Biology Faculty Publications and Presentations   Biology

4-2018

Origin, Evolution, and Loss of Bacterial Small RNAs

H. Auguste Dutcher
Portland State University

Rahul Raghavan
Portland State University, rahul.raghavan@pdx.edu

Citation Details

Origin, evolution, and loss of bacterial small RNAs

H. Auguste Dutcher[1] and **Rahul Raghavan**[1,*]

[1]Department of Biology and Center for Life in Extreme Environments, Portland State University, Portland, Oregon, 97201, USA

Abstract

Despite the central role of bacterial non-coding small RNAs (sRNAs) in posttranscriptional regulation, little is understood about their evolution. Here, we compile what has been studied to date and trace a life cycle of sRNAs—from their mechanisms of emergence, through processes of change and frequent neofunctionalization, to their loss from bacterial lineages. Because they possess relatively unrestrictive structural requirements, we find that sRNA origins are varied, and include *de novo* emergence as well as formation from pre-existing genetic elements via duplication events and horizontal gene transfer. The need for only partial complementarity to their mRNA targets facilitates apparent rapid change, which also contributes to significant challenges in tracing sRNAs across broad evolutionary distances. We document that recently-emerged sRNAs in particular evolve quickly, mirroring dynamics observed in microRNAs, their functional analogs in eukaryotes. Mutations in mRNA-binding regions, transcriptional regulator or sigma factor binding sites, and protein-binding regions are all likely sources of shifting regulatory roles of sRNAs. Finally, using examples from the few evolutionary studies available, we examine cases of sRNA loss, and describe how these may be the result of adaptive in addition to neutral processes. We highlight the need for more comprehensive analyses of sRNA evolutionary patterns as a means to improve novel sRNA detection, enhance genome annotation, and deepen our understanding of regulatory networks in bacteria.

INTRODUCTION

As our understanding of the transcriptional landscape of bacteria continues to expand, it has become clear that non-coding small RNAs (sRNAs) play a pivotal regulatory role (1–3). Typically 50–400 nucleotides in length, sRNAs posttranscriptionally regulate gene expression, usually by base-pairing with one or more messenger RNA (mRNA) targets (4). sRNAs likely provide certain advantages over protein regulators because they act quickly, are relatively metabolically inexpensive, and provide an additional way to respond to environmental signals (1). Beyond these basic characteristics, however, the roles of bacterial sRNAs are extremely diverse: they are capable of upregulating or downregulating translation, stabilizing mRNAs or targeting them for degradation, sharing targets and/or targeting multiple mRNAs. Variability in their sequence, structure, and how and when they are transcribed allow them to meet a wide range of nuanced regulatory needs based on the diverse environments to which bacteria must adapt.

[*]Author for correspondence: Rahul Raghavan | Phone: 503-725-3872 | rahul.raghavan@pdx.edu.

Despite their regulatory importance, we lack a full understanding of how sRNAs originate and evolve. This topic was recently reviewed with particular attention to the feature requirements that shape sRNA evolution (5). These requirements typically include an environmentally-regulated promoter, a rho-independent terminator, double-stranded regions that allow for stable secondary structure, and an unstructured seed region, where the sRNA base-pairs with its target (5). These features do impose limitations on sequence evolution, but the marked divergence observed among sRNAs also suggests that these constraints are not so limiting as to prevent rapid and significant change (2).

Indeed, the fast pace at which sRNA sequences change is one of several factors that make systematic studies of bacterial sRNA evolution particularly challenging. High degrees of both intra- and inter-species polymorphism yield low sequence similarity, especially as compared to protein-coding genes, and sRNAs on the whole are known to be poorly conserved across broad evolutionary distances (6). Comparing the contents of Rfam (7)—a database that houses sequence and structural information for known sRNA families—to Pfam (8), an analogous database for protein families, only 60% of RNA families were found to be conserved between species of the same taxonomic family, as opposed to 90% for proteins (9). In evolutionary analyses, where a clear picture of relationships between gene sequences through time is critical in order to draw meaningful conclusions, this presents the unique challenge of distinguishing between an sRNA's absence in a given lineage and the possibility that the sequence or structure of the gene has simply diverged too drastically to accurately trace its ancestry.

Not only do sRNA genes evolve rapidly; they also are arguably more difficult to identify than protein-coding genes. Whereas open reading frames (ORFs) serve as signposts for coding sequences, sRNAs have no such singular reliable identifier and, in fact, many of the features they do require are found readily throughout bacterial genomes. In *E. coli,* for example, sigma-70 promoter-like sequences are widely distributed and highly redundant (10); more broadly, secondary structure alone is not enough to reliably identify sRNAs (11), and algorithms that predict non-coding RNAs often have high false-positive rates (12). Moreover, without codons as guides, homologous sRNAs are notoriously difficult to align (13), presenting obstacles to accurate quantification of their sequence evolution.

sRNAs are known to play key roles in regulatory networks (2, 3), and so are crucial to our understanding of the evolution of metabolic pathways, environmental adaptation, virulence, speciation, and countless other processes in bacteria. Furthermore, their evolutionary trajectories are inextricably linked with that of their targets, and as such, our understanding of the evolution of protein-coding genes is not complete without inquiries into that of these key regulators. Given the existence of their regulatory analogs in eukaryotes and archaea, insights into bacterial sRNA evolution could also lead to a better understanding of gene regulation in all domains of life. Finally, since posttranscriptional controls are now understood to be more than a peripheral consideration when it comes to regulating gene expression—rather, they are a key part of it—successful identification of sRNAs in bacterial genomes has become all the more critical. A better understanding of sRNA evolution could lead to improved methodologies for identifying novel sRNAs and facilitate enhanced annotation of bacterial genomes to reflect these key genetic components.

With these goals in mind, this review explores the evolutionary considerations associated with sRNA origin, functional divergence, and loss from bacterial genomes. While many of the dynamics discussed here are broadly applicable to non-coding RNAs in general, this article will focus on *trans*-acting sRNAs, which are most notably characterized by their requirement for only partial mRNA complementarity and their being encoded in genetic loci often distant from that of their targets. Other works, including articles that accompany this one, provide ample examples of sRNAs and their full range of diverse characteristics and functions. Consequently, we will focus less on detailing the features of known sRNAs, and instead turn our attention to broad evolutionary patterns and the resulting questions that emerge. Tracing a 'life cycle' of sRNA evolution, we will begin by documenting their origins, both *de novo* and from pre-existing genetic elements, detail what is known about their functional evolution, and end by discussing the dynamics associated with sRNA loss from bacterial genomes.

sRNA ORIGINS

Despite the challenges associated with identifying sRNAs and tracking them over evolutionary time, several mechanisms for their emergence have been investigated. It is possible that sRNA genes have, on the whole, more diverse origins as compared to protein-coding genes, for which duplication events and horizontal gene transfer (HGT) are considered to be the dominant forces (14, 15), though *de novo* origination of coding sequences is becoming increasingly well studied (16, 17). For microRNAs (miRNAs), which are the functional analogs of sRNAs in eukaryotes, it was recently shown that the requirements for formation are somewhat less stringent than protein-coding genes in part due to their small size: protein-coding genes are significantly longer than miRNA genes and require open reading frames in order to be functional, whereas the hairpins that form potential miRNAs are pervasive in mammalian transcriptomes (18). Similarly, provided that certain structural requirements are met, a bacterial sRNA requires only a short seed sequence in order to acquire a viable mRNA target (19, 20). Particularly since the seed sequence need only be partially complementary to the target mRNA, the existence (or emergence by chance) of such a target is likely (21). Thus, it is highly likely that functional sRNAs may emerge more readily and evolve more quickly than their protein-coding counterparts.

Documented bacterial sRNA origins include duplication events, HGT, and *de novo* emergence (Fig. 1), mirroring known eukaryotic gene origins (22), but altogether less well investigated to date. The mechanisms briefly discussed below, though by no means entirely discrete, raise new questions regarding the selection pressures that result in sRNA emergence. Additionally, gene origination source has been shown to correlate with trends in network integration and functional gain in eukaryotes. For instance, in *Saccharomyces*, genes that originated *de novo* gained function more quickly, were more likely to reflect responses to environmental changes, and formed new interactions faster than genes that arose via duplication events (23). Hence, the characterization of sRNAs by their evolutionary origins may help illuminate patterns in their highly varied interactions and expression levels within bacterial genomes.

Emergence of new sRNAs through duplication

Since duplication events are a major driver behind protein origination and evolution, it comes as no surprise that homologous sRNAs, with one or more copies thought to have arisen via gene duplication, are relatively common in bacterial genomes. There are several examples of bacterial sRNAs that have persisted as closely-related but discrete copies (24), including PrrF1 and PrrF2 in *Pseudomonas*, which regulate iron metabolism via the ferric uptake regulator (Fur) repressor (25), the functionally analogous RyhB1 and RyhB2 in *Salmonella* (26), a suite of Qrr sRNAs in *Vibrio* that control quorum-sensing (27), and AbcR1 and AbcR2 in Rhizobiales, involved in ABC transporter regulation (28–30). Duplicated sRNAs exhibit varying degrees of functional overlap. Some, such as OmrA and OmrB in *E. coli*, show apparent target redundancy (31); others appear to have significantly diverged since their emergence—either in sequence, target, regulation, or a combination of these characteristics (5). The evolutionary considerations regarding sRNA functional divergence will be discussed later in this chapter.

For those sRNA duplicates that share mRNA targets, an interesting question emerges regarding the utility of such redundancy (24, 32), and thus the selection to maintain it. It has been suggested that multi-copy sRNAs enable more efficient or nuanced regulatory control (27, 32), or ensure proper regulation in the case of particularly critical functions (24). Redundancy also enables differential regulation of the sRNAs themselves (2), perhaps allowing their regulatory cascades to be activated under an additional set of environmental conditions (24). This scenario is easy to envision, as a small sequence change in the course of an sRNA's rapid evolution might allow it to be regulated by a different transcription factor (5).

Acquisition of new sRNAs via HGT

Another common source of sRNAs in bacterial genomes is HGT via plasmids, transposable elements, and bacteriophages. Some of the most well-studied plasmid-derived sRNAs are associated with virulence, and include the anti-toxins of many toxin-antitoxin systems (33) and the island-encoded sRNAs in *Salmonella* Typhimurium and *Staphylococcus aureus* (26, 34–36). For example, the HGT-derived sRNA IsrJ in *Salmonella* enables the use of a Type III secretion system to invade intestinal epithelial cells (26), and IsrM likewise regulates proteins required for intracellular invasion and replication (35). Prophage-origin IsrK is also encoded in a *Salmonella* pathogenicity island, and is notable as an example of dual-function RNAs (37), which are the subject of an accompanying review. The short IsrK isoform activates the longer isoform, an inactive mRNA, ultimately causing expression of anti-terminator protein AntQ. The dual-function phenomenon warrants further study in the context of sRNA evolution; for example, did mRNAs and sRNAs with shared genetic loci co-emerge? Given that 'overprinting' is relatively common in viral genomes (38), might there be a relationship between viral origination of an sRNA gene and the likelihood that it is associated with an overlapping mRNA? Other evolutionary implications of dual functionality will be discussed later in this chapter.

In addition to the now well-documented phenomenon of sRNAs encoded on pathogenicity islands, there may also be structural considerations related to the likelihood of sRNA

emergence from bacteriophages or transposable elements. Two sRNA genes in *Coxiella burnetii*, the causative agent of Q fever, were found to be encoded by families of a transposon class called miniature inverted-repeat transposable elements (MITEs) (Mike Minnick, personal communication). Transposons, including MITES in particular, are a known source of microRNAs in eukaryotes (39). Given that these elements are known to readily form stable stem-loop structures (40), it is possible that MITEs along with other types of transposons are particularly suited for sRNA origination.

De novo origination of sRNAs

As more bacterial sRNAs are identified and characterized, and their origins elucidated, *de novo* origination from apparently nonfunctional transcripts has also been observed. Genome rearrangements, such as those that occur via homologous recombination, can form new intergenic regions (IGRs), from which new sRNAs can emerge (41). A survey of novel IGRs in *E. coli* and *Salmonella* Typhimurium identified one such rearrangement in the latter species that resulted in the origination of the SesR2 sRNA (41). The formation of this IGR is believed to have been mediated by horizontal gene transfer, as evidenced by plasmid-derived flanking genes and a nearby prophage gene. The sRNA encoded in that IGR likely arose via a point mutation that created a functional sigma-70 promoter region (Fig. 1), which explains the significant transcription observed in *Salmonella,* but not in the corresponding regions in *E. coli.* This example highlights the non-discrete nature of sRNA origination sources, in that the emergence of SesR2 was ultimately facilitated both by HGT and *de novo* promoter emergence. For the purposes of this discussion, we consider it to have originated *de novo* given that the horizontally-transferred IGR from which it was born was presumed to be nonfunctional prior to the point mutation that created a promoter region.

The same study that identified SesR2's origins found that an *E. coli*-specific IGR gave rise to the sRNA EcsR2 (41), and this sRNA was later determined to have originated from a degraded bacteriophage gene (21). EcsR2 is also believed to have originated *de novo* via promoter-like sequence emergence, and exhibits low expression compared to older sRNAs in the same species. In reconstructing the likely evolution of EcsR2, it appears that its mRNA-binding region has become less structured and its terminator more structured over time, suggesting that secondary structure changes are a key component of bacterial sRNA evolution, as has been observed in eukaryotes (42) and suggested in bacteria (43). Though EcsR2 is one of only a handful of bacterial sRNAs for which a source has been determined, the phenomena of noncoding genes emerging *de novo* and originating from selfish genetic elements have been observed in eukaryotes (22, 42, 44, 45). Further comparative genomics and transcriptomics studies in bacteria are required in order to identify all modes of sRNA origination.

GAIN OF FUNCTION AND FUNCTIONAL DIVERGENCE

Although there are few comprehensive surveys of related sRNAs, with most focusing on well-studied enteric bacteria (21, 32, 46), there is evidence that newly-born sRNAs evolve quickly (21). This, coupled with the observation that constraints on sRNA evolution appear to be loose enough to accommodate rapid structure and sequence change (2), point to

significant functional divergence in relatively short spans of evolutionary time. This corroborates the most likely evolutionary scenario of bacterial sRNAs: emergence of the initial sRNA-mRNA interaction induces selection to maintain the sRNA, followed by the accumulation of additional mRNA binding sites (21, Fig. 1). Of course, this initial selection to maintain the sRNA is predicated on the beneficial nature of the interaction. The same short-seed sequence and partial-complementarity requirements that promote the emergence of functional sRNAs also create many opportunities for deleterious effects. In a recent study, Kacharia et al. showed that recently emerged sRNAs in *E. coli* and *Salmonella* were expressed at significantly lower levels than older sRNAs, indicating that low expression of incipient sRNAs might help to mitigate the risk of any fitness costs while maintaining the sRNA long enough for beneficial interactions to arise (21). This pattern of low expression coupled with rapid evolution has been observed in young miRNAs as well (18, 47, 48), suggesting a possible mechanism by which newly-emerged non-coding RNAs can gain function and be successfully integrated into regulatory networks.

A high rate of evolution among sRNAs genes would help to explain the now well-documented observation that even closely-related sRNAs often exhibit distinct characteristics in expression patterns, targets, and transcriptional regulation (5, 24). For example, the sRNA GcvB is expressed at high levels during exponential phase and low levels during stationary phase in *Salmonella*; in *Vibrio*, this expression pattern is reversed (49). Given that the abundance of mRNA targets as well as the consequence of mRNA regulation will be different based on the organism's stage of growth, it is highly likely that differential expression patterns indicate diverged sRNA function. This is supported by the observation that differences in gene expression affect bacterial fitness and adaptation during exposure to stress (50).

Homologous sRNAs are likewise known to be involved in different regulatory circuits; for example, RyhB regulates the expression of different genes in *Vibrio* as compared to its homolog in *E. coli* (51). Perhaps the starkest examples of divergence come from examination of intra-organismal homologs, which show varying levels of functional redundancy (24). One such example is GlmY and GlmZ in *Enterobacteriaceae*, which act hierarchically to regulate the enzyme GlmS as part of the cell wall synthesis pathway (52). Despite sharing 66% sequence identity in *E. coli* along with extremely similar secondary structures, only GlmZ has a high affinity for Hfq, allowing it to directly activate *glmS*. However, due to their structural similarity, GlmY is able to act as a decoy for the endonuclease-recruiting adapter protein RapZ, reducing the chances of GlmZ degradation and thereby indirectly regulating *glmS* (53).

In addition to being an interesting example of the regulatory complexity that can result from sRNA gene duplication and subsequent divergence, GlmY/GlmZ also underscore the importance of examining sRNA functional divergence in the context of the other regulatory elements, such as proteins, with which they interact. Understanding the nuances of protein binding may aid in our understanding of highly specific RNA degradation mechanisms (53), which in turn could shed light on how such closely-related sRNAs come to occupy such distinct regulatory niches. Interestingly, GlmY/GlmZ is also an example of differential transcriptional control, in that this pair of sRNA genes is transcribed via different

combinations of sigma-54 and sigma-70 promoter sequences in closely related species: sigma-54 promoter for *glmY* and *glmZ* in *Yersinia tuberculosis*, overlapping sigma-54 and sigma-70 promoters for *glmY* and *glmZ* in *Salmonella* Typhimurium and *glmY* in *E. coli*, and sigma-70 promoters for *glmZ* in *E. coli* (52). This observation identifies the evolution of transcription factor or sigma factor binding site as yet another means by which the regulatory roles of sRNAs can diverge.

Conversely, examining sRNAs across large phylogenetic distances has yielded some likely examples of parallel evolution, in which discrete sRNAs have independently evolved nearly identical functions. PrrF1 and PrrF2 in *Pseudomonas aeruginosa* and RyhB in *E. coli* provide one such example. PrrF sRNAs and RyhB are both regulated by the Fur repressor and thus transcribed when iron concentrations are low, and all act by inhibiting ribosome binding on an overlapping set of mRNA targets (5, 25, 54). Similarly, MicA in *E. coli* and VrrA in *Vibrio cholerae* both regulate OmpA (outer membrane protein A), although Hfq appears to be required for MicA function but not for VrrA (5, 55, 56). The lack of sequence similarity between functional analogs in each of these cases points to likely convergent evolution; however, we cannot rule out the possibility that these sRNAs shared a common ancestor but have diverged so significantly that this relationship is obscured.

Despite ample anecdotal evidence for diversification of sRNA expression and function, the mechanisms by which these changes take place have yet to be systematically investigated. Because an sRNA requires only short, partial complementarity to its mRNA target in order to be functional, it seems likely that both target gain and target loss—perhaps especially the former—could occur fairly readily via point mutations. *E. coli* sRNAs, for example, appear to have gradually accumulated mRNA binding sites since their emergence (46). Aside from binding sites, other mutations within sRNA genes could increase sRNA stability, indirectly allowing for acquisition of new targets by prolonged sRNA survival into a different stage of growth where alternative mRNAs are available (Fig. 1). A study examining the relationships between sRNA structure, function, and stability found that mutations in structural elements could result in increased sRNA abundance and induce regulation of non-specific targets (43). The same study found that in sRNAs with short seed sequences, mutations in this region could be used to fine-tune an sRNA's effect on its target, whereas sRNAs with longer seed sequences were less susceptible to these mutation-induced changes. This finding begs a more thorough investigation into the relationship between seed sequence characteristics and evolutionary patterns among sRNAs. Additionally, given that mRNA-binding sequences are the most conserved regions of sRNAs (19, 21, Fig. 2), this characteristic can be exploited to identify potential seed sequences in novel sRNAs and aid in mRNA target prediction (21).

Just as they promote target gain, sequence changes can also catalyze target loss. The loss of mRNA targets in turn could cause complete loss of sRNA function, however, it is difficult to clearly draw the line between functional loss and functional divergence unless a given sRNA is unequivocally deemed nonfunctional. For example, sequence changes were the probable cause for the loss of original function observed in two Pxr paralogs, which typically regulate fruiting body development in *Cystobacter* species (57), but whether the Pxr sRNAs in question have functionally diverged or become nonfunctional entirely is unclear. This question is further complicated by the fact that sRNA pseudogenes may be more likely than

most protein-coding pseudogenes to reacquire function via the mechanisms previously described. Not constrained by the requirement of an ORF, even 'decommissioned' sRNA genes may not permanently remain so.

Since the initial discovery of bacterial sRNAs, increasing numbers of what we may consider non-canonical regulatory mechanisms have been elucidated, which highlights the varied means by which sRNAs can gain (or regain) function. For example, the *E. coli* sRNAs SgrS and DicF were recently found to bind the mannose transporter *manX* mRNA target outside regions that would allow for direct translational repression, with Hfq instead blocking the ribosomal binding site (RBS) (58). Evolutionary analysis suggests that the SgrS-*manX* mRNA interaction—facilitated by a relatively less well-conserved region of SgrS and a coding region of the *manX* mRNA—evolved much later than the sRNA's association with the sugar transporter *ptsG* mRNA, whose RBS binds with the most highly conserved region of SgrS (46). This illustrates the capacity of sRNAs to gain targets over evolutionary time. In another example, Qrr3, one of the quorum-sensing sRNAs in *Vibrio* species, has four known targets and can bind mRNAs in two regions, which in turn allows the sRNA to act via four discrete mechanisms (59). mRNA targets that bind the first stem-loop of Qrr3 induce degradation of the sRNA, resulting in either mRNA activation or repression; targets that bind the second stem-loop sequester Qrr3, with downstream effects dependent in part on binding strength, which in turn is affected by even small sequence changes. While there has not been an analogous analysis of Qrr evolution, it seems likely that only one stem-loop structure was critical for an ancestral Qrr, with additional functions emerging later via sequence mutations. The evolutionary implications of sRNA mechanism diversity and divergence warrant further investigation.

Clearly, the evolutionary trajectories of sRNAs must be considered along with those of the other genetic elements with which they interact, including sRNA-binding proteins, target mRNAs, and 'sponge' sRNAs (2). The evolutionary relationships between sRNAs and their mRNA targets were the subject of a survey of the relatively well-studied sRNAs in enteric bacteria (46), which linked the accumulation of new sRNAs in *Enterobacteriales* with the evolution of a longer C-terminal region in Hfq, a protein known to bind and stabilize many sRNAs in this lineage. This is significant given that the C-terminal region of Hfq, which differs between Gram-negative and Gram-positive bacteria, likely modulates access to ribosomal binding sites (60). While Hfq is the most well-known, it is not the only sRNA-binding protein that may be critical to sRNA regulatory networks (4). For example, RapZ protein in *E. coli* acts as an adapter, recruiting an endoribonuclease and thereby inducing sRNA degradation (53). Expression of ProQ, a protein that regulates osmolarity via interactions with the ProP transporter, significantly affects sRNA abundance in *Salmonella*, and has been shown to aid in translation inhibition both by increasing sRNA stability and by blocking ribosome binding on target mRNAs (61, 62). Moreover, potential coevolutionary considerations are not limited to sRNA-binding proteins: a recent comparison of *Listeria* genomes identified 12 sRNAs that appear to have coevolved with protein-coding genes linked to virulence (63). Further study is needed in order to better understand these coevolutionary dynamics, but their import is becoming increasingly clear.

MODES AND CONSEQUENCES OF sRNA LOSS

From the few studies that have examined sRNA evolution more broadly across species, strains, or other closely-related taxa, it appears that sRNA loss in a given lineage is a fairly common, but understudied, phenomenon. It is worth noting here that given the aforementioned relative ease with which sRNAs should be able to emerge in bacteria, and the high likelihood that their initial interactions will be deleterious, sRNAs are very likely born and lost at a rapid rate. The 'death rate' of microRNAs in eukaryotes has been quantified by comparing an estimated birth rate with the number of persisting miRNAs (18); for bacterial sRNAs, no such estimate is feasible, as the rate of sRNA emergence has yet to be quantified. The discussion of loss here will thus be limited to those sRNAs determined through phylogenetic analysis to be ancestral to a given lineage, but lost in one or more clades, and the considerations that come to bear on selective pressures to retain or lose sRNA genes.

One such consideration stems from the aforementioned phenomenon of dual-function sRNAs. Logically, the encoding of sRNAs and mRNAs at intersecting loci should affect the retention and evolution of these genetic elements in the genome. For example, TnpA, a transposable element-derived sRNA in *Salmonella*, shares its promoter with an mRNA, which may have resulted in increased selective pressure to maintain the transposon within which it is encoded (64). It stands to reason that even if its target mRNA is lost, the sRNA may persist long enough to acquire a new target. Additionally, when it comes to sequence diversity, we would expect the evolution of dual-encoded sRNAs to be more restricted as compared to regions that encode a single RNA, but this question has yet to be systematically investigated. These dynamics regarding conservation and retention would also apply to sRNAs that are the product of mRNA processing of $3'$ untranslated regions (UTRs) and the so-called intraRNAs whose promoters are found within protein-coding genes (65–67).

Despite the varied selective pressures that may act to maintain them, sRNA loss has been documented in groups of highly related, well-studied organisms such as the *E. coli-Shigella* clade. Pathogenic strains of *E. coli,* categorized as *Shigella* and enteroinvasive *E. coli* (EIEC), are believed to have evolved multiple times from non-pathogenic ancestors via the acquisition of a large virulence plasmid, pINV (68–70). The transition to an intracellular, pathogenic lifestyle was accompanied by loss of protein-coding genes from the chromosomes of *Shigella* and EIEC strains (71, 72). For example, the independent loss of *cadA* from multiple *Shigella* species is believed to have increased the virulence of these pathogens (72). This suggests that gene loss could be an adaptive process in which beneficial pseudogenizations or deletions are the product of positive selection rather than a neutral process associated with relaxed selection and genetic drift typically observed among organisms with a newly-acquired intracellular lifestyle (73). The mechanism by which gene loss contributes to bacterial fitness is thought to be through the 'rewiring' of regulatory networks, perhaps resetting metabolic regulation in order to develop a system more suited to the new environment (66, Fig. 1). Even in cases when existing genes would hypothetically provide a fitness advantage in the new environment, existing regulatory network structure might limit the ability of those genes to confer the advantage (74). Given that sRNAs play a key role in these regulatory networks, it follows that sRNA gene loss could likewise be part

of this rewiring, and thus an adaptive process in certain scenarios. There is possible evidence of this phenomenon in the *E. coli-Shigella* clade, where approximately one-third of sRNAs were found to have variable distribution across strains, with sRNA loss observed to be most frequent in *Shigella* and *E. coli* strains with a restricted host range (32).

Due to the paucity of studies on the topic, it is difficult to draw many broad conclusions about sRNA loss within particular bacterial lineages, though other discrete examples of absence could perhaps be explained using an adaptive framework. For example, the IGR encoding EcsR1 in *E. coli* is absent in *Salmonella* due to a genome rearrangement event that fragmented the sRNA gene and caused it to be effectively lost from the bacterium (41). EcsR1 was shown to inhibit biofilm formation, a function that would likely not be advantageous to a pathogen like *Salmonella*, and it is therefore plausible to consider that perhaps this sRNA was lost due to positive selection. Although ectopic expression of EscsR1 in *Salmonella* did not cause a significant difference in biofilm formation, it was correlated with increased expression of genes associated with pathogenicity, which could ultimately be disruptive to the highly specialized virulence-related regulatory networks *Salmonella* has evolved since its split from *E. coli* (41). McaS, another sRNA that regulates biofilm formation in *E. coli*, is also missing from *Salmonella* and a few other enteric bacteria, but the evolutionary significance of its absence is not known (75, 76).

CONCLUDING REMARKS

There is clearly much left to be investigated regarding the origination, evolution, and loss of bacterial sRNAs; however, some patterns emerge from what has been studied to date. sRNAs arise from diverse origins and appear to evolve quickly, acquiring targets and integrating into regulatory circuits via mutations that form regulatory motifs, seed sequences, and protein-binding regions. They share broad structural characteristics, but are varied in their expression patterns, distribution across bacterial lineages, and network interactions. In particular, sRNAs of different ages or at different stages of network integration are likely regulated differently and/or expressed at different levels. On the whole, sRNAs likely evolve faster than protein-coding genes, but a majority are still under purifying selection (21), indicative of their critical role in bacterial fitness. Similarly, at least a few cases of sRNA loss or pseudogenization is likely adaptive, as opposed to being a consequence of genome reduction or neutral processes.

Thorough, systematic investigation of these dynamics is necessary if we are to fully understand sRNA evolution. This includes continued improvement of our ability to identify and characterize novel sRNAs—particularly in groups other than well-studied enteric bacteria—and enhanced capacities to detect homology across species and genera. Furthermore, it is essential that we examine the evolution of sRNAs along with that of their targets. The observation that seed sequences change at lower rates than other sRNA gene regions can be put to work in batch identification of such sequences, helping both to trace the evolution of sRNA-mRNA interactions and to discern sRNA functions and network interactions. Drawing guidance from the study of eukaryotic microRNAs, it would be useful to more broadly characterize sRNA emergence and loss dynamics in a large number of bacterial genomes, enabling us to quantify the rates at which they emerge, evolve and

persist. Studying sRNAs in the context of these evolutionary forces will help us better understand not only sRNA emergence and change, but also how these elements fit in the larger context of bacterial posttranscriptional regulation and adaptation to environmental stress.

References

1. Beisel CL, Storz G. Base pairing small RNAs and their roles in global regulatory networks. FEMS Microbiol Rev. 2010; 34:866–882. [PubMed: 20662934]

2. Gottesman S, Storz G. Bacterial small RNA regulators: Versatile roles and rapidly evolving variations. Cold Spring Harb Perspect Biol. 2011; 3:a003798. [PubMed: 20980440]

3. Nitzan M, Rehani R, Margalit H. Integration of Bacterial Small RNAs in Regulatory Networks. Annu Rev Biophys. 2017; 46:131–48. [PubMed: 28532217]

4. Storz G, Vogel J, Wassarman KM. Regulation by Small RNAs in Bacteria: Expanding Frontiers. Mol Cell. 2011; 43:880–891. [PubMed: 21925377]

5. Updegrove TB, Shabalina SA, Storz G. How do base-pairing small RNAs evolve? FEMS Microbiol Rev. 2015; 39:379–91. [PubMed: 25934120]

6. Hoeppner MP, Gardner PP, Poole AM. Comparative Analysis of RNA Families Reveals Distinct Repertoires for Each Domain of Life. PLoS Comput Biol. 2012; 8:1–8. [PubMed: 22629235]

7. Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, Floden EW, Gardner PP, Jones TA, Tate J, Finn RD. Rfam 12.0: updates to the RNA families database. Nucleic Acids Res. 2015; 43:D130–D137. [PubMed: 25392425]

8. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016; 44:D279–D285. [PubMed: 26673716]

9. Lindgreen S, Umu SU, Lai ASW, Eldai H, Liu W, McGimpsey S, Wheeler NE, Biggs PJ, Thomson NR, Barquist L, Poole AM, Gardner PP. Robust Identification of Noncoding RNA from Transcriptomes Requires Phylogenetically-Informed Sampling. PLoS Comput Biol. 2014; 10:e1003907. [PubMed: 25357249]

10. Huerta AM, Francino MP, Morett E, Collado-Vides J. Selection for unequal densities of sigma-70 promoter-like signals in different regions of large bacterial genomes. PLoS Genet. 2006; 2:1740–1750.

11. Rivas E, Eddy SR. Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs. Bioinformatics. 2000; 16:583–605. [PubMed: 11038329]

12. Zhang Y, Huang H, Zhang D, Qiu J, Yang J, Wang K, Zhu L, Fan J, Yang J. A Review on Recent Computational Methods for Predicting Noncoding RNAs. Biomed Res Int. 2017; 2017:1–14.

13. Gardner PP, Wilm A, Washietl S. A benchmark of multiple sequence alignment programs upon structural RNAs. Nucleic Acids Res. 2005; 33:2433–2439. [PubMed: 15860779]

14. Zhang J. Evolution by gene duplication: an update. Trends Ecol Evol. 2003; 18:292–298.

15. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. Nature. 2000; 405:299–304. [PubMed: 10830951]

16. Schlötterer C. Genes from scratch—the evolutionary fate of de novo genes. Trends Genet. 2015; 31:215–9. [PubMed: 25773713]

17. Tautz D, Domazet-Lošo T. The evolutionary origin of orphan genes. Nat Rev Genet. 2011; 12:692–702. [PubMed: 21878963]

18. Lyu Y, Shen Y, Li H, Chen Y, Guo L, Zhao Y, Hungate E, Shi S, Wu C-I, Tang T. New MicroRNAs in Drosophila—Birth, Death and Cycles of Adaptive Evolution. PLoS Genet. 2014; 10:e1004096. [PubMed: 24465220]

19. Peer A, Margalit H. Accessibility and evolutionary conservation mark bacterial small-rna target-binding regions. J Bacteriol. 2011; 193:1690–701. [PubMed: 21278294]

20. Richter AS, Backofen R. General features of bacterial small RNA-mRNA interactions? RNA Biol. 2012; 9:954–965. [PubMed: 22767260]

21. Kacharia FR, Millar JA, Raghavan R. Emergence of New sRNAs in Enteric Bacteria is Associated with Low Expression and Rapid Evolution. J Mol Evol. 2017; 84:204–213. [PubMed: 28405712]

22. Kaessmann H. Origins, evolution, and phenotypic impact of new genes. Genome Res. 2010; 20:1313–26. [PubMed: 20651121]

23. Capra JA, Pollard KS, Singh M. Novel genes exhibit distinct patterns of function acquisition and network integration. Genome Biol. 2010; 11:R127. [PubMed: 21187012]

24. Caswell CC, Oglesby-Sherrouse AG, Murphy ER. Sibling rivalry: related bacterial small RNAs and their redundant and non-redundant roles. Front Cell Infect Microbiol. 2014; 4:1–13. [PubMed: 24478989]

25. Wilderman PJ, Sowa NA, Fitzgerald DJ, Fitzgerald PC, Gottesman S, Ochsner UA, Vasil ML. Identification of tandem duplicate regulatory small RNAs in Pseudomonas aeruginosa involved in iron homeostasis. Proc Natl Acad Sci USA. 2004; 101:9792–9797. [PubMed: 15210934]

26. Padalon-Brauch G, Hershberg R, Elgrably-Weiss M, Baruch K, Rosenshine I, Margalit H, Altuvia S. Small RNAs encoded within genetic islands of *Salmonella typhimurium* show host-induced expression and role in virulence. Nucleic Acids Res. 2008; 36:1913–1927. [PubMed: 18267966]

27. Lenz DH, Mok KC, Lilley BN, Kulkarni RV, Wingreen NS, Bassler BL, Way I. The Small RNA Chaperone Hfq and Multiple Small RNAs Control Quorum Sensing in *Vibrio harveyi* and *Vibrio cholerae*. Cell. 2004; 118:69–82. [PubMed: 15242645]

28. del Val C, Rivas E, Torres-Quesada O, Toro N, Jiménez-Zurdo JI. Identification of differentially expressed small non-coding RNAs in the legume endosymbiont *Sinorhizobium meliloti* by comparative genomics. Mol Microbiol. 2007; 66:1080–91. [PubMed: 17971083]

29. Wilms I, Voss B, Hess WR, Leichert LI, Narberhaus F. Small RNA-mediated control of the *Agrobacterium tumefaciens* GABA binding protein. Mol Microbiol. 2011; 80:492–506. [PubMed: 21320185]

30. Torres-Quesada O, Millán V, Nisa-Martínez R, Bardou F, Crespi M, Toro N, Jiménez-Zurdo JI. Independent Activity of the Homologous Small Regulatory RNAs AbcR1 and AbcR2 in the Legume Symbiont *Sinorhizobium meliloti*. PLoS One. 2013; 8:e68147. [PubMed: 23869210]

31. Papenfort K, Vogel J. Multiple target regulation by small noncoding RNAs rewires gene expression at the post-transcriptional level. Res Microbiol. 2009; 160:278–287. [PubMed: 19366629]

32. Skippington E, Ragan MA. Evolutionary dynamics of small RNAs in 27 *Escherichia coli* and *Shigella* genomes. Genome Biol Evol. 2012; 4:330–345. [PubMed: 22223756]

33. Brantl S, Jahn N. sRNAs in bacterial type I and type III toxin-antitoxin systems. FEMS Microbiol Rev. 2015; 39:413–427. [PubMed: 25808661]

34. Chabelskaya S, Gaillot O, Felden B. A *Staphylococcus aureus* small RNA is required for bacterial virulence and regulates the expression of an immune-evasion molecule. PLoS Pathog. 2010; 6:e1000927. [PubMed: 20532214]

35. Gong H, Vu G-P, Bai Y, Chan E, Wu R, Yang E, Liu F, Lu S. A *Salmonella* Small Non-Coding RNA Facilitates Bacterial Invasion and Intracellular Replication by Modulating the Expression of Virulence Factors. PLoS Pathog. 2011; 7:e1002120. [PubMed: 21949647]

36. Lee YH, Kim S, Helmann JD, Kim B-H, Park YK. RaoN, a small RNA encoded within *Salmonella* pathogenicity island-11, confers resistance to macrophage-induced stress. Microbiology. 2013; 159:1366–78. [PubMed: 23657681]

37. Hershko-Shalev T, Odenheimer-Bergman A, Elgrably-Weiss M, Ben-Zvi T, Govindarajan S, Seri H, Papenfort K, Vogel J, Altuvia S. Gifsy-1 Prophage IsrK with Dual Function as Small and Messenger RNA Modulates Vital Bacterial Machineries. PLoS Genet. 2016; 12:1–31.

38. Sabath N, Wagner A, Karlin D. Evolution of viral proteins originated de novo by overprinting. Mol Biol Evol. 2012; 29:3767–3780. [PubMed: 22821011]

39. Piriyapongsa J, Jordan IK. A family of human microRNA genes from miniature inverted-repeat transposable elements. PLoS One. 2007; 2:e203. [PubMed: 17301878]

40. Chen Y, Zhou F, Li G, Xu Y. A recently active miniature inverted-repeat transposable element, *Chunjie*, inserted into an operon without disturbing the operon structure in *Geobacter uraniireducens* Rf4. Genetics. 2008; 179:2291–7. [PubMed: 18660544]

41. Raghavan R, Kacharia FR, Millar JA, Sislak CD, Ochman H. Genome rearrangements can make and break small RNA genes. Genome Biol Evol. 2015; 7:557–566. [PubMed: 25601101]

42. Heinen TJAJAJ, Staubach F, Häming D, Tautz D. Emergence of a New Gene from an Intergenic Region. Curr Biol. 2009; 19:1527–1531. [PubMed: 19733073]

43. Peterman N, Lavi-Itzkovitz A, Levine E. Large-scale mapping of sequence-function relations in small regulatory RNAs reveals plasticity and modularity. Nucleic Acids Res. 2014; 42:12177–12188. [PubMed: 25262352]

44. Piriyapongsa J, Mariño-Ramírez L, Jordan IK. Origin and evolution of human microRNAs from transposable elements. Genetics. 2007; 176:1323–37. [PubMed: 17435244]

45. Smalheiser NR, Torvik VI. Mammalian microRNAs derived from genomic repeats. Trends Genet. 2005; 21:322–326. [PubMed: 15922829]

46. Peer A, Margalit H. Evolutionary patterns of *Escherichia coli* small RNAs and their regulatory interactions. RNA. 2014; 20:994–1003. [PubMed: 24865611]

47. Chen K, Rajewsky N. The evolution of gene regulation by transcription factors and microRNAs. Nat Rev Genet. 2007; 8:93–103. [PubMed: 17230196]

48. Jovelin R, Cutter AD. Microevolution of nematode miRNAs reveals diverse modes of selection. Genome Biol Evol. 2014; 6:3049–63. [PubMed: 25355809]

49. Papenfort K, Förstner KU, Cong J-P, Sharma CM, Bassler BL. Differential RNA-seq of *Vibrio cholerae* identifies the VqmR small RNA as a regulator of biofilm formation. Proc Natl Acad Sci USA. 2015; 112:E766–75. [PubMed: 25646441]

50. Otoupal PB, Erickson KE, Escalas-Bordoy A, Chatterjee A. CRISPR Perturbation of Gene Expression Alters Bacterial Fitness under Stress and Reveals Underlying Epistatic Constraints. ACS Synth Biol. 2017; 6:94–107. [PubMed: 27529436]

51. Massé E, Vanderpool CK, Gottesman S. Effect of RyhB small RNA on global iron use in *Escherichia coli*. J Bacteriol. 2005; 187:6962–71. [PubMed: 16199566]

52. Göpel Y, Lüttmann D, Heroven AK, Reichenbach B, Dersch P, Görke B. Common and divergent features in transcriptional control of the homologous small RNAs GlmY and GlmZ in *Enterobacteriaceae*. Nucleic Acids Res. 2011; 39:1294–1309. [PubMed: 20965974]

53. Göpel Y, Papenfort K, Reichenbach B, Vogel J, Görke B. Targeted decay of a regulatory small RNA by an adaptor protein for RNase E and counteraction by an anti-adaptor RNA. Genes Dev. 2013; 27:552–64. [PubMed: 23475961]

54. Massé E, Gottesman S. A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. Proc Natl Acad Sci USA. 2002; 99:4620–5. [PubMed: 11917098]

55. Udekwu KI, Darfeuille F, Vogel J, Reimegård J, Holmqvist E, Wagner EGH. Hfq-dependent regulation of OmpA synthesis is mediated by an antisense RNA. Genes Dev. 2005; 19:2355–66. [PubMed: 16204185]

56. Song T, Mika F, Lindmark B, Liu Z, Schild S, Bishop A, Zhu J, Camilli A, Johansson J, Vogel J, Wai SN. A new *Vibrio cholerae* sRNA modulates colonization and affects release of outer membrane vesicles. Mol Microbiol. 2008; 70:100–11. [PubMed: 18681937]

57. Chen I-CK, Velicer GJ, Yu Y-TN. Divergence of functional effects among bacterial sRNA paralogs. BMC Evol Biol. 2017; 17:199. [PubMed: 28830343]

58. Azam, MS., Vanderpool, CK. Translational regulation by bacterial small RNAs via an unusual Hfq-dependent mechanism; Nucleic Acids Res. 2017. p. gkx1286 https://doi.org/10.1093/nar/gkx1286

59. Feng L, Rutherford ST, Papenfort K, Bagert JD, van Kessel JC, Tirrell DA, Wingreen NS, Bassler BL. A Qrr noncoding RNA deploys four different regulatory mechanisms to optimize quorum-sensing dynamics. Cell. 2015; 160:228–40. [PubMed: 25579683]

60. Robinson KE, Orans J, Kovach AR, Link TM, Brennan RG. Mapping Hfq-RNA interaction surfaces using tryptophan fluorescence quenching. Nucleic Acids Res. 2014; 42:2736–49. [PubMed: 24288369]

61. Smirnov A, Förstner KU, Holmqvist E, Otto A, Günster R, Becher D, Reinhardt R, Vogel J. Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. Proc Natl Acad Sci USA. 2016; 113:11591–11596. [PubMed: 27671629]

62. Smirnov A, Wang C, Drewry LL, Vogel J. Molecular mechanism of mRNA repression in trans by a ProQ-dependent small RNA. EMBO J. 2017; 36:1029–1045. [PubMed: 28336682]

63. Cerutti F, Mallet L, Painset A, Hoede C, Moisan A, Bécavin C, Duval M, Dussurget O, Cossart P, Gaspin C, Chiapello H. Unraveling the evolution and coevolution of small regulatory RNAs and coding genes in *Listeria*. BMC Genomics. 2017; 18:1–15. [PubMed: 28049423]

64. Ellis MJ, Trussler RS, Charles O, Haniford DB. A transposon-derived small RNA regulates gene expression in *Salmonella* Typhimurium. Nucleic Acids Res. 2017; 45:5470–5486. [PubMed: 28335027]

65. Lybecker M, Bilusic I, Raghavan R. Pervasive transcription: detecting functional RNAs in bacteria. Transcription. 2014; 5:e944039. [PubMed: 25483405]

66. Bilusic I, Popitsch N, Rescheneder P, Schroeder R, Lybecker M. Revisiting the coding potential of the *E. coli* genome through Hfq co-immunoprecipitation. RNA Biol. 2014; 11:641–54. [PubMed: 24922322]

67. Chao Y, Papenfort K, Reinhardt R, Sharma CM, Vogel J. An atlas of Hfq-bound transcripts reveals 3′ UTRs as a genomic reservoir of regulatory small RNAs. EMBO J. 2012; 31:4005–19. [PubMed: 22922465]

68. Rolland K, Lambert-Zechovsky N, Picard B, Denamur E. *Shigella* and enteroinvasive *Escherichia coli* strains are derived from distinct ancestral strains of E. coli. Microbiology. 1998; 144:2667–2672. [PubMed: 9782516]

69. Van Den Beld MJC, Reubsaet FAG. Differentiation between *Shigella*, enteroinvasive *Escherichia coli* (EIEC) and noninvasive *Escherichia coli*. Eur J Clin Microbiol Infect Dis. 2012; 31:899–904. [PubMed: 21901636]

70. Lan R, Alles MC, Donohoe K, Marina B, Reeves PR, Martinez MB. Molecular Evolutionary Relationships of Enteroinvasive *Escherichia coli* and *Shigella* spp. Infect Immun. 2004; 72:5080–5088. [PubMed: 15322001]

71. Parsot C. *Shigella* spp. and enteroinvasive *Escherichia coli* pathogenicity factors. FEMS Microbiol Lett. 2005; 252:11–18. [PubMed: 16182469]

72. Day J, Fernández RE, Maurelli AT. Pathoadaptive mutations that enhance virulence: Genetic organization of the *cadA* regions of *Shigella* spp. Infect Immun. 2001; 69:7471–7480. [PubMed: 11705922]

73. Koskiniemi S, Sun S, Berg OG, Andersson DI. Selection-driven gene loss in bacteria. PLoS Genet. 2012; 8:1–7.

74. Hottes AK, Freddolino PL, Khare A, Donnell ZN, Liu JC, Tavazoie S. Bacterial Adaptation through Loss of Function. PLoS Genet. 2013; 9:e1003617. [PubMed: 23874220]

75. Jørgensen MG, Thomason MK, Havelund J, Valentin-hansen P, Storz G. Dual function of the McaS small RNA in controlling biofilm formation. Genes Dev. 2013; 27:1132–1145. [PubMed: 23666921]

76. Thomason MK, Fontaine F, Lay N, De Storz G. A small RNA that regulates motility and biofilm formation in response to changes in nutrient availability in *Escherichia coli*. Mol Microbiol. 2012; 84:17–35. [PubMed: 22289118]
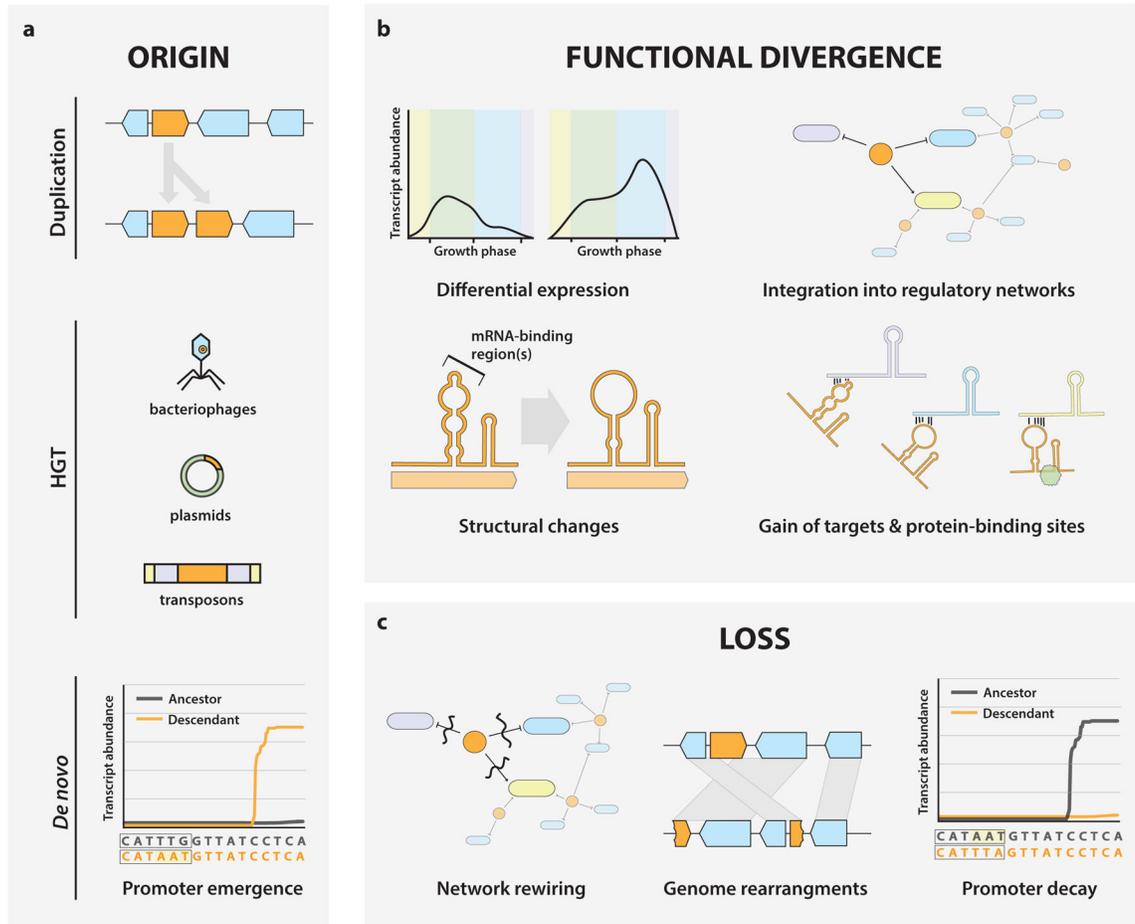
Figure 1. sRNA origin, functional divergence, and loss
(a) sRNA sources include duplication events, horizontal gene transfer, and *de novo* origination via promoter emergence. (b) Sequence and structural changes are often accompanied by differential sRNA gene expression and accumulation of mRNA targets and/or protein-binding regions, causing the sRNA to become fully integrated into regulatory networks. (c) sRNA loss occurs through mutations that erode promoter sequences, genome rearrangements that split sRNA-containing intergenic regions (IGRs), and selective pressures that prompt shifts in network interactions.
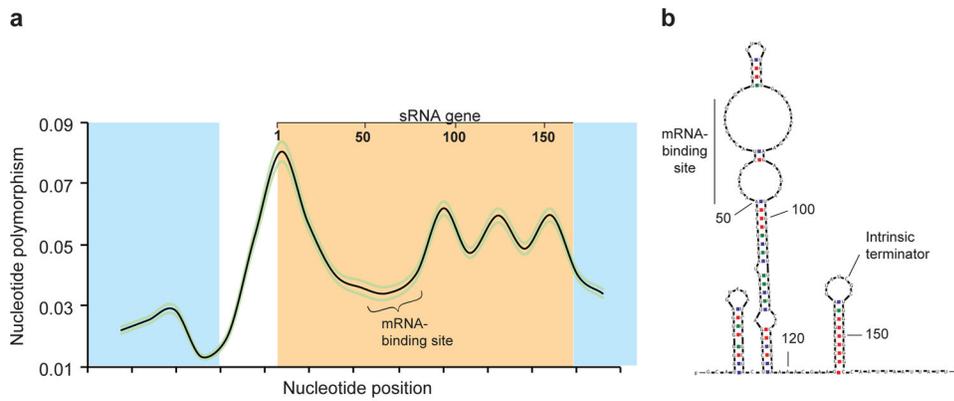
Figure 2. Sequence conservation and structure of an sRNA gene
(a) Sequence conservation within an sRNA gene (orange) and flanking protein-coding genes (blue). The black line represents nucleotide diversity index, π, calculated using a sliding window analysis; the flanking green lines indicate the 95% confidence interval. Lowest nucleotide polymorphism within sRNA genes is observed in mRNA-binding regions. (b) Predicted structure of an sRNA, showing a single-stranded mRNA-binding site and a terminator hairpin. Adapted from Kacharia et al. 2017.