

Spring 6-8-2016

Investigating the Origin and Functions of a Novel Small RNA in *Escherichia coli*

Fenil Rashmin Kacharia
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds



Part of the [Bacteria Commons](#), [Biology Commons](#), and the [Genetics and Genomics Commons](#)

Let us know how access to this document benefits you.

Recommended Citation

Kacharia, Fenil Rashmin, "Investigating the Origin and Functions of a Novel Small RNA in *Escherichia coli*" (2016). *Dissertations and Theses*. Paper 3106.

<https://doi.org/10.15760/etd.3103>

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

Investigating the Origin and Functions of a Novel Small RNA in *Escherichia coli*

by

Fenil Rashmin Kacharia

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Science
in
Biology

Thesis Committee:
Rahul Raghavan, Chair
Michael Bartlett
Niles Lehman

Portland State University
2016

ABSTRACT

Non-coding small RNAs (sRNAs) regulate various cellular processes in bacteria. They bind to a chaperone protein Hfq for stability and regulate gene expression by base-pairing with target mRNAs. Although the importance of sRNAs in bacteria has been well established, the mode of origination of novel sRNA genes is still elusive, mainly because the rapid rate of evolution of sRNAs obscures their original sources. To overcome this impediment, we identified a recently formed sRNA (EcsR2) in *E. coli*, and show that it evolved from a degraded bacteriophage gene. Our analyses also revealed that young sRNAs such as EcsR2 are expressed at low levels and evolve at a rapid rate in comparison to older sRNAs, thereby uncovering a novel process that potentially facilitates newly emerging (and probably mildly deleterious) sRNAs to persist in bacterial genomes. We also show that even though EcsR2 is slightly deleterious to *E. coli*, it could bind to Hfq and mRNAs to regulate the expression of several genes. Interestingly, while EcsR2 expression is induced by glucose, the expression of its putative targets are regulated by the transcription factor CRP in response to glucose, indicating that EcsR2 has been incorporated into the carbon regulatory network in *E. coli*. Collectively, this work provides evidence for the emergence, evolution and functions of a novel 'young' sRNA in bacteria.

ACKNOWLEDGEMENTS

I would like to start by saying that this work wouldn't have been possible without the confidence and foresight of Rashmin Laldas Kacharia, Kanan Rashmin Kacharia, Mitsul Rashmin Kacharia and most importantly my thesis adviser Dr. Rahul Raghavan. When I was at the lowest of my academic career, Dr. Raghavan took a gamble by accepting me as his first graduate student. For bestowing this opportunity on me and investing time and effort in me, I don't think there are words in the dictionary to express my gratitude for him. The boundaries of day and night and weekdays and weekend have blurred over the past three years while working in this lab. The only reason for working with such diligence was to do science of impeccable quality, thereby not letting down Dr. Raghavan and fulfilling my father's aspiration of seeing his son getting a masters degree conferred upon.

The journey thus far has not been an easy one, but I knew what I signed up for and I would be lying if I said that I was always confident and enthusiastic about achieving the goals of my project. During hard times Dr. Raghavan, his wife Misha Dawn and my family always cheered me up and reminded me that it wont be easy but it will be worth it in the end. And I am proud to say that it has been.

Lastly, my committee members, Dr. Michael Bartlett and Dr. Niles Lehman gave good feedback and helped me connect the dots to finish my work. With their support and guidance I am very confident and proud to present my research. I would also like to extend my thanks to the Raghavan Laboratory members specifically, Abraham Shepherd

Moses, Jess Millar, Todd Smith, Jim Archuleta and all the current and past volunteers. I hope my work will have contributed to the scientific community.

TABLE OF CONTENTS

	PAGE
ABSTRACT	i
ACKNOWLEDGEMENTS	ii
LIST OF TABLES	v
LIST OF FIGURES	vi
INTRODUCTION	1
MATERIALS AND METHODS	7
RESULTS	16
DISCUSSION	35
APPENDICES	
A. Role of MgrR in polymyxin B resistance.	50
B. Function of a novel sRNA.	52

LIST OF TABLES

1:	List of plasmids and strains used in this study.	9
2:	List of primers and oligonucleotides used in this study.	10

LIST OF FIGURES	PAGE
1: Impact of EcsR1 on biofilm formation.	4
2: Detection of EcsR2 in <i>yagU-ykgJ</i> IGR in <i>E. coli</i> .	17
3: Conservation and formation of <i>yagU-ykgJ</i> IGR in <i>E. coli</i> .	18
4: Origination of EcsR2 from a protein-coding gene.	20
5: Expression of EcsR2 at different growth phases.	22
6: Mildly deleterious effect of EcsR2 in minimal medium.	23
7: Young sRNAs have low expression and high SNPs.	24
8: Hfq stabilizes EcsR2.	27
9: Crosslink-Seq.	28
10: Predicted EcsR2-target mRNA interactions.	30
11: Down regulation of target genes by EcsR2.	31
12: Glucose induces EcsR2 expression.	33
13: Mlc binding site.	34
14: Predicted secondary structure of EcsR2.	36

INTRODUCTION

Phenotypic diversity and response to environmental signals have long been attributed to proteins in all domains of life. However, the discovery of non-protein coding RNAs (ncRNAs) that bind with proteins, DNA and/or mRNAs has changed our understanding of gene regulation. As more sophisticated methods were developed for genetic studies, strong evidence for transcripts from intergenic regions (IGR) started to accumulate, leading to the identification of several classes of ncRNAs with diverse modes of mechanisms and functions in Eukaryotes, Archaea and Bacteria. In Eukaryotes, major classes of ncRNA include microRNA (miRNA) and small-interfering RNA (siRNA) that bind to target mRNAs and modulate translation (Filipowicz et al. 2005; Murchison et al. 2004; Bartel 2004). In Archaea, ncRNAs such as tRFs (tRNA derived fragments) have been found to regulate many biological roles such as adaptation to extreme temperature, metabolic regulation, stress response, and regulation of cellular morphology (Babski et al. 2014). In bacteria, post-transcriptional regulation of gene expression is carried out by small RNAs (sRNA), which are typically 50-500 nucleotides (nt) long and are encoded in IGRs. Other important classes of bacterial ncRNAs are anti-sense RNA (asRNA), which are transcribed from the opposite strand of the protein-coding genes (Georg and Hess 2011), riboswitches and RNA thermometers that are located in the untranslated regions of certain mRNAs (Breaker 2011; Kortmann and Narberhaus 2012), and intraRNAs that originate from within protein-coding genes (Miyakoshi et al. 2015).

Bacterial sRNAs control a wide range of cellular processes, including biofilm production, lipopolysaccharide modification, motility and survival in harsh conditions

(Raghavan et al. 2015; Moon et al. 2009; Thomason et al. 2012; Jones et al. 2006). For example, an sRNA EcsR1 that is a part of CRP and FNR regulons impacts biofilm production in *E. coli* (**Figure 1**; Raghavan et al. 2015). In addition to biofilm production, sRNAs also control motility by sensing changes in nutrient availability. When cells reach stationary growth phase and the nutrient availability is scarce, sRNA McaS is highly expressed. The sRNA binds to and activates genes involved in enhancing biofilm and flagella synthesis thereby helping *E. coli* to obtain a sessile lifestyle (Jorgensen et al. 2013). Another sRNA in *E. coli* that mediates cellular survival by responding to environmental conditions is DsrA. When cells experience osmotic shock and low temperatures, DsrA is expressed and binds to RpoS, a general stress and stationary phase sigma factor. The binding of DsrA to *rpoS* enhances the translation of the protein thereby helping the cells to mediate stress (McCullen et al. 2010).

Most sRNAs are expressed in response to environmental cues such as availability of certain carbon source, nutrient concentration and cell density (Večerek et al. 2007; Negrete et al. 2010; Hoe et al. 2013). There are several advantages of gene regulation via sRNAs over gene regulation by proteins. The ability of sRNAs to directly bind to mRNAs causes the degradation of both sRNAs and mRNAs leading to precise control of gene expression. Controlling these activities through sRNA-mRNA interaction is faster as no translation is required, and degradation of sRNA and mRNA helps in faster recycling of nucleotides. sRNAs are also energetically cheaper to synthesize compared to proteins (Beisel and Storz 2010). Moreover, nucleotide substitutions in sRNAs can alter

the strength of sRNA-mRNA interaction, which is advantageous for fine-tuning gene expression. Such modifications are difficult to achieve in proteins, as substantial

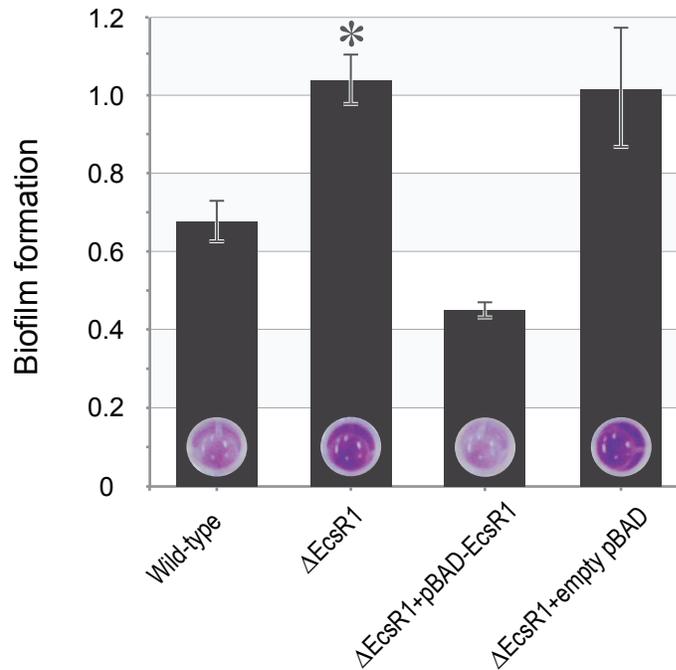


Figure 1. Impact of EcsR1 on biofilm formation. Biofilm formation is influenced by sRNA EcsR1 (Raghavan et al. 2015). A wild-type strain, an EcsR1-deleted strain (Δ EcsR1), a EcsR1 strain containing pBAD with cloned EcsR1 (Δ EcsR1+pBAD-EcsR1), and a EcsR1 strain containing empty pBAD (Δ EcsR1+empty pBAD) were tested. Asterisks indicate a statistically significant difference between wild-type and EcsR1 strains ($P < 0.0001$). Biofilm formation was measured at OD_{600} . The blue circles show the retention of crystal violet in the biofilm.

nucleotide changes are required to change a protein's function. Additionally, sRNAs can alter protein activity by binding directly to them. For example 6S sRNA, one of the first sRNAs to be discovered, accumulates in *E. coli* in stationary phase and binds to RNA polymerase to repress transcription from a σ^{70} -dependent promoter (Wassarman and Storz 2000). This helps *E. coli* to survive when the nutrients are scarce by saving energy. Thus there is a huge array of known sRNA functional mechanisms, and as more sRNAs are discovered, this spectrum is bound to expand.

As diverse mechanisms of sRNA functionality have become apparent, novel methods to detect sRNAs have also been developed. Examples of such methods are functional genetic screenings, microarrays, co-purification with proteins, ortholog gene searches between related species, northern blots and computational predictions. Although these techniques have been fruitful in discovering large number of sRNAs, they also have a lot of shortcomings. For example, genetic screening relies highly on strong phenotypic features of bacteria, is labor intensive, and many sRNAs that are expressed under special conditions may not be identified. Computational analysis is another powerful approach in finding new sRNAs but it requires a list of experimentally confirmed sRNAs in related species. Detection by co-purification with proteins requires sRNAs to be bound to proteins throughout the purification process, and for northern blots to work, the expression of the sRNA has to be high enough to show up on the blot (Vogel and Sharma 2005). Such shortcomings make these methods less reliable for detecting novel sRNAs. Additionally, because traditional sRNA detection methods rely on high

expression and conservation across species, they overlooked most novel sRNAs that have low expression levels and/or are present only in few species.

With the recent development of high-throughput sequencing technologies (RNA-seq), bacterial transcriptomes could be interrogated at great depth, and novel sRNAs could be identified without requiring any prior information about sRNA sequence or structure (Wilderman et al. 2004). Using RNA-seq, our lab was able to identify several novel sRNAs in enteric bacteria (Raghavan et al. 2011, 2012, 2015).

While the importance of sRNAs to bacterial gene regulation is well established, the mechanisms through which sRNAs originate remain largely unknown. One of the main impediments in studying sRNA origination is that, unlike protein-coding genes, sRNA genes evolve at a rapid rate, which has made it difficult to identify their original sources (Gottesman and Storz 2011). One approach to circumvent this impediment is to utilize recently evolved i.e., ‘young’ sRNAs; however, because such sRNAs are phylogenetically restricted, they have been largely overlooked by previous studies. By combining structural and evolutionary genomics approaches with transcriptomics data, we were able to identify several ‘young’ sRNAs in *E. coli* and *Salmonella enterica* (Raghavan et al. 2015). One of the sRNAs present only in *E. coli* is EcsR2. The goal of this project is to (1) characterize the origination of EcsR2, (2) define the expression pattern of EcsR2 and (3) identify potential mRNA targets of EcsR2.

MATERIALS AND METHODS

Bacterial strains and plasmids

The plasmids and strains used in this study are described in **Table 1**. When present, antibiotics were used at following concentrations: ampicillin 100 µg/mL; kanamycin 25 µg/mL; chloramphenicol 25 µg/mL. For pC2 plasmid construction, *EcsR2* gene was amplified from *E. coli* wild-type MG1655 chromosome using primer set *yagU3'*-*nheI_F* and *yagU3'*-*HindIII_R* containing *NheI* and *HindIII* restriction sites respectively. *EcsR2* PCR product and pBAD24 were digested with *NheI* and *HindIII* (ThermoFisher) and gel purified. The purified digestion products were ligated at 1:1 ratio at room temperature using Rapid DNA Ligation kit (ThermoFisher). The ligation reaction was chemically transformed in to *E. coli* JM109 strain (Promega) by following manufacturers instructions and plated on LB agar plates supplemented with ampicillin. Colonies with correct plasmid constructs were confirmed by PCR and DNA sequencing. For pBH plasmid construction, primer set BH_F and BH_R was used to amplify entire pC2 plasmid except nucleotides at positions 50-80 of *EcsR2* by using Phusion High-Fidelity PCR kit (ThermoFisher). The amplified product was ligated using Rapid DNA Ligation Kit (ThermoFisher) and transformed in to *E. coli* JM109. Colonies with correct plasmid constructs were confirmed by PCR and DNA sequencing. Plasmids pC2, pBH and pBAD24 were extracted from *E. coli* JM109 strains by Midiprep (ThermoFisher) and electroporated into *EcsR2*-deletion (RL_001) strain. *EcsR2*-deletion (RL_001) and *hfg*-deletion (RL_005) MG1655 strains were constructed using λ Red-mediated

recombination (Datsenko and Wanner 2000). **Table 2** lists the sequences of all the primers used in this study.

Table 1: List of plasmids and strains used in this study.

Plasmids	Description	Reference or source
pBAD24	Ampr; araBAD promoter-based expression vector with a pACYC origin	Guzman <i>et al.</i> (1995)
pC2	EcsR2 coding region (158 bp) cloned into the <i>NheI</i> and <i>HindIII</i> sites behind the arabinose-inducible promoter on pBAD24	This study
pBH	EcsR2 missing 50-80 nts (96 bp) cloned into the <i>NheI</i> and <i>HindIII</i> sites behind the arabinose-inducible promoter on pBAD24	This study
Strains		
MG1655	Wild-type <i>E. coli</i>	Courcelle Lab, PSU
RL_001	MG1655 $\Delta EcsR2::cm$	This study
RL_002	RL_001, pBAD24	This study
RL_003	RL_001, pC2	This study
RL_004	RL_001, pBH	This study
RL_005	MG1655 $\Delta hfg::cm$	This study
BW25113	F-, $\Delta(\text{araD-araB})567$, $\Delta\text{lacZ}4787(::\text{rrnB-3})$, λ -, rph-1, $\Delta(\text{rhaD-rhaB})568$, hsdR514	CGSC Yale
JW5720-4	F-, $\Delta(\text{araD-araB})567$, $\Delta\text{lacZ}4787(::\text{rrnB-3})$, λ -, $\Delta\text{crp-765}::\text{kan}$, rph-1, $\Delta(\text{rhaD-rhaB})568$, hsdR514	CGSC Yale
JW1328-1	F-, $\Delta(\text{araD-araB})567$, $\Delta\text{lacZ}4787(::\text{rrnB-3})$, λ -, $\Delta\text{fnr-771}::\text{kan}$, rph-1, $\Delta(\text{rhaD-rhaB})568$, hsdR514	CGSC Yale
JW0662-2	F-, $\Delta(\text{araD-araB})567$, $\Delta\text{lacZ}4787(::\text{rrnB-3})$, $\Delta\text{nagC}725::\text{kan}$, λ -, rph-1, $\Delta(\text{rhaD-rhaB})568$, hsdR514	CGSC Yale
JW1586-1	F-, $\Delta(\text{araD-araB})567$, $\Delta\text{lacZ}4787(::\text{rrnB-3})$, λ -, $\Delta\text{dgsA}742::\text{kan}$, rph-1, $\Delta(\text{rhaD-rhaB})568$, hsdR514	CGSC Yale

Table 2: List of primers and oligonucleotides used in this study.

Primers	Sequence (5' to 3')	Description	Sites in blue
yagU3'-nheI_F	AT GCTAGC GCAGATAGT CAGTGAGTATATC	Amplify EcsR2 from MG1655 wild-type	<i>NheI</i> restriction site
yagU3'-aatII_F	GACGTC GCAGATAGTCA GTGAGTATATC	Amplify EcsR2 from MG1655 wild-type	<i>AatII</i> restriction site
yagU3'-HindIII_R	AAGCTT ATTGCACTGCC CCGCCAGCG	Amplify EcsR2 from MG1655 wild-type	<i>HindIII</i> restriction site
pBAD_F	CTGTTTCTCCATACCCG TT	Verify inserts cloned into pBAD24 plasmid	
pBAD_R1	CTCATCCGCCAAAACAG	Verify inserts cloned into pBAD24 plasmid	
sRNA2-KO-new_F	CAAATACAGCGCCATTT TTATAGGTTAAAAACAT TGCTTT CATATGAATAT CCTCCTTAG	Region flanking EcsR2, to amplify <i>cat</i> gene from pKD3 for deleting EcsR2	Binds to pKD3
sRNA2-KO-new_R	ATACGGGCTGACACCGT TATAAACATACAATAAT TAATT GGTGTAGGCTGG AGCTGCTTC	Region flanking EcsR2, to amplify <i>cat</i> gene from pKD3 for deleting EcsR2	Binds to pKD3
sRNA2-flank_F	GCGGCTGTTTATACCTT TGC	Amplifies region on MG1655 gDNA flanking EcsR2 IGR	
sRNA2-flank_R	GACGATGCTGGCGGTAC TAT	Amplifies region on MG1655 gDNA flanking EcsR2 IGR	
hfq-KO_F	AAGGTTCAAAGTACAAA TAAGCATATAAGGAAAA GAGAG CATATGAATAT CCTCCTTAG	Region flanking <i>hfq</i> gene, to amplify <i>cat</i> gene from pKD3 for deleting hfq	Binds to pKD3
hfq-KO_R	AGGATCGCTGGCTCCCC GTGTAACAAAAACAGCCC GAAACCG TGTAGGCTGG AGCTGCTTC	Region flanking hfq, to amplify <i>cat</i> gene from pKD3 for deleting hfq	Binds to pKD3
hfq-flank_F	TTCGTTGCGTGGGTTAT CGC	Amplifies region on MG1655 gDNA flanking <i>hfq</i> gene	
hfq-flank_R	ACTGCTTTACCTTCACC TAC	Amplifies region on MG1655 gDNA flanking <i>hfq</i> gene	
BH_F	CGGATCTACATCATCCT GAA	Amplifies C2 without 50 to 80 nts of EcsR2	
BH_R	GGCAAATAGAAGAAGTA TCG	Amplifies C2 without 50 to 80 nts of EcsR2	
sRNA2-Biotintarget-oligo	mGmCmGmCmGmAmUmAm UmAmCmUmCmAmCmUmG /iBiodT//iBiodT// iBiodT//3iBiodT/	Biotin T's binds to the neutravidin beads; complementary to 6-22 nucleotides of EcsR2	
EcsR2_qPCR_F	ATCGCGCTACTTCAGGA TGA	Quantify EcsR2 expression levels from cDNA by qPCR	
EcsR2_qPCR_R	CAGCGATAATAGCGGGG CTT	Quantify EcsR2 expression levels from cDNA by qPCR	

16s_F	GCAGCCATGCCGCGTGT ATG	Quantify 16s expression levels from cDNA by qPCR	
16s_R	AATGAGCAAAGGTATTA ACT	Quantify 16s expression levels from cDNA by qPCR	
frdA-Ec- qPCR_F	GGCCCGGAAACTCCGCT GGG	Quantify frdA expression levels from cDNA by qPCR	
frdA-Ec- qPCR_R	GTGACGCAAGTCGAGAT AAA	Quantify frdA expression levels from cDNA by qPCR	
ansB-Ec- qPCR_F	TCCTCTGGGTACATTC ACA	Quantify ansB expression levels from cDNA by qPCR	
ansB-Ec- qPCR_R	GAAGATCGGATGCGTTA GCG	Quantify ansB expression levels from cDNA by qPCR	
ygiB-Ec- qPCR_F	GTTCCGCAAAAACCTGGA GCG	Quantify ygiB expression levels from cDNA by qPCR	
ygiB-Ec- qPCR_R	CTTTTGCCTGGGTTTGC AGC	Quantify ygiB expression levels from cDNA by qPCR	

Growth Assay

For growth analysis, *E. coli* MG1655 and RL_001 were grown overnight in M9 minimal medium (Amresco) supplemented with 0.2% glycerol or in Lysogeny Broth (LB).

Overnight growths were diluted 100 fold in fresh media and growth was measured every hour at OD₆₀₀ using a Bio-Rad SmartSpec 3000.

Crosslink-seq

E. coli strains RL_002 and RL_003 were grown aerobically to OD₆₀₀ value of 0.5.

Cultures were supplemented with arabinose (0.2%) for 10 minutes to transiently induce the expression of EcsR2, and washed twice with phosphate buffered saline (PBS). Cells were resuspended in 8 mL PBS and incubated on ice. 4'-aminomethyl-trioxsalen (AMT) (Cayman Chemicals) was added to the cells at a concentration of 0.2mg/mL and cells were incubated on ice for 10 minutes. 3.2 mL cells treated with AMT were irradiated with long wavelength ultraviolet (UV) light at 365 nm for 1 hour on ice. The cells were then washed once with PBS and total RNA was isolated using TRI reagent (Ambion). RNA was treated with DNase I (Fermentas) to eliminate any DNA contamination. DNase treated RNA was mixed in hybridization buffer (20 mM HEPES pH 8, 5 mM MgCl₂, 300 mM KCl, 0.01% NP-40, 1 mM DTT) and heated at 80°C for 2 minutes followed by immediate cooling on ice. Biotinylated antisense oligo (10 nmol) that binds to EcsR2 (Table 2) was added to the RNA and incubated at room temperature overnight. 150 µL of NeutrAvidin agarose resin (50% slurry) was washed twice in WB100 buffer (20 mM HEPES pH 8, 10 mM MgCl₂, 100 mM KCl, 0.01% NP-40 1 mM DTT)

followed by blocking the beads for 2 hours (blocking buffer: WB100, 50 μ L BSA [10mg/mL], 40 μ L tRNA [10 mg/mL], 10 μ L glycogen [20 mg/mL]). The blocked beads were once again washed with fresh blocking buffer and added to the hybridized RNAs bound to the biotinylated oligo. The sample was incubated for 4 hours at 4°C. Next, the samples were washed five times with WB400 buffer (20 mM HEPES pH 8, 10 mM MgCl₂, 400 mM KCl, 0.01% NP40, 1 mM DTT). The hybridized RNAs bound to the beads were isolated using TRI reagent (Ambion). EcsR2-mRNA hybrid RNAs were un-crosslinked by short wavelength UV light at 254 nm on ice for 15 minutes. The RNA samples were deep-sequenced at Oregon Health and Science University Massively Parallel Sequencing Shared Resource (Illumina HiSeq, 100 cycles, single-end), and the trimmed reads were mapped to *E. coli* MG1655 genome to determine the genes that were enriched in test samples (expressing EcsR2) in comparison to controls (no EcsR2).

Conservation of EcsR2 in *E. coli*

A local BLAST database of genome sequences of all *E. coli* strains from NCBI was created. Local megablast of the intergenic region (IGR) that contains EcsR2 along with 30 bp flanking sequence was performed. The resulting *yagU-ykgJ* IGR sequences from all the genomes were aligned using Clustal Omega (Goujon et al. 2010). For investigating the conservation of *yagU-ykgJ* IGR, *yagU* and *ykgJ* orthologs were identified in closely related enteric species using reciprocal BLAST best-hit approach. The genomic locations of the orthologs were determined on NCBI graphical user interface to confirm presence or absence of the IGR.

EcsR2 expression assay and target verification

For measuring EcsR2 expression, *E. coli* MG1655 was grown aerobically in M9 minimal medium (Amresco) supplemented with 0.2% glycerol to OD₆₀₀ value of 0.2, 0.5, 1.0 and 1.5. Total RNA was extracted from 20 mL of culture using TRI reagent (Ambion). 5 µg of RNA was treated with DNaseI (Fermentas) to eliminate DNA contamination. 500 ng of DNase treated RNA was used to generate cDNA using RevertAid First Strand cDNA Synthesis Kit (ThermoFisher), and semi-quantitative PCR (22 cycles) was done to measure EcsR2 expression. For measuring the effect of *hfq* deletion on the expression of EcsR2, *E. coli* MG1655 and RL_005 strains were grown aerobically in LB to OD₆₀₀ value of 0.5. For testing the impact of glucose on expression of EcsR2, BW25113 strain was grown aerobically in LB and LB+glucose (0.2%) to OD₆₀₀ value of 0.5. For EcsR2 target validation, strains RL_002 (control), RL_003 and RL_004 were grown aerobically to OD₆₀₀ value of 0.5. Cultures were supplemented with arabinose (0.2%) for 10 minutes to transiently induce the expression of EcsR2 and cDNA was generated as described above. EcsR2 and target gene expression levels were quantified by qRT-PCR and were normalized to 16s rRNA gene expression levels.

Evolution and expression of sRNAs

Using blastn (≥30% identity, ≥60% length) we identified the homologs of 81 sRNAs (Raghavan et al. 2011) in 85 *E. coli* genomes. Sequences were aligned using Clustal Omega (Sievers et al. 2011), and single nucleotide polymorphisms (SNPs) were

quantified using the nucleotide polymorphism index π (Nei 1987) with DnaSP 5.10 (Librado and Rozas 2009). To determine whether the sRNAs are present in other enteric bacteria, all sRNAs were searched (blastn, $\geq 30\%$ identity, $\geq 60\%$ length) against the following genomes: *Yersinia enterocolitica* (NC_008800.1), *Klebsiella pneumoniae* (NC_011283.1), *Enterobacter aerogenes* (NC_015663.1), *Citrobacter freundii* (NZ_CP007557.1), *Serratia marcescens* (NZ_HG326223.1), and *Salmonella enterica* Typhimurium SL1344 (NC_016810.1). sRNAs were assigned to one of three groups based on presence in the seven genomes, and Kruskal-Wallis test (non-parametric 1-way ANOVA) and Dunn's test for multiple comparisons were conducted in SAS Studio 3.5 to test for differences in the three groups in sRNA expression and nucleotide polymorphism index π .

RESULTS

EcsR2 is a novel sRNA unique to *E. coli* that originated from a degraded prophage gene.

We performed a directional RNA-seq analysis on *E. coli* to determine whether any of the species-specific IGRs contained highly transcribed regions. After mapping sequencing reads onto the *E. coli* genome, we detected a transcriptional peak, that indicated the presence of an sRNA, in *yagU-ykgJ* IGR in *E. coli* (**Figure 2**) (Raghavan et al. 2015). Transcripts mapping to the corresponding location in the *E. coli* genome have been observed in previous microarray based studies (Tjaden et al. 2002; Chen et al. 2002) further verifying their transcriptional status, and there were no potential open reading frames (ORFs) of substantial length within the transcript, indicating that they represent an sRNA. A modified 3'-RACE experiment was performed to confirm the size of the EcsR2 transcript (genomic location 302905–303070) (Raghavan et al. 2015).

We investigated the uniqueness of *yagU-ykgJ* using BLAST in related bacteria and found that *yagU-ykgJ* IGR is present in all strains of *E. coli* but is absent in any closely related enteric bacteria such as *Salmonella*, *Citrobacter* or *Klebsiella* (**Figure 3A**). The arrangement in which *yagU* neighbors *ykgJ* is found only in *E. coli*. An alternate gene order (*yciC-ykgJ-ompW*) is detected in other enteric bacteria, including in *Escherichia albertii*, *E. coli*'s closest relative, indicating that *ykgJ* moved to its current location in *E. coli* after the two bacteria split from a common ancestor (**Figure 3B**). Additionally, *ykgJ* ORF is smaller in *E. coli* than in *E. albertii*, and a ~90 bp remnant of the gene's 3' end is still recognizable in the *yciC-ompW* IGR in *E. coli*, confirming that *ykgJ* was translocated recently to its current location in *E. coli* to create the unique *yagU-ykgJ* IGR (**Figure 3B**).

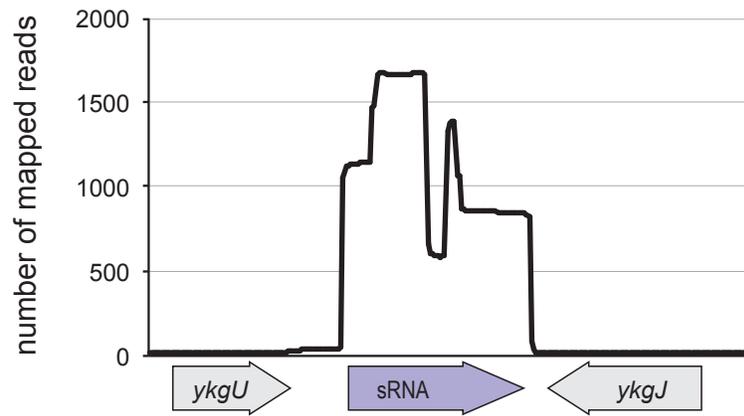


Figure 2. Detection of EcsR2 in *yagU-ykgJ* IGR in *E. coli*. RNA-seq analysis of EcsR2 (sRNA) transcript levels in *yagU-ykgJ* IGR (Raghavan et al. 2015).

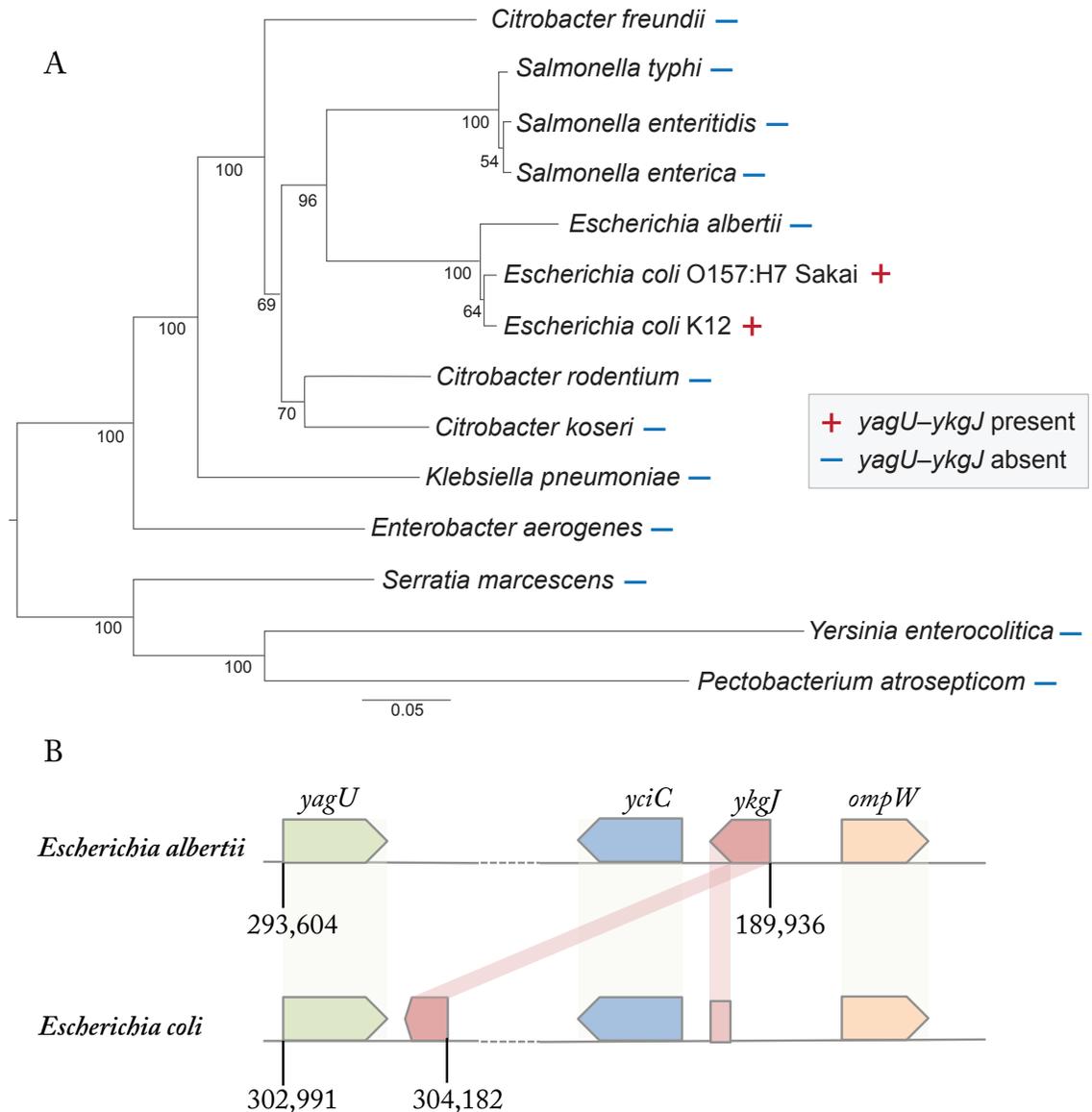


Figure 3. Conservation and formation of *yagU-ykgJ* IGR in *E. coli*. (A) Phylogenetic tree showing presence or absence of *yagU-ykgJ* IGR among enteric bacteria. (B) Arrangement of *yciC-ykgJ-ompW* in *E. albertii*. Relocation of *ykgJ* leading to formation of novel *yagU-ykgJ* IGR in *E. coli*. Genomic locations of *yagU* and *ykgJ* in both bacteria are also shown.

It is usually difficult to trace the ancestry of sRNAs due to their rapid rate of evolution (Gottesman and Storz 2011). However, because EcsR2 emerged in an IGR that was formed recently, we were able to identify through sequence alignment that the sRNA evolved from the 3' end of a vestigial bacteriophage gene. We determined that a major portion of the 5' end of EcsR2 (107/166 nts) shared very high similarity (~76%) with a prophage tail fiber assembly gene *tfaR*, while the 3' end of the sRNA, where the intrinsic terminator is located is only ~50% similar to *tfaR* (Figure 4A). Interestingly, two other pseudogenized copies (*tfaD* and *tfaX*) of the same phage tail fiber gene that consists of varying lengths of its 3' end are also present in *E. coli*, indicating that EcsR2 originated from a similarly truncated non-functional version of the phage gene through the accumulation of point mutations (Figure 4B). We also located a putative sigma-70 promoter associated with EcsR2. As shown in Figure 13, nucleotides centered around -10 and -35 positions are very close to the sigma 70 consensus sequence TATAAT and TTGACA respectively (Malhotra et al. 1996). From this data, it is clear that EcsR2 originated by combining the remnants of a degraded phage gene with a sigma-70 promoter. To investigate the conservation of EcsR2, its promoter and terminator, we scanned all the available genomes of *E. coli* in the RefSeq database. This analyses showed that EcsR2, its promoter and terminator are conserved across all strains except in Enterotoxigenic *E. coli* (ETEC) H10407. In this strain, a putative transposase has disrupted the sRNA.

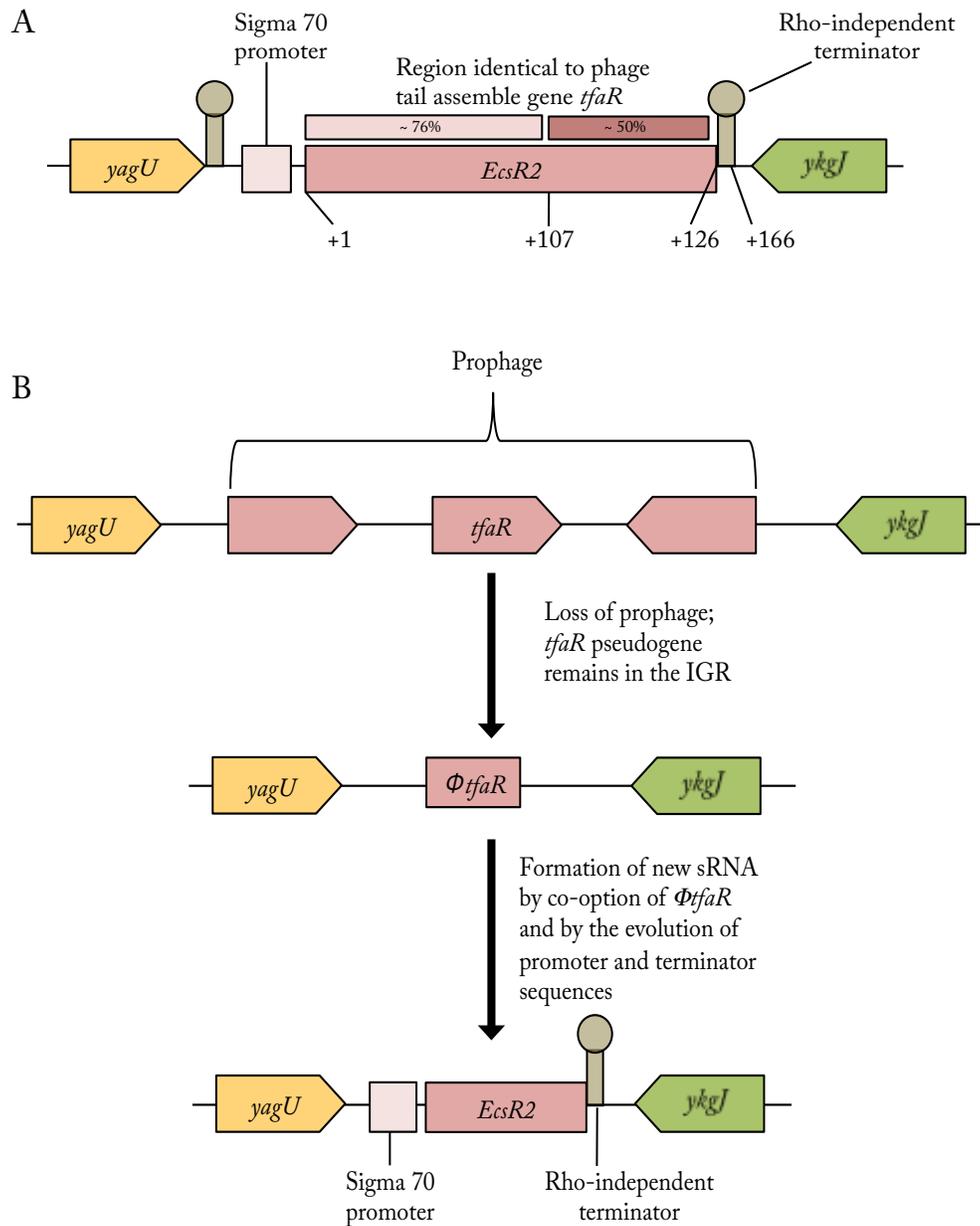


Figure 4. Model for origination of EcsR2 from a protein-coding gene. (A) IGR with encoded EcsR2, sigma-70 promoter and terminator. Region of sRNA similar to the phage tail assemble gene highlighted in red. **(B)** Representation of origination of EcsR2 from a vestigial protein-coding gene.

EcsR2 is a ‘young’ and fast-evolving sRNA that is slightly deleterious to *E. coli*.

EcsR2 is expressed highest at exponential phase (OD₆₀₀ values of 0.2 and 0.5) and its expression decreases as the cells reach stationary phase (OD₆₀₀ values of 1.0 and 1.5) (Figure 5). To test whether EcsR2 has any physiological effect on *E. coli*, we measured the growth of wild type and EcsR2 deletion strains in LB and in M9 minimal media supplemented with 0.2% glycerol as a carbon source. Interestingly, while both wild-type and EcsR2-deletion strains grew at comparative rates in LB, the wild-type strain displays a delay in growth in minimal medium (Figure 6). This data signifies that EcsR2, a newly evolved sRNA, is slightly deleterious to *E. coli*. Additionally, the low expression of EcsR2 might compensate for its slight toxicity, as observed for newly originated miRNAs in eukaryotes (Jovelin 2014).

To test whether low expression is a feature that is common to all newly evolved sRNAs, we compared the expression of 82 sRNAs in *E. coli* (Raghavan et al. 2011), to their phylogenetic distribution. As shown in Figure 7A, phylogenetically restricted i.e. ‘young’ sRNAs have significantly lower expression than sRNAs that are conserved in several species of enteric bacteria i.e. ‘old’ sRNAs. In addition, we also determined that ‘young’ sRNAs are evolving at a significantly faster rate than ‘old’ sRNAs (Figure 7B). Collectively, our data indicate that low expression and rapid evolution could delay the elimination of slightly deleterious ‘young’ sRNAs, thereby enabling their integration into regulatory networks.

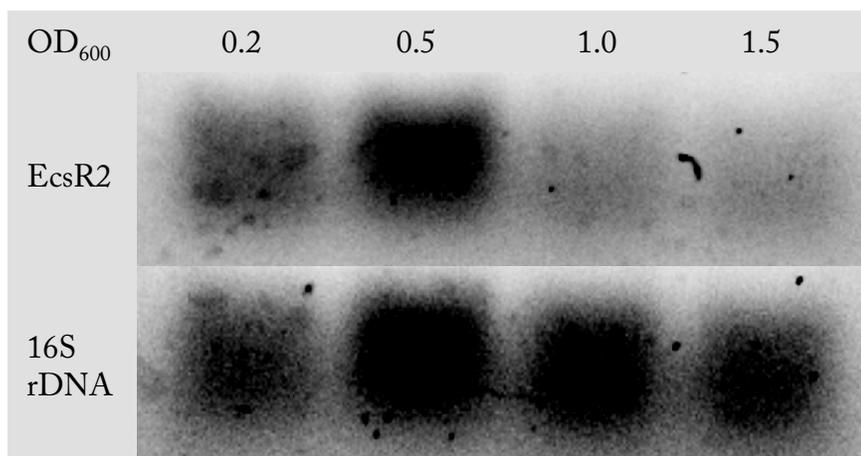


Figure 5. Expression of EcsR2 at different growth phases. Expression of EcsR2 during *E. coli* growth in M9 minimal medium supplemented with 0.2% glycerol. cDNA was amplified by PCR (20 cycles) and ran on 3% agarose gel.

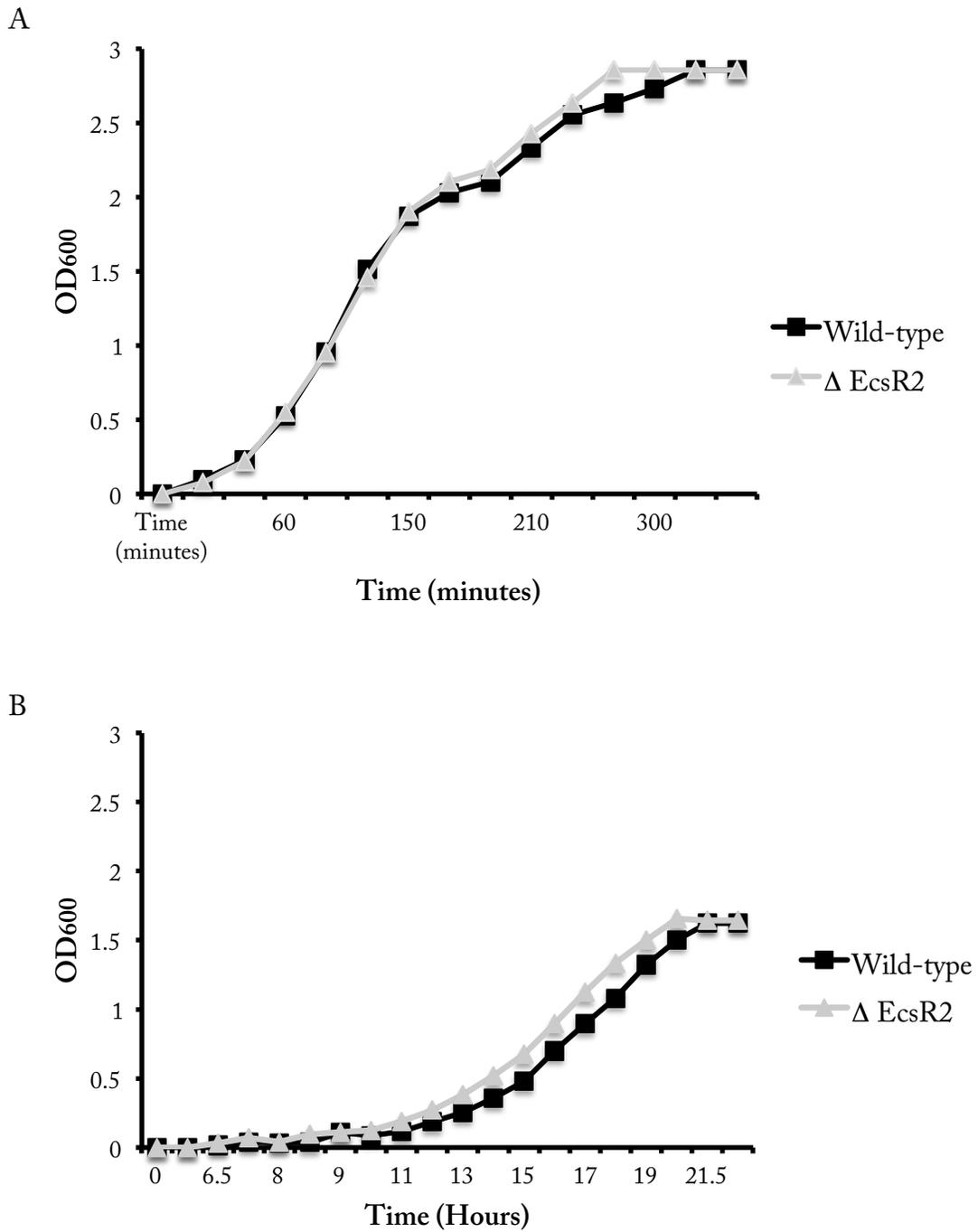


Figure 6. Mildly deleterious effect of EcsR2 in minimal medium. (A) *E. coli* wild-type and *EcsR2*-deletion strains grown in LB. **(B)** *E. coli* wild-type and *EcsR2*-deletion strains grown in M9 minimal medium.

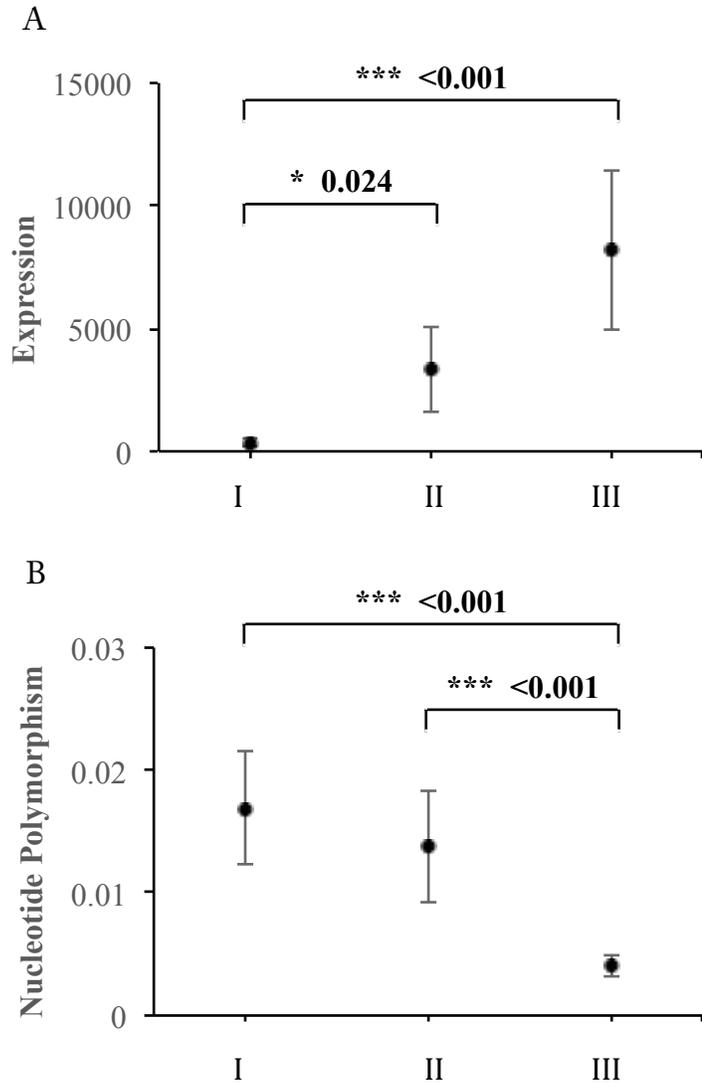


Figure 7. Young sRNAs have low expression and high SNPs. Phylogenetically restricted sRNAs have low expression (A) and rapid evolution (B). X-axes: I = sRNAs conserved in 1-3 genomes, II = sRNAs conserved in 4-5 genomes, III = sRNAs conserved in 6-7 genomes.

EcsR2 is a functional base-pairing sRNA that is integrated into the carbon metabolism regulatory network.

In enteric bacteria, majority of the functional sRNAs requires an RNA chaperone protein Hfq for its stability and optimal function (Delay et al. 2013). Hfq is an RNA-binding protein with structural similarities to eukaryotic Sm proteins (Zhang et al. 2002; Møller et al. 2002). To investigate whether Hfq plays a role in the functioning of EcsR2, we measured its expression in *E. coli* MG1655 and *hfq*-deletion strains at exponential phase and found that EcsR2 expression decreases around seven fold in *hfq*-deletion strain as compared to in wild-type strain (**Figure 8**). This suggests that EcsR2 has gained the ability to bind to Hfq, an important milestone in sRNA evolution (Peer and Margalit 2014).

Identification of mRNA targets of sRNAs is one of the biggest challenges in the field. Several target prediction software are available (Kery et al. 2014; Wright et al. 2013), but most of them are not very reliable due to the inherent challenges of predicting RNA-RNA interactions. A GFP-based reporter system is also available to experimentally determine the targets of an sRNA; however, this method is very labor intensive and hence is not suitable for high-throughput analysis (Urban and Vogel 2007). To overcome these difficulties, we developed a novel genome-scale technique, that we call Crosslink-seq, to identify mRNAs targeted by an sRNA (**Figure 9**; see Materials and Methods). To distinguish mRNAs that bind to EcsR2, we identified genes that were significantly enriched in the test samples (*E. coli* expressing EcsR2) when compared to the control

samples (*E. coli* without EcsR2). We identified three genes, *ansB*, *frdA* and *ygiB* that were enriched greater than three times in the test samples. These potential targets were

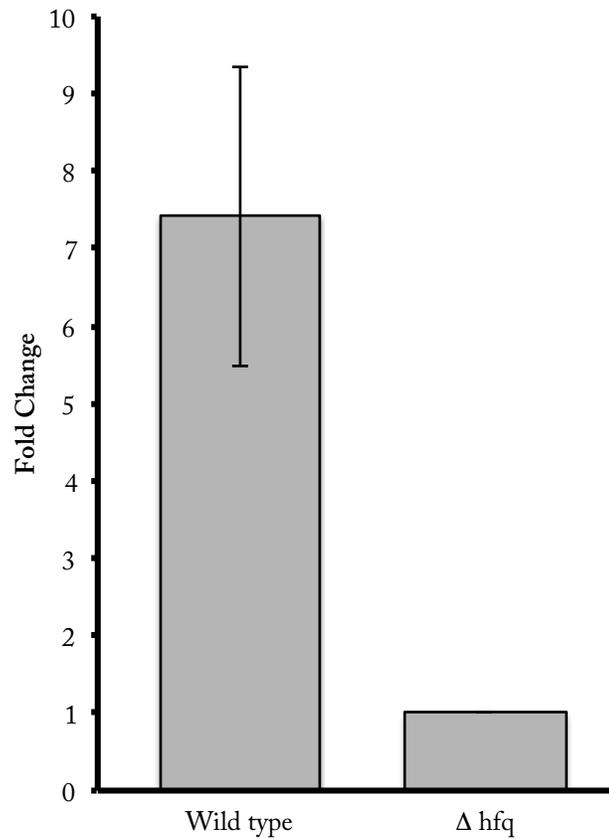


Figure 8. Hfq stabilizes EcsR2. Expression of EcsR2 in *E. coli* MG1655 (Wild type) and *hfq*-deletion strains measured in fold change. Wild type and *hfq*-deletion strains were grown to OD₆₀₀ value of 0.5 in LB. cDNA was synthesized from 500 ng of total RNA and EcsR2 expression was measured by qPCR.

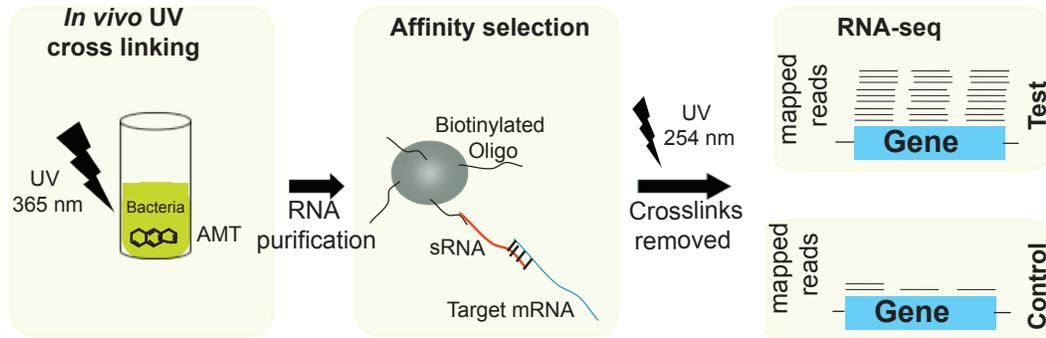


Figure 9. Crosslink-seq. AMT crosslinks EcsR2 to mRNAs in the presence of UV (365 nm), which were purified using affinity selection. mRNAs were uncrosslinked using UV (254 nm) and quantified using RNA-seq to identify those enriched in strains containing EcsR2 (test) in comparison to strains without EcsR2 (control).

validated by qPCR. We predicted the potential sRNA-mRNA binding regions using IntaRNA (Wright et al. 2014), and to confirm that the predicted interaction sites are indeed functional, we transiently expressed a mutant version of the sRNA that did not contain a region (nucleotide positions 50- 80) that showed complementarity with the targets (**Figure 10**). Expression of all three targets increased around two fold in the presence of the mutant Ecsr2 in comparison to the full-length EcsR2 (**Figure 11**).

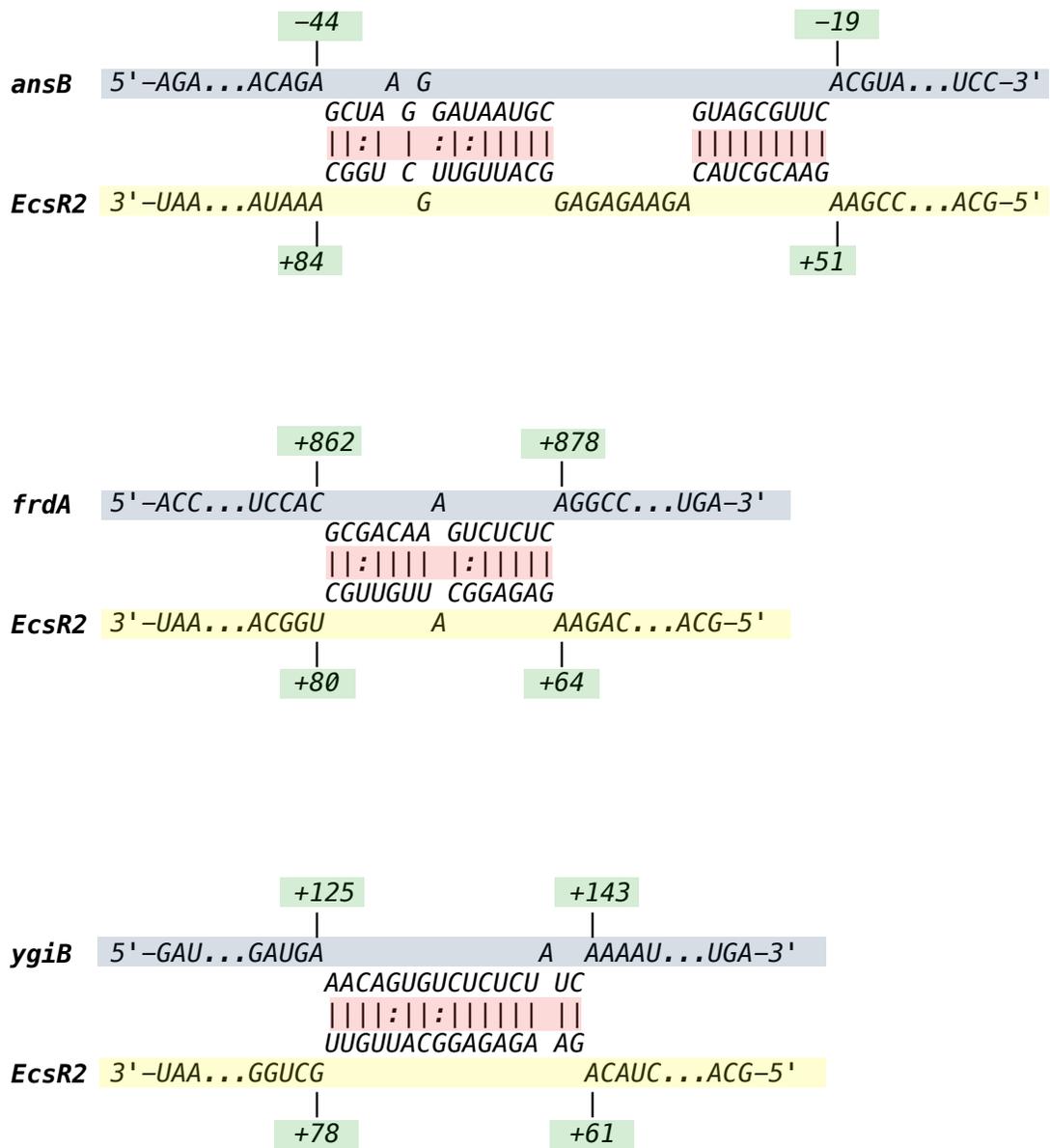


Figure 10. Predicted EcsR2-target mRNA interactions. Green denotes the nucleotide positions relative to the start sites of genes, blue represents the target genes, yellow denotes EcsR2 and red shows the potential base pairing between EcsR2 and target mRNA.

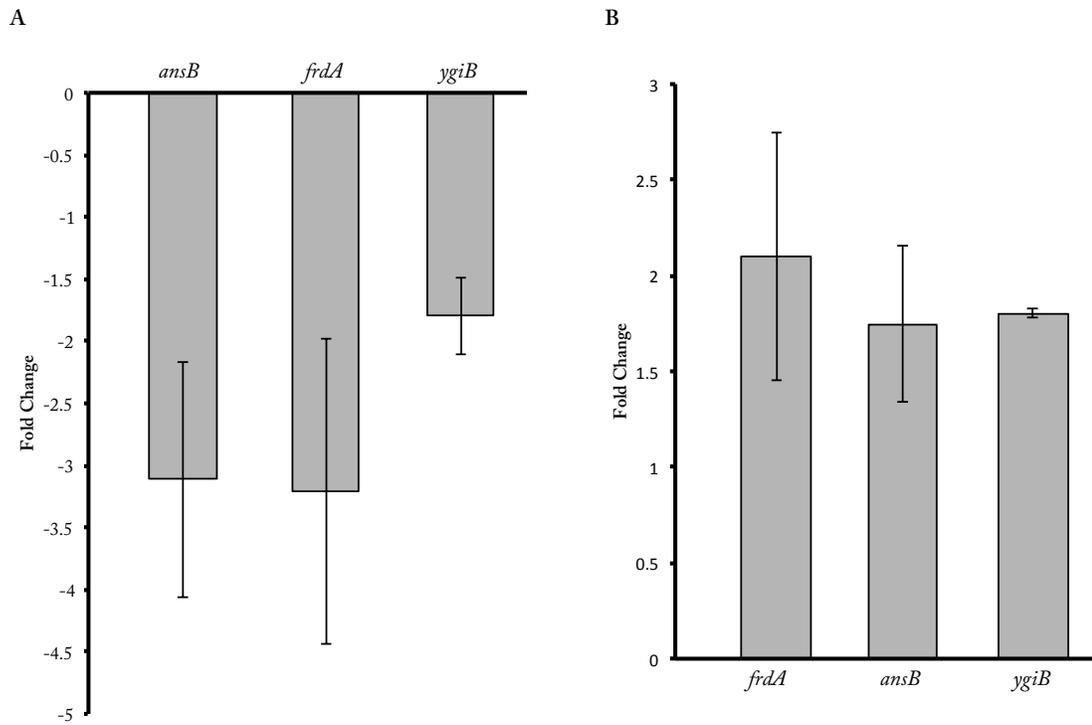


Figure 11. Down regulation of target genes by EcsR2. Expression of *ansB*, *frdA* and *ygiB* is repressed by EcsR2. **(A)** Fold change in gene expression in *E. coli* expressing EcsR2 compared to *E. coli* without EcsR2. **(B)** Fold change in gene expression in *E. coli* expressing EcsR2 without the seed region (nucleotide positions 50-80) in comparison to *E. coli* expressing full-length EcsR2.

CRP and FNR are transcriptional regulators that affect the expression of *ansB* and *frdA* depending on glucose levels in the medium (Tseng et al. 1994). We tested the effect of glucose on EcsR2 and found that its expression was significantly induced (~27 fold) when LB was supplemented with 0.2% glucose (**Figure 12**). We also found a putative binding site in the promoter region of EcsR2 for Mlc (**Figure 13**), which is a glucose dependent transcription factor that is controlled by CRP (Plumbridge 2001; Perrenoud and Sauer 2005). As shown in **Figure 13**, the putative Mlc binding region (shown in red) is very similar to the consensus Mlc binding sequence aTtaTTTcgctgcgcgAAAttaa (Plumbridge 2001). The regulation of *ansB* and *frdA* by CRP and FNR, the induction of EcsR2 expression by glucose, and the presence of Mlc binding site in its promoter region, collectively indicate that EcsR2 has been integrated in to the carbon regulatory network in *E. coli*.

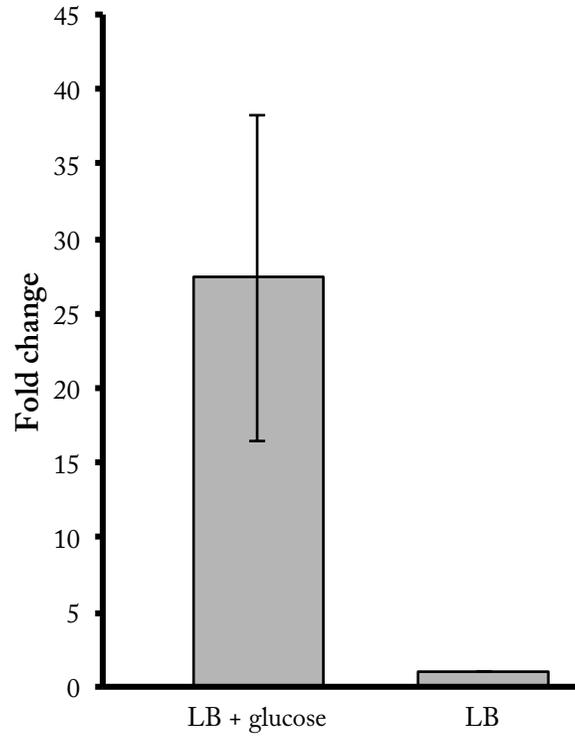


Figure 12. Glucose induces EcsR2 expression. Induction of EcsR2 sRNA in wild type *E. coli* in Lysogeny broth supplemented with 0.2% glucose.



Figure 13. Mlc binding site. Putative Mlc binding site is shown in red. Sigma-70 promoter region is shown in green. The numbers are labeled corresponding to the transcription start site of EcsR2. EcsR2 transcription start nucleotide is shown in bold.

DISCUSSION

One of the fundamental gaps in knowledge in bacteriology is the understanding of how new genes originate. Gene duplication and horizontal transfer are two known paths that amplify and improve upon already formed genes (Wilderman et al. 2004; Pichon and Felden 2005); but the source and mechanism of origination of novel genes in bacteria is still elusive. In this study, we identified a degraded protein-coding gene as a source for the origination of a new non-coding RNA in *E. coli*. We also show that ‘young’ sRNAs such as EcsR2 are expressed at low levels and evolve at a rapid rate, thereby explaining how newly emerged sRNAs could gain enough time to evolve functions that enable their retention in bacterial genomes.

Many sRNAs function in conjunction with transcription factors. Their activity can enhance or repress expression of various genes at transcriptional and posttranscriptional levels (Sharma et al. 2007). Thus sRNAs have been established as key players in large regulatory networks. Environment signals often trigger responses that require transcription factors. Fine-tuning these responses with variable speed and strength is possible through sRNAs. For example, sRNAs can modulate one gene by direct base-pairing or multiple genes by controlling a transcription factor. To function with such versatility, there are certain features that an sRNA must possess. Presence of a Rho-independent terminator is one of the signature features of base-pairing sRNAs (Chen et al. 2002). EcsR2 has a Rho-independent terminator (**Figure 14**), which in addition to terminating the transcript, could bind to Hfq (Chen et al. 2002) and provide resistance to degradation by ribonucleases (Ishikawa et al. 2012).

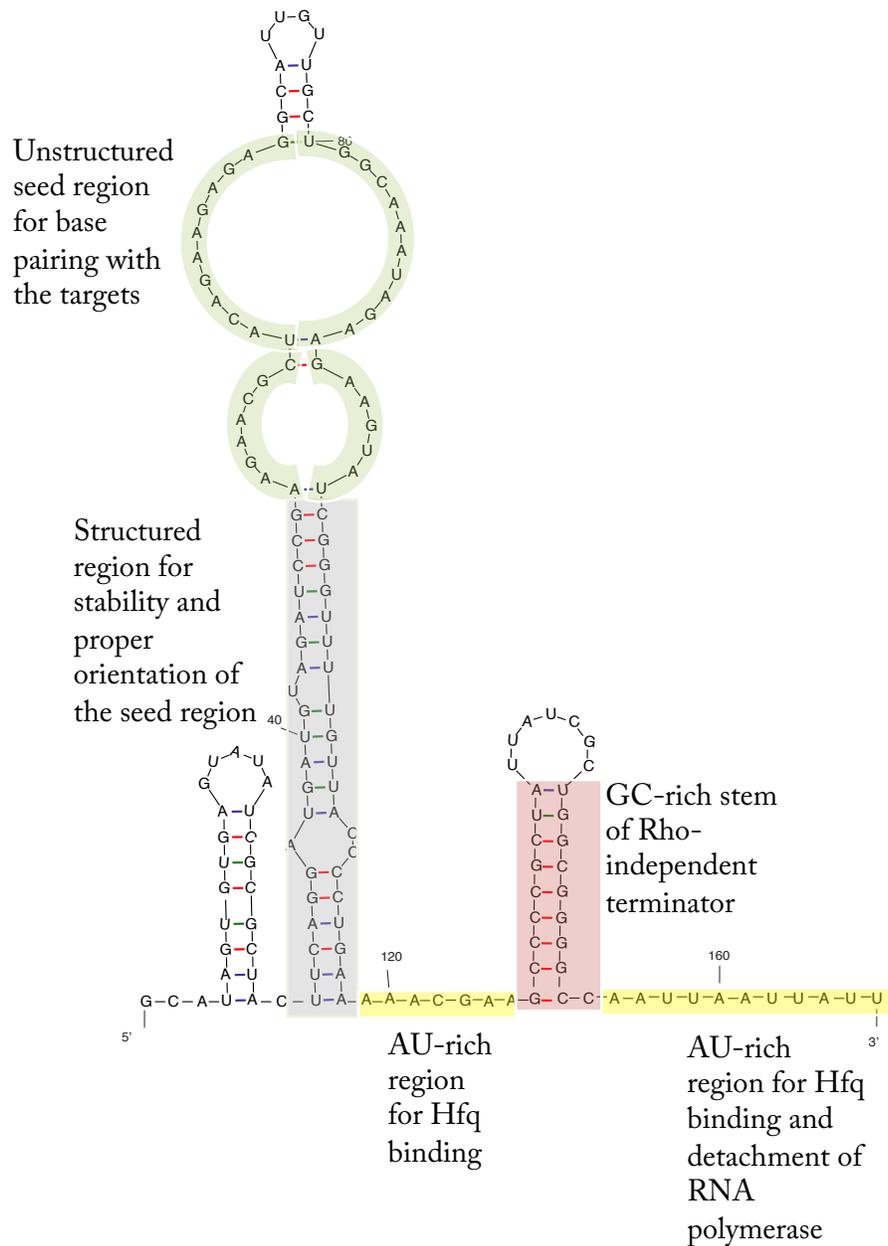


Figure 14. Predicted secondary structure of EcsR2. Green shows the seed region, gray represents double stranded region, yellow denotes the AU-rich region that binds to Hfq and red denotes the GC-rich stem of the Rho-independent terminator region. The secondary structure was predicted by using mFold software.

Another characteristic of a functional sRNA is the region where binding with target mRNA occurs. This sequence called the 'seed region' is usually single stranded (Fröhlich et al. 2013; Shao et al. 2013; Srikhanta et al. 2013). In **Figure 14**, the 50-80 nts of EcsR2 that binds to AnsB and FrdA is predicted to be a highly unstructured region. This mostly single stranded sequence is an ideal seed region. On mRNAs, sRNAs are known to bind either to 5' UTR, coding region or 3' UTR. In two of the three targets, EcsR2 binds to the coding region (*frdA* and *ygiB*) and for the third target (*ansB*), the sRNA seed region binds in the 5' UTR (**Figure 10**). Additionally, sRNAs also contain double-stranded regions that provide structural stability and proper orientation of the seed region. This proper orientation allows the Hfq binding region and the seed region to be easily accessible for binding to the target mRNAs or proteins (Updegrove, Shabalina and Storz 2015). **Figure 13** shows that EcsR2 has three stem-loops with the tallest stem-loop consisting of the seed region followed by the double stranded region which might allow the seed region to interact with mRNAs (Chen et al. 2002). Moreover, the mRNA binding region of EcsR2 has become more unstructured i.e. more accessible, and the region that contains the Rho-independent terminator has become more structured, in comparison to the original *tfaR* mRNA. This pattern of sRNA structural change is similar to what has been observed in eukaryotes, where, the evolution of a spurious transcript into a functional ncRNA has been associated with changes in the RNA structure (Heinen et al. 2009). Collectively these characteristics confirm that EcsR2 is a functional sRNA.

Two of the three genes that are down regulated by EcsR2 belong to the carbon metabolism and anaerobic pathways. The *ansB* gene codes for L-asparaginase II that permits *E. coli* to grow on glycerol and asparagine under anaerobic condition to synthesize fumarate. Fumarate acts as an electron acceptor during growth in absence of oxygen (Russell and Yamazaki 1978). Expression of *ansB* is under catabolite repression and is induced under anaerobic conditions (Cedar and Schwartz 1968; Jennings and Beacham 1990; Tseng et al. 1994). It was also reported that L-asparaginase II synthesis was inhibited by glucose, and since EcsR2 is significantly induced by glucose, we suspect that repression of L-asparaginase II by glucose is mediated through EcsR2. Another EcsR2 target *frdA* is also expressed optimally under anaerobic conditions. It was reported that under aerobic conditions, *frdA* was expressed lowest during rapid growth and the expression increased as the growth rate decreased (Tseng et al. 1994). The reason for this differential expression was that the product of *frdABCD* operon is functional only under anaerobic conditions. Interestingly, the expression pattern of EcsR2 is completely opposite to that of *frdA* in aerobic conditions, suggests that *E. coli* might have evolved a mechanism to shut down *ansB* and *frdA* transcription under aerobic conditions using EcsR2. Furthermore, Hfq stabilizes EcsR2 and a glucose-sensitive transcriptional factor (Mlc) putatively regulates this sRNA, indicating that EcsR2 has been integrated into the carbon metabolism regulatory network. In total, our data show that a vestigial bacteriophage protein-coding gene has evolved to into a functional sRNA in *E. coli*.

Because EcsR2 evolved from an erstwhile prophage gene, it is not clear whether the whole evolutionary process occurred in its current genomic location or whether it

occurred while the nascent sRNA was still part of a bacteriophage genome. Nevertheless, phages seem to have had a large impact on shaping sRNA repertoires in bacteria, with several sRNAs located within prophages, pathogenicity islands, or next to phage insertion sites (Pichon and Felden 2005; Sittka et al. 2008; De Lay and Gottesman 2009). This observation is not surprising because high rate of recombination, rapid evolution, and profuse genetic exchange make bacteriophages powerful engines of genetic innovation (Daubin and Ochman 2004; Raghavan and Minnick 2009).

EcsR2 and other young sRNAs are expressed at low levels in comparison to evolutionarily older sRNAs probably because their promoters are not yet fully functional. In bacteria, promoter-like sequences arise spontaneously through point mutations, especially in IGRs (Stone and Wray 2001; Mendoza-Vargas et al. 2009), and inefficient transcription from these promoters are one of the causes for pervasive transcription i.e., RNAs originating from all across the genome (Dornenburg et al. 2010; Raghavan et al. 2012; Thomason et al. 2015). The functions, if any, of these genome-wide transcripts are not yet understood, however, they could serve as the raw material for the emergence of new functional RNAs (Gottesman and Storz 2011; Lybecker et al. 2014; Wade and Grainger 2014). Pervasive transcription has been observed in all domains of life, and recently it was shown that new functional RNAs could evolve from such transcripts in humans (Ruiz-Orera et al. 2015). In the current study, our data also points towards such a scenario where the emergence of a promoter-like sequence resulted in the transcription of a stable RNA that evolved into EcsR2 by gaining Hfq-binding sites and an intrinsic terminator. Furthermore, low expression and rapid evolution have also been observed for

young eukaryotic miRNAs (Chen and Rajewsky 2007; Jovelin and Cutter 2014; Lyu et al. 2014), indicating that this is a universal phenomenon that facilitates the emergence of new functional RNAs in all domains of life.

REFERENCES

- Babitzke P, Romeo T. 2007. CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Curr Opin Microbiol.* 10:156-63.
- Babski J, Maier LK, Heyer R, Jaschinski K, Prasse D, Jäger D, Randau L, Schmitz RA, Marchfelder A, Soppa J. 2014. Small regulatory RNAs in Archaea. *RNA Biol.* 11:484-93.
- Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell.* 116:281-97.
- Beisel CL, Storz G. 2010. Base pairing small RNAs and their roles in global regulatory networks. *FEMS Microbiol Rev.* 34:866-82.
- Breaker RR. 2011. Prospects for riboswitch discovery and analysis. *Mol Cell.* 43:867-79.
- Cedar H, Schwartz JH. 1968. Production of L-asparaginase II by *Escherichia coli*. *J Bacteriol.* 96:2043-8.
- Chen Q, Crosa JH. 1996. Antisense RNA, fur, iron, and the regulation of iron transport genes in *Vibrio anguillarum*. *J Biol Chem.* 271:18885-91.
- Chen S, Lesnik EA, Hall TA, Sampath R, Griffey RH, Ecker DJ, Blyn LB. 2002. A bioinformatics based approach to discover small RNA genes in the *Escherichia coli* genome. *Biosystems.* 65:157-77.
- Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA.* 97:6640-5.

- Daubin V, Ochman H. 2004. Bacterial genomes as new gene homes: the genealogy of ORFans in *E. coli*. *Genome Res.* 14:1036-42.
- De Lay N, Schu DJ, Gottesman S. 2013. Bacterial small RNA-based negative regulation: Hfq and its accomplices. *J Biol Chem.* 288:7996-8003.
- Dornenburg JE, DeVita AM, Palumbo MJ, Wade JT. 2010. Widespread antisense transcription in *Escherichia coli*. *MBio.* 1:e00024-10.
- Filipowicz W, Jaskiewicz L, Kolb FA, Pillai RS. 2005. Post-transcriptional gene silencing by siRNAs and miRNAs. *Curr Opin Struct Biol.* 15:331-41.
- Fröhlich KS, Papenfort K, Fekete A, Vogel J. 2013. A small RNA activates CFA synthase by isoform-specific mRNA stabilization. *EMBO J.* 32:2963-79.
- Georg J, Hess WR. 2011. cis-antisense RNA, another level of gene regulation in bacteria. *Microbiol Mol Biol Rev.* 75:286-300.
- Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R. 2010. A new bioinformatics analysis tools framework at EMBL–EBI. *Nucleic Acids Res.* 38:W695-9.
- Heinen TJ, Staubach F, Häming D, Tautz D. 2009. Emergence of a new gene from an intergenic region. *Curr Biol.* 19:1527-31.
- Hoe CH, Raabe CA, Rozhdestvensky TS, Tang TH. 2013. Bacterial sRNAs: regulation in stress. *Int J Med Microbiol.* 303:217-29.
- Ishikawa H, Otaka H, Maki K, Morita T, Aiba H. 2012. The functional Hfq-binding module of bacterial sRNAs consists of a double or single hairpin preceded by a U-rich sequence and followed by a 3' poly (U) tail. *RNA.* 18:1062-74.

- Jennings MP, Beacham IR. 1990. Analysis of the *Escherichia coli* gene encoding L-asparaginase II, ansB, and its regulation by cyclic AMP receptor and FNR proteins. *J Bacteriol.* 172:1491-8.
- Jones AM, Goodwill A, Elliott T. 2006. Limited role for the DsrA and RprA regulatory RNAs in rpoS regulation in *Salmonella enterica*. *J Bacteriol.* 188:5077-88.
- Jorgensen MG, Thomason MK, Havelund J, Valentin-Hansen P, Storz G. 2013. Dual function of the McaS small RNA in controlling biofilm formation. *Genes Dev.* 27:1132-45.
- Jovelin R, Cutter AD. 2014. Microevolution of nematode miRNAs reveals diverse modes of selection. *Genome Biol Evol.* 6:3049-63.
- Kawano M, Aravind L, Storz G. 2007. An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin. *Mol Microbiol.* 64:738-54.
- Kazantsev AV, Pace NR. 2006. Bacterial RNase P: a new view of an ancient enzyme. *Nat Rev Microbiol.* 4:729-40.
- Kery MB, Feldman M, Livny J, Tjaden B. 2014. TargetRNA2: identifying targets of small regulatory RNAs in bacteria. *Nucleic Acids Res.* 42:W124-9.
- Kortmann J, Narberhaus F. 2012. Bacterial RNA thermometers: molecular zippers and switches. *Nat Rev Microbiol.* 10:255-65.
- Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics.* 25:1451-2.

- Lustig Y, Wachtel C, Safro M, Liu L, Michaeli S. 2010. 'RNA walk' a novel approach to study RNA–RNA interactions between a small RNA and its target. *Nucleic Acids Res.* 38:e5.
- Lybecker M, Bilusic I, Raghavan R. 2014. Pervasive transcription: detecting functional RNAs in bacteria. *Transcription.* 5:e944039.
- Malhotra A, Severinova E, Darst SA. 1996. Crystal structure of a σ 70 subunit fragment from *E. coli* RNA polymerase. *Cell.* 87:127-36.
- McCullen CA, Benhammou JN, Majdalani N, Gottesman S. 2010. Mechanism of positive regulation by DsrA and RprA small noncoding RNAs: pairing increases translation and protects rpoS mRNA from degradation. *J Bacteriol.* 192:5559-71.
- Mendoza-Vargas A, Olvera L, Olvera M, Grande R, Vega-Alvarado L, Taboada B, Jimenez-Jacinto V, Salgado H, Juárez K, Contreras-Moreira B, Huerta AM. 2009. Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli*. *PLoS One.* 4:e7526.
- Miyakoshi M, Chao Y, Vogel J. 2015. Cross talk between ABC transporter mRNAs via a target mRNA-derived sponge of the GcvB small RNA. *EMBO J.* e201490546.
- Møller T, Franch T, Højrup P, Keene DR, Bächinger HP, Brennan RG, Valentin-Hansen P. 2002. Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. *Mol Cell.* 9:23-30.

- Moon K, Gottesman S. 2009. A PhoQ/P-regulated small RNA regulates sensitivity of *Escherichia coli* to antimicrobial peptides. *Mol Microbiol.* 74:1314-30.
- Moore SD, Sauer RT. 2007. The tmRNA system for translational surveillance and ribosome rescue. *Annu. Rev. Biochem.* 76:101-24.
- Murchison EP, Hannon GJ. 2004. miRNAs on the move: miRNA biogenesis and the RNAi machinery. *Curr Opin Cell Biol.* 16:223-9.
- Negrete A, Ng WI, Shiloach J. 2010. Glucose uptake regulation in *E. coli* by the small RNA SgrS: comparative analysis of *E. coli* K-12 (JM109 and MG1655) and *E. coli* B (BL21). *Microb Cell Fact.* 9:1.
- Nei M. 1987. *Molecular evolutionary genetics.* New York (NY): Columbia
- Opdyke JA, Kang JG, Storz G. 2004. GadY, a small-RNA regulator of acid response genes in *Escherichia coli*. *J Bacteriol.* 186:6698-705.
- Peer A, Margalit H. 2014. Evolutionary patterns of *Escherichia coli* small RNAs and their regulatory interactions. *RNA.* 20:994-1003.
- Perrenoud A, Sauer U. 2005. Impact of global transcriptional regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr, and Mlc on glucose catabolism in *Escherichia coli*. *J Bacteriol.* 187:3171-9.
- Pichon C, Felden B. 2005. Small RNA genes expressed from *Staphylococcus aureus* genomic and pathogenicity islands with specific expression among pathogenic strains. *Proc Natl Acad Sci USA.* 102:14249-54.

- Plumbridge, J. 2001. DNA binding sites for the Mlc and NagC proteins: regulation of *nagE*, encoding the *N*-acetylglucosamine-specific transporter in *Escherichia coli*. *Nucleic Acids Res.* 29:506–514.
- Raghavan R, Groisman EA, Ochman H. 2011. Genome-wide detection of novel regulatory RNAs in *E. coli*. *Genome Res.* 21:1487-97.
- Raghavan R, Kacharia FR, Millar JA, Sislak CD, Ochman H. 2015. Genome rearrangements can make and break small RNA genes. *Genome Biol Evol.* 7:557-66.
- Raghavan R, Minnick MF. 2009. Group I introns and inteins: disparate origins but convergent parasitic strategies. *J Bacteriol.* 191:6193-202.
- Raghavan R, Sloan DB, Ochman H. 2012. Antisense transcription is pervasive but rarely conserved in enteric bacteria. *MBio.* 3:e00156-12.
- Ruiz-Orera J, Hernandez-Rodriguez J, Chiva C, Sabidó E, Kondova I, Bontrop R, Marqués-Bonet T, Albà MM. 2015. Origins of de novo genes in human and chimpanzee. *PLoS Genet.* 11:e1005721.
- Russell LV, Yamazaki H. 1978. The dependence of *Escherichia coli* asparaginase II formation on cyclic AMP and cyclic AMP receptor protein. *Can J Microbiol.* 24:629-31.
- Shao Y, Feng L, Rutherford ST, Papenfort K, Bassler BL. 2013. Functional determinants of the quorum-sensing non-coding RNAs and their roles in target regulation. *EMBO J.* 32:2158-71.

- Sharma CM, Darfeuille F, Plantinga TH, Vogel J. 2007. A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes Dev.* 21:2804-17.
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* 7(1).
- Sittka A, Lucchini S, Papenfort K, Sharma CM, Rolle K, Binnewies TT, Hinton JC, Vogel J. 2008. Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet.* 4:e1000163.
- Sledjeski DD, Whitman C, Zhang A. 2001. Hfq is necessary for regulation by the untranslated RNA DsrA. *J Bacteriol.* 183:1997-2005.
- Srikhanta YN, Atack JM, Beacham IR, Jennings MP. 2013. Distinct physiological roles for the two L-asparaginase isozymes of *Escherichia coli*. *Biochem Biophys Res Commun.* 436:362-5.
- Stone JR, Wray GA. 2001. Rapid evolution of cis-regulatory sequences via local point mutations. *Mol Biol Evol.* 18:1764-70.
- Stork M, Di Lorenzo M, Welch TJ, Crosa JH. 2007. Transcription termination within the iron transport-biosynthesis operon of *Vibrio anguillarum* requires an antisense RNA. *J Bacteriol.* 189:3479-88.

- Thomason MK, Bischler T, Eisenbart SK, Förstner KU, Zhang A, Herbig A, Nieselt K, Sharma CM, Storz G. 2015. Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in *Escherichia coli*. *J Bacteriol.* 197:18-28.
- Thomason MK, Fontaine F, De Lay N, Storz G. 2012. A small RNA that regulates motility and biofilm formation in response to changes in nutrient availability in *Escherichia coli*. *Mol Microbiol.* 84:17-35.
- Tjaden B, Saxena RM, Stolyar S, Haynor DR, Kolker E, Rosenow C. 2002. Transcriptome analysis of *Escherichia coli* using high-density oligonucleotide probe arrays. *Nucleic Acids Res.* 30:3732-8.
- Tseng CP, Hansen AK, Cotter P, Gunsalus RP. 1994. Effect of cell growth rate on expression of the anaerobic respiratory pathway operons *frdABCD*, *dmsABC*, and *narGHJI* of *Escherichia coli*. *J Bacteriol.* 176:6599-605.
- Updegrave TB, Shabalina SA, Storz G. 2015. How do base-pairing small RNAs evolve? *FEMS Microbiol Rev.* 39:379-91.
- Urban JH, Vogel J. 2007. Translational control and target recognition by *Escherichia coli* small RNAs in vivo. *Nucleic Acids Res.* 35:1018-37.
- Večerek B, Moll I, Bläsi U. 2007. Control of Fur synthesis by the non-coding RNA RyhB and iron-responsive decoding. *EMBO J.* 26:965-75.
- Vogel J, Sharma CM. 2005. How to find small non-coding RNAs in bacteria. *Biol Chem.* 386:1219-38.

- Wade JT, Grainger DC. 2014. Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. *Nature Rev Microbiol.* 12:647-53.
- Wassarman KM, Storz G. 2000. 6S RNA regulates *E. coli* RNA polymerase activity. *Cell.* 101:613-23.
- Wilderman PJ, Sowa NA, FitzGerald DJ, FitzGerald PC, Gottesman S, Ochsner UA, Vasil ML. 2004. Identification of tandem duplicate regulatory small RNAs in *Pseudomonas aeruginosa* involved in iron homeostasis. *Proc Natl Acad Sci USA.* 101:9792-7.
- Wright PR, Georg J, Mann M, Sorescu DA, Richter AS, Lott S, Kleinkauf R, Hess WR, Backofen R. 2014. CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. *Nucleic Acids Res.* 42:W119-23.
- Wright PR, Richter AS, Papenfort K, Mann M, Vogel J, Hess WR, Backofen R, Georg J. 2013. Comparative genomics boosts target prediction for bacterial small RNAs. *Proc Natl Acad Sci USA.* 110:E3487-96.
- Zhang A, Altuvia S, Tiwari A, Argaman L, Hengge-Aronis R, Storz G. 1998. The OxyS regulatory RNA represses rpoS translation and binds the Hfq (HF-I) protein. *EMBO J.* 17:6061-8.
- Zhang A, Wassarman KM, Ortega J, Steven AC, Storz G. 2002. The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. *Mol Cell.* 9:11-22.

APPENDICES

A. Role of MgrR in polymyxin B resistance. MgrR is an sRNA that modulates resistance against the antibiotic polymyxin B in *E. coli*. It was found that an *mgrR*-deletion strain was significantly resistant to the antibiotic polymyxin B (Moon and Gottesman 2009).

We identified a homolog of *mgrR* in *E. fergusonii*, which contained a 50 bp insertion sequence. To study the role of MgrR in polymyxin B resistance, we generated an *mgrR*-deletion *E. fergusonii* strain. Wild-type and *mgrR*-deletion *E. fergusonii* strains were grown in LB to an OD₆₀₀ of 0.5,

challenged with final concentration of 4 µg/mL of polymyxin B for 30 minutes, and colony forming units were counted. Three independent trials of the experiment were

conducted. We found that *E.*

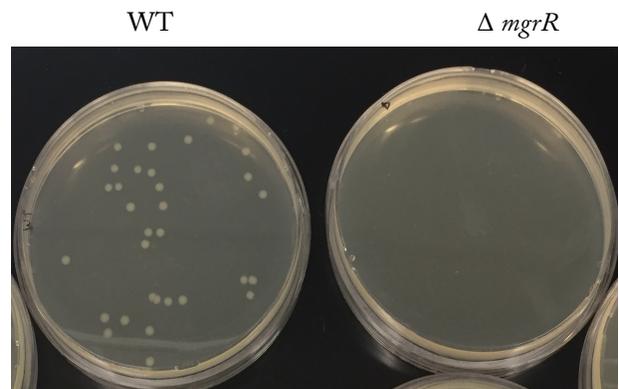


Figure 1

fergusonii wild type strain was resistant to polymyxin B whereas the *mgrR*-deletion strain was susceptible to polymyxin B (Figure 1). The additional 50 nt in MgrR of *E. fergusonii* reverses the effect of polymyxin B in *E. fergusonii* compared to its effect in *E. coli*.

REFERENCES

Moon K, Gottesman S. 2009. A PhoQ/P-regulated small RNA regulates sensitivity of *Escherichia coli* to antimicrobial peptides. *Mol Microbiol.* 74:1314-30.

B. Function of a novel sRNA. We

discovered a novel sRNA in *E. coli* called EcsR1, which influences biofilm formation (Raghavan et al. 2015). Due to genome rearrangement, the sRNA was lost in *Salmonella* by splitting into two pieces, making EcsR1 unique to *E.*

coli. To study the effect of EcsR1

on biofilm formation, the sRNA was cloned into the *NheI* and *HindIII* sites behind the arabinose-inducible promoter on plasmid pBAD using PCR. *EcsR1*-deletion strain of *E. coli* was constructed using λ Red-mediated recombination (Datsenko and Wanner 2000). *E. coli* or *Salmonella* strains grown overnight at 37°C in LB (or LB with 100 μ g/mL ampicillin) were diluted 1:100 in fresh media and grown in 96-well microtiter plates for 48 h at 28°C without shaking. Planktonic growth (OD_{600}) of *E. coli* and *Salmonella* strains measured on a Victor X5 microplate reader (Perkin Elmer) did not significantly differ from each other. Supernatants containing non-adhered cells were discarded, and samples were washed twice with distilled water and the attached biofilm in each well was stained with 0.1% crystal violet for 30 minutes. Unbound stain was removed by washing with distilled water. To quantify biofilm production, the crystal violet associated with biofilms was dissolved in 100% ethanol and absorbance (A_{600}) was measured, and normalized to the OD_{600} value of each strain, as described previously (Gualdi et al. 2008). Average

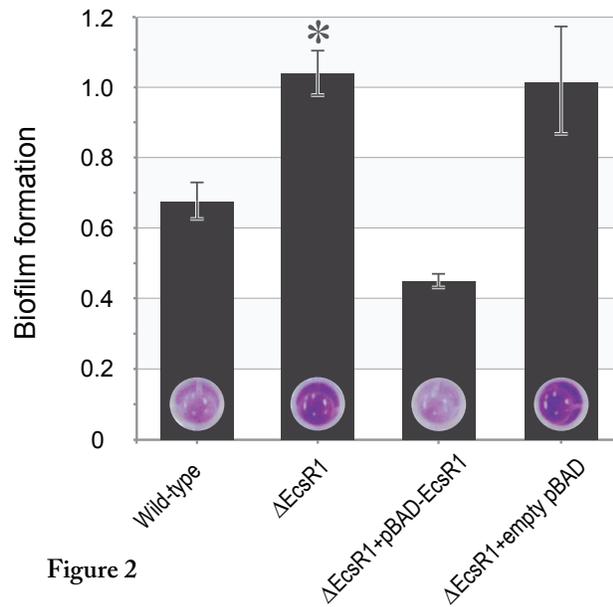


Figure 2

intensity of biofilm formation for each strain was generated from at least four replicate experiments. As shown in **Figure 2**, biofilm production is significantly reduced ($p < 0.0001$) in wild-type *E. coli* when compared to the EcsR1-deleted strain. Reintroduction of a plasmid-borne copy of EcsR1 into the deletion strain restored biofilm formation to the same level as that of the wild-type strain (**Figure 2**), indicating that biofilm-inhibition is an sRNA-specific phenotype. Biofilm production is important to virulence of enteric pathogens, so we tested the effects of EcsR1 on biofilm production in *Salmonella* by reintroducing the sRNA in an expression vector. There was no significant difference in biofilm production between the wild type and EcsR1-overexpression strains.

REFERENCES

- Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA*. 97:6640-5.
- Gualdi L, Tagliabue L, Bertagnoli S, Ieranò T, De Castro C, Landini P. 2008. Cellulose modulates biofilm formation by counteracting curli-mediated colonization of solid surfaces in *Escherichia coli*. *Microbiol* **154**: 2017-2024.
- Raghavan R, Kacharia FR, Millar JA, Sislak CD, Ochman H. 2015. Genome rearrangements can make and break small RNA genes. *Genome Biol Evol*. 7:557-66.