A Peroxide-Responding sRNA Evolved from a Peroxidase mRNA

Madeline C. Krieger,[†,1,2] H. Auguste Dutcher,[†,1,3] Andrew J. Ashford,[1,4] and Rahul Raghavan [*,1,5]

[1]Department of Biology, Portland State University, Portland, OR, USA

[2]Department of Restorative Dentistry, School of Dentistry, Oregon Health and Science University, Portland, OR, USA

[3]Laboratory of Genetics and Center for Genomic Science Innovation, University of Wisconsin-Madison, Madison, WI, USA

[4]Department of Molecular and Medical Genetics, Oregon Health and Science University, Portland, OR, USA

[5]Department of Molecular Microbiology and Immunology, The University of Texas at San Antonio, San Antonio, TX, USA

[†]These authors contributed equally to this work.

***Corresponding author:** E-mail: rahul.raghavan@utsa.edu.

Associate editor: Harmit Malik

Abstract

Small RNAs (sRNAs) are important gene regulators in bacteria, but it is unclear how new sRNAs originate and become part of regulatory networks that coordinate bacterial response to environmental stimuli. Using a covariance modeling-based approach, we analyzed the presence of hundreds of sRNAs in more than a thousand genomes across Enterobacterales, a bacterial order with a confluence of factors that allows robust genome-scale sRNA analyses: several well-studied organisms with fairly conserved genome structures, an established phylogeny, and substantial nucleotide diversity within a narrow evolutionary space. We discovered that a majority of sRNAs arose recently, and uncovered protein-coding genes as a potential source from which new sRNAs arise. A detailed investigation of the emergence of OxyS, a peroxide-responding sRNA, revealed that it evolved from a fragment of a peroxidase messenger RNA. Importantly, although it replaced the ancestral peroxidase, OxyS continues to be part of the ancestral peroxide-response regulon, indicating that an sRNA that arises from a protein-coding gene would inherently be part of the parental protein's regulatory network. This new insight provides a fresh framework for understanding sRNA origin and regulatory integration in bacteria.

***Key words:* OxyS, OxyR, sRNA, sRNA evolution, peroxidase, peroxide.**

Introduction

Bacterial small RNAs (sRNAs) control gene expression by modulating translation or by altering the stability of messenger RNAs (mRNAs). sRNAs allow precise and efficient control of gene expression because they are produced quickly, regulate multiple genes simultaneously, and could degrade along with target mRNAs (Hör et al. 2020). These qualities are especially beneficial under conditions such as oxidative stress that require abrupt reprogramming of regulatory networks (Holmqvist and Wagner 2017). In bacteria, oxidative stress caused by hydrogen peroxide (H_2O_2) is mitigated mainly by peroxidases (Imlay 2008). For instance, a peroxidase system encoded by *ahpCF* genes is induced by the regulator OxyR when *Escherichia coli* is exposed to H_2O_2; OxyR simultaneously upregulates the expression of several other genes, including the sRNA OxyS that together assuage H_2O_2 toxicity (Altuvia et al. 1997; Zheng et al. 1998; González-Flecha and Demple 1999; Imlay 2015). OxyS is one of the most well-studied sRNAs. More than two decades of research on this sRNA has revealed many of the foundational details about sRNA-mediated gene regulation (Altuvia et al. 1997; Zhang et al. 2002; Barshishat et al. 2018). In *E. coli* and *Salmonella enterica*, OxyS is encoded by a gene located in the intergenic region (IGR) between *oxyR* and *argH* genes. Similar to OxyS, most sRNAs in bacteria are transcribed from genes present in IGRs; however, in recent years, numerous sRNAs that are encoded within protein-coding genes and 3′ untranslated regions (UTRs) have also been identified (Miyakoshi et al. 2015).

Despite the discovery of hundreds of sRNAs, we do not fully understand how new sRNAs originate in bacteria (Dutcher and Raghavan 2018). One of the main impediments to elucidating the evolutionary histories of sRNAs is the difficulty in tracing sRNAs across large phylogenetic distances (Barquist et al. 2016). Unlike proteins that are fairly easy to identify in distant bacteria, sRNAs can only be reliably detected within clusters of related microbes (Lindgreen et al. 2014). This difficulty is due to a combination of factors, including their small size (50–400 nt), rapid turnover, and lack of open reading frames (ORFs) or other features that serve as signposts (Lindgreen et al. 2014; Updegrove et al. 2015; Barquist et al. 2016; Kacharia et al. 2017; Dutcher and Raghavan 2018). Given these constraints, an ideal group of bacteria to study sRNA evolution is the order Enterobacterales (Lindgreen et al. 2014), which has an established phylogeny, substantial nucleotide diversity within a narrow evolutionary space, and contains well-characterized organisms with diverse lifestyles but enough similarity in genome structure to enable meaningful comparative genomics.

Here, by analyzing the prevalence of hundreds of sRNAs in more than a thousand Enterobacterales genomes, we show that most sRNAs arose recently, and that mRNAs are a potential source for the generation of new sRNAs. One sRNA that originated from an mRNA is OxyS, which evolved from a 3′-end fragment of a peroxidase mRNA. Interestingly, both the parental peroxidase and OxyS are regulated by OxyR, suggesting a novel paradigm for understanding how new sRNAs arise and are recruited into preexisting regulatory networks: Transformation of a protein-coding gene into an sRNA gene could give rise to a new sRNA that is under the control of the parental protein's regulatory network.

Results

Most sRNAs in Enteric Bacteria Arose Recently

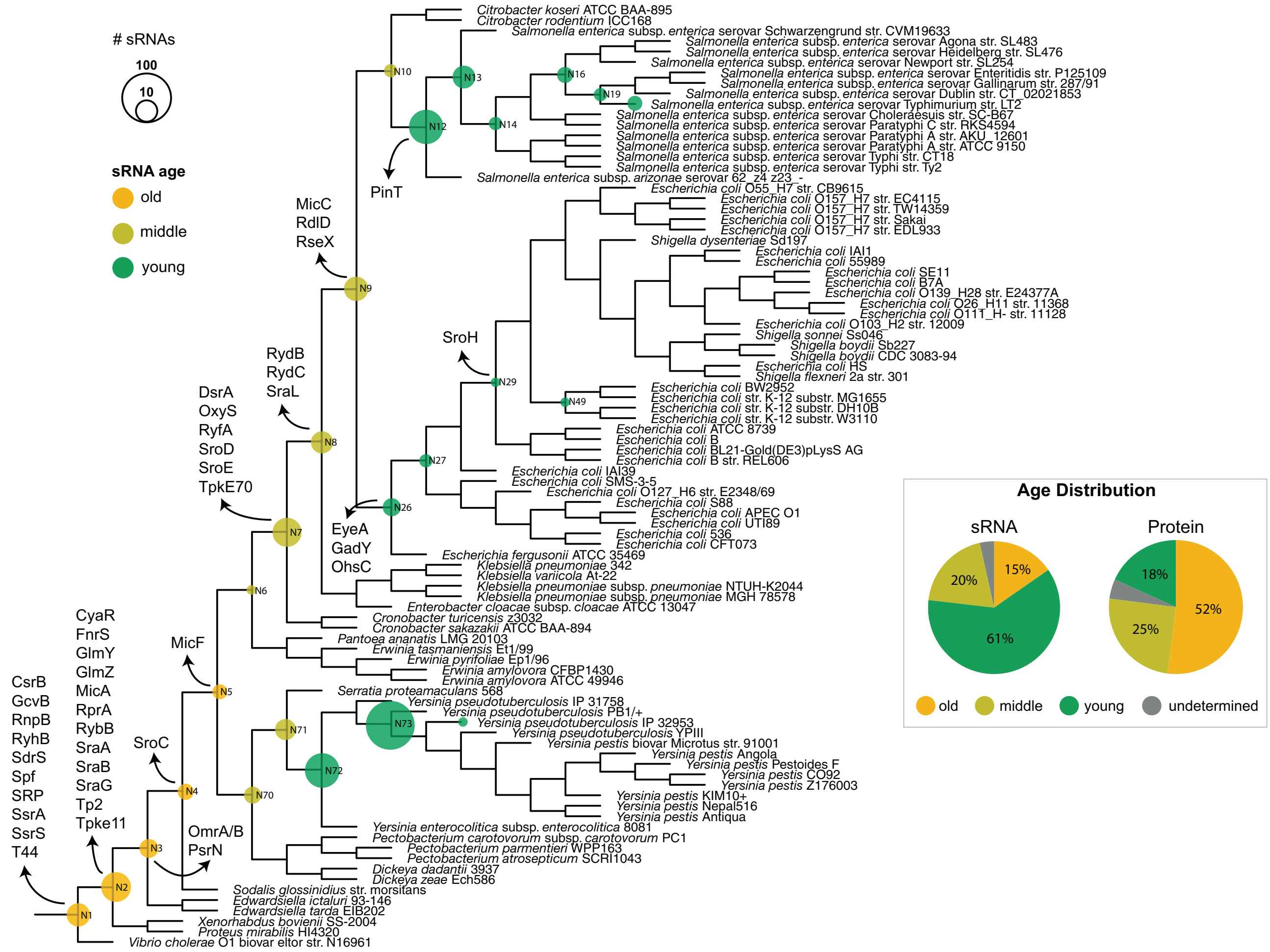
We built covariance models for 371 sRNAs described in *E. coli* K-12 MG1655, *S. enterica* Typhimurium SL1344, and *Yersinia pseudotuberculosis* IP32953, and located their homologs across 1105 Enterobacterales genomes. The ensuing phyletic patterns of sRNA presence and absence was used to perform an evolutionary reconstruction of ancestral states using a maximum likelihood approach (fig. 1 and supplementary table S1, Supplementary Material online). This order-wide analysis showed that 61% of sRNAs (228/371) emerged at the root of a genus or more recently (categorized as "young"). In comparison, among 148 proteins that function as gene regulators in *E. coli* and *S. enterica*, only 18% fall in this category (fig. 1 inset and supplementary fig. S1 and table S2, Supplementary Material online). The overrepresentation of recently evolved sRNAs in our data set indicates that most sRNAs probably arose in response to lineage-specific selection pressures. It should be noted however that the functions, if any, of most recently emerged sRNAs have not been determined, and that nearly all sRNAs with known functions have putative origins ancestral to the root of their respective genera ("middle" and "old" categories) (fig. 1 and supplementary table S1, Supplementary Material online).

Protein-Coding Genes Are Potential Progenitors of sRNA Genes

Our covariance modeling-based search identified 62 sRNAs that were located in IGRs in the hub genomes (*E. coli* K-12 MG1655, *S. enterica* Typhimurium SL1344, or *Y. pseudotuberculosis* IP32953) but mapped to the coding strands of protein-coding genes in other Enterobacterales members (supplementary table S3, Supplementary Material online). A majority of the overlaps were at the 3′-ends of genes (34/62), whereas 18 were at 5′-ends and 10 within gene boundaries. The sRNA-ORF overlaps suggest that some of the sRNAs were originally part of mRNAs, and later evolved into independent sRNAs when the protein-coding genes decayed, leaving behind only the sRNA-encoding segments.

To better understand their evolutionary histories, we further examined several sRNAs that overlapped protein-coding genes with known functions. This analysis revealed that OxyS, a ∼110-nt sRNA produced in response to peroxide stress in *E. coli* and *S. enterica*, overlapped the 3′-end of a peroxidase gene in *Serratia* and *Dickeya* (fig. 2). The peroxidase gene is located in the same genetic context—divergent from *oxyR*, as OxyS is in *E. coli* and *S. enterica*, denoting that the sRNA likely evolved from the peroxidase gene. In addition, the promoter regions of both *oxyS* and peroxidase genes contain OxyR-binding sites, indicating that the expression of the peroxidase gene is controlled by OxyR, as shown for OxyS (Altuvia et al. 1997; Zheng et al. 1998). Another sRNA that seems to be part of its parental protein's regulatory circuit is StyR-3. This sRNA of unknown function is highly abundant in *S. enterica* (Chinni et al. 2010). It shares sequence homology with the 5′-end of an MBL-fold metallohydrolase (MMH) gene in *Citrobacter* and *Klebsiella*, and both StyR-3 and MMH are located divergently from the transcriptional regulator gene *ramR*. In addition, the IGR between *ramR* and StyR-3/MMH contains a RamR-binding site (fig. 2). The sequence similarity, homologous genetic location, and conservation of RamR-binding site suggest that StyR-3 evolved from the 5′-end of the MMH gene and continues to be under the regulatory control of the divergently encoded RamR.

A third example of an sRNA that likely evolved from a protein-coding gene is STnc240, an sRNA with unknown function in *S. enterica*. The gene for this sRNA is located between *yeeY* and *yoeI* genes in *Salmonella* species, but in *Cronobacter*, the *yoeI-yeeY* IGR contains a 4-aminobutyrate-2-oxoglutarate transaminase (*gabT*) gene whose 3′-end contains a sequence that is very similar to that of STnc240 (fig. 2). Transcriptional regulation of *gabT* and STnc240 are not well defined, but sequence homology and conservation of genetic location suggest that the sRNA arose from the remnants of the *gabT* gene. An sRNA that seems to have evolved recently in *Salmonella* from a protein-coding gene is STnc3230. This "young" sRNA likely emerged from the 3′-end of a 1,3-1,4-beta-glucanase sugar-binding protein (SBP) (fig. 2). Although both *S. bongori* and *S. enterica* Arizonae contain a gene for SBP between *dapB* and *carA* genes, STnc3230, which shares sequence similarity with 3′-end of the SBP gene, is located in this IGR in *S. enterica* Typhi and *S. enterica* Typhimurium. Lastly, an sRNA that seems to have evolved from within a protein-coding gene is IsrK. This prophage-encoded sRNA likely originated from the ASH domain of a bacteriophage protein-coding gene (Iyer et al. 2002) and evolved to regulate the expression of a prophage-encoded anti-terminator protein AntQ (Hershko-Shalev et al. 2016). Similar to the origin of IsrK from a degenerated prophage gene, we have shown previously that EcsR2, an sRNA present in *E. coli*, evolved from a degraded phage tail fiber gene (Kacharia et al. 2017). Additionally, three sRNAs (Esr2, Esr4, and Ysr232) overlap genes that encode transposases and integrases (supplementary table S3, Supplementary Material online), suggesting that they arose in transposons or insertion sequences, as we showed recently for sRNAs in the pathogen *Coxiella burnetii* (Wachter et al. 2018). Of these



# sRNAs
100
10
sRNA age
old
middle
young
N1
N2
N3
N4
N5
N6
N7
N8
N9
N10
N12
N13
N14
N16
N19
N26
N27
N29
N49
N70
N71
N72
N73
CsrB
GcvB
RnpB
RyhB
SdrS
Spf
SRP
SsrA
SsrS
T44
CyaR
FnrS
GlmY
GlmZ
MicA
RprA
RybB
SraA
SraB
SraG
Tp2
Tpke11
SroC
OmrA/B
PsrN
MicF
DsrA
OxyS
RyfA
SroD
SroE
TpkE70
RydB
RydC
SraL
MicC
RdlD
RseX
EyeA
GadY
OhsC
SroH
PinT
Citrobacter koseri ATCC BAA-895
Citrobacter rodentium ICC168
Salmonella enterica subsp. enterica serovar Schwarzengrund str. CVM19633
Salmonella enterica subsp. enterica serovar Agona str. SL483
Salmonella enterica subsp. enterica serovar Heidelberg str. SL476
Salmonella enterica subsp. enterica serovar Newport str. SL254
Salmonella enterica subsp. enterica serovar Enteritidis str. P125109
Salmonella enterica subsp. enterica serovar Gallinarum str. 287/91
Salmonella enterica subsp. enterica serovar Dublin str. CT_02021853
Salmonella enterica subsp. enterica serovar Typhimurium str. LT2
Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67
Salmonella enterica subsp. enterica serovar Paratyphi C str. RKS4594
Salmonella enterica subsp. enterica serovar Paratyphi A str. AKU_12601
Salmonella enterica subsp. enterica serovar Paratyphi A str. ATCC 9150
Salmonella enterica subsp. enterica serovar Typhi str. CT18
Salmonella enterica subsp. enterica serovar Typhi str. Ty2
Salmonella enterica subsp. arizonae serovar 62_z4 z23_-
Escherichia coli O55_H7 str. CB9615
Escherichia coli O157_H7 str. EC4115
Escherichia coli O157_H7 str. TW14359
Escherichia coli O157_H7 str. Sakai
Escherichia coli O157_H7 str. EDL933
Shigella dysenteriae Sd197
Escherichia coli IAI1
Escherichia coli 55989
Escherichia coli SE11
Escherichia coli B7A
Escherichia coli O139_H28 str. E24377A
Escherichia coli O26_H11 str. 11368
Escherichia coli O111_H- str. 11128
Escherichia coli O103_H2 str. 12009
Shigella sonnei Ss046
Shigella boydii Sb227
Shigella boydii CDC 3083-94
Escherichia coli HS
Shigella flexneri 2a str. 301
Escherichia coli BW2952
Escherichia coli str. K-12 substr. MG1655
Escherichia coli str. K-12 substr. DH10B
Escherichia coli str. K-12 substr. W3110
Escherichia coli ATCC 8739
Escherichia coli B
Escherichia coli BL21-Gold(DE3)pLysS AG
Escherichia coli B str. REL606
Escherichia coli IAI39
Escherichia coli SMS-3-5
Escherichia coli O127_H6 str. E2348/69
Escherichia coli S88
Escherichia coli APEC O1
Escherichia coli UTI89
Escherichia coli 536
Escherichia coli CFT073
Escherichia fergusonii ATCC 35469
Klebsiella pneumoniae 342
Klebsiella variicola At-22
Klebsiella pneumoniae subsp. pneumoniae NTUH-K2044
Klebsiella pneumoniae subsp. pneumoniae MGH 78578
Enterobacter cloacae subsp. cloacae ATCC 13047
Cronobacter turicensis z3032
Cronobacter sakazakii ATCC BAA-894
Pantoea ananatis LMG 20103
Erwinia tasmaniensis Et1/99
Erwinia pyrifoliae Ep1/96
Erwinia amylovora CFBP1430
Erwinia amylovora ATCC 49946
Serratia proteamaculans 568
Yersinia pseudotuberculosis IP 31758
Yersinia pseudotuberculosis PB1/+
Yersinia pseudotuberculosis IP 32953
Yersinia pseudotuberculosis YPIII
Yersinia pestis biovar Microtus str. 91001
Yersinia pestis Angola
Yersinia pestis Pestoides F
Yersinia pestis CO92
Yersinia pestis Z176003
Yersinia pestis KIM10+
Yersinia pestis Nepal516
Yersinia pestis Antiqua
Yersinia enterocolitica subsp. enterocolitica 8081
Pectobacterium carotovorum subsp. carotovorum PC1
Pectobacterium parmentieri WPP163
Pectobacterium atrosepticum SCRI1043
Dickeya dadantii 3937
Dickeya zeae Ech586
Sodalis glossinidius str. morsitans
Edwardsiella ictaluri 93-146
Edwardsiella tarda EIB202
Xenorhabdus bovienii SS-2004
Proteus mirabilis HI4320
Vibrio cholerae O1 biovar eltor str. N16961
Age Distribution
sRNA
Protein
15%
20%
61%
18%
52%
25%
old
middle
young
undetermined

Fig. 1. sRNA nodes of origin. sRNAs that arose at each node is depicted by circles. Size and color of each circle corresponds, respectively, to the number of sRNAs and their ages, as shown in the side panel. Nodes of origin of a few well-studied sRNAs are also marked. Inset: Comparison of age distributions of sRNAs and regulatory proteins (supplementary fig. S1, Supplementary Material online).

eight potentially ORF-derived sRNAs, we focus on the origin of OxyS in the rest of the article.

A Peroxidase Gene Was Replaced by *oxyS* Gene in Enterobacteriaceae

In the order Enterobacterales, *oxyS* gene is present only in the family Enterobacteriaceae (e.g., *E. coli*, *S. enterica*), where it is located divergently from the *oxyR* gene in the *oxyR-argH* IGR (fig. 3). In contrast, a peroxidase (peroxiredoxin-glutaredoxin hybrid) gene occupies the same locus in families Erwiniaceae, Pectobacteriaceae, Yersiniaceae, Hafniaceae, and Budviciaceae. Bacteria belonging to orders Pasteurellales and Vibrionales also contain orthologous peroxidase genes at this location (fig. 3). The most parsimonious explanation for this phylogenetic profile is that the peroxidase gene was present in the common ancestor of all Enterobacterales and that it was subsequently replaced by the *oxyS* gene in Enterobacteriaceae.

We identified OxyS-like sequences at the 3′-ends of peroxidase genes in Pectobacteriaceae and Yersiniaceae (fig. 2), but not in other families. Curiously, although Erwiniaceae is more closely related to Enterobacteriaceae, no matches to OxyS were found in this family, probably because peroxidase 3′-ends have diverged substantially in Erwiniaceae. A closer examination of the *oxyS*-like sequence in the peroxidase gene of *Serratia* (Yersiniaceae), which had the best match to our OxyS covariance model outside of Enterobacteriaceae, suggests that the last ∼65 nt of the peroxidase coding sequence, ∼25 nt of the 3′-UTR, and the downstream intrinsic terminator collectively transformed into the *oxyS* gene (fig. 4). Based on these data, we conclude that *oxyS* gene present in Enterobacteriaceae is the remnant of the 3′-end of the ancestral peroxidase gene present in the rest of the members of the order Enterobacterales.

Exposure to H_2O_2 Induced Peroxidase Expression and Production of mRNA Fragments

Similar to *oxyS*, the peroxidase gene is located divergently from *oxyR*, and the IGR between the two genes contain putative OxyR-binding sites (fig. 3 and supplementary fig. S2, Supplementary Material online). To test whether the expression of the peroxidase gene is induced by H_2O_2, we grew two Enterobacterales members (*Serratia marcescens* [family Yersiniaceae], *Edwardsiella hoshinae* [family Hafniaceae])

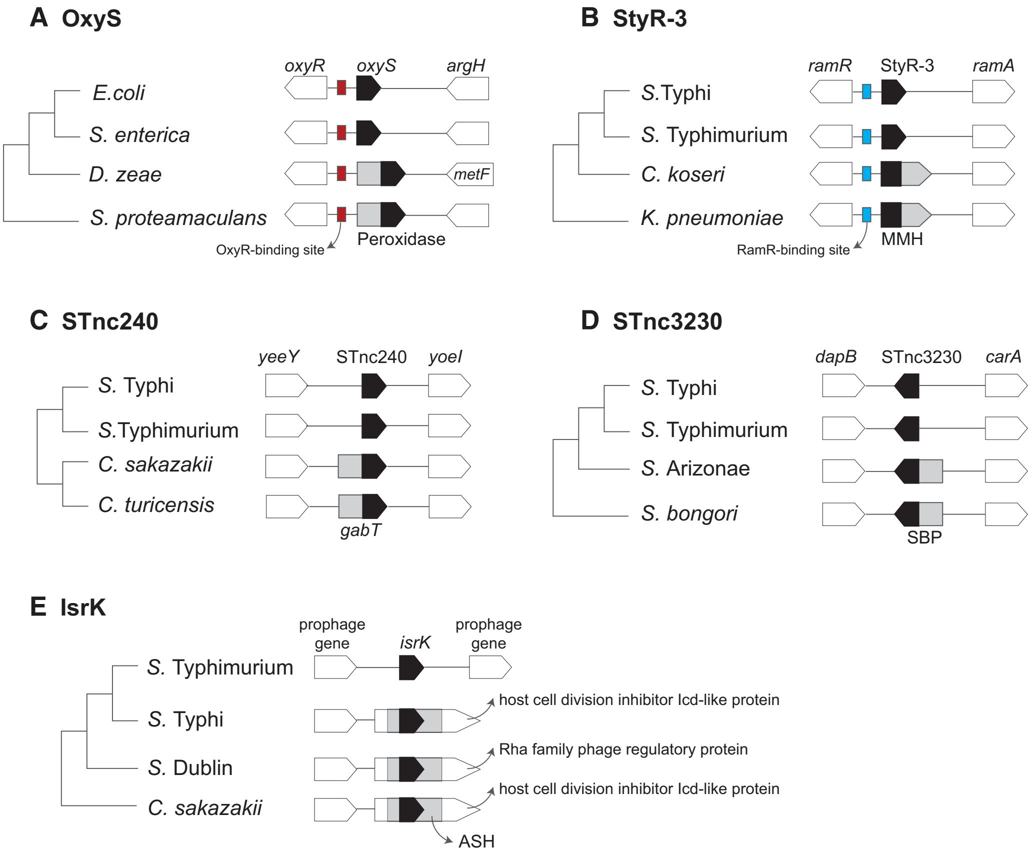


Fig. 2. sRNAs genes replaced protein-coding genes. Several examples of sRNAs that likely originated from protein-coding genes are shown. (*A*) *oxyS* genes in *Escherichia coli* and *Salmonella enterica* share sequence homology with the 3′ of a peroxidase gene in *Serratia proteamaculans* and *Dickeya zeae*. Black arrows represent *oxyS* and its homologous sequences in peroxidase genes (gray arrows). Peroxidase and *oxyS* genes are located divergently from *oxyR*, and OxyR-binding sites (red boxes) are present upstream of both *oxyS* and peroxidase genes. (*B*) StyR-3 in *Salmonella enterica* Typhi and *S. enterica* Typhimurium share sequence homology with the 5′-end of an MBL-fold metallohydrolase (MMH) gene in *Citrobacter Koseri* and *Klebsiella pneumoniae*. Black arrows represent StyR-3 and its homologous sequences in MMH genes (gray arrows). StyR-3 and MMH genes are located between *ramR* and *ramA* genes, and RamR-binding sites (blue boxes) are present upstream of both StyR-3 and MMH genes. (*C*) STnc240 in *S. enterica* Typhi and *S. enterica* Typhimurium share sequence homology with the 3′-end of *gabT* gene in *Cronobacter sakazakii*, and *C. turicensis*. Both STnc240 and *gabT* genes are located between *yeeY* and *yoeI* genes. (*D*) STnc3230 in *S. enterica* Typhimurium and *S. enterica* Typhi share sequence homology with the 3′-end of a SBP gene in *S. enterica* Arizonae and *Salmonella bongori*. Both STnc3230 and SBP genes are located between *dapB* and *carA* genes. DapZ, an sRNA transcribed from within *dapB*, is not shown. (*E*) IsrK, a prophage-encoded sRNA, shares sequence homology with a region within ASH domains present in several prophage genes. Black arrows represent *isrK* and its homologous sequences in ASH domains (gray boxes). Genes and IGRs are not drawn to scale.

and *Vibrio harveyi* (order Vibrionales, family Vibrionaceae), a representative from outside of Enterobacterales, to an OD600 of ∼0.5, and exposed them to 1 mM of H_2O_2 for 10 min. We investigated peroxidase gene expression in H_2O_2-exposed and nonexposed bacteria using RNA-seq, and as shown in figure 5, large transcriptional peaks that correspond to high peroxidase expression was observed in H_2O_2-exposed bacteria but not in nonexposed controls. qRT-PCR assays confirmed the induction of peroxidase gene expression by H_2O_2 in the three bacteria (fig. 5), indicating that the peroxidase gene is regulated by OxyR, as observed for *oxyS* gene (Altuvia et al. 1997).

In addition to induction by H_2O_2, peroxidase genes in *Serratia*, *Edwardsiella*, and *Vibrio* produced small mRNA 3′ fragments that correspond to the region from where OxyS likely emerged (fig. 6). We could not detect promoter-like sequences within peroxidase genes, suggesting that the smaller fragments are not primary transcripts and are probably generated by RNase digestion of peroxidase mRNAs. Although the cleavage products are not identical in length in the three bacteria—perhaps because RNase cleavage sites are located at slightly different regions of the mRNAs, the production of stable 3′ fragments appears to be an ancestral trait conserved across Enterobacterales and Vibrionales. Based on these data, we surmise that the expression of the ancestral peroxidase gene is induced by H_2O_2 and degradation products generated from the peroxidase mRNA provided the raw material from which OxyS eventually evolved in Enterobacteriaceae (fig. 7).

Discussion

Despite their importance to bacterial physiology and virulence, the evolutionary processes that produce new sRNAs are not well understood. One of the main reasons for this lack of clarity about sRNA origination is that unlike protein-coding

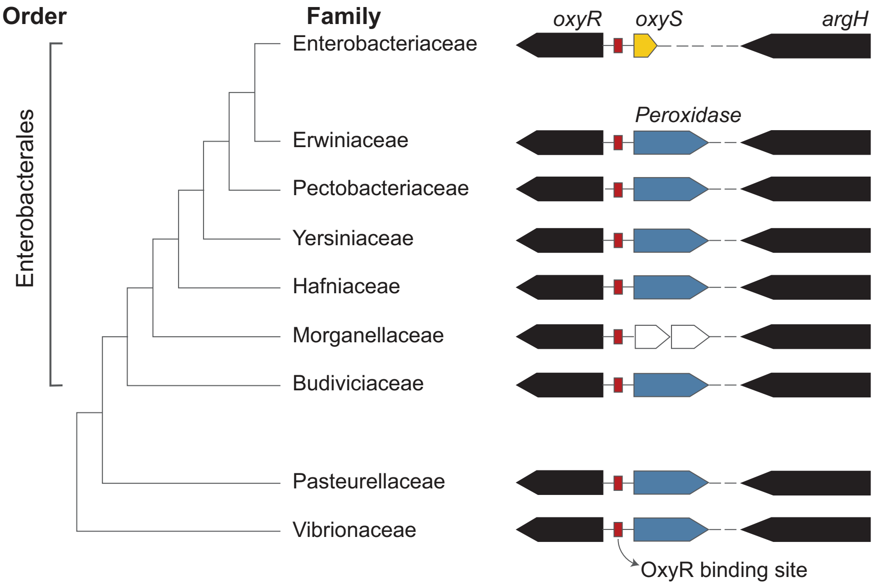
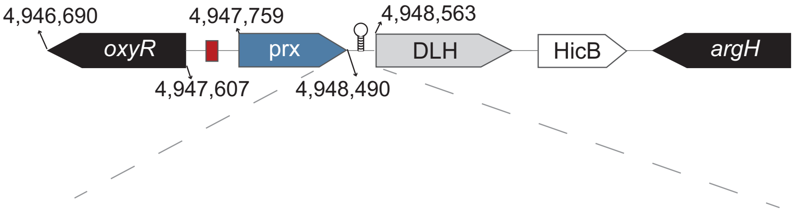


Fig. 3. OxyS arose from a peroxidase gene. Arrangement of *oxyR*, peroxidase and *argH* genes in bacterial families within orders Enterobacterales, Pasteurellales, and Vibrionales is shown. In Enterobacteriaceae family, *oxyS* (yellow arrow) is found in place of the peroxidase gene (blue arrow). The IGR between peroxidase gene and *argH* varies between families. In Morgenallaceae, transposon-associated genes (white arrows) are located in this region. The cladogram is based on Adeolu et al. (2016).



```
    E. coli oxyS  GAAACGGAGCGGCACCTCTTTTAACCCTTGAAGTCACTGCCCGTTTCGAGAGTTTCTCAA 60
S.marcescens prx  TCATCGGCG-GTCGTCACATCGGCGGCAGCGACGATCTGGAAACCTTCCTGTCAGCTTAA 59
                    * *** * * *  * * *      *    *    ***      *         ** **

    E. coli oxyS  CTCGAATAACTAAAGCCAACGTGAACTTTTGCGGATCTCCAGGATCCGCTTTTTTTT 117
S. marcescens prx TAAGTAATAACAAAGCCAACGTGAACTTTGGCGGGCTACGGCCCGCCCTTTTTTTAC 116
                     * *  *  ****************** ****    *      **  ******
```

Fig. 4. Alignment of *E. coli* OxyS with 3′-end of peroxidase gene in *Serratia*. The peroxidase (prx) gene (blue arrow) in *Serratia marcescens* ATCC13880 (CP041233) is flanked by *oxyR* (black arrow) and a dihydrolipoyl dehydrogenase (DLH) gene (gray arrow). A HicB family antitoxin gene (white arrow) is located between DLH and *argH* genes. A ClustalW alignment of *oxyS* gene in *E. coli* MG1655 (NC_000913.3) with 3′-end of peroxidase (prx) gene in *S. marcescens* ATCC13880 is shown at the bottom. Nucleotides in blue are part of the peroxidase coding sequence, the stop codon is in red, and the predicted Rho-independent terminator sequence is underlined. The IGR between peroxidase and *oxyR* genes contains putative OxyR-binding sites (red square).

genes, sRNA genes are difficult to trace across large phylogenetic distances (Barquist et al 2016). Following up on previous research that showed that enteric bacteria are at optimum distances from one another to effectively investigate sRNA prevalence (Lindgreen et al. 2014), we traced the presence of hundreds of sRNAs across Enterobacterales and show that a majority emerged recently. This observation fits with earlier findings that sRNAs evolve rapidly in bacteria and are typically genus- or species-specific (Skippington and Ragan 2012; Raghavan et al. 2015; Kacharia et al. 2017). Interestingly, we found that most well-studied sRNAs belong to "middle" and "old" age groups. A similar observation was made by a study that examined the evolutionary histories of 58 experimentally validated sRNAs in *E. coli* (Peer and Margalit 2014). Although specific age categories were not assigned in that study, when we classified the sRNAs into three age groups based on the distance of gain-node from *E. coli* (old: >0.1; middle: 0.001—0.009; young: 0.0001—0.0009), 51/58 sRNAs were deemed to be old or middle aged (supplementary table S4, Supplementary Material online). This biased representation is probably due to the propensity of older sRNAs to be expressed at high levels, thereby making them more amenable to discovery and experimental validation (Raghavan et al. 2011; Kacharia et al. 2017).

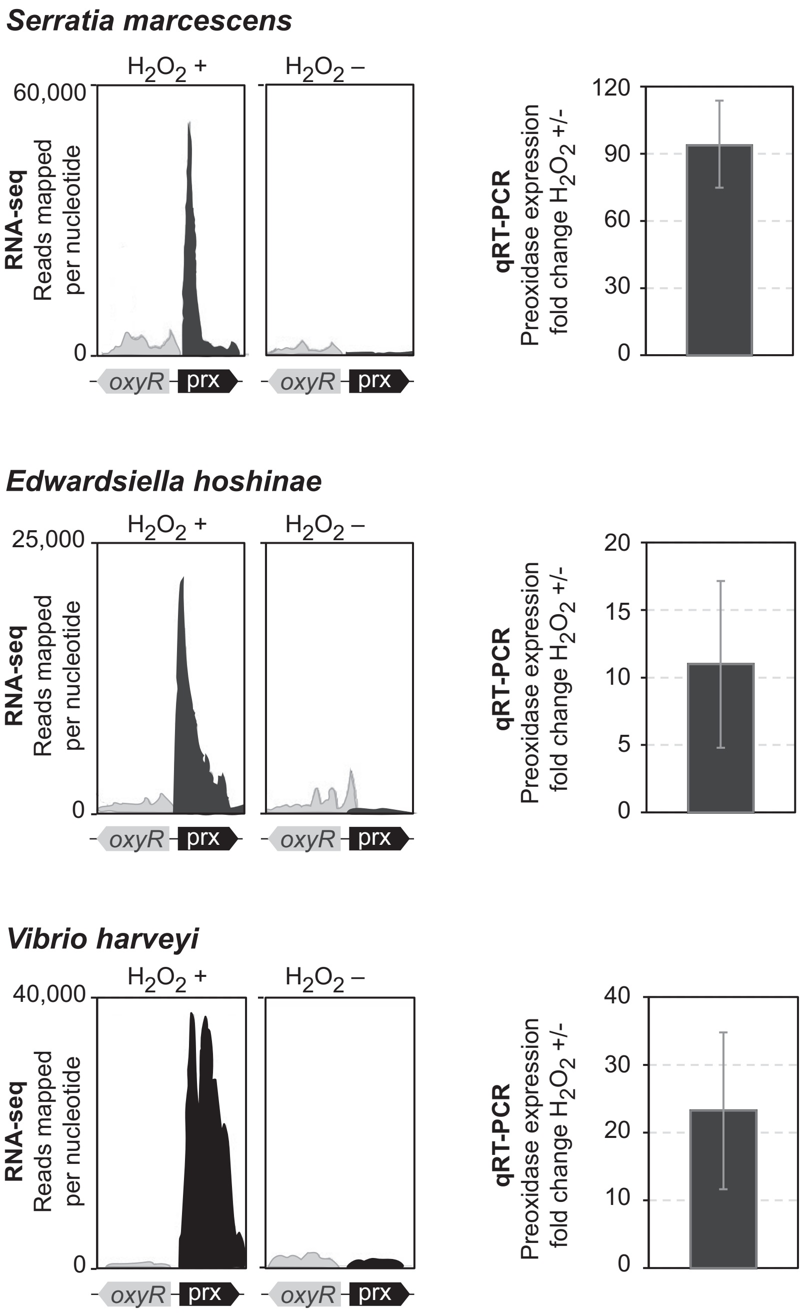


Fig. 5. Peroxidase expression in *Serratia*, *Edwardsiella*, and *Vibrio* is induced by H_2O_2. RNA-seq expression profiles of peroxidase (prx) and *oxyR* genes in *S. marcescens*, *E. hoshinae*, and *V. harveyi* exposed to H_2O_2 (H_2O_2 +) in comparison with no-exposure controls (H_2O_2 −) are shown on the left. Induction of peroxidase expression was confirmed using qRT-PCR (right panels). Peroxidase expression fold-change values (mean ± SD) were calculated from two independent growth experiments.

We have previously shown that new sRNAs arise de novo and from degraded bacteriophage- and transposon-associated genes (Raghavan et al. 2015; Kacharia et al 2017; Wachter et al. 2018). In this study, we report that protein-coding genes could serve as a raw material for sRNA biogenesis and support this conclusion by showing that OxyS, a peroxide-responding sRNA, originated from a peroxidase gene. OxyS was first noticed by researchers because it is transcribed divergently from the *oxyR* gene that encodes a transcriptional regulator that orchestrates *E. coli*'s antioxidant response (Altuvia et al. 1997). We show that the *oxyR*–*oxyS* gene arrangement is present only in the family Enterobacteriaceae, whereas a peroxidase gene, whose expression is also induced by H_2O_2, occupies the genetic locus next to *oxyR* in other members of the order Enterobacterales. The evolutionary process that led to the replacement of the ancestral peroxidase gene by *oxyS* gene in Enterobacteriaceae could have occurred through two routes (fig. 7). In one, the mRNA 3′-end fragment gained a regulatory function, which resulted in its retention when the rest of the peroxidase gene

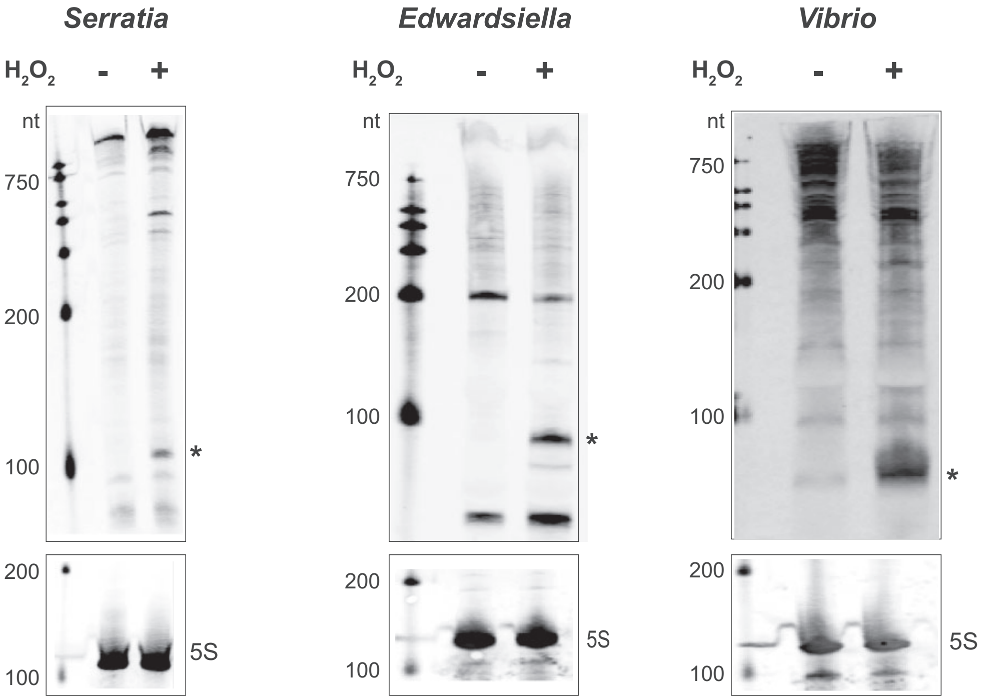


Fig. 6. Peroxidase mRNA fragmentation. 3′-end fragments, including ones that are similar in size to OxyS (marked with *) were cleaved from peroxidase mRNAs in *S. marcescens*, *E. hoshinae*, and *V. harveyi*. "+" indicates samples exposed to 1 mM of H_2O_2 for 10 min, and nonexposure controls are shown with "−." Northern blotting was performed with probes that bind to the 3′-ends of peroxidase mRNAs. A single-stranded RNA ladder was used to estimate the size of the transcripts. 5S rRNA was used as loading controls (bottom panels).

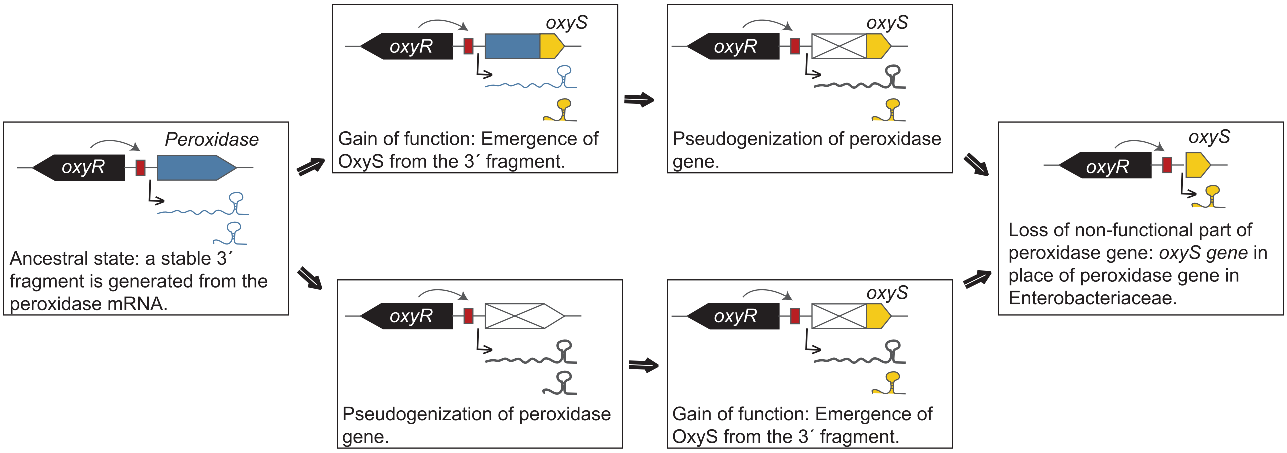

Peroxidase
oxyR
Ancestral state: a stable 3´ fragment is generated from the peroxidase mRNA.
oxyS
Gain of function: Emergence of OxyS from the 3´ fragment.
Pseudogenization of peroxidase gene.
Loss of non-functional part of peroxidase gene: oxyS gene in place of peroxidase gene in Enterobacteriaceae.

Fig. 7. Two possible routes of OxyS evolution. A stable 3′-end fragment was produced from the ancestral peroxidase mRNA. Top path: functional OxyS emerged as a 3′-end-derived sRNA prior to the pseudogenization of the peroxidase gene. Bottom path: OxyS emerged at the 3′-end of an RNA transcribed from a pseudogenized peroxidase gene. Ultimately, the nonfunctional part of the peroxidase gene was deleted from the genome, resulting in the formation of the *oxyS* gene in Enterobacteriaceae. OxyR binds to sites (red boxes) located in the IGR and regulates the expression of peroxidase and *oxyS* genes.

was pseudogenized and later deleted. Alternatively, the peroxidase gene continued to be transcribed even after being pseudogenized, thus producing the 3′-end fragment that gained a regulatory function and was retained when the rest of the gene was deleted. In either case, the low sequence conservation observed between *E. coli*'s OxyS and the peroxidase coding region in *S. marcescens* (fig. 4) suggests that the sRNA accumulated a large number of mutations, which was likely aided by the loss of selection pressure to maintain codons when the peroxidase gene was pseudogenized. An evolutionary scenario in which OxyS evolved rapidly as it emerged from peroxidase mRNA is supported by previous studies that showed that newly emerged noncoding RNAs in both bacteria and eukaryotes evolve at significantly higher rates than older ones (Jovelin and Cutter 2014; Kacharia et al. 2017).

An unresolved question in the field of sRNA biology is how new sRNAs become incorporated into regulatory networks. This study provides a novel explanation: an sRNA arising from an mRNA would inherently be part of the parental protein's regulon. For instance, in either scenario in figure 7, the mRNA fragment that gave rise to OxyS would have been produced as part of the OxyR regulon even before it gained any function. Thus, when functional OxyS later emerged from that fragment, it was already part of the OxyR regulatory network. Unlike OxyS, it is possible that other sRNAs evolved from protein-coding genes without the involvement of mRNAs. For instance, a new promoter sequence could have arisen near or within a protein-coding gene and produced a novel transcript that later evolved into an sRNA independent of the mRNA. In this case, the new sRNA would be integrated into the parental protein's regulatory network if the new promoter were also controlled by the same regulator.

Similar to the origin of OxyS, 3′-ends of mRNAs appear to be the most favorable location for sRNA genesis (34/62 in our data set) probably due to the presence of intrinsic terminators that improve RNA stability and promote Hfq binding, two factors that are critical to sRNA evolution and function (Updegrove et al. 2015; Jose et al. 2019). The next best location seems to be 5′-ends (18/62), likely due to proximity to promoter regions that regulate transcription, with middle region being the least likely to contribute to sRNA evolution (10/62). Irrespective of this apparent difference, any part of an mRNA could potentially evolve into a regulatory RNA, as shown in a recent study that demonstrated the generation of sRNA-like transcripts from 5′, middle, and 3′ segments of mRNAs in *E. coli* (Dar and Sorek 2018). Furthermore, transformation of protein-coding genes into noncoding RNA genes has occurred multiple times in Eukaryotes (Kaessmann 2010), indicating that this is a universal mechanism of regulatory RNA origin. In sum, protein-coding genes, which are present in various stages of decay in bacterial genomes (Ochman and Davalos 2006), have the potential to function as a rich resource from which new sRNA genes could arise in response to environmental pressures.

Materials and Methods

Determining sRNA Presence across the Order Enterobacterales

A list of candidate-sRNAs in *E. coli* K-12 MG1655 (NC_000913.3), *S. enterica* Typhimurium SL1344 (NC_016810.1), and *Yersinia pseudotuberculosis* IP32953 (NC_006155.1) were compiled from previously published studies (Raghavan et al. 2011; Kröger et al. 2013; Nuss et al. 2015). Several exclusion criteria were used to remove spurious and redundant sRNAs: 1) sRNAs under 60 nucleotides in length, 2) sRNAs that overlapped each other by more than ten nucleotides, 3) sRNAs that were present in multiple copies, 4) RNAs that were identified as cis-acting in the Rfam database (Kalvari et al. 2018), and 5) sRNAs that shared 95% or more nucleotide identity over at least 60 nucleotides of their lengths. For each sRNA of interest, a representative sequence from each hub genome was used as the query for a BLASTn (wordsize 7, maxdbsize 100 kb, dbsize normalized to 4 Mb, *e*-value $\leq$1e-5) using BLAST v2.7.1 against a database of 1105 Enterobacterales genomes (supplementary table S5 and fig. S3, Supplementary Material online) that met the following criteria: 1) full genome sequence was available on GenBank, and 2) the genome was within 0.08 16S rDNA pairwise distance from the hub species (supplementary table S6, Supplementary Material online). Drawing on guidance from previous studies (Lindgreen et al. 2014; Barquist et al. 2016), hits with pident >65% covering at least 95% of the length of the original query served as seed sequences from which to construct a covariance model. Candidate hits were next binned by percent identity, and a randomly selected set of sequences (one from each percent identity bin) were chosen to serve as a seed sequence for the covariance model. These sequences were aligned using ClustalW, and the Infernal suite of tools (v1.1.2) was used for subsequent covariance model construction (cmbuild), calibration (cmcalibrate), and homolog searches (cmsearch) (Nawrocki and Eddy 2013). Models were constructed from the BLAST-derived seed sequences using cmbuild, whereas cmscan was used to identify sRNAs already represented by existing Rfam models. These newly constructed models, plus the existing Rfam models, were then used in parallel to search the 1105-genome database for homologs (supplementary table S7, Supplementary Material online). For sRNAs that were represented by an existing Rfam model, cmsearch results from this model were compared with that from the newly constructed model, and the model that yielded more hits was selected for continued iteration. Results from cmsearch with an *e*-value <1e-5 were used to add unrepresented sequences to the query model, which was then refined, recalibrated, and used for another round of cmsearch. This process was repeated for each sRNA until a cmsearch with its corresponding model failed to yield new unrepresented sequences. In order to ensure that any two models were not yielding the same set of hits, results from cmsearch with the finalized models were compared across sRNAs; models with redundant hits were omitted, as were any models that yielded >1e4 hits. An sRNA gene was considered present in a given organism if a hit of *e*-value <1e-5 was found on its chromosome and/or plasmid. All resultant hits were cross-checked by genome location to ensure that a given hit was not represented more than once in the final results. Presence/absence data for all 1105 organisms were collected, but only data for 89 Enterobacteriaceae, plus *Vibrio cholerae* El Tor str. N16961 as an outgroup was used for downstream sRNA-ORF overlap and phylogenetic analyses.

Evolutionary Reconstruction

Enterobacterales phylogenetic tree was downloaded from MicrobesOnLine, and node of origin for each sRNA was determined using the Gain and Loss Mapping Engine (GLOOME) as described previously (Cohen et al. 2010; Peer and Margalit 2014). For sRNAs present in a single hub genome (*E. coli* K-12 MG1655, *S. enterica* Typhimurium SL1344, or *Y. pseudotuberculosis* IP32953), the determined gain node was the most ancestral node with a posterior probability of $\geq$0.6, where all nodes leading from this ancestor to the hub genome had a posterior probability $\geq$0.6. If an sRNA was present in more than one hub genome, and the most recent last common ancestor (LCA) of the hub genomes in which it was present had a posterior probability $\geq$0.6, the gain node was the most ancestral node with a posterior probability $\geq$0.6, where all nodes leading from this ancestor to the aforementioned LCA had a posterior probability $\geq$0.6. sRNAs that emerged at the root of a genus or more recently were classified as "young" ($n = 228$), those present at the LCA of all three hub genomes were deemed to be "old" ($n = 57$), and sRNAs that emerged in between the two age groups were considered "middle" ($n = 73$). If the LCA of the hub genomes in which the sRNA was present did not have a posterior probability of $\geq$0.6, ages of these sRNAs were considered undetermined ($n = 13$).

Regulatory Protein Tree

Using the QuickGO annotation table (Binns et al. 2009), GO : 0010629 (negative regulation of gene expression) and its child terms were utilized to identify regulatory proteins in *S. enterica* Typhimurium LT2 and *E. coli* K-12 MG1655. Protein sequences (UP00000104, UP000000625) were downloaded from Uniport. Redundancies among the regulatory proteins from the two species were identified using a BLASTp of the regulatory protein sequence set against itself (pident >80, *e*-value <1e-10). Homologs of this final set of query proteins were identified using BLASTp (*e*-value <1e-10) against a database of protein sequences from the genomes used for phylogenetic analysis, yielding a presence/absence matrix. Node of origin was determined using GLOOME as described above.

Finding *oxyR-argH* IGRs

To find *oxyR* and *argH* orthologs, tBLASTn searches were carried out with OxyR and ArgH sequences from *E. coli* K-12 MG1655 against Enterobacterales genomes (*e*-value $\leq$10e-10, percent positive $\geq$60%, percentage alignment length $\geq$60%). We corroborated the tBLASTn hits by confirming that they contain Pfam domains PF03466.20 and PF00126.27 (for OxyR), and PF00206.20 and PF14698.6 (for ArgH) (Finn et al. 2014). We then compiled a list of bacteria that have both *oxyR* and *argH* genes and obtained the nucleotide sequences between the two genes using the Entrez E-utilities tool.

Identifying 5′ Neighbors of *oxyR* and OxyR-Binding Sites

Using BioPython (Cock et al. 2009), we first determined the direction of the gene next to *oxyR*'s 5′-end. If the neighboring gene was oriented divergently, we determined the identity of the encoded protein using Pfam as described above. All Enterobacterales and Vibrionales (except Morganellaceae) included in this study contained a peroxidase gene with Glutaredoxin (PF00462) and Redoxin (PF08534) domains in this locus. To identify putative OxyR-binding sites located in the IGR between *oxyR* and its neighbor, we extracted 50 bp at the 5′-end of *oxyR* along with 100 bp of the adjoining IGR from each bacterium. Sequences were aligned with MUSCLE and the 50 bp *oxyR* sequence was trimmed from each sequence (Edgar 2004). Using the Multiple Em for Motif Elicitation (MEME) tool (Bailey and Elkan 1994), we first detected ~37 bp palindromic sequences in IGRs, and then used the output from MEME in Find Individual Motif Occurrences (FIMO) program to identify putative OxyR-binding site in each bacterium (Grant et al. 2011). Sequence logos were generated using WebLogo 3 (Crooks et al. 2004).

Bacterial Growth, RNA-seq, and qRT-PCR

We procured *S. marcescens* ATCC 13880, *E. hoshinae* ATCC 35051, and *V. harveyi* ATCC 43516 from American Type Culture Collection (ATCC). *Serratia marcescens* was grown at 37 °C in Lysogeny broth (LB), *E. hoshinae* at 26 °C in LB, and *V. harveyi* at 26 °C in Marine Broth, all shaking at 200 rpm. All cultures were grown to an OD600 of 0.4-0.5 (supplementary fig. S4, Supplementary Material online) and split into two. One half was allowed to grow under the same conditions for 10 min, whereas the other half was exposed to 1 mM of H_2O_2 for 10 min. RNA Stop Solution (5% phenol, 95% ethanol) was added to the cultures at the end of 10 min incubation and total RNA was extracted using TRI reagent (Thermo Fisher Scientific). RNA was treated with TURBO DNase (Thermo Fischer Scientific), depleted of ribosomal RNA (rRNA) using Ribo-Zero Bacteria kit (Illumina Inc.) and RNA sequencing (RNA-seq) was performed at the Yale Center for Genome Analysis using Illumina NovaSeq (paired end, 150 bp). Low-quality RNA-seq reads and adapters were removed using Trimmomatic (Bolger et al. 2014) and CLC Genomics workbench was used to map the reads to the respective genomes. A custom Perl script (available at https://github.com/rahul-rna/RNA-seq_scripts, last accessed January 29, 2022) was used to convert the read-mapping information (SAM files) into text files that could be read by Artemis genome browser (Carver et al. 2012) to generate coverage plots (fig. 5).

Induction of peroxidase expression in the three bacteria by H_2O_2 was confirmed using quantitative Reverse Transcription PCR (qRT-PCR), as described previously (Wright et al. 2021). Briefly, two independent cultures of each bacterium were grown to an OD600 of ~0.5 in media and conditions described above. Cultures were split into two, and one half was allowed to grow under the same conditions for 10 min, whereas the other half was exposed to 1 mM of H_2O_2 for 10 min. Total RNA was extracted using TRI reagent, treated with TURBO DNase, and 500 ng of DNA-free RNA along with random hexamer primers and a High-Capacity cDNA Reverse Transcription Kit (Thermo Fischer Scientific) were used to generate cDNA. qRT-PCR was performed using SYBR Green master mix (Thermo Fischer Scientific) and primers listed in supplementary table S8, Supplementary Material online on a Mx3005P qRT-PCR system (Stratagene). Expression values of peroxidase genes were normalized to 16S rRNA gene expression values to calculate the fold change in peroxidase expression between H_2O_2-exposed and nonexposed samples.

Northern Blot

RNA samples were loaded onto either 6% or 10% TBE-Urea Gel (Thermo Fischer Scientific) with a biotinylated RNA ladder (Kerafast). Gels were run in 1× TBE buffer at 180 V, 60 min (6% gels) or 180 V, 80 min (10% gels). Membranes and filter paper were presoaked and RNA was transferred to a Biodyne B Nylon Membrane (Thermo Fischer Scientific) overnight. Membranes were UV crosslinked using a Stratalinker 2400 UV Crosslinker (1200 mj) and RNA probes (supplementary table S8, Supplementary Material online) were hybridized overnight at 45 °C with rotation. Hybridization solution was removed and membranes were washed, blocked in Licor Intercept Blocking Buffer and treated with Streptavidin-IRDye 800 CW and examined on a Licor Odyssey scanner.

Acknowledgments

We thank Samantha Fancher for technical assistance and Jim Archuleta for bioinformatics support. This work was supported in part by National Institutes of Health grants AI133023 and DE028409 to R.R.

Data Availability

RNA-seq reads are available in NCBI BioProject under the accession PRJNA665492.

References

Adeolu M, Alnajar S, Naushad S, Gupta R. 2016. Genome-based phylogeny and taxonomy of the "Enterobacteriales": proposal for Enterobacterales ord. nov. divided into the families Enterobacteriaceae, Erwiniaceae fam. nov., Pectobacteriaceae fam. nov., Yersiniaceae fam. nov., Hafniaceae fam. nov., Morganellaceae fam. nov., and Budviciaceae fam. nov. *Int J Syst Evol Microbiol.* 66(12):5575–5599.

Altuvia S, Weinstein-Fischer D, Zhang A, Postow L, Storz G. 1997. A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. *Cell* 90(1):43–53.

Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol.* 2:28–36.

Barquist L, Burge SW, Gardner PP. 2016. Studying RNA homology and conservation with infernal: from single sequences to RNA families. *Curr Protoc Bioinformatics.* 54:12.13.1–12.13.25.

Barshishat S, Elgrably-Weiss M, Edelstein J, Georg J, Govindarajan S, Haviv M, Wright PR, Hess WR, Altuvia S. 2018. OxyS small RNA induces cell cycle arrest to allow DNA damage repair. *Embo J.* 37(3):413–426.

Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. 2009. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* 25(22):3045–3046.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.

Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28(4):464–469.

Chinni SV, Raabe CA, Zakaria R, Randau G, Hoe CH, Zemann A, Brosius J, Tang T-H, Rozhdestvensky TS. 2010. Experimental identification and characterization of 97 novel npcRNA candidates in *Salmonella enterica* serovar Typhi. *Nucleic Acids Res.* 38(17):5893–5908.

Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, et al. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25(11):1422–1423.

Cohen O, Ashkenazy H, Belinky F, Huchon D, Pupko T. 2010. GLOOME: gain loss mapping engine. *Bioinformatics* 26(22):2914–2915.

Crooks GE, Hon G, Chandonia J-M, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res.* 14(6):1188–1190.

Dar D, Sorek R. 2018. Bacterial noncoding RNAs excised from within protein-coding transcripts. *mBio* 9(5):e01730–18.

Dutcher HA, Raghavan R. 2018. Origin, evolution, and loss of bacterial small RNAs. *Microbiol Spectr.* 6(2). 10.1128/microbiolspec.RWR-0004-2017.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32(5):1792–1797.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, et al. 2014. Pfam: the protein families database. *Nucleic Acids Res.* 42(Database issue):D222–D230.

González-Flecha B, Demple B. 1999. Role for the *oxyS* gene in regulation of intracellular hydrogen peroxide in *Escherichia coli. J Bacteriol.* 181(12):3833–3836.

Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27(7):1017–1018.

Hershko-Shalev T, Odenheimer-Bergman A, Elgrably-Weiss M, Ben-Zvi T, Govindarajan S, Seri H, Papenfort K, Vogel J, Altuvia S. 2016. Gifsy-1 prophage IsrK with dual function as small and messenger RNA modulates vital bacterial machineries. *PLoS Genet.* 12(4):e1005975.

Holmqvist E, Wagner EGH. 2017. Impact of bacterial sRNAs in stress responses. *Biochem Soc Trans.* 45(6):1203–1212.

Hör J, Matera G, Vogel J, Gottesman S, Storz G. 2020. Trans-acting small RNAs and their effects on gene expression in *Escherichia coli* and *Salmonella enterica*. *EcoSal Plus.* 9(1). 10.1128/ecosalplus.ESP-0030-2019.

Imlay JA. 2008. Cellular defenses against superoxide and hydrogen peroxide. *Annu Rev Biochem.* 77:755–776.

Imlay JA. 2015. Transcription factors that defend bacteria against reactive oxygen species. *Annu Rev Microbiol.* 69:93–108.

Iyer LM, Koonin EV, Aravind L. 2002. Extensive domain shuffling in transcription regulators of DNA viruses and implications for the origin of fungal APSES transcription factors. *Genome Biol.* 3(3):research0012.1.

Jose BR, Gardner PP, Barquist L. 2019. Transcriptional noise and exaptation as sources for bacterial sRNAs. *Biochem Soc Trans.* 47(2):527–539.

Jovelin R, Cutter AD. 2014. Microevolution of nematode miRNAs reveals diverse modes of selection. *Genome Biol Evol.* 6(11):3049–3063.

Kacharia FR, Millar JA, Raghavan R. 2017. Emergence of new sRNAs in enteric bacteria is associated with low expression and rapid evolution. *J Mol Evol.* 84(4):204–213.

Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. *Genome Res.* 20(10):1313–1326.

Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, Bateman A, Finn RD, Petrov AI. 2018. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* 46(D1):D335–D342.

Kröger C, Colgan A, Srikumar S, Händler K, Sivasankaran SK, Hammarlöf DL, Canals R, Grissom JE, Conway T, Hokamp K, et al. 2013. An infection-relevant transcriptomic compendium for *Salmonella enterica* Serovar Typhimurium. *Cell Host Microbe.* 14(6):683–695.

Lindgreen S, Umu SU, Lai AS-W, Eldai H, Liu W, McGimpsey S, Wheeler NE, Biggs PJ, Thomson NR, Barquist L, et al. 2014. Robust identification of noncoding RNA from transcriptomes requires phylogenetically-informed sampling. *PLoS Comput Biol.* 10(10):e1003907.

Miyakoshi M, Chao Y, Vogel J. 2015. Regulatory small RNAs from the 3' regions of bacterial mRNAs. *Curr Opin Microbiol.* 24:132–139.

Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics.* 29(22):2933–2935.

Nuss AM, Heroven AK, Waldmann B, Reinkensmeier J, Jarek M, Beckstette M, Dersch P. 2015. Transcriptomic profiling of *Yersinia pseudotuberculosis* reveals reprogramming of the Crp regulon by temperature and uncovers Crp as a master regulator of small RNAs. *PLoS Genet.* 11(3):e1005087.

Ochman H, Davalos LM. 2006. The nature and dynamics of bacterial genomes. *Science* 311(5768):1730–1733.

Peer A, Margalit H. 2014. Evolutionary patterns of *Escherichia coli* small RNAs and their regulatory interactions. *RNA* 20(7):994–1003.

Raghavan R, Groisman EA, Ochman H. 2011. Genome-wide detection of novel regulatory RNAs in *E. coli*. *Genome Res.* 21(9):1487–1497.

Raghavan R, Kacharia FR, Millar JA, Sislak CD, Ochman H. 2015. Genome rearrangements can make and break small RNA genes. *Genome Biol Evol*. 7(2):557–566.

Skippington E, Ragan MA. 2012. Evolutionary dynamics of small RNAs in 27 *Escherichia coli* and *Shigella* genomes. *Genome Biol Evol*. 4(3):330–345.

Updegrove TB, Shabalina SA, Storz G. 2015. How do base-pairing small RNAs evolve? *FEMS Microbiol Rev*. 39(3):379–391.

Wachter S, Raghavan R, Wachter J, Minnick MF. 2018. Identification of novel MITEs (miniature inverted-repeat transposable elements) in *Coxiella burnetii*: implications for protein and small RNA evolution. *BMC Genomics*. 19(1):247.

Wright AP, Dutcher HA, Butler B, Nice TJ, Raghavan R. 2021. A small RNA is functional in *Escherichia fergusonii* despite containing a large insertion. *Microbiology*. 167(10): 001099.

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(1):11–22.

Zheng M, Aslund F, Storz G. 1998. Activation of the OxyR transcription factor by reversible disulfide bond formation. *Science* 279(5357):1718–1721.