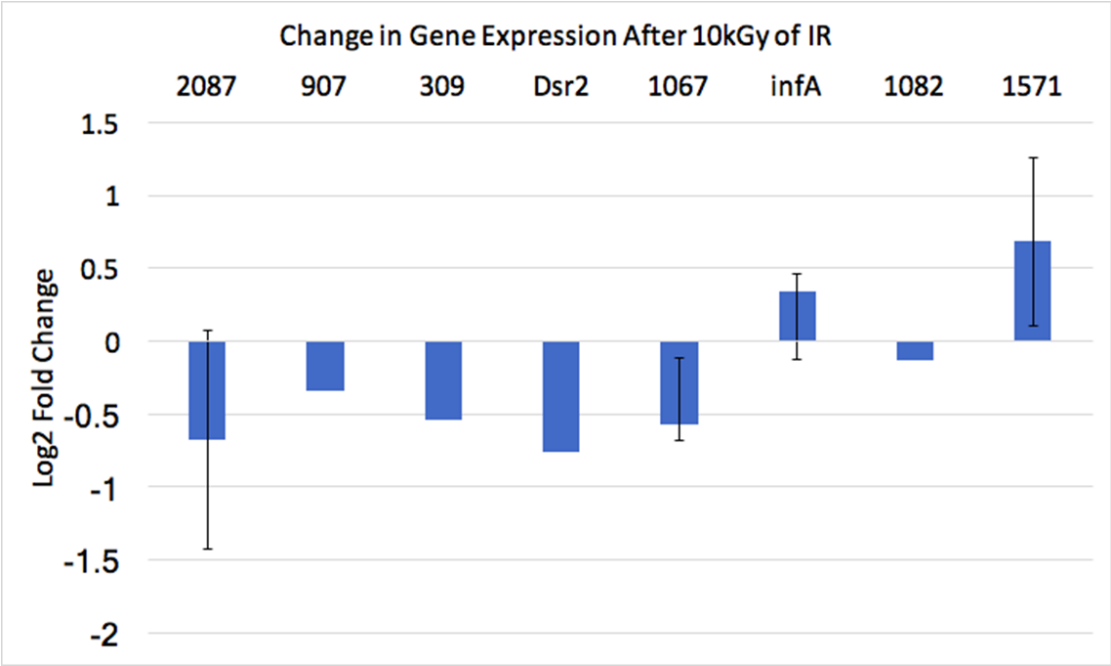


Figure C.4. Quantitative RTPCR of Dsr2 and EMSA verified in-vitro targets under acute irradiation

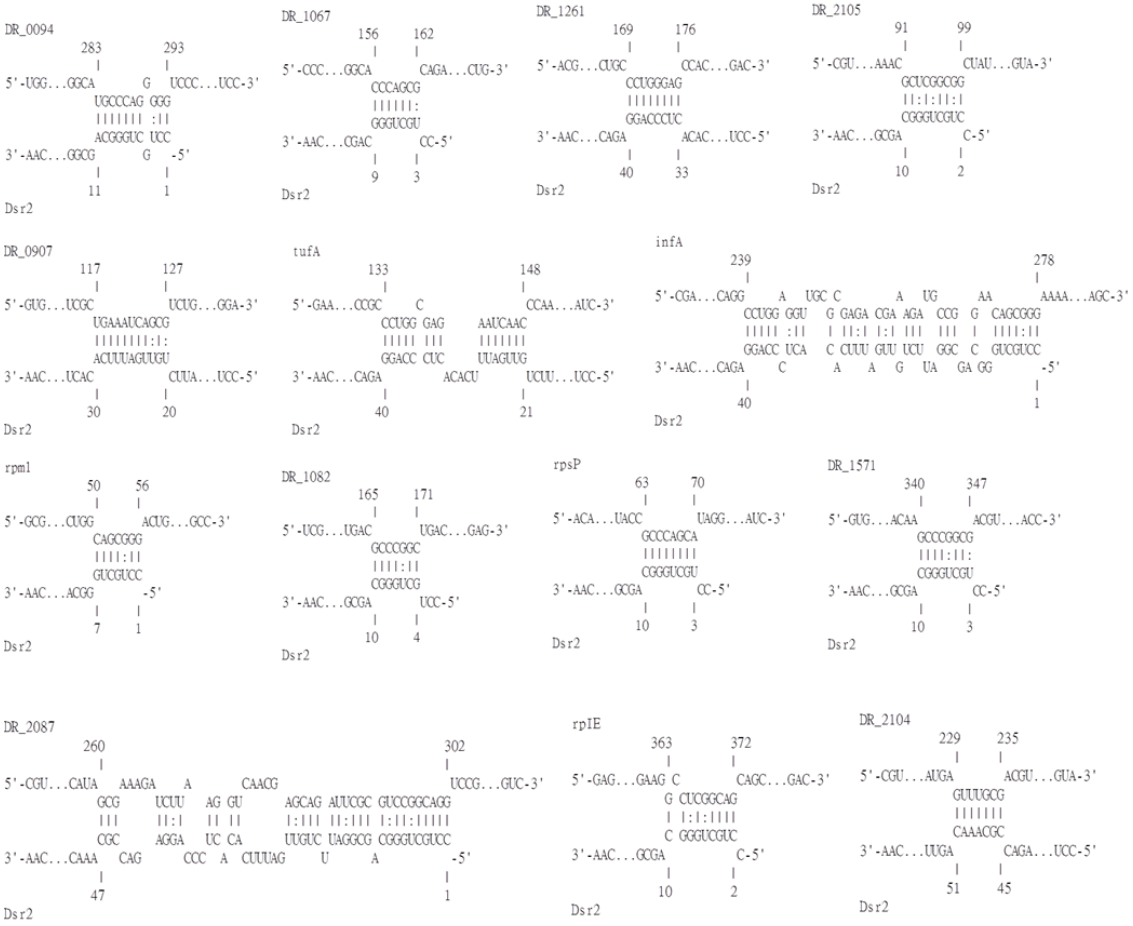


Figure C.5. Predictions of Dsr2-mRNA binding sites by IntaRNA

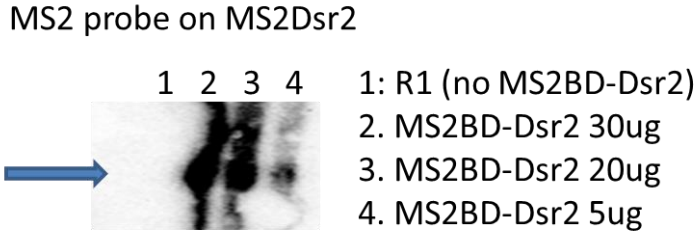


Figure C.6. Northern blotting analysis confirmation of MS2BD-Dsr2 transcript

Supplementary figures for Chapter Four

Table C.1. Differential expressed sRNA under acute irradiation from previous study

sRNA		Size	5' end	3' end	Co-transcription
Dsr2		75	139835	139760	
Dsr5		70	1133980	1133910	
Dsr7		110/150	1678925	1679035/1679075	yes
Dsr12		90/400	99987	100077/100387	yes
Dsr18		200	2615779	2615579	
Dsr27		400	364790	364390	
Dsr30		82/200	1948885	1949103	
Dsr39		250	1029203	1028950	

Table C.2. Differential expressed genes in Dsr2KD and Dsr2OE

This oversize Table was uploaded to Texas Digital Library.

Table C.3. Potential mRNA candidates of Dsr2 verified with MS2 pull-down (HITS-CLIP)

		pull down FC	Dsr2KD FC	Dsr2OE FC
groES	10 kDa chaperonin	4.3258	0.4741	3.9639

rpmF	50S ribosomal protein L32	2.8665	0.6173	7.7491
DR_1571	ABC transporter-binding protein	2.7482	0.3241	19.6356
rpmA	50S ribosomal protein L27	2.7212	0.5794	15.0007
infA	Translation initiation factor IF-1	2.6414	0.5796	12.0576
DR_1245	DdrB partner	2.5027	0.7560	19.0516
DR_1314	heat shock response	2.4003	3.2993	27.1591
DR_t31	tRNA	2.3752	1.9969	
tufA	Elongation factor Tu	2.3517	0.4105	12.4596
DR_t32	tRNA	2.3222		
DR_1172	desiccation resistance protein	2.3091	1.4346	10.6358
rpmE	50S ribosomal protein L31	2.2793	0.6157	6.8524
DR_1082	Ribosome hibernation promotion factor	2.2462	0.6719	6.3068
DR_0662	mRNA interferase	2.2243	0.4381	3.7757
rpsP	30S ribosomal protein S16	2.2230	0.2753	9.3066
rpmI	50S ribosomal protein L35	2.2059	0.9090	17.5708
rplE	50S ribosomal protein L5	2.1964	0.3074	18.0111
DR_0099	Single-stranded DNA-binding protein	2.1962	0.5693	8.7954
DR_0094	Uncharacterized protein	2.1895	0.5194	2.9516
DR_1844	Uncharacterized protein	2.1859	2.1208	5.3818
DR_1317	Uncharacterized protein	2.1816	0.6221	4.5493
DR_1539	Uncharacterized protein	2.1790	1.0893	4.3081
DR_1261	Uncharacterized protein	2.1700	2.5228	18.3527
DR_2105	Uncharacterized protein	2.1651	1.7427	11.3723
DR_t15	tRNA	2.1593	0.0923	
DR_2104	Uncharacterized protein	2.1574		3.7643
groEL	60 kDa chaperonin	2.1479	0.5794	19.6983
rpsF	30S ribosomal protein S6	2.1429	0.4916	20.9339
DR_1840	Uncharacterized protein	2.1159	0.9106	5.0838
DR_1432	Uncharacterized protein	2.1158	0.9325	
DR_2344	Uncharacterized protein	2.1134	1.5128	19.2615
rpsT	30S ribosomal protein S20	2.1133	0.2685	13.0079
DR_2527	Uncharacterized protein	2.1100	0.3712	
DR_0907	Cold shock protein	2.1090	0.6624	29.8056
DR_1754	Uncharacterized protein	2.0955	0.1190	8.7172
rpsQ	30S ribosomal protein S17	2.0848	0.4739	12.9723
DR_0547	Uncharacterized protein	2.0811	1.0829	
DR_1296	Transposase	2.0747	0.2419	6.0898
DR_0781	Response regulator, OmpR/PhoB family	2.0724	0.7041	5.0859

DR_1067	Uncharacterized protein	2.0663	0.4962	3.7695
rplX	50S ribosomal protein L24	2.0629	0.3199	8.4448
DR_2287	Transcriptional regulator	2.0418	2.1804	
DR_t43	tRNA	2.0383	5.4638	
DR_0221	Uncharacterized protein	2.0200	11.0857	
DR_2389	Transglycosylase	2.0176	0.6945	5.3570
DR_2087	Translation initiation factor IF-3	2.0115	0.3638	6.7939
DR_t41	tRNA	2.0045	8.1107	

* pull down FC: Fold changes in MS2 pull-down analysis. Dsr2KD FC: Fold changes in Dsr2KD compared with R1 under sham condition. Dsr2OE FC: Fold changes in Dsr2OE compared with R1 under sham condition.

Table C.4. Proteomics analysis of the 47 MS2-pull down mRNAs.

Gene Number	ANOVA (Total Spectrum Count), Comparison Level: Category (WT,OE,KO)	t-test WTKO (Total Spectrum Count), Comparison Level: Category (WT and KO)	t-test WTOE(Total Spectrum Count), Comparison Level: Category (WT and OE)	t-test WTIR (Total Spectrum Count), Comparison Level: Category (WT versus WTIR10)	WT normalized counts	Dsr2KD normalized counts	Dsr2OE normalized counts	WT-IR10kGy normalized counts	WT/Dsr2KD FC	WT/Dsr2OE FC	WT/WT-IR10kGy FC	EMSA Binding?
DR_0309	0.004453	0.435863	0.007958	0.497717	1418.871	1487.398	1898.812	1260.058	1.048297	1.338255	0.888071	yes
groEL	0.004637	0.362365	6.57E-04	0.155018	925.2369	1136.275	1768.829	499.501	1.228091	1.911758	0.539863	no
DR_1571	0.551342	0.475604	0.297677	0.751586	651.7722	553.5073	516.4207	641.3346	0.849234	0.792333	0.983986	yes
DR_0099	0.569191	0.654854	0.597449	8.71E-04	126.6849	117.854	130.8617	242.556	0.930292	1.03297	1.91464	no
rplE	0.579939	0.681898	0.557353	0.03927	125.8113	119.4795	136.1313	176.7781	0.949673	1.082028	1.405106	yes
DR_1314	0.508163	0.558797	0.288148	0.033647	114.4533	103.2238	88.70492	224.056	0.901886	0.775032	1.95762	no
DR_2087	0.880433	0.81981	0.644215	0.15943	72.51621	72.33795	70.26132	127.4447	0.997542	0.968905	1.757465	yes
DR_1172	0.521227	0.8575	0.44127	0.309815	75.13728	75.5891	91.33971	57.55567	1.006013	1.215638	0.766007	no
DR_1245	0.380756	0.199986	0.840153	0.614024	77.75834	67.46124	74.65266	80.16682	0.867576	0.96006	1.030974	no
DR_1067	0.778476	0.612182	0.777436	0.144158	71.64252	79.65302	64.99172	45.22231	1.111812	0.907167	0.631222	yes
DR_2527	0.990259	0.893633	0.966805	0.104127	65.5267	65.83566	66.74825	43.16675	1.004715	1.018642	0.658766	no
DR_1432	0.969	0.884596	0.925624	0.303896	54.16873	56.89502	53.57426	65.77791	1.05033	0.989026	1.214315	N/A
DR_0094	0.001748	0.006065	0.004415	0.416192	20.09485	40.6393	44.79159	28.77784	2.022374	2.229008	1.4321	yes
DR_0781	0.007433	0.218268	0.001198	0.949101	32.3265	24.38358	14.05226	30.8334	0.754291	0.434698	0.953812	no
rplX	0.659023	0.655003	0.479489	0.79266	18.34747	21.13244	26.348	18.50004	1.15179	1.436056	1.008315	N/A
rpsT	0.084649	0.10382	0.084894	0.040521	10.48427	18.69408	43.91333	0	1.783059	4.188496	0	N/A
groES	0.048022	0.598673	0.032349	0.25421	13.97903	15.44293	25.46973	20.5556	1.104722	1.821996	1.47046	no
infA	0.580062	0.828228	0.344835	0.082231	17.47379	17.06851	15.8088	22.61116	0.976806	0.904715	1.294004	yes
DR_2287	0.560102	0.343025	0.879332	0.390612	16.6001	13.81736	16.68706	20.5556	0.832366	1.005239	1.238282	no
rpsF	0.238352	0.147262	0.299013	0.626886	15.72641	8.940646	10.5392	20.5556	0.568512	0.670159	1.307075	N/A
rpsQ	0.268896	0.255355	0.365577	0.331981	11.35796	8.12786	18.4436	8.222239	0.715609	1.623848	0.723919	N/A
rpmA	0.864241	0.719992	0.657767	0.474016	10.48427	11.379	11.41746	14.38892	1.085341	1.089009	1.372429	no
DR_1261	0.127455	0.591413	0.048214	0.002133	10.48427	8.12786	1.756533	39.05563	0.775243	0.16754	3.725164	yes
DR_0907	0.002	0.01081	0.01081	4.34E-04	8.736893	0	0	41.11119	0	0	4.70547	yes
rpsP	0.839				27.958	26.009	28.105	28.778	0.930288	1.005258	1.02933	yes
DR_1082	0.083				6.116	0	1.757	0				yes
DR_0662	0.219				3.495	5.69	5.27	4.11				no
rpmF	0.531				0	1.626	2.635	0				no
rpmE					not detected							no
rpmI					not detected							yes
DR_1844	0.144				5.242	9.752	6.148	6.167				no
DR_1317					not detected							no
DR_1539					not detected							no
DR_2105					not detected							yes
DR_2104					not detected							yes

*Highlight in green: p-value is lower than 0.05. Pink: unable to detect protein.

Table C.5. Predicted targets of Dsr2 by CopraRNA

Gene_ID	p-value	FDR	Annotation
dr_1460	0.00019	0.490604	
dr_0279	0.00223	0.891574	similar to GP:2995392 percent identity: 60.45; identified by sequence similarity; putative
dr_2612	0.002759	0.891574	DedA family protein
dr_a0202	0.00319	0.891574	Cu/Zn family superoxide dismutase
dr_0863	0.003494	0.891574	
dr_0768	0.005189	0.891574	UDP-N-acetylmuramoylalanyl-D-glutamyl-2 6-diaminopimelate--D-alanyl-D-alanyl ligase
dr_0781	0.00535	0.891574	response regulator
dr_1442	0.005553	0.891574	similar to GB:AL009126 percent identity: 47.96; identified by sequence similarity; putative
dr_0403	0.005745	0.891574	inosine-uridine preferring nucleoside hydrolase
dr_0907	0.005839	0.891574	CSD family cold shock protein
dr_0409	0.00628	0.891574	
dr_1560	0.006826	0.891574	similar to GB:AL123456 percent identity: 44.81; identified by sequence similarity; putative
dr_0423	0.007283	0.891574	
dr_1788	0.007517	0.891574	
dr_0368	0.008948	0.891574	similar to GB:AE000666 percent identity: 47.12; identified by sequence similarity; putative
dr_1649	0.009409	0.891574	immunogenic protein
dr_b0044	0.009435	0.891574	GGDEF family protein
dr_2489	0.010571	0.891574	
dr_2237	0.01079	0.891574	
dr_0544	0.01107	0.891574	similar to GB:Pyro_h percent identity: 56.68; identified by sequence similarity; putative
dr_0638	0.011089	0.891574	
dr_a0251	0.011401	0.891574	NADPH quinone oxidoreductase
dr_1108	0.011401	0.891574	
dr_2263	0.011438	0.891574	Dps family DNA-binding stress response protein
dr_a0143	0.012102	0.891574	3-hydroxybutyryl-CoA dehydrogenase
dr_1000	0.012707	0.891574	
dr_1201	0.01282	0.891574	
dr_1557	0.013222	0.891574	
dr_1063	0.013657	0.891574	peptidyl-prolyl cis-trans isomerase C
dr_1332	0.013665	0.891574	
dr_b0131	0.013729	0.891574	identified by sequence similarity

dr_1311	0.013782	0.891574	methionine aminopeptidase
dr_2059	0.013851	0.891574	glycyl-tRNA synthetase
dr_1779	0.014005	0.891574	
dr_a0233	0.014037	0.891574	oxidoreductase iron-sulfur subunit
dr_0179	0.014135	0.891574	putative deoxyribonucleotide triphosphate pyrophosphatase
dr_0324	0.014447	0.891574	serine cycle enzyme
dr_a0157	0.014576	0.891574	phosphate ABC transporter periplasmic phosphate-binding protein
dr_1786	0.015005	0.891574	
dr_1906	0.015229	0.891574	L-lactate permease
dr_0645	0.017322	0.946843	molybdopterin-guanine dinucleotide biosynthesis protein A
dr_1842	0.017634	0.946843	
dr_1225	0.017697	0.946843	mannosyltransferase
dr_2261	0.017727	0.946843	aldo/keto reductase
dr_1007	0.018689	0.946843	MutT/nudix family protein
dr_2505	0.019618	0.946843	
dr_0745	0.021753	0.946843	periplasmic serine protease
dr_2451	0.023359	0.946843	similar to PID:1001780 PID:1001829 percent identity: 55.47; identified by sequence similarity; putative
dr_0490	0.023725	0.946843	pyrrolidone-carboxylate peptidase
dr_1060	0.025189	0.946843	arginine/ornithine transport system ATPase
dr_a0046	0.025486	0.946843	
dr_1965	0.025618	0.946843	similar to SP:P54452 PID:1303787 GB:AL009126 percent identity: 60.49; identified by sequence similarity; putative
dr_2514	0.025723	0.946843	
dr_a0337	0.025821	0.946843	glutaryl-CoA dehydrogenase
dr_0108	0.025831	0.946843	similar to GP:3334797 percent identity: 56.76; identified by sequence similarity; putative
dr_0356	0.026844	0.946843	BioC family methyltransferase
dr_0554	0.028036	0.946843	
dr_b0086	0.028181	0.946843	potassium-transporting ATPase subunit A
dr_a0066	0.028456	0.946843	N-acetylglucosamine-6-phosphate deacetylase
dr_0954	0.028915	0.946843	succinate dehydrogenase cytochrome subunit
dr_1229	0.029331	0.946843	
dr_0959	0.029575	0.946843	peptide ABC transporter permease
dr_1355	0.030233	0.946843	phosphatidylglycerophosphatase B-like protein
dr_2121	0.031177	0.946843	branched-chain amino acid ABC transporter permease
dr_1976	0.031745	0.946843	DNA mismatch repair protein MutS

dr_a0048	0.031805	0.946843	mannose-6-phosphate isomerase
dr_0216	0.031887	0.946843	similar to GB:AL123456 percent identity: 63.45; identified by sequence similarity; putative
dr_a0006	0.032094	0.946843	similar to PID:1653488 percent identity: 48.46; identified by sequence similarity; putative
dr_2303	0.032955	0.946843	chloramphenicol acetyltransferase
dr_2574	0.033378	0.946843	transcriptional regulator
dr_1209	0.033938	0.946843	bacterioferritin comigratory protein
dr_2312	0.033987	0.946843	carbohydrate kinase
dr_0634	0.034183	0.946843	
dr_1871	0.034236	0.946843	chloromuconate cycloisomerase
dr_1388	0.035082	0.946843	
dr_0898	0.035336	0.946843	similar to GB:AL009126 percent identity: 56.70; identified by sequence similarity; putative
dr_1646	0.035692	0.946843	nitrogen regulator
dr_2586	0.03599	0.946843	
dr_1392	0.036045	0.946843	
dr_1671	0.037456	0.946843	
dr_0694	0.037731	0.946843	
dr_a0182	0.037857	0.946843	
dr_0349	0.037928	0.946843	ATP-dependent protease LA
dr_1037	0.038602	0.946843	branched-chain amino acid ABC transporter permease
dr_1328	0.039312	0.946843	similar to PID:1651948 percent identity: 69.01; identified by sequence similarity; putative
dr_1940	0.03941	0.946843	similar to PID:1001216 PID:1001280 percent identity: 48.05; identified by sequence similarity; putative
dr_1708	0.040196	0.946843	
dr_1875	0.041175	0.946843	similar to GB:AL009126 percent identity: 59.56; identified by sequence similarity; putative
dr_a0282	0.041298	0.946843	
dr_0426	0.04348	0.946843	imidazole glycerol phosphate synthase subunit HisH
dr_b0118	0.043739	0.946843	dessication-associated protein
dr_2143	0.043961	0.946843	similar to GB:U00096 SP:P77367 PID:1773171 PID:1786697 percent identity: 86.80; identified by sequence similarity; putative

Table C.6. Oligonucleotides used in this study

Dsr2 Northern primer	TTT GCG GTC TCC TGG GAG TGT
Dsr1 Northern primer	TTT CCT TCA GAG TAT TAC TCC A
Dsr2 PCR primers F	ACGCTGGCTCAGTAGGTGAC
Dsr2 PCR primers R	TCATCGCCAGGAGGACATCG
Dsr1 PCR primers F	TGCCGGAGCAGGAACAACG
Dsr1 PCR primers R	CGGCGTCGTTTCCAAAGTTCTG
DR_2087 qPCR F	AGT CAA GGC GAT CAA GTT CC
DR_2087 qPCR R	ATG GTG ACC TTG ACC TTG TG
DR_0907 qPCR F	TGG TGA GAG CAA GGA GAT TTG
DR_0907 qPCR R	CTG CCT TCC GTC TCG ATA AAG
rpsP qPCR F	CAC TAC CGC ATC GTC GTC

rpsP qPCR R	ATC GAT CTT CAG GAA GTT CTC G
DR_1571 qPCR F	CTT CGA CTA CTG GAC CAT CAA C
DR_1571 qPCR R	AGG AAG ATA CCG GAC AGG AA
DR_1067 qPCR F	CAG GAC ACC AAC AAC ACC A
DR_1067 qPCR R	GAG ACC CAG CAG ACC GA
DR_0309 qPCR F	CCG ACA AGA CCT TCC TGA TG
DR_0309 qPCR R	ACG ATT TCG ACT TCG TCC TG
Dsr2 qPCR F	CTG CTG GGC AGC GGA TT
Dsr2 qPCR R	GGG TAA CTG TTT GCG GTC TCC
infA qPCR F	GAA GAA GGA TGA GTC CGA CAG
infA qPCR R	CGC TGA TAT AAG CCA GGA TGT
DR_1082 qPCR F	TGG AAG TGC AGC TCA ATG T
DR_1082 qPCR R	CAT GTA GCG GGT CTT GAA TTT G
RplE qPCR F	CTT GAT GCC GAG GTT GTA GTT
RplE qPCR R	GCA TGT ACG TGT TCC TCG AA
RpmI qPCR F	AGG TCA TGG CGT TCA AGA G
RpmI qPCR R	TCG GGA GCA TGA GTT TCA TC
rpsP qPCR F	AAA TTC GTC TGT CCC GCT TC
rpsP qPCR	GGA AGT TCT CGC TGG TCT TG

R	
pprA qPCR F	GCAGTATGGCAAGGGCTAAA
pprA qPCR R	GCGATCTGGCTGTCCAC
T7-1571-F	GAA TTC AAT TAA TAC GAC TCA CTA TAG GGA GAG CCA GAG ACA TAA AGA GCA G
T7-0907-F	GAA TTC AAT TAA TAC GAC TCA CTA TAG GGA GAT GTT GAA CGA GAG GCA A
T7-0907-R	AGA TTT TCG ACG CGC T
T7-Gro-F	GAA TTC AAT TAA TAC GAC TCA CTA TAG GGA GAA GCT AAC AGC TGG CAA
T7-Gro-R	TTT CCT TGG CGG AAT C
1067-T7-f	GAA TTC AAT TAA TAC GAC TCA CTA TAG GGA GAA AGG GTC CGC CTC CT
1067-T7-r	GAC GCC GCG ACA GGC A
0309-T7-F	GAA TTC AAT TAA TAC GAC TCA CTA TAG GGA GAC GGC GCT CTT GCC CT
0309-T7-r	GG TCT TGC CGT GGT CGA
rpsp-T7-f	GAA TTC AAT TAA TAC GAC TCA CTA TAG GGA GAC TAA TCT CGC TGA TGA CTT
rpsp-T7-r	GTG GGC CGA GCC GAA
infA-T7-f	GAA TTC AAT TAA TAC GAC TCA CTA TAG GGA GAA AGC AGA GCA CCT GAA GTC AAG
infA-T7-r	TGT TCG GCA GCG CCT
1245-T7-F	GAA TTC AAT TAA TAC GAC TCA CTA TAG GGA GAA GAC GGT TAC GAG GTG CT
1245-T7-r	GAA CTT CTT TTT CTT GCA GGT ACT T
DR_2087 T7 F	GAA TTC AAT TAA TAC GAC TCA CTA TAG GGA GAG GTC CTG ACC GCG CAT T
DR_2087 T7 R	GGG AAA GAG AGA AGG TTC AGG
tuf_T7_F	GAA TTC AAT TAA TAC GAC TCA CTA TAG GGA GAT CAC ACC ACC AGA ACC T
tuf_T7_R	AGC CAC GCA GTT TGA TAC
DR_0907_T7_F	GAA TTC AAT TAA TAC GAC TCA CTA TAG GGA GAA ATC TTT AGA CTC TGT CCA AGC
DR_0907_T7_R	TTT ACG GTC TTG CGA TTA CC
rpsP_T7_F	GAA TTC AAT TAA TAC GAC TCA CTA TAG GGA GAG GCC TTC TTG TTG GTT CTT
rpsP_T7_R	CGA TAC CCT CGA AAT CAA CC
DR_0099_	GAA TTC AAT TAA TAC GAC TCA CTA TAG GGA GAG AAG ACC AAG AAG GCC TGA G

T7_F	
DR_0099_T7_R	TTG CCG AGG ACC AGG AC
DR_2527_T7_F	GAA TTC AAT TAA TAC GAC TCA CTA TAG GGA GAT GCT ATG TCA AAT GCT CAG G
DR_2527_T7_R	GCT GGC CTT CTT GTT GAA
DR_1067_T7_F	GAA TTC AAT TAA TAC GAC TCA CTA TAG GGA GAG CGC AGC CGT CAA ACT G
DR_1067_T7_R	GGG AAA GCT TCT AAG CAG GTC
MS2-Dsr2	TCTAGAGGCGCCACGCGTCCTGAGGTAATTATAACCCGGGCCCTATATATGGATCCTAAGGTACCTAATTG CCTAGAAAACATGAGGATCACCCATGTCTGCAGGTCGACAAACATGAGGATCACCCATGTCTGCAGTATTC CCGGGTTCATTAGATCCTAAGGTACCTAAGGAGCGTGTCGATGCGGAGGTGACTGGCGTGCGTTTTGAAGT GCCGGGTGAGTGCGGTAGCCTGCTGGGCAGCGGATTCTGTTGATTTCACACTCCCAGGAGACCGCAAACA GTTACCCCCATCAAGCTTGGTGGGGGCAACTGTTTGTTTTTGCCGTCTGGGACAGCATCGAACTGAAGTTG TCACCTACTGAGCCAGCGAACCGGGAGCGCTGTGAATACAGTGCTCCCTTTTTTTTTATTTCTAGA

Table C.7. Sequences of mRNA used for EMSA analysis

DR_0094	GAATTCAATTAATACGACTCACTATAGGGAGACTGGCCTGATACGGTTTCAAATTGAGTC CCGGACATGTCCGGGCGCAATTTTGCCGGGAGCGCATGGAAAAAATACGGTTTTAAGGA GATAGACGGGCATCCGGCGCCTCTTTCGGATGTTCGGGAATCGGATTAAAGCAGTATGA AACGTCGGTGGCTAGAACACTGGCAGTTCCCGGCGGCAGCCCCCGGCCCCACCATTGACT GAAAAGGTGAGAAATATTACAGTAGGGCATGCCCAGGGGGTCCCCGGGCGTCCCGGAG AATCGAAAGGAGGTCCTATGAAGCGTTCTGCTGCTCT
DR_0099	GAATTCAATTAATACGACTCACTATAGGGAGACCATCGCCAGCAGCCTGCGCCTGCGCGA CAACGTCCGCCGCGTCCTGGTGGTCAAGGACCGCCCGGAGTGGAAGACCAAGAAGGCCT GAGCCTTTTATGTCATTGACATAATTGACTCTGCTTGTTACTATCTAGTGAACCCGCAAGG GCCGTCCGCCAGCAACATCGAACTCTAGTTATTTTGCCAGCTACCCAAGGAGA
DR_0309	GAATTCAATTAATACGACTCACTATAGGGAGACCCGGCGCTCTTGCCCTTCTAGGCCTTTT CATGTAAGCTAAGCAGTCGGTGCGCCGTGGCCTCTCAGCCCGGTGCAACGTGCCAGAGA GCAGGCGGGACACCGCCCTGGCGAGAATCAACCCAATGTGGGTCACACCACCAGAACCT CGCGGCATGCGCGGGGGAAGCCCGATCAAAGGGACGTTTTTACGTGTGAATCCACCGCT TGGAGGGAGTAAGACAATGGCAAAAGGAACGTTCGAGCGCACCAAGCCCCACGTGAAC ATCGGCACCATCGGTCACGTCGACCACGGCAAGACCA
DR_0662	GAATTCAATTAATACGACTCACTATAGGGAGAACTTGGACTTCAGTTCATGACCTTCGACA CGCACTTGAGAACTGTGGCCGAGCAGGTGTTACCGGGACAGGTATGGACGCCGTAAAAG GAACAGGGACGAGCTACCCCGTCCCCCCGCCCCGCTTTTCCCTGCCAAAGCAGAAGCCTG AGCCAGAATTCAATTCTCGTAAGCCAGAACGTATGAATGCTAGTATGAAGGCATGACTTA CCAGAACGCCGAACGCATGACCATCAGCCTGCCGCCCGATATCGCCCGTTACATCAAGGA CTACCAGCAAACGCACGGACTGGAAAGCCGCAG

DR_0781	GAATTCAATTAATACGACTCACTATAGGGAGACAGCCGCGTCCCAGGGGCGAAGCGCGG CGTCTACGATACTTGGCAAAGATTTGGTAGGGGACGCACGCTGCGGGGCTGGGCGCGGT CTCGATCTTTGAAGGGCATCGCATGGAACAACGTATTCTCCTGATCGAAGACAATCCGGA TATCACCCGCGTGGTGCAGTACGAGCTTGAACAGGCAGGGTACCGCGTCATTACGGCGC CAGACGGCATCACCGGCCTGACGAGCGCCCGCGAAAACACCCCTGACCTGGTCATTCTCG ACCTCGGTCTGCCGGACTTCGACGGGGCGGAGGTGG
DR_0907	GAATTCAATTAATACGACTCACTATAGGGAGACGAGATTTTCGACGCGCTGATGGGTTCG GAAGTCGCGCCGCGCAAGGACTTTATCCGCGAGAACGCCCGCTTCGCTGAAATCAGCGTC TGAGCCTATCCGCATGAAGGCCCGCTTCCCATCCGGAGGCGGGCTTTTTCGTGTTCTCCGC GTTCTGCTCGGGGCAGAGAAGGCCCCGAATCTTTAGACTCTGTCCAAGCGCCGGCCAAAT GTGTGCCGAGTTTCGATTGTTGGTGGCGGTGGCCCCTATCTTGCCTCTCGTTCAACACCAG
DR_1067	GAATTCAATTAATACGACTCACTATAGGGAGACTGAGCCCCGGCTGCGTTCGAGCTGCCT GGAAAGGGTCCGCCTCCTCCCCTGGGGGCGGCTTTTTTGTGCCTGCCACTCGCCTGGCAC CCAGCGCAGATGAGACGGACACAAATGTCAGACCCATGACCGTCTAAAGGCTCCATCAA GGACAAGCGGCGGCGCAGCCGTCAAACTGCGGAGCATGAAACTGCTCAAGACCGTTGCC GTCGTCGCCGCCCTCGCGCTGCCTGTCGCGGCGTCCGCCCAGGACACCAACAACACCACC GGCACGA
DR_1082	GAATTCAATTAATACGACTCACTATAGGGAGAGCAGCGCTGGGGCGGTGGTCTGGTTCG GGAAACGTGCCTGACTTTAGATTTTTCTGACCCGTCCTGAAGCATAGCTCGCCAGGACGG GCGGGCGCTGTGAGCCTTTTCCCTGACGCCCGGCTGACGCAGGGGCTAGAGTGCGGGCA CCGGAAAACCGCACACCATTTCTGAGGTCTTTCCGGCGCGAAGGAGCGTGAACTGTGCA GATTTATCAGCTCTCTGGCCGCAACGTCGAAGTCACCGAACCCATGCGCGAGTACGTGGA AGAAAAGCTCTCGCGCCTCGACCGGTACACCGACCA
DR_1172	GAATTCAATTAATACGACTCACTATAGGGAGAGTCAGCGAACAAAAAAGCGCGGCGGGG ACCGGAGCCGCTCCGGGCAGGGACAGAGGCGGACGGGTCGGGGGCGGACACATTCACA GTCACAGTGTAGGGGCCGGGCGCCCGGAGCGGCCCGCGCACGAAACGGCGGGAGCCCA AAGCCTTCCCTAATCTGGCCTCATCCTTTCATGAGAGCCGCAGCGTAGGCTCCTGACATGT TTGAACGCGATGAACATCACTTTCCGGTTAAGCGTCTGTTGCTGCTCGGTGCCCTCGTCGG GGCCGGCGCCTACTACCTGAGCCGCGAGCAAAACCG
DR_1245	GAATTCAATTAATACGACTCACTATAGGGAGATGCTGAGCCACTCGGCCAAAGAAAGCGT GTGAAGGGCTGGCTCCGTTAACCCTCAGAATTCTGCCCCAGCCTTGAGACTCCCTGGCCC GGCCCGAACATGAGAGCGCAGGTGGTAGAATCGTTCGGGTCTCCCGACCGGGCCTGAGC TTTTGCGGCCCGCCTTTACTTCCCAGGAGGAAAACACCATGGAAACTGCTCTGCTTACCCT CGACACCCTCGCCAAGTACCTGCAAGAAAAAGAAGTTCAGCTCGACATCGAAGAAAACG GCGGCCAGCGCTTTATTCGCATGGGCTGGCGCTTCGAGATGGGCGACG

DR_1261	GAATTCAATTAATACGACTCACTATAGGGAGAGGCGAGGTTTGAGACCTGCCCTGGGAG CCACGCACCATCTCTAACGACGCCTTGATGTTGCTATCACGCCGGGGCTGCTGACGGGAC TTAGGGTAACTTCCATGACGAACGACATCAGTGGACGCATTGCCGAAGACCTCAAGAACC GTCTGGCCCAGGAAGGCGAGCACCTGCAAGTGAAGGACAAGGATGGCGAGCACGTCGG CACCGTGGACCACTTCGACGGCAACCAGCTCAAGCTGACCCGCGACGACAGCACCGACG GCCAGCACCACTACGTCGCCCTCTCGCAAGTCGAGAG
DR_1314	GAATTCAATTAATACGACTCACTATAGGGAGATCGGCGGTCGCCATCGCCTCACGCATGG CGGGAGTGGGCTTGGTGACGGTATCGGAGCGCAGGTCGGCGATCAGCGGCGGGCGGGT CATGGCGCACACCATAGCGGGCAGGGAAGGCCAGAATGGAAAAAAGACTCACCTAGCAC ACGTCCCGCACCTCTGGCAAGAACCGTGACCTCCATAACGGTCCAGTTAAGGTCAGGCTT GGATGGTGTTTACTTTGCACTCAAACAAATGTCCGGGCTGCGGGCCAAACTGCTTCTCGTC GGGAACGAAAACCCGGTCCCAACAGTCCTTCCCGC
DR_1317	GAATTCAATTAATACGACTCACTATAGGGAGACACGGGAACATCCCGCTCTGAGCGACGC TCGCCGAGTCTAAGGGTGCGTTGCTCACGTTCTCATGCACGCTCCCTGGCGGCTTTCGACA ATGGCGCCATGAAAAAAGCTGTCCTCGCCGTTCCTGCCCTGCTGCTCGCCCTGTCGCTCTC CGGTTGCCAGAAGCAGGCCGACAGCAACACCTCGACTTCCACCACCACGACCAAGAGCAC CGACTCCACCGGTCAGAACACTGGCACGACCACGACCAGCACGACGACCACCGACACCA ACAACAAGTAAGTCAGGCGTTTTTCAGTGCTG
DR_1539	GAATTCAATTAATACGACTCACTATAGGGAGAAGGATTTCTGCGCCGGGCCGCAGGTGG GCTTTTTTCGCCAACTCTGAAGGCTCCCTGACCCCCTGTTCATGTGCTGGGGACAAAGCAG GCACATGCTTGACCCTGACCGCAGGCCGCGCGCCGGGCAGAACTCCCTTAGGTTTGCCGT GCCGGTCCGGGAATTCAACTTCAGGAGGTCAATTCATGAGTGAGAAAACGACGCTGGAC CATCTGGCCGATGCCGCAGGTGCCAAGCTGAACGAAGTCGCCGACCGTGCCCGCGCCGC CGGGCATGAGGTGGCGGCCCGCGTGTCGGACAGCC
DR_1571	GAATTCAATTAATACGACTCACTATAGGGAGAGCCAGAGACATAAAGAGCAGCATAGCG CGTCGTGGCCGCTGAACCGCTCTCAGAGAAGTGAAACGGCAGGTCGAGTCACTGAAGAA CGTTGCATACGCGCGCCTTTTATCTGGAGTGGCCTGCAAGACGTGAGGTCAGGTCATTAT ATGAGGATCACATGACAAAGGCTTGACCTTCTCTAGGCCAGTCCATATATTGACCTCGCTG GAAATCCCTGACTTTTTCTACCGACGGCCCCCTGCTGGGAGCGTTGGTCTTTTCCAGAACC GTTTGGAGGTTCCATGAAGAAAGTGATGATGTTGGCCCTGGCCCTCGGCGCGAGCACCTC GCTCGCCGCCCCCTTCGTCTACCCTGCCAACTGGACCAGCAACAAGCCCGGCGA
DR_1844	GAATTCAATTAATACGACTCACTATAGGGAGACTCGGCAGCGGCGGCGGCCTCACGGGT CATCTGAACGCGGGCCTTTTCGTACTCCGCGCCGTTGACCAGACCCCCGAACTCGGGCGG GCTGACCTGCACCCACGACGCCACGCCGCACACGCCTTCCATATCCACGCTGATAACGACT TTCATAAGACCGAGGGTAAAGGCTCGGAGACTGCCCGTGCCTACAATGACGTATGCCAG AACCCCGCAAGAACGCAGACCAGGGCCGCTACGAACTCACCGACGGCGGCCAGATCGTC GGTTTCGCTGAGTACCAGGACCAGGGCGACACCGT

DR_2087	GAATTCAATTAATACGACTCACTATAGGGAGACAACATCGGCTTCTGAAGTCCTAACATTT TTCTAGTGCTTTTCGGCGGGAAAAGGGCGCGGGCGTGGTAGACTACTTTCCTGAACTGCC CGGGACACGGCGCAGTTGCAAGCCGCCCGGCACAGGTCCTGACCGCGCATTCGCGCCGC AGCCTGTGGGGCGCGGCGGAGGTGATAACCATAGCGAAAGATCTTAAGGTCAACGAGCA GATTCGCGTCCGGCAGGTCCGCCTGATCGGCGCGGAAGGGGAGCAGATCGGGATTATCG ATACCCGCGAGGCCATGAACATGGCCCGTGAAAAG
DR_2104	GAATTCAATTAATACGACTCACTATAGGGAGACGGTGGAACTCAACCGCCTGATCGAGCT GGAAATGGAAGGTTCGGTGGGCTGAGGTAGGAGATGGAAATGCCCAGCGTCAATGCTG TATGTCGTGAGTTGGAAGCTGTATTTCCTGACTTTGAGAAGTGGCTGCTTCACGACGCCAT CTTCTGCAAGTCTGAGGGCAGCTTCACTTACTGCGGCGTATTTCTTGAGCTGGTGCATTTC CTCTCTGATAGACCGGATCAGCCTGGAGAGGTCTGGAGAAAACTTGGAAACGTCGTGGA AGACATGGCCAAGATAAGTTCTCCAGCGAAGATTCCCCGGCAAGATGACCTCTCTGAAGC GGTTGAACTCTGCTTCCTTGAAACGCTCGGCGGCTATCCGGAAATTTCCGAGAACTTAAG GCCTTTTCTAAGTCAGGAAAGTTTGAGGATGCTGGATTTGGATATCGGCTAACTCCTAGG CACCGCTTGGTGCTTTGAGGCTGGCCTGATTCAAGCAGATGAGTTTGCGACGTA
DR_2105	GAATTCAATTAATACGACTCACTATAGGGAGAGAAGCTGTATTTCCTGACTTTGAGAAGT GGCTGCTTCACGACGCCATCTTCTGCAAGTCTGAGGGCAGCTTCACTTACTGCGGCGTATT TCTTGAGCTGGTGCATTTCCTCTCTGATAGACCGGATCAGCCTGGAGAGGTCTGGAGAAA ACTTGGAAACGTCGTGGAAGACATGGCCAAGATAAGTTCTCCAGCGAAGATTCCCCGGCA AGATGACCTCTCTGAAGCGGTTGAACTCTGCTTCCTTGAAACGCTCGGCGGCTATCCGGA AATTTCCGAGAACTTAAGGCCTTTTCTAAGTC
DR_2287	GAATTCAATTAATACGACTCACTATAGGGAGAAGGGCCCCGGGCCTCAGCTCTCCATGGT GTTCAGCTTGGCTAGAGAGGACAAACAGCTCACTCGTCCTCCCAGCAATGCATAAGCCAC AGTCTGAAGGCTCAGGCCCGCAGAATCAAACCGGCAGCGGCGAAAAGGCAGCCTGAGTA AGCTAATTTTACTGAGCATAAGCCGCCGCACGTCTCCAGCGCGTACCATGTCGGGCATGG TGACAGCAATCGTGATGGTTCAGGCCGAGCGCCACCGCATTCAGGAAACCGCCGA
DR_2527	GAATTCAATTAATACGACTCACTATAGGGAGACCTGCAGTCCTCTGACATCCTGGGGCAC CTGCGGCAGACACTTGCAGGTGTCCCTTCTTCCGTTAGTGCCTAGAGCGTTTGACAAGAG AATTATCTGGCGCTGTTTGACCCTCTACCCAGGGGAGAGGGCCTGCCAAAGGCAGGGGT GAGGGGTCTTTGCTATGTCAAATGCTCAGGCTTCCCCCGCCCAATGTGCGAGCGAAGGAT AAAGGAAGCGGCCCTTGTCCGGCCCTGAGGGCATTTGTAGCCTGAAGATCATTCCGGGC ACCCGCCCCAACACCCTGGACCCGGGCCGCGCTTC
groES	GAATTCAATTAATACGACTCACTATAGGGAGAAGCTAACAGCTGGCAAGGGGATACCCCC ATTCCCCGTCCCAGCGCCCCTTGAGCGTCATAGACTCAGATTGTCAGCTTCGGTCAGTTGA CATTTTTCTTATCGGCGCTCTACCATCCGTGACGGATTGAAGGCGCTGGGCGGGAAAAAG CTCGCCGGCACGACTCTCCGCCATTCCATCTCACTCACAGGAGGACCCCACATGCTGAAAC CTTTAGGCGACCGCGTTCTGGTTGAAATTATCGAAGAAGCCGAGCAGAAGACTGCCGGC GGCCTGTACGTCCCCGATTCCGCCAAGGAAAA

infA	GAATTCAATTAATACGACTCACTATAGGGAGAAAAAGCAGAGCACCTGAAGTCAAGCCG CTGCGCCGACCGTTCTAAACGTTGTCTGGAATGGTCGTGTCCCGTATAATGAGACACTGA CCCGTGAAACAGCACAGGCACGCGCATCTCCCCGCAAAATGTGCTGGTCAAGGTTGCCG GAACAGCCGATGGGTGGTACTTTTATGTTTGGCTTGTATCAGGCCTGGAGGTTGCGTGGC GAGACGAAAGATGCCGGAACAGCGGGAAAAGAAGAAGAAGGATGAGTCCGACAGCGT GCGGGCCGAGGGCGTGGTGGAAGAGGCGCTGCCGAACA
rplE	GAATTCAATTAATACGACTCACTATAGGGAGACCCATGACGCCCCAGCTCTTGGGGCTTC AACGTCTCAATACGGGGGTGACCTGAGCGGCGCTGGCTGAAACACTGCCAGACCGCTCG AAGGAAAAATGCCATGCAACAGCTCAAGACGAAGTACAACGACCAGGTGCGCCCCGCCC TGATGCAGCAGTTCGGTTACTCCAGCGTGATGGCCGTTCCCCGCATCGAAAAGATCGTGG TCAACGAAGGCCTCGGCAGCAGCAAGGAAGACTCCAAGGCGATCGACAAGGCCGCCAA GGAACTCGCGCTGATCACCCTGCAAAAGCCGATCATC
rpmA	GAATTCAATTAATACGACTCACTATAGGGAGACGAGGCGCCGAAGAGCGACCCTGCCTG ACCCCACAGCGCACTGCTAGGAGCCAACCTGCCGTCTGCTATAGTGGCCGGCAGTCGCGG CCCCGGCAGATGTCCGAGGTCGCCCGGAGGCTTTAGCAACCGACTCGCCCAGGCGCCAC CCCACCGCGAGTCCATGTGTACCGCGAGGACACATAGATCAGAACGGGAACATGCACGC GGTCCACTTCTTATTCGCCCCAACCTGGGGCCAGGACTGTCTGCTTTTGCCCGTTCCGGTC AAACTTCCGGTTCGGGCTTGTTCCCGTGGTCAGGT
RpmE	GAATTCAATTAATACGACTCACTATAGGGAGATGAGCCGCGACTTGTGATACACTGTCTC GGTTAGCCCGCTACGCACAGTGACGTGGCAAAAGGAGTCAAAAGATGCAAAAAGACCTG CACCCCAAGGCCGTGCCCTGCAAGATCATCTACCAGGGTCAAGTCGTGATGGAAACCATG AGCACCCGCCCCGAAATCCACGTGGACGTGTGGAGCGGCGTGCACCCCTTCTGGACCGG CGAAGAACGCTTCCTCGACACCGAAGGCCGCGTGGACAAGTTCAACAAGCGCTTCGGCG ACAGCTACCGCCGCGGCAGCAAGAAGTAAACCGCAA
rpmF	GAATTCAATTAATACGACTCACTATAGGGAGACCGCAGCCTGGAACCGGGCGAGGCCTG GGCCTACTGCTATGTTCACGACCGCACAGCTCCCCGCTGAAGCGTGACGGCCCTTTTGTCA GAGTGCCGCCCAGCGTGGTATACTCCCCGCTGTTGTCTCGAACCGGCCTCCCGTCCGTGCT GGGGGCCACGTTCCGAGAAGAAGGCCCCGGCGAGACGCGCCGTGTGGCTCCCCGAGGA GAAAGATCATGGCCAAACACCCCGTTCCCAAGAAGAAGACCAGCAAGAGCAAGCGCGAC ATGCGCCGCAGCCACCACGCGCTGACCGCCCCCAA
RpmI	GAATTCAATTAATACGACTCACTATAGGGAGATGGGGCAGGGCGACCAACGGCTCTTTG AGAACTTTCTCAACAAGAGAGCGGCCCGAACCCCAGACAGAAGCCTCCCCAGGGAGCGC GACTGGACTACGAAGGAGGGCTCTCCATGCCCAAGATGAAGACTCACAAGATGGCCAAG CGCCGGATCAAGATCACCGGCACCGGCAAGGTCATGGCGTTCAAGAGTGGCAAGCGCCA CCAGAACACCGGCAAGAGCGGCGACGAAATTCGCGGCAAGGGCAAGGGCTTCGTGCTC GCCAAAGCCGAATGGGCCCGGATGAAACTCATGCTCCCG
rpsP	GAATTCAATTAATACGACTCACTATAGGGAGACTAATCTCGCTGATGACTTGCCCGTGCCC CCTTGCCCAGGCTCTCCGCTTTCTGATAGGGTTGTTCGGTGCGCTCCGATTCTGGCTGCTG ACGGGCGCATGTTCTTCGGGCACCTACTCCGGAGACGATACCCTCGAAATCAACCCCTGA GGTCTACTGCACATGGTGAAAATTCGTCTGTCCCGCTTCGGCTCGGCCCACAACCCCCACT ACCGCATCGTCGTCGCCGACGTGCGCCGTCCCCGCGACGGTGGCTACATA

References

[1] K.VMorris andJ. S.Mattick, “The rise of regulatory RNA.,” *Nat. Rev. Genet.*, vol. 15, no. 6, pp. 423–37, 2014.

[2] G.Desnoyers, M.-P.Bouchard, andE.Massé, “New insights into small RNA-dependent translational regulation in prokaryotes.,” *Trends Genet.*, vol. 29, no. 2, pp. 92–8, Feb.2013.

[3] S.Gottesman, C. aMcCullen, M.Guillier, C. K.Vanderpool, N.Majdalani, J.Benhammou, K. M.Thompson, P. C.FitzGerald, N. aSowa, andD. J.FitzGerald, “Small RNA regulators and the bacterial response to stress.,” *Cold Spring Harb. Symp. Quant. Biol.*, vol. 71, pp. 1–11, Jan.2006.

[4] J.Vazquez-Anderson andL. M.Contreras, “Regulatory RNAs: Charming gene management styles for synthetic biology applications,” *RNA Biol.*, vol. 10, no. 12, pp. 1778–1797, 2013.

[5] L. S.Waters andG.Storz, “Regulatory RNAs in bacteria.,” *Cell*, vol. 136, no. 4, pp. 615–28, Feb.2009.

[6] Thomas R. Cech andJoan A. Steitz, “The Noncoding RNA Revolution— Trashing Old Rules to Forge New Ones,” *Cell*, vol. 157, no. 1, pp. 77–94, 2014.

[7] R. C.Lee, “The C. elegans Heterochronic Gene lin-4 Encodes Small RNAs with Antisense Complementarity to &II-14,” *Cell*, vol. 75, pp. 843–854, 1993.

[8] A.Shinhara, M.Matsui, K.Hiraoka, W.Nomura, R.Hirano, K.Nakahigashi, M.Tomita, H.Mori, andA.Kanai, “Deep sequencing reveals as-yet-undiscovered small RNAs in Escherichia coli.,” *BMC Genomics*, vol. 12, no. 1, p. 428, Jan.2011.

[9] C.-H.Tsai, R.Liao, B.Chou, andL. M.Contreras, “Transcriptional analysis of Deinococcus radiodurans reveals novel small RNAs that are differentially expressed under ionizing radiation.,” *Appl. Environ. Microbiol.*, vol. 81, no. 5, pp. 1754–64, 2015.

[10] M.Albrecht, C. M.Sharma, R.Reinhardt, J.Vogel, andT.Rudel, “Deep sequencing-based discovery of the Chlamydia trachomatis transcriptome.,” *Nucleic Acids Res.*, vol. 38, no. 3, pp. 868–77, Jan.2010.

[11] Z.Wang, M.Gerstein, andM.Snyder, “RNA-Seq: a revolutionary tool for transcriptomics.,” *Nat. Rev. Genet.*, vol. 10, no. 1, pp. 57–63, 2009.

[12] J.Livny andM. K.Waldor, “Identification of small RNAs in diverse bacterial species.,” *Curr. Opin. Microbiol.*, vol. 10, no. 2, pp. 96–101, Apr.2007.

[13] B.Wiedenheft, S. H.Sternberg, andJ. aDoudna, “RNA-guided genetic silencing systems in bacteria and archaea.,” *Nature*, vol. 482, no. 7385, pp. 331–8, Feb.2012.

[14] E.Loh, O.Dussurget, J.Gripenland, K.Vaitkevicius, T.Tiensuu, P.Mandin, F.Repoila, C.Buchrieser, P.Cossart, andJ.Johansson, “A trans-Acting Riboswitch Controls Expression of the Virulence Regulator PrfA in Listeria monocytogenes,” *Cell*, vol. 139, no. 4, pp. 770–9, Nov.2009.

[15] J. R.Mellin, T.Tiensuu, C.Bécavin, E.Gouin, J.Johansson, andP.Cossart, “A riboswitch-regulated antisense RNA in Listeria monocytogenes,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 32, pp. 13132–13137, 2013.

[16] A.Wachter, “Riboswitch-mediated control of gene expression in eukaryotes,” *RNA Biol.*, vol. 7, no. 1, pp. 67–76, 2010.

[17] G.André, S.Even, H.Putzer, P.Burguière, C.Croux, A.Danchin, I.Martin-Verstraete, andO.Soutourina, “S-box and T-box riboswitches and antisense RNA control a sulfur metabolic operon of Clostridium acetobutylicum,” *Nucleic Acids Res.*, vol. 36, no. 18, pp. 5955–5969, 2008.

[18] G.Storz, J.Vogel, andK. M.Wassarman, “Regulation by small RNAs in bacteria: expanding frontiers.,” *Mol. Cell*, vol. 43, no. 6, pp. 880–91, Sep.2011.

[19] N.Delihas, “Regulation of gene expression by trans-encoded antisense RNAs.,” *Mol. Microbiol.*, vol. 15, no. 3, pp. 411–4, Feb.1995.

[20] J.Vogel andC. M.Sharma, “How to find small non-coding RNAs in bacteria.,” *Biol. Chem.*, vol. 386, no. 12, pp. 1219–38, Dec.2005.

[21] B.Tjaden, R. M.Saxena, S.Stolyar, D. R.Haynor, E.Kolker, andC.Rosenow, “Transcriptome analysis of Escherichia coli using high-density oligonucleotide probe arrays.,” *Nucleic Acids Res.*, vol. 30, no. 17, pp. 3732–8, Sep.2002.

[22] I. M.Axmann, P.Kensche, J.Vogel, S.Kohl, H.Herzel, andW. R.Hess, “Identification of cyanobacterial non-coding RNAs by comparative genome analysis,” *Genome Biol.*, vol. 6, no. 1465–6914 (Electronic), p. R73, 2005.

[23] P.Acebo, A. J.Martin-Galiano, S.Navarro, A.Zaballos, andM.Amblar, “Identification of 88 regulatory small RNAs in the TIGR4 strain of the human pathogen Streptococcus pneumoniae.,” *RNA*, vol. 18, no. 3, pp. 530–46, Mar.2012.

[24] E.Rivas andS. R.Eddy, “Noncoding RNA gene detection using comparative sequence analysis.,” *BMC Bioinformatics*, vol. 2, p. 8, Jan.2001.

[25] M. M.Cox andJ. R.Battista, “Deinococcus radiodurans - the consummate survivor.,” *Nat. Rev. Microbiol.*, vol. 3, no. 11, pp. 882–92, Nov.2005.

[26] M. J.Daly, “A new perspective on radiation resistance based on Deinococcus radiodurans,” *Nat Rev Microbiol*, vol. 7, pp. 237–45, 2009.

[27] D.Anderson, A. W., Nordan, H. C., Cain, R. F., Parrish, G., & Duggan, “Studies on a radio-resistant micrococcus. 1. Isolation, morphology, cultural characteristics,

and resistance to gamma radiation," *Food Technol.*, vol. 10, no. 12, pp. 75–78, 1956.

[28] K. S.Makarova, L.Aravind, Y. I.Wolf, L.Roman, K. W.Minton, E.VKoonin, J.Daly, andR. L.Tatusov, "Genome of the Extremely Radiation-Resistant Bacterium Deinococcus radiodurans Viewed from the Perspective of Comparative Genomics," *Microbiol. Mol. Biol. Rev.*, vol. 65, no. 1, p. 44, 2001.

[29] M. J.Daly, E. K.Gaidamakova, V. Y.Matrosova, J. G.Kiang, R.Fukumoto, D.-Y.Lee, N. B.Wehr, G. aViteri, B. S.Berlett, andR. L.Levine, "Small-molecule antioxidant proteome-shields in Deinococcus radiodurans.," *PLoS One*, vol. 5, no. 9, p. e12570, Jan.2010.

[30] I.Narumi, K.Satoh, S.Cui, T.Funayama, S.Kitayama, andH.Watanabe, "PprA: A novel protein from Deinococcus radiodurans that stimulates DNA ligation," *Mol. Microbiol.*, vol. 54, no. 1, pp. 278–285, 2004.

[31] R. C.Friedman, S.Kalkhof, O.Doppelt-Azeroual, S.Mueller, M.Chovancova, M.vonBergen, andB.Schwikowski, "Common and phylogenetically widespread coding for peptides by bacterial small RNAs," *bioRxiv*, p. 30619, 2015.

[32] S.Gottesman andG.Storz, "Bacterial small RNA regulators: versatile roles and rapidly evolving variations.," *Cold Spring Harb. Perspect. Biol.*, vol. 3, no. 12, p. a003798, Dec.2011.

[33] D. H.Lenz, K. C.Mok, B. N.Lilley, R.VKulkarni, N. S.Wingreen, andB. L.Bassler, "The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in Vibrio harveyi and Vibrio cholerae.," *Cell*, vol. 118, no. 1, pp. 69–82, Jul.2004.

[34] A.Zhang, K. M.Wassarman, C.Rosenow, B. C.Tjaden, G.Storz, andS.Gottesman, "Global analysis of small RNA and mRNA targets of Hfq," *Mol. Microbiol.*, vol. 50, no. 4, pp. 1111–1124, Oct.2003.

[35] B. L.Tianyan Song, Franziska Mika, J. Z.Zhi Liu, Stefan Schild, Anne Bishop, J. V.Andrew Camilli, Jörgen Johansson, and and S. N.Wai, "A new Vibrio cholerae sRNA modulates colonization and affects release of outer membrane vesicles," *Mol. Microbiol.*, vol. 70, no. 1, pp. 100–111, 2008.

[36] C. S.Baker, L. A.Eöry, H.Yakhnin, J.Mercante, T.Romeo, andP.Babitzke, "CsrA inhibits translation initiation of Escherichia coli hfq by binding to a single site overlapping the Shine-Dalgarno sequence.," *J. Bacteriol.*, vol. 189, no. 15, pp. 5472–81, Aug.2007.

[37] K. M.Wassarman, "6S RNA: a small RNA regulator of transcription.," *Curr. Opin. Microbiol.*, vol. 10, no. 2, pp. 164–8, Apr.2007.

[38] J. E.Barrick, N.Sudarsan, Z.Weinberg, andW. L.Ruzzo, "6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter," *RNA*, vol. 11, no. 5, pp. 774–784, 2005.

[39] K. M. W.Amy T. Cavanagh, Jamie M. Sperger, “Regulation of 6S RNA by pRNA synthesis is required for efficient recovery from stationary phase in E. coli and B. subtilis,” *Nucleic Acids Res.*, vol. 40, no. 5, pp. 34–46, 2011.

[40] B.Reichenbach, A.Maes, F.Kalamorz, E.Hajnsdorf, andB.Görke, “The small RNA GlmY acts upstream of the sRNA GlmZ in the activation of glmS expression and is subject to regulation by polyadenylation in Escherichia coli,” *Nucleic Acids Res.*, vol. 36, no. 8, pp. 2570–2580, 2008.

[41] Z.Kang, X.Wang, Y.Li, W.Qian, andQ.Qingsheng, “Small RNA RyhB as a potential tool used for metabolic engineering in Escherichia coli,” *Biotechnol Lett*, vol. 34, no. 3, pp. 527–531, 2012.

[42] F.Li, Y.Wang, K.Gong, Q.Wang, Q.Liang, andQ.Qi, “Constitutive expression of RyhB regulates the heme biosynthesis pathway and increases the 5-aminolevulinic acid accumulation in Escherichia coli,” *FEMS Microbiol Lett*, vol. 350, no. 2, pp. 209–15, 2013.

[43] E.Massé, C. K.Vanderpool, andS.Gottesman, “Effect of RyhB small RNA on global iron use in Escherichia coli,” *J. Bacteriol.*, vol. 187, no. 20, pp. 6962–6971, 2005.

[44] S.Gottesman, “The small RNA regulators of Escherichia coli: roles and mechanisms.,” *Annu. Rev. Microbiol.*, vol. 58, pp. 303–28, Jan.2004.

[45] P. J.Pomposiello andB.Demple, “Redox-operated genetic switches: The SoxR and OxyR transcription factors,” *Trends Biotechnol.*, vol. 19, no. 3, pp. 109–114, 2001.

[46] J. A.Opdyke, E. M.Fozo, M. R.Hemm, andG.Storz, “RNase III participates in gadY-dependent cleavage of the gadX-gadW mRNA,” *J. Mol. Biol.*, vol. 406, no. 1, pp. 29–43, 2011.

[47] A.Boysen, J.Møller-Jensen, B.Kallipolitis, P.Valentin-Hansen, andM.Overgaard, “Translational regulation of gene expression by an anaerobically induced small non-coding RNA in Escherichia coli,” *J. Biol. Chem.*, vol. 285, no. 14, pp. 10690–10702, 2010.

[48] R.Barrangou andP.Horvath, “CRISPR: New Horizons in Phage Resistance and Strain Identification,” *Annu. Rev. Food Sci. Technol.*, vol. 3, no. 1, pp. 143–162, 2012.

[49] S.Altuvia, “Identification of bacterial small non-coding RNAs: experimental approaches.,” *Curr. Opin. Microbiol.*, vol. 10, no. 3, pp. 257–61, Jun.2007.

[50] C. M.Sharma, “Experimental approaches for the discovery and characterization of regulatory small RNA,” *Curr. Opin. Microbiol.*, vol. 12, no. 5, pp. 536–546, 2009.

[51] E.Sonnleitner, T.Sorger-Domenigg, M. J.Madej, S.Findeiss, J.Hackermüller, A.Hüttenhofer, P. F.Stadler, U.Bläsi, andI.Moll, “Detection of small RNAs in

Pseudomonas aeruginosa by RNomics and structure-based bioinformatic tools.," *Microbiology*, vol. 154, no. Pt 10, pp. 3175–87, Oct.2008.

[52] J.Vogel, V.Bartels, T.Tang, G.Churakov, andJ.Slagter-Jäger, "RNomics in Escherichia coli detects new sRNA species and indicates parallel transcriptional output in bacteria," *Nucleic Acids Res.*, vol. 31, no. 22, pp. 6435–6443, Nov.2003.

[53] J.Sridhar andP.Gunasekaran, "Computational small RNA prediction in bacteria.," *Bioinform. Biol. Insights*, vol. 7, pp. 83–95, Jan.2013.

[54] L.Argaman, R.Hershberg, J.Vogel, G.Bejerano, E. G.Wagner, H.Margalit, andS.Altuvia, "Novel small RNA-encoding genes in the intergenic regions of Escherichia coli.," *Curr. Biol.*, vol. 11, no. 12, pp. 941–950, Jun.2001.

[55] J.Livny, M. A.Fogel, B. M.Davis, andM. K.Waldor, "sRNAPredict: an integrative computational approach to identify sRNAs in bacterial genomes," *Nucleic Acids Res.*, vol. 33, no. 13, pp. 4096–4105, 2005.

[56] J.Livny, M. A.Fogel, B. M.Davis, andM. K.Waldor, "sRNAPredict: An integrative computational approach to identify sRNAs in bacterial genomes," *Nucleic Acids Res.*, vol. 33, no. 13, pp. 4096–4105, 2005.

[57] W.Li, X.Ying, Q.Lu, andL.Chen, "Predicting sRNAs and Their Targets in Bacteria," *Genomics. Proteomics Bioinformatics*, vol. 10, no. 5, pp. 276–284, 2012.

[58] Z.Wang, M.Gerstein, andM.Snyder, "RNA-Seq: a revolutionary tool for transcriptomics.," *Nat. Rev. Genet.*, vol. 10, no. 1, pp. 57–63, Jan.2009.

[59] H.Li andR.Durbin, "Fast and accurate short read alignment with Burrows-Wheeler transform.," *Bioinformatics*, vol. 25, no. 14, pp. 1754–1760, Jul.2009.

[60] B.Langmead, C.Trapnell, M.Pop, andS. L.Salzberg, "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.," *Genome Biol.*, vol. 10, no. 3, p. R25, Jan.2009.

[61] M.Gómez-Lozano, R. L.Marvig, S.Molin, andK. S.Long, "Genome-wide identification of novel small RNAs in Pseudomonas aeruginosa.," *Environ. Microbiol.*, vol. 14, no. 8, pp. 2006–16, Aug.2012.

[62] A.Sittka, S. Lucchini, K.Papenfort, C. M.Sharma, K.Rolle, T. T.Binnewies, J. C. D.Hinton, andJ.Vogel, "Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq.," *PLoS Genet.*, vol. 4, no. 8, p. e1000163, Jan.2008.

[63] CH.Tsai, R.Liao, B.Chou, and L. M.Contreras, "Transcriptional Analysis of Deinococcus radiodurans Reveal Novel sRNAs that are Differentially Expressed under Ionizing Radiation," *Appl. Environ. Microbiol.*, vol. 81, no. 5, pp. 54–64, 2015.

[64] A.Conesa, P.Madrigal, S.Tarazona, D.Gomez-Cabrero, A.Cervera, A.McPherson, M. W.Szczesniak, D. J.Gaffney, L. L.Elo, X.Zhang, andA.Mortazavi, “A survey of best practices for RNA-seq data analysis,” *Genome Biol.*, vol. 17, no. 1, pp. 13–19, 2016.

[65] B.Tjaden, “TargetRNA: a tool for predicting targets of small RNA action in bacteria.,” *Nucleic Acids Res.*, vol. 36, no. Web Server issue, pp. W109--13, Jul.2008.

[66] A.Busch, A. S.Richter, R.Backofen, B.Group, A.Freiburg, andF.D-, “IntaRNA : efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions,” *bioinformatics*, vol. 24, no. 24, pp. 2849–2856, 2008.

[67] F.Eggenhofer, H.Tafer, P. F.Stadler, andI. L.Hofacker, “RNApredator : fast accessibility-based prediction of sRNA targets,” *Nucleic Acids Res*, vol. 39, no. June, pp. 149–154, 2011.

[68] A. A.Agapov andA.VKulbachinskiy, “Mechanisms of stress resistance and gene regulation in the radioresistant bacterium Deinococcus radiodurans,” *Biochem.*, vol. 80, no. 10, pp. 1201–1216, Oct.2015.

[69] K.-W.Jung, S.Lim, andY.-S.Bahn, “Microbial radiation-resistance mechanisms,” *J. Microbiol.*, vol. 55, no. 7, pp. 499–507, Jul.2017.

[70] I.Narumi, K.Satoh, S.Cui, T.Funayama, S.Kitayama, andH.Watanabe, “PprA: A novel protein from Deinococcus radiodurans that stimulates DNA ligation,” *Mol. Microbiol.*, vol. 54, no. 1, pp. 278–285, 2004.

[71] Y.Hua, I.Narumi, G.Gao, B.Tian, K.Satoh, S.Kitayama, andB.Shen, “PprI: A general switch responsible for extreme radioresistance of Deinococcus radiodurans,” *Biochem. Biophys. Res. Commun.*, vol. 306, no. 2, pp. 354–360, 2003.

[72] P.Zhao, Z.Zhou, W.Zhang, M.Lin, M.Chen, andG.Wei, “Global transcriptional analysis of Escherichia coli expressing IrrE , a regulator from Deinococcus radiodurans , in response to NaCl shock,” *Mol. Biosyst.*, vol. 11, pp. 1165–1171, 2015.

[73] H.Ohba, K.Satoh, H.Sghaier, T.Yanagisawa, andI.Narumi, “Identification of PprM: A modulator of the PprI-dependent DNA damage response in Deinococcus radiodurans,” *Extremophiles*, vol. 13, no. 3, pp. 471–479, 2009.

[74] E.Bentchikou, P.Servant, G.Coste, andS.Sommer, “A major role of the RecFOR pathway in DNA double-strand-break repair through ESDSA in Deinococcus radiodurans,” *PLoS Genet.*, vol. 6, no. 1, 2010.

[75] M. J.Daly, “Death by protein damage in irradiated cells.,” *DNA Repair (Amst).*, vol. 11, no. 1, pp. 12–21, Jan.2012.

[76] M. J.Daly, “Accumulation of Mn (II) in Deinococcus radiodurans Facilitates Gamma-Radiation Resistance,” *Science (80-.).*, vol. 306, pp. 1025–28, 2004.

[77] J. T.Lett, I.Caldwell, andJ. G.Little, “Repair of X-ray damage to the DNA in Micrococcus radiodurans: the effect of 5-bromodeoxyuridine,” *J. Mol. Biol.*, vol. 48, no. 3, pp. 395–408, 1970.

[78] B.Vukovic-Nagy, B. W.Fox, andM.Fox, “The Release of a Deoxyribonucleic Acid Fragment after X-irradiation of Micrococcus Radiodurans,” *Int. J. Radiat. Biol. Relat. Stud. Physics, Chem. Med.*, vol. 25, no. 4, pp. 329–337, 1974.

[79] C. H.Tsai, R.Liao, B.Chou, M.Palumbo, andL. M.Contreras, “Genome-wide analyses in bacteria show small-RNA enrichment for long and conserved intergenic regions,” *J. Bacteriol.*, vol. 197, no. 1, pp. 40–50, 2015.

[80] C. H.Tsai, C.Baranowski, J.Livny, K. aMcDonough, J. T.Wade, andL. M.Contreras, “Identification of Novel sRNAs in Mycobacterial Species.,” *PLoS One*, vol. 8, no. 11, p. e79411, Jan.2013.

[81] G.Storz andD.Haas, “A guide to small RNAs in microorganisms,” *Curr. Opin. Microbiol.*, vol. 10, no. 2, pp. 93–95, Apr.2007.

[82] D.Pellin, P.Miotto, A.Ambrosi, D. M.Cirillo, andC.DiSerio, “A Genome-Wide Identification Analysis of Small Regulatory RNAs in Mycobacterium tuberculosis by RNA-Seq and Conservation Analysis,” *PLoS One*, vol. 7, no. 3, p. e32723, Mar.2012.

[83] L.Ji andX.Chen, “Regulation of small RNA stability: methylation and beyond.,” *Cell Res.*, vol. 22, no. 4, pp. 624–36, Apr.2012.

[84] S.Gottesman andG.Storz, “Bacterial Small RNA Regulators : Versatile Roles and Rapidly Evolving Variations,” *Cold Spring Harb Perspect Biol*, vol. 3, no. 12, p. a003798, 2011.

[85] T.Suzuki, C.Ueguchi, andT.Mizuno, “H-NS regulates OmpF expression through micF antisense RNA in Escherichia coli,” *J. Bacteriol.*, vol. 178, no. 12, pp. 3650–3, 1996.

[86] B.Xiao, W.Li, G.Guo, B.Li, Z.Liu, K.Jia, Y.Guo, X.Mao, andQ.Zou, “Identification of small noncoding RNAs in Helicobacter pylori by a bioinformatics-based approach.,” *Curr. Microbiol.*, vol. 58, no. 3, pp. 258–63, Mar.2009.

[87] M.Guillier andS.Gottesman, “Remodelling of the Escherichia coli outer membrane by two small regulatory RNAs.,” *Mol. Microbiol.*, vol. 59, no. 1, pp. 231–47, Jan.2006.

[88] G.Wagner andR. W.Simons, “Anitisense RNA control in bacteria, phages, and plasmids,” *Annu. Rev. Microbial.*, vol. 48, pp. 713–42, 1994.

[89] N.Majdalani, C.Cunning, D.Sledjeski, T.Elliott, andS.Gottesman, “DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription.,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 95, no. 21, pp. 12462–7, Oct.1998.

[90] R.Backofen andW. R.Hess, “Computational prediction of sRNAs and their targets in bacteria.,” *RNA Biol.*, vol. 7, no. 1, pp. 33–42, 2010.

[91] F. J.Isaacs, D. J.Dwyer, andJ. J.Collins, “RNA synthetic biology.,” *Nat. Biotechnol.*, vol. 24, no. 5, pp. 545–54, May2006.

[92] J. C.Liang, R. J.Bloom, andC. D.Smolke, “Engineering biological systems with synthetic RNA molecules.,” *Mol. Cell*, vol. 43, no. 6, pp. 915–26, Sep.2011.

[93] K.Papenfort andJ.Vogel, “Multiple target regulation by small noncoding RNAs rewires gene expression at the post-transcriptional level.,” *Res. Microbiol.*, vol. 160, no. 4, pp. 278–87, May2009.

[94] L.Liu, Y.Li, S.Li, N.Hu, Y.He, R.Pong, D.Lin, L.Lu, andM.Law, “Comparison of next-generation sequencing systems.,” *J. Biomed. Biotechnol.*, vol. 2012, p. 251364, Jan.2012.

[95] D.MacLean, V.Moulton, andD. J.Studholme, “Finding sRNA generative locales from high-throughput sequencing data with NiBLS.,” *BMC Bioinformatics*, vol. 11, p. 93, Jan.2010.

[96] J.Sridhar andP.Gunasekaran, “Computational small RNA prediction in bacteria.,” *Bioinform Biol Insights*, vol. 7, pp. 83–95, Jan.2013.

[97] J.Livny, “Bioinformatic Discovery of Bacterial Regulatory RNAs Using SIPHT,” in *Bacterial Regulatory RNA: Methods and Protocols*, vol. 905, K. C.Keiler, Ed.Totowa, NJ: Humana Press, 2012, pp. 3–14.

[98] X.Lu, H.Goodrich-blair, andB.Tjaden, “Assessing computational tools for the discovery of small RNA genes in bacteria,” *RNA*, vol. 17, pp. 1635–1647, 2011.

[99] C.Pichon, L.duMerle, M. E.Caliot, P.Trieu-Cuot, andC.LeBouguénec, “An in silico model for identification of small RNAs in whole bacterial genomes: characterization of antisense RNAs in pathogenic Escherichia coli and Streptococcus agalactiae strains.,” *Nucleic Acids Res.*, vol. 40, no. 7, pp. 2846–61, Apr.2012.

[100] S.Washietl andI. L.Hofacker, “Consensus folding of aligned sequences as a new measure for the detection of functional RNAs by comparative genomics.,” *J. Mol. Biol.*, vol. 342, no. 1, pp. 19–30, Sep.2004.

[101] A.Marchais, M.Naville, C.Bohn, P.Bouloc, andD.Gautheret, “Single-pass classification of all noncoding sequences in a bacterial genome using phylogenetic profiles.,” *Genome Res.*, vol. 19, no. 6, pp. 1084–92, Jun.2009.

[102] T.Akama, K.Suzuki, K.Tanigawa, A.Kawashima, H.Wu, N.Nakata, Y.Osana, Y.Sakakibara, andN.Ishii, “Whole-genome tiling array analysis of Mycobacterium leprae RNA reveals high expression of pseudogenes and noncoding regions.,” *J. Bacteriol.*, vol. 191, no. 10, pp. 3321–7, May2009.

[103] Z.Hu, A.Zhang, G.Storz, S.Gottesman, andS. H.Leppla, “An antibody-based microarray assay for small RNA detection.,” *Nucleic Acids Res.*, vol. 34, no. 7, p. e52, Jan.2006.

[104] I.Wilms, A.Overlöper, M.Nowrousian, C. M.Sharme, andF.Narberhaus, “Deep sequencing uncovers numerous small RNAs on all four replicons of the plant pathogen Agrobacterium tumefaciens,” *RNA Biol.*, vol. 9, no. 4, pp. 446–57, 2012.

[105] G.Gelderman andL. M.Contreras, “Discovery of Posttranscriptional Regulatory RNAs Using Next Generation Sequencing Technologies,” in *Systems Metabolic Engineering: Methods and Protocols, Methods in Molecular Biology*, vol. 985, H. S.Alper, Ed.New York, NY: Humana Press, 2013, pp. 269–295.

[106] K. C.Pang, M. C.Frith, andJ. S.Mattick, “Rapid evolution of noncoding RNAs : lack of conservation does not mean lack of function,” *Genome Anal.*, vol. 22, no. 1, pp. 1–5, 2006.

[107] B.Voß, J.Georg, V.Schön, S.Ude, andW. R.Hess, “Biocomputational prediction of non-coding RNAs in model cyanobacteria,” *BMC Genomics*, vol. 15, pp. 1–15, 2009.

[108] S.Washietl, I. L.Hofacker, andP. F.Stadler, “Fast and reliable prediction of noncoding RNAs,” *PNAS*, vol. 102, no. 7, pp. 2454–59, 2005.

[109] S. W.Burge, J.Daub, R.Eberhardt, J.Tate, L.Barquist, E. P.Nawrocki, S. R.Eddy, P. P.Gardner, andA.Bateman, “Rfam 11.0: 10 years of RNA families.,” *Nucleic Acids Res.*, vol. 41, no. Database issue, pp. D226-32, Jan.2013.

[110] J. M.Liu, J.Livny, M. S.Lawrence, M. D.Kimball, M. K.Waldor, andA.Camilli, “Experimental discovery of sRNAs in Vibrio cholerae by direct cloning, 5S/tRNA depletion and parallel sequencing.,” *Nucleic Acids Res.*, vol. 37, no. 6, p. e46, Apr.2009.

[111] B. M.Davis andM. K.Waldor, “RNase E-dependent processing stabilizes MicX, a Vibrio cholerae sRNA.,” *Mol. Microbiol.*, vol. 65, no. 2, pp. 373–85, Jul.2007.

[112] E. S.Bradley, K.Bodi, A. M.Ismail, andA.Camilli, “A genome-wide approach to discovery of small RNAs involved in regulation of virulence in Vibrio cholerae.,” *PLoS Pathog.*, vol. 7, no. 7, p. e1002126, Jul.2011.

[113] P.Mandin, F.Repoila, M.Vergassola, T.Geissmann, andP.Cossart, “Identification of new noncoding RNAs in Listeria monocytogenes and prediction of mRNA targets.,” *Nucleic Acids Res.*, vol. 35, no. 3, pp. 962–74, Jan.2007.

[114] A.Toledo-Arana, O.Dussurget, G.Nikitas, N.Sesto, H.Guet-Revillet, D.Balestrino, E.Loh, J.Gripenland, T.Tiensuu, K.Vaitkevicius, M.Barthelemy, M.Vergassola, M.-A.Nahori, G.Soubigou, B.Régnault, J.-Y.Coppée, M.Lecuit, J.Johansson, andP.Cossart, "The Listeria transcriptional landscape from saprophytism to virulence.," *Nature*, vol. 459, no. 7249, pp. 950–6, Jun.2009.

[115] M. aMraheil, A.Billion, W.Mohamed, K.Mukherjee, C.Kuenne, J.Pischimarov, C.Krawitz, J.Retey, T.Hartsch, T.Chakraborty, andT.Hain, "The intracellular sRNA transcriptome of Listeria monocytogenes during growth in macrophages.," *Nucleic Acids Res.*, vol. 39, no. 10, pp. 4235–48, May2011.

[116] C.Bohn, C.Rigoulay, S.Chabelskaya, C. M.Sharma, A.Marchais, P.Skorski, E.Borezée-Durant, R.Barbet, E.Jacquet, A.Jacq, D.Gautheret, B.Felden, J.Vogel, andP.Bouloc, "Experimental discovery of small RNAs in Staphylococcus aureus reveals a riboregulator of central metabolism.," *Nucleic Acids Res.*, vol. 38, no. 19, pp. 6620–36, Oct.2010.

[117] T.Geissmann, C.Chevalier, M.-J.Cros, S.Boisset, P.Fechter, C.Noirot, J.Schrenzel, P.François, F.Vandenesch, C.Gaspin, andP.Romby, "A search for small noncoding RNAs in Staphylococcus aureus reveals a conserved sequence motif for regulation.," *Nucleic Acids Res.*, vol. 37, no. 21, pp. 7239–57, Nov.2009.

[118] L. F.Abu-Qatouseh, S.VChinni, J.Seggewiss, R. A.Proctor, J.Brosius, T. S.Rozhdestvensky, G.Peters, C.vonEiff, andK.Becker, "Identification of differentially expressed small non-protein-coding RNAs in Staphylococcus aureus displaying both the normal and the small-colony variant phenotype.," *J. Mol. Med.*, vol. 88, no. 6, pp. 565–75, Jun.2010.

[119] M.Beaume, D.Hernandez, L.Farinelli, C.Deluen, P.Linder, C.Gaspin, P.Romby, J.Schrenzel, andP.Francois, "Cartography of methicillin-resistant S. aureus transcripts: detection, orientation and temporal expression during growth phase and stress conditions.," *PLoS One*, vol. 5, no. 5, p. e10725, Jan.2010.

[120] C.Pichon andB.Felden, "Small RNA genes expressed from Staphylococcus aureus genomic and pathogenicity islands with specific expression among pathogenic strains," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 102, no. 40, pp. 14249–54, 2005.

[121] S.Saito, H.Kakeshita, andK.Nakamura, "Novel small RNA-encoding genes in the intergenic regions of Bacillus subtilis.," *Gene*, vol. 428, no. 1–2, pp. 2–8, Jan.2009.

[122] S.Rasmussen, H. B.Nielsen, andH.Jarmer, "The transcriptionally active regions in the genome of Bacillus subtilis.," *Mol. Microbiol.*, vol. 73, no. 6, pp. 1043–57, Sep.2009.

[123] R.Kumar, P.Shah, E.Swiatlo, S. C.Burgess, M. L.Lawrence, andB.Nanduri, "Identification of novel non-coding small RNAs from Streptococcus pneumoniae

TIGR4 using high-resolution genome tiling arrays.," *BMC Genomics*, vol. 11, p. 350, Jan.2010.

[124] B.Mann, T.vanOpijnen, J.Wang, C.Obert, Y.-D.Wang, R.Carter, D. J.McGoldrick, G.Ridout, A.Camilli, E. I.Tuomanen, andJ. W.Rosch, "Control of virulence by small RNAs in Streptococcus pneumoniae.," *PLoS Pathog.*, vol. 8, no. 7, p. e1002788, Jan.2012.

[125] N.Perez, J.Treviño, Z.Liu, S. C. M.Ho, P.Babitzke, andP.Sumby, "A genome-wide analysis of small regulatory RNAs in the human pathogen group A Streptococcus.," *PLoS One*, vol. 4, no. 11, p. e7668, Jan.2009.

[126] K.Shioya, C.Michaux, C.Kuenne, T.Hain, N.Verneuil, A.Budin-Verneuil, T.Hartsch, A.Hartke, andJ.-C.Giard, "Genome-wide identification of small RNAs in the opportunistic pathogen Enterococcus faecalis V583.," *PLoS One*, vol. 6, no. 9, p. e23948, Jan.2011.

[127] A.Fouquier d'Hérouel, F.Wessner, D.Halpern, J.Ly-Vu, S. P.Kennedy, P.Serror, E.Aurell, andF.Repoila, "A simple and efficient method to search for selected primary transcripts: non-coding and antisense RNAs in the human pathogen Enterococcus faecalis.," *Nucleic Acids Res.*, vol. 39, no. 7, p. e46, Apr.2011.

[128] J. M.DiChiara, L. M.Contreras-Martinez, J.Livny, D.Smith, K. A.McDonough, andM.Belfort, "Multiple small RNAs identified in Mycobacterium bovis BCG are also expressed in Mycobacterium tuberculosis and Mycobacterium smegmatis.," *Nucleic Acids Res.*, vol. 38, no. 12, pp. 4067–78, Jul.2010.

[129] R.Hershberg, S.Altuvia, andM.Hanah, "A survey of small RNA-encoding genes in Escherichia coli," *Nucleic Acids Res.*, vol. 31, no. 7, pp. 1813–1820, Apr.2003.

[130] E.Rivas, R. J.Klein, T. aJones, andS. R.Eddy, "Computational identification of noncoding RNAs in E. coli by comparative genomics.," *Curr. Biol.*, vol. 11, no. 17, pp. 1369–73, Sep.2001.

[131] L.Argaman, R.Hershberg, J.Vogel, G.Bejerano, E. G.Wagner, H.Margalit, andS.Altuvia, "Novel small RNA-encoding genes in the intergenic regions of Escherichia coli.," *Curr. Biol.*, vol. 11, no. 12, pp. 941–50, Jun.2001.

[132] M.Macvanin, R.Edgar, F.Cui, A.Trostel, V.Zhurkin, andS.Adhya, "Noncoding RNAs binding to the nucleoid protein HU in Escherichia coli.," *J. Bacteriol.*, vol. 194, no. 22, pp. 6046–55, Nov.2012.

[133] M.Castillo-keller, P.Vuong, andR.Misra, "Novel Mechanism of Escherichia coli Porin Regulation," *J. Bacteriol.*, vol. 188, no. 2, pp. 576–586, 2006.

[134] M.Kawano, T.Oshima, H.Kasai, andH.Mori, "Molecular characterization of long direct repeat (LDR) sequences expressing a stable mRNA encoding for a 35-amino-acid cell-killing peptide and a cis-encoded small antisense RNA in Escherichia coli.," *Mol. Microbiol.*, vol. 45, no. 2, pp. 333–49, Jul.2002.

[135] S.Chen, E. aLesnik, T. aHall, R.Sampath, R. H.Griffey, D. J.Ecker, andL. B.Blyn, "A bioinformatics based approach to discover small RNA genes in the Escherichia coli genome.," *BioSystems*, vol. 65, no. 2–3, pp. 157–77, 2002.

[136] A.Boysen, J.Møller-Jensen, B.Kallipolitis, P.Valentin-Hansen, andM.Overgaard, "Translational regulation of gene expression by an anaerobically induced small non-coding RNA in Escherichia coli.," *J. Biol. Chem.*, vol. 285, no. 14, pp. 10690–702, Apr.2010.

[137] M.Kawano, A. A.Reynolds, J.Miranda-Rios, andG.Storz, "Detection of 5'- and 3'-UTR-derived small RNAs and cis-encoded antisense RNAs in Escherichia coli.," *Nucleic Acids Res*, vol. 33, no. 3, pp. 1040–50, Jan.2005.

[138] C. M.Sharma, S.Hoffmann, F.Darfeuille, J.Reignier, S.Findeiss, A.Sittka, S.Chabas, K.Reiche, J.Hackermüller, R.Reinhardt, P. F.Stadler, andJ.Vogel, "The primary transcriptome of the major human pathogen Helicobacter pylori.," *Nature*, vol. 464, no. 7286, pp. 250–5, Mar.2010.

[139] Y.Chao, K.Papenfort, R.Reinhardt, C. M.Sharma, andJ.Vogel, "An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs.," *EMBO J.*, vol. 31, no. 20, pp. 4005–19, Oct.2012.

[140] S.Ferrara, M.Brugnoli, A.DeBonis, F.Righetti, F.Delvillani, G.Dehò, D.Horner, F.Briani, andG.Bertoni, "Comparative profiling of Pseudomonas aeruginosa strains reveals differential expression of novel unique and conserved small RNAs.," *PLoS One*, vol. 7, no. 5, p. e36553, Jan.2012.

[141] J.Livny, A.Brencic, S.Lory, andM. K.Waldor, "Identification of 17 Pseudomonas aeruginosa sRNAs and prediction of sRNA-encoding genes in 10 diverse pathogens using the bioinformatic tool sRNAPredict2.," *Nucleic Acids Res.*, vol. 34, no. 12, pp. 3484–93, Jan.2006.

[142] S.VChinni, C. A.Raabe, R.Zakaria, G.Randau, C. H.Hoe, A.Zemann, J.Brosius, andT.Tang, "Experimental identification and characterization of 97 novel npcRNA candidates in Salmonella enterica serovar Typhi," vol. 38, no. 17, pp. 5893–5908, 2010.

[143] I.Irnov, C. M.Sharma, J.Vogel, andW. C.Winkler, "Identification of regulatory RNAs in Bacillus subtilis.," *Nucleic Acids Res.*, vol. 38, no. 19, pp. 6637–51, Oct.2010.

[144] L.Li, D.Huang, M. K.Cheung, W.Nong, Q.Huang, andH. S.Kwan, "BSRD: a repository for bacterial small regulatory RNA.," *Nucleic Acids Res.*, vol. 41, no. Database issue, pp. D233-8, Jan.2013.

[145] S.Li, P. K.Ng, H. A. O.Qin, J. K.Lau, andT. C.Lau, "Identification of small RNAs in Mycobacterium smegmatis using heterologous Hfq Identification of small RNAs in Mycobacterium smegmatis using heterologous Hfq," *RNA*, vol. 19, no. 1, pp. 74–84, 2013.

[146] P.Miotto, F.Forti, A.Ambrosi, D.Pellin, D. F.Veiga, G.Balazsi, M. L.Gennaro, C.DiSerio, D.Ghisotti, andD. M.Cirillo, "Genome-wide discovery of small RNAs in Mycobacterium tuberculosis.," *PLoS One*, vol. 7, no. 12, p. e51950, Jan.2012.

[147] G.Lamichhane, K. B.Arnvig, andK. A.McDonough, "Definition and annotation of (myco)bacterial non-coding RNA.," *Tuberculosis (Edinb).*, vol. 93, no. 1, pp. 26–9, Jan.2013.

[148] S.Olivarius, C.Plessy, andP.Carninci, "High-throughput verification of transcriptional starting sites by Deep-RACE.," *Biotechniques*, vol. 46, no. 2, pp. 130–2, Feb.2009.

[149] A.Beauregard, E.Smith, B.Petrone, N.Singh, C.Karch, K.McDonough, andJ.Wade, "Identification and characterization of small RNAs in Yersinia pestis," *RNA Biol.*, vol. 10, no. 3, pp. 397–405, 2013.

[150] A.Peer andH.Margalit, "Accessibility and evolutionary conservation mark bacterial small-rna target-binding regions.," *J. Bacteriol.*, vol. 193, no. 7, pp. 1690–701, Apr.2011.

[151] L. T.Stauffer andG.V.Stauffer, "Multiple Roles for the sRNA GcvB in the Regulation of Slp Levels in Escherichia coli," *ISRN Bacteriol.*, vol. 2013, pp. 1–8, 2013.

[152] P. H.Degnan, H.Ochman, andN. aMoran, "Sequence conservation and functional constraint on intergenic spacers in reduced genomes of the obligate symbiont Buchnera.," *PLoS Genet.*, vol. 7, no. 9, p. e1002252, Sep.2011.

[153] D.Hupalo andA. D.Kern, "Conservation and functional element discovery in 20 angiosperm plant genomes.," *Mol. Biol. Evol.*, vol. 30, no. 7, pp. 1729–44, Jul.2013.

[154] J. A.Willment, D. P.Martin, K. E.Palmer, W. H.Schnippenkoetter, D. N.Shepherd, andE. P.Rybicki, "Identification of long intergenic region sequences involved in maize streak virus replication," *J. Gen. Virol.*, vol. 88, pp. 1831–1841, 2007.

[155] C.Zhang, S. C.Tintó, G.Li, N.Lin, M.Chung, E.Moreno, K. H.Moberg, andL.Zhou, "An intergenic regulatory region mediates Drosophila Myc-induced apoptosis and blocks tissue hyperplasia.," *Oncogene*, no. April, pp. 1–13, Jun.2014.

[156] D. J.Lipman, A.Souvorov, E.VKoonin, A. R.Panchenko, andT. aTatusova, "The relationship of protein conservation and sequence length.," *BMC Evol. Biol.*, vol. 2, p. 20, Nov.2002.

[157] K.Tamura, D.Peterson, N.Peterson, G.Stecher, M.Nei, andS.Kumar, "MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.," *Mol. Biol. Evol.*, vol. 28, no. 10, pp. 2731–9, Oct.2011.

[158] J.Livny, H.Teonadi, M.Livny, andM. K.Waldor, “High-throughput, kingdom-wide prediction and annotation of bacterial non-coding RNAs.,” *PLoS One*, vol. 3, no. 9, p. e3197, Jan.2008.

[159] G.Postic, E.Frapy, M.Dupuis, I.Dubail, J.Livny, A.Charbit, andK. L.Meibom, “Identification of small RNAs in Francisella tularensis.,” *BMC Genomics*, vol. 11, p. 625, Jan.2010.

[160] J.-S.Khoo, S.-F.Chai, R.Mohamed, S.Nathan, andM.Firdaus-Raih, “Computational discovery and RT-PCR validation of novel Burkholderia conserved and Burkholderia pseudomallei unique sRNAs.,” *BMC Genomics*, vol. 13 Suppl 7, no. Suppl 7, p. S13, Jan.2012.

[161] L.Xia, W.Xia, S.Li, W.Li, J.Liu, H.Ding, J.Li, H.Li, Y.Chen, X.Su, W.Wang, L.Sun, C.Wang, N.Shao, andB.Chu, “Identification and expression of small non-coding RNA, L10-Leader, in different growth phases of Streptococcus mutans.,” *Nucleic Acid Ther.*, vol. 22, no. 3, pp. 177–86, Jun.2012.

[162] J.Galagan, A.Lyubetskaya, andA.Gomes, “ChIP-Seq and the complexity of bacterial transcriptional regulation,” *Curr Top Microbiol Immunol*, vol. 363, pp. 43–68, 2013.

[163] K. B.Arnvig andD. B.Young, “Identification of small RNAs in Mycobacterium tuberculosis.,” *Mol Microbiol*, vol. 73, no. 3, pp. 397–408, Aug.2009.

[164] H.Li andR.Durbin, “Fast and accurate short read alignment with Burrows-Wheeler transform.,” *Bioinformatics*, vol. 25, no. 14, pp. 1754–60, Jul.2009.

[165] S.Gottesman, “Micros for microbes: non-coding regulatory RNAs in bacteria.,” *Trends Genet.*, vol. 21, no. 7, pp. 399–404, Jul.2005.

[166] K. M.Wassarman, “Small RNAs in bacteria: diverse regulators of gene expression in response to environmental changes.,” *Cell*, vol. 109, no. 2, pp. 141–4, Apr.2002.

[167] N.Majdalani, C. K.Vanderpool, andS.Gottesman, “Bacterial Small RNA Regulators,” *Crit Rev Biochem Mol Biol*, vol. 40, pp. 93–113, 2005.

[168] K.Papenfort andJ.Vogel, “Regulatory RNA in bacterial pathogens.,” *Cell Host Microbe*, vol. 8, no. 1, pp. 116–27, Jul.2010.

[169] M. R.Garcia-silva, F.Cabrera-cabrera, M. C.Güida, andA.Cayota, “Novel aspects of tRNA-derived small RNAs with potential impact in infectious diseases,” *Adv Biosci Biotechnol*, vol. 4, pp. 17–25, 2013.

[170] S.Debroy, M.Gebbie, A.Ramesh, J. R.Goodson, M. R.Cruz, A.VanHoof, W. C.Winkler, andD. A.Garsin, “A riboswitch-containing sRNA controls gene expression by sequestration of a response regulator,” *Science (80-.).*, vol. 345, no. 6199, pp. 937–940, 2014.

[171] V.Sharma, A.Yamamura, andY.Yokobayashi, “Engineering Artificial Small RNAs for Conditional Gene Silencing in Escherichia coli,” *ACS Synth Biol*, vol. 1, pp. 6–13, 2012.

[172] D.Na, S. M.Yoo, H.Chung, H.Park, J. H.Park, andS. Y.Lee, “Metabolic engineering of Escherichia coli using synthetic small regulatory rNAs,” *Nat. Biotechnol.*, vol. 31, no. 2, pp. 170–174, 2013.

[173] R.Raghavan, E. A.Groisman, andH.Ochman, “Genome-wide detection of novel regulatory RNAs in E. coli.,” *Genome Res*, vol. 21, no. 9, pp. 1487–97, Sep.2011.

[174] F.Lecointe, G.Coste, S.Sommer, andA.Bailone, “Vectors for regulated gene expression in the radioresistant bacterium Deinococcus radiodurans.,” *Gene*, vol. 336, no. 1, pp. 25–35, Jul.2004.

[175] H.Brim, S. C.McFarlan, J. K.Fredrickson, K. W.Minton, M.Zhai, L. P.Wackett, andM. J.Daly, “Engineering Deinococcus radiodurans for metal remediation in radioactive mixed waste environments.,” *Nat. Biotechnol.*, vol. 18, no. 1, pp. 85–90, Jan.2000.

[176] D.Appukuttan, A. S.Rao, andS. K.Apte, “Engineering of Deinococcus radiodurans R1 for bioprecipitation of uranium from dilute nuclear waste.,” *Appl. Environ. Microbiol.*, vol. 72, no. 12, pp. 7873–8, Dec.2006.

[177] C. C.Lange, L. P.Wackett, K. W.Minton, andM. J.Daly, “Engineering a recombinant Deinococcus radiodurans for organopollutant degradation in radioactive mixed waste environments,” *Nat. Biotechnol.*, vol. 16, pp. 929–933, 1998.

[178] K. W.Minton, “DNA repair in the extremely radioresistant bacterium Deinococcus radiodurans.,” *Mol Microbiol*, vol. 13, no. 1, pp. 9–15, Jul.1994.

[179] D.Slade andM.Radman, “Oxidative stress resistance in Deinococcus radiodurans.,” in *Microbiology and molecular biology reviews : MMBR*, vol. 75, no. 1, 2011, pp. 133–91.

[180] M.Tanaka, A. M.Earl, H. aHowell, M.-J.Park, J. aEisen, S. N.Peterson, andJ. R.Battista, “Analysis of Deinococcus radiodurans’s transcriptional response to ionizing radiation and desiccation reveals novel proteins that contribute to extreme radioresistance.,” *Genetics*, vol. 168, no. 1, pp. 21–33, Sep.2004.

[181] B.Basu andS. K.Apte, “Gamma radiation-induced proteome of Deinococcus radiodurans primarily targets DNA repair and oxidative stress alleviation.,” *Mol. Cell. Proteomics*, vol. 11, no. 1, p. M111.011734, Jan.2012.

[182] G.Xu, L.Wang, H.Chen, H.Lu, N.Ying, B.Tian, andY.Hua, “RecO is essential for DNA damage repair in Deinococcus radiodurans.,” *J. Bacteriol.*, vol. 190, no. 7, pp. 2624–8, Apr.2008.

[183] C.Bonacossa de Almeida, G.Coste, S.Sommer, andA.Bailone, “Quantification of RecA protein in Deinococcus radiodurans reveals involvement of RecA, but not LexA, in its regulation.,” *Mol. Genet. Genomics*, vol. 268, no. 1, pp. 28–41, Sep.2002.

[184] J.Repar, S.Cvjetan, D.Slade, M.Radman, D.Zahradka, andK.Zahradka, “RecA protein assures fidelity of DNA repair and genome stability in Deinococcus radiodurans.,” *DNA Repair (Amst).*, vol. 9, no. 11, pp. 1151–61, Nov.2010.

[185] D.Slade, A. B.Lindner, G.Paul, andM.Radman, “Recombination and replication in DNA repair of heavily irradiated Deinococcus radiodurans.,” *Cell*, vol. 136, no. 6, pp. 1044–55, Mar.2009.

[186] T.Bordelon, S. P.Wilkinson, A.Grove, andM. E.Newcomer, “The crystal structure of the transcriptional regulator HucR from Deinococcus radiodurans reveals a repressor preconfigured for DNA binding.,” *J. Mol. Biol.*, vol. 360, no. 1, pp. 168–77, Jun.2006.

[187] X.Chen, D. W.Taylor, C. C.Fowler, J. E.Galan, H.-W.Wang, andS. L.Wolin, “An RNA degradation machine sculpted by Ro autoantigen and noncoding RNA.,” *Cell*, vol. 153, no. 1, pp. 166–77, Mar.2013.

[188] X.Chen, A. M.Quinn, andS. L.Wolin, “Ro ribonucleoproteins contribute to the resistance of Deinococcus radiodurans to ultraviolet irradiation,” *Genes Dev.*, vol. 14, pp. 777–782, 2000.

[189] C.Tsai, R.Liao, B.Chou, M.Palumbo, andM.Contreras, “Genome-Wide Analyses in Bacteria Show Small-RNA Enrichment for Long and Conserved Intergenic Regions,” *J. Bacteriol.*, vol. 197, no. 1, pp. 40–50, 2015.

[190] B.Langmead andS. L.Salzberg, “Fast gapped-read alignment with Bowtie 2,” *Nat. Methods*, vol. 9, no. 4, pp. 357–359, 2012.

[191] P.Gueneau de Novoa andK. P.Williams, “The tmRNA website: reductive evolution of tmRNA in plastids and other endosymbionts.,” *Nucleic Acids Res.*, vol. 32, no. Database issue, pp. D104-8, Jan.2004.

[192] S.Luban andD.Kihara, “Comparative genomics of small RNAs in bacterial genomes.,” *OMICS*, vol. 11, no. 1, pp. 58–73, Jan.2007.

[193] J.Kim, A. K.Sharma, S. N.Abbott, A.Wood, D. W.Dwyer, A.Jambura, W.Kenneth, R. B.Inman, M. J.Daly, M. M.Cox, E. A.Wood, andK. W.Minton, “RecA Protein from the Extremely Radioresistant Bacterium Deinococcus radiodurans : Expression , Purification , and Characterization,” *J. Bacteriol.*, vol. 184, no. 6, pp. 1649–60, 2002.

[194] J. D.Carroll, M. J.Daly, andK. W.Minton, “Expression of recA in Deinococcus radiodurans.,” *J. Bacteriol.*, vol. 178, no. 1, pp. 130–5, Jan.1996.

[195] M.VOmelchenko, Y. I.Wolf, E. K.Gaidamakova, V. Y.Matrosova, A.Vasilenko, M.Zhai, M. J.Daly, E.VKoonin, andK. S.Makarova, “Comparative genomics of Thermus thermophilus and Deinococcus radiodurans: divergent routes of adaptation to thermophily and radiation resistance.,” *BMC Evol. Biol.*, vol. 5, p. 57, Jan.2005.

[196] C.Bauche andJ.Laval, “Repair of oxidized bases in the extremely radiation-resistant bacterium Deinococcus radiodurans.,” *J. Bacteriol.*, vol. 181, no. 1, pp. 262–9, Jan.1999.

[197] R. G. E.Murray, “The family Deinococcaceae,” in *The Prokaryotes*, A.Balows, H. G.Trüper, M.Dworkin, W.Harder, andK.-H.Schleifer, Eds.Springer-Verlag, New York, N.Y., 1992, pp. 3732–3744.

[198] Y.Liu, J.Zhou, M.VOmelchenko, A. S.Beliaev, A.Venkateswaran, J.Stair, L.Wu, D. K.Thompson, D.Xu, I. B.Rogozin, E. K.Gaidamakova, M.Zhai, K. S.Makarova, E.VKoonin, andM. J.Daly, “Transcriptome dynamics of Deinococcus radiodurans recovering from ionizing radiation.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 7, pp. 4191–6, Apr.2003.

[199] H.Luan, N.Meng, J.Fu, X.Chen, X.Xu, Q.Feng, H.Jiang, J.Dai, X.Yuan, Y.Lu, A. aRoberts, X.Luo, M.Chen, S.Xu, J.Li, C. J.Hamilton, C.Fang, andJ.Wang, “Genome-wide transcriptome and antioxidant analyses on gamma-irradiated phases of deinococcus radiodurans R1.,” *PLoS One*, vol. 9, no. 1, p. e85649, Jan.2014.

[200] O.White, “Genome Sequence of the Radioresistant Bacterium Deinococcus radiodurans R1,” *Science (80-.).*, vol. 286, no. 5444, pp. 1571–1577, Nov.1999.

[201] B. E.Moseley andD. M.Evans, “Isolation and properties of strains of Micrococcus (Deinococcus) radiodurans unable to excise ultraviolet light-induced pyrimidine dimers from DNA: evidence for two excision pathways.,” *J. Gen. Microbiol.*, vol. 129, no. 8, pp. 2437–45, Aug.1983.

[202] K. S.Makarova, M.VOmelchenko, E. K.Gaidamakova, V. Y.Matrosova, A.Vasilenko, M.Zhai, A.Lapidus, A.Copeland, E.Kim, M.Land, K.Mavrommatis, S.Pitluck, P. M.Richardson, C.Detter, T.Brettin, E.Saunders, B.Lai, B.Ravel, K. M.Kemner, Y. I.Wolf, A.Sorokin, A.VGerasimova, M. S.Gelfand, J. K.Fredrickson, E.VKoonin, andM. J.Daly, “Deinococcus geothermalis: the pool of extreme radiation resistance genes shrinks.,” *PLoS One*, vol. 2, no. 9, p. e955, Jan.2007.

[203] A. C.Ferreira, M. F.Nobre, F. A.Rainey, M. T.Silva, R.Wait, J.Burghardt, A. P.Chung, andM. S.Costa, “Deinococcus geothermalis sp. nov. and Deinococcus murrayi sp. nov.,Two Extremely Radiation-Resistant and Slightly Thermophilic Species from Hot Springs,” *Int J Syst Bacteriol*, vol. 47, no. 4, pp. 939–947, 1997.

[204] B.Tjaden, “TargetRNA: a tool for predicting targets of small RNA action in bacteria.,” *Nucleic Acids Res.*, vol. 36, no. Web Server issue, pp. W109-13, Jul.2008.

[205] D.Rath, S. H.Mangoli, A. R.Pagedar, andN.Jawali, “Involvement of pnp in survival of UV radiation in Escherichia coli K-12.,” *Microbiology*, vol. 158, no. Pt 5, pp. 1196–205, May2012.

[206] A. J.Potter, S. P.Kidd, A. G.McEwan, andJ. C.Paton, “The MerR/NmlR family transcription factor of Streptococcus pneumoniae responds to carbonyl stress and modulates hydrogen peroxide production.,” *J. Bacteriol.*, vol. 192, no. 15, pp. 4063–6, Aug.2010.

[207] J. L.Ramos, M.Martínez-bueno, J.Antonio, W.Terán, K.Watanabe, M. T.Gallegos, R.Brennan, M.Martı, A. J.Molina-henares, andW.Tera, “The TetR Family of Transcriptional Repressors,” *Microbiol. Mol. Biol. Rev.*, vol. 69, no. 2, pp. 326–356, 2005.

[208] R.Khankal, J. W.Chin, D.Ghosh, andP. C.Cirino, “Transcriptional effects of CRP* expression in Escherichia coli.,” *J. Biol. Eng.*, vol. 3, p. 13, Jan.2009.

[209] M. J.Daly, “Death by protein damage in irradiated cells.,” *DNA Repair (Amst).*, vol. 11, no. 1, pp. 12–21, Jan.2012.

[210] S. S.Sukhi, R.Shashidhar, S. aKumar, andJ. R.Bandekar, “Radiation resistance of Deinococcus radiodurans R1 with respect to growth phase.,” *FEMS Microbiol. Lett.*, vol. 297, no. 1, pp. 49–53, Aug.2009.

[211] S.Litthauer, S.Gargiulo, E.vanHeerden, F.Hollmann, andD. J.Opperman, “Heterologous expression and characterization of the ene-reductases from Deinococcus radiodurans and Ralstonia metallidurans,” *J. Mol. Catal. B Enzym.*, vol. 99, pp. 89–95, Jan.2014.

[212] L. I.Wei, M. A.Yun, X.Fangzhu, andH. E.Shuya, “Ionizing Radiation Resistance in Deinococcus Radiodurans,” *Adv. Nat. Sci.*, vol. 7, no. 2, pp. 6–14, 2014.

[213] D.Slade andM.Radman, “Oxidative stress resistance in Deinococcus radiodurans.,” in *Microbiology and molecular biology reviews : MMBR*, vol. 75, no. 1, 2011, pp. 133–191.

[214] V.Fallico, R. P.Ross, G. F.Fitzgerald, andO.McAuliffe, “Genetic response to bacteriophage infection in Lactococcus lactis reveals a four-strand approach involving induction of membrane stress proteins, D-alanylation of the cell wall, maintenance of proton motive force, and energy conservation.,” *J. Virol.*, vol. 85, no. 22, pp. 12032–12042, 2011.

[215] K. A.Spriggs, M.Bushell, andA. E.Willis, “Translational Regulation of Gene Expression during Conditions of Cell Stress,” *Mol. Cell*, vol. 40, no. 2, pp. 228–237, 2010.

[216] A.Boysen, J.Møller-Jensen, B.Kallipolitis, P.Valentin-Hansen, andM.Overgaard, "Translational regulation of gene expression by an anaerobically induced small non-coding RNA in Escherichia coli," *J. Biol. Chem.*, vol. 285, no. 14, pp. 10690–10702, 2010.

[217] N.Delihas andS.Forst, "MicF: an antisense RNA gene involved in response of Escherichia coli to global stress factors.," *J. Mol. Biol.*, vol. 313, no. 1, pp. 1–12, 2001.

[218] J.Andersen, S. A.Forst, K.Zhao, M.Inouye, andN.Delihas, "The function of micF RNA. micF RNA is a major factor in the thermal regulation of OmpF protein in Escherichia coli," *J. Biol. Chem.*, vol. 264, no. 30, pp. 17961–17970, 1989.

[219] J. M.Harms, D. N.Wilson, F.Schluenzen, S. R.Connell, T.Stachelhaus, Z.Zaborowska, C. M. T.Spahn, andP.Fucini, "Translational Regulation via L11: Molecular Switches on the Ribosome Turned On and Off by Thiostrepton and Micrococcin," *Mol. Cell*, vol. 30, no. 1, pp. 26–38, 2008.

[220] C.Hoe, C. A.Raabe, T. S.Rozhdestvensky, andT.Tang, "Bacterial sRNAs: Regulation in stress," *Int. J. Med. Microbiol.*, vol. 303, no. 5, pp. 217–229, 2013.

[221] S. W.Sowa, G.Gelderman, A. N.Leistra, A.Buvanendiran, S.Lipp, A.Pitaktong, C. A.Vakulskas, T.Romeo, M.Baldea, andL. M.Contreras, "Integrative FourD omics approach profiles the target network of the carbon storage regulatory system," *Nucleic Acids Res.*, vol. 45, no. 4, pp. 1673–1686, 2017.

[222] M.Ashburner, C. A.Ball, J. A.Blake, D.Botstein, H.Butler, J. M.Cherry, A. P.Davis, K.Dolinski, S. S.Dwight, J. T.Eppig, M. A.Harris, D. P.Hill, L.Issel-Tarver, A.Kasarskis, S.Lewis, J. C.Matese, J. E.Richardson, M.Ringwald, G. M.Rubin, G.Sherlock, andG. O.Consortium, "Gene Ontology: Tool for The Unification of Biology," *Nat. Genet.*, vol. 25, no. 1, pp. 25–29, 2000.

[223] J. M.Eggington, N.Haruta, E. A.Wood, andM. M.Cox, "The single-stranded DNA-binding protein of Deinococcus radiodurans," *BMC Microbiol.*, vol. 4, no. 1, p. 2, Jan.2004.

[224] A. K.Schmid andM. E.Lidstrom, "Involvement of Two Putative Alternative Sigma Factors in Stress Response of the Radioresistant Bacterium Deinococcus radiodurans," *J. Bacteriol.*, vol. 184, no. 22, pp. 6182–6189, Nov.2002.

[225] B.Basu andS. K.Apte, "Gamma radiation-induced proteome of Deinococcus radiodurans primarily targets DNA repair and oxidative stress alleviation.," *Mol. {&} Cell. proteomics MCP*, vol. 11, no. 1, p. M111.011734, Jan.2012.

[226] P. R.Wright, J.Georg, M.Mann, D. A.Sorescu, A. S.Richter, S.Lott, R.Kleinkauf, W. R.Hess, andR.Backofen, "CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains," *Nucleic Acids Res.*, vol. 42, no. Web Server issue, pp. W119–W123, Jul.2014.

[227] H.Lu, G.Gao, G.Xu, L.Fan, L.Yin, B.Shen, andY.Hua, “Deinococcus radiodurans PprI switches on DNA damage response and cellular survival networks after radiation damage.,” *Mol. {&} Cell. proteomics MCP*, vol. 8, no. 3, pp. 481–494, 2009.

[228] T. R.Cech, J. A.Steitz, S.Stickel, N.Gomes, T. T.Su, A.Giuffrè, V. B.Borisov, M.Arese, P.Sarti, E.Forte, J.Livny, M. A.Fogel, B. M.Davis, M. K.Waldor, K. M.Wassarman, F.Repoila, C.Rosenow, G.Storz, S.Gottesman, K.Prévost, H.Salvail, G.Desnoyers, J. F.Jacques, É. Phaneuf, E.Massé, L. M.Markillie, S. M.Varnum, P.Hradecky, K. K.Wong, H.Kawamoto, Y.Koide, T.Morita, H.Aiba, J.Vogel, V.Bartels, T. H.Tang, G.Churakov, J. G.Slagter-Jäger, A.Hüttenhofer, E. G. H.Wagner, E.Bentchikou, P.Servant, G.Coste, S.Sommer, S. H.Park, H.Singh, D.Appukuttan, S.Jeong, Y. J.Choi, J. H.Jung, I.Narumi, S.Lim, H.Ohba, K.Satoh, H.Sghaier, T.Yanagisawa, I.Narumi, Y.Zeng, Y.Ma, F.Xiao, W. W.Wang, S.He, L. Y.Wang, L. F.Yin, G. Z.Xu, M. F.Li, H.Zhang, B.Tian, Y. J.Hua, H.Lu, G.Gao, G. Z.Xu, L.Fan, L. F.Yin, B.Shen, Y. J.Hua, P.Zhao, Z.Zhou, W.Zhang, M.Lin, M.Chen, G.Wei, Y. J.Hua, I.Narumi, G.Gao, B.Tian, K.Satoh, S.Kitayama, B.Shen, P.Shan, G.Fan, L.Sun, J.Liu, W. W.Wang, C.Hu, X.Zhang, Q.Zhai, X.Song, L.Cao, Y.Cui, S.Zhang, C.Wang, P. J.Pomposiello, B.Demple, S.Stickel, N.Gomes, T. T.Su, D.Biology, N.Delihas, S. A.Forst, J.Andersen, S. A.Forst, K.Zhao, M.Inouye, N.Delihas, B.Wightman, I.Ha, G.Ruvkun, N.Shahmolky, D. L.Lefebvre, R.Poon, Y.Bai, M.Sharma, C. F.Rosen, J. M.Harms, D. N.Wilson, F.Schluenzen, S. R.Connell, T.Stachelhaus, Z.Zaborowska, C. M. T.Spahn, P.Fucini, V.Fallico, R. P.Ross, G. F.Fitzgerald, O.McAuliffe, A. M.Giuliodori, A.Brandi, M.Giangrossi, C. O.Gualerzi, C. L.Pon, S.Soto, J.Vila, M. M.Tavío, P.Acebo, A. J.Martin-galiano, S.Navarro, J.Vogel, C. M.Sharma, T.Song, F.Mika, B.Lindmark, Z.Liu, S.Schild, A.Bishop, J.Zhu, A.Camilli, J.Johansson, J.Vogel, S. N.Wai, A.Airo, S. L.Chan, Z.Martinez, M. O.Platt, J. D.Trent, B.Reichenbach, A.Maes, F.Kalamorz, E.Hajnsdorf, B.Görke, A.Manuscript, J. A.Opdyke, E. M.Fozo, M. R.Hemm, G.Storz, A.Boysen, J.Møller-Jensen, B.Kallipolitis, P.Valentin-Hansen, M.Overgaard, C. L.Squires, F.Krasin, F.Hutchinson, L. I.Wei, M. A.Yun, X.Fangzhu, H. E.Shuya, D.Slade, M.Radman, A.Conesa, P.Madrigal, S.Tarazona, D.Gomez-Cabrero, A.Cervera, A.McPherson, M. W.Szczesniak, D. J.Gaffney, L. L.Elo, X.Zhang, A.Mortazavi, M.VOmelchenko, Y. I.Wolf, E. K.Gaidamakova, V. Y.Matrosova, A.Vasilenko, M.Zhai, M. J.Daly, E.VKoonin, K. S.Makarova, L.Oakford, R.Barrangou, P.Horvath, J.Wyman, J. P.Changeux, D.Filmer, T. M.Jovin, W.Baehr, J. J.Holbrook, N.Dattagupta, D. M.Crothers, G. W.Hatfield, R.Bruinsma, T.Maniatis, S. C.Harrison, A. J.Spakowitz, P. C.Blainey, C. M.Schroeder, X. S.Xie, T.Strzelecka, L. F.Dorner, I.Schildkraut, A. K.Aggarwal, S.Bailey, T. A.Steitz, L.Finzi, C.Bustamante, C. T.Martin, S. S.Patel, A.Kumar, S. S.Patel, S.Oehler, A. K.Aggarwal, S.Stayrook, A.Rosenberg, M.Lewis, J.Widom, J.Hynes, A.Szabo, C.Bustamante, M.Citartan, C. A.Raabe, C. H.Hoe, T. S.Rozhdestvensky, T. H.Tang, P.Wan, H.Lu, G.Gao, G. Z.Xu, L.Fan, L. F.Yin, B.Shen, Y. J.Hua, A. K.Schmid, H. aHowell, J. R.Battista, N.Peterson, M.

E.Lidstrom, S. N.Peterson, A. T.Cavanagh, J. M.Sperger, K. M.Wassarman, L.Argaman, R.Hershberg, J.Vogel, G.Bejerano, E. G. H.Wagner, H.Margalit, S.Altuvia, M.Tanaka, A. M.Earl, H. aHowell, M.-J.Park, J. aEisen, S. N.Peterson, J. R.Battista, L. S.Waters, andG.Storz, "Knockout of pprM Decreases Resistance to Desiccation and Oxidation in Deinococcus radiodurans," *Cell*, vol. 11, no. 1, pp. 272–290, Sep.2009.

[229] S. H.Park, H.Singh, D.Appukuttan, S.Jeong, Y. J.Choi, J. H.Jung, I.Narumi, andS.Lim, "PprM, a cold shock domain-containing protein from deinococcus radiodurans, confers oxidative stress tolerance to escherichia coli," *Front. Microbiol.*, vol. 7, no. JAN, pp. 1–11, 2017.

[230] S.-W.Jeong, H. S.Seo, M.-K.Kim, J.-I.Choi, H.-M.Lim, andS.Lim, "PprM is necessary for up-regulation of katE1, encoding the major catalase of Deinococcus radiodurans, under unstressed culture conditions," *J. Microbiol.*, vol. 54, no. 6, pp. 426–431, Jun.2016.

[231] A.Airo, S. L.Chan, Z.Martinez, M. O.Platt, andJ. D.Trent, "Heat shock and cold shock in Deinococcus radiodurans.," *Cell Biochem. Biophys.*, vol. 40, no. 3, pp. 277–288, 2004.

[232] A. K.Schmid, H. A.Howell, J. R.Battista, N.Peterson, M. E.Lidstrom, andS. N.Peterson, "Global Transcriptional and Proteomic Analysis of the Sig1 Heat Shock Regulon of Deinococcus radiodurans Global Transcriptional and Proteomic Analysis of the Sig1 Heat Shock Regulon of Deinococcus radiodurans †," vol. 187, no. 10, pp. 3339–3351, 2005.

[233] L. Y.Wang, L. F.Yin, G. Z.Xu, M. F.Li, H.Zhang, B.Tian, andY. J.Hua, "Cooperation of PprI and DrRRA in response to extreme ionizing radiation in Deinococcus radiodurans," *Chinese Sci. Bull.*, vol. 57, no. 1, pp. 98–104, 2012.

[234] A. M.Giuliodori, A.Brandi, M.Giangrossi, C. O.Gualerzi, andC. L.Pon, "Cold-stress-induced de novo expression of infC and role of IF3 in cold-shock translational bias," *RNA*, vol. 13, no. 8, pp. 1355–1365, Aug.2007.

[235] P.Wan, "Mechanisms of Radiation Resistance in Deinococcus radiodurans R1 Revealed by the Reconstruction of Gene Regulatory Network Using Bayesian Network Approach," *J. Proteomics {&} Bioinforma.*, vol. 1, no. S6, pp. 6–10, 2013.

[236] H. S.Misra, N. P.Khairnar, S.Kota, S.Shrivastava, V. P.Joshi, andS. K.Apte, "An exonuclease I-sensitive DNA repair pathway in Deinococcus radiodurans : a major determinant of radiation resistance," *Mol Microbiol*, vol. 59, no. 4, pp. 1308–1316, 2006.

[237] J. K.Villa, P.Amador, J.Janovsky, A.Bhuyan, R.Saldanha, T. J.Lamkin, andL. M.Contreras, "A genome-wide search for ionizing-radiationresponsive elements in Deinococcus radiodurans reveals a regulatory role for the DNA gyrase subunit A

gene's 5' untranslated region in the radiation and desiccation response," *Appl. Environ. Microbiol.*, vol. 83, no. 12, pp. 1–19, 2017.

[238] J.Vazquez-Anderson, M. K.Mihailovic, K. C.Baldridge, K. G.Reyes, K.Haning, S. H.Cho, P.Amador, W. B.Powell, andL. M.Contreras, "Optimization of a novel biophysical model using large scale in vivo antisense hybridization data displays improved prediction capabilities of structurally accessible RNA regions," *Nucleic Acids Res.*, vol. 45, no. 9, pp. 5523–5538, 2017.

[239] M. I.Love, W.Huber, andS.Anders, "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2," *Genome Biol*, vol. 15, 2014.