



**HELMHOLTZ**

Centre for Infection Research

**Transcriptional noise and exaptation  
as sources for bacterial sRNAs.**

Item Type	Article
Authors	Jose, Bethany R; Gardner, Paul P; Barquist, Lars
Citation	Biochem Soc Trans. 2019 Apr 30;47(2):527-539. doi: 10.1042/BST20180171. Epub 2019 Mar 5.
DOI	<a href="https://doi.org/10.1042/BST20180171">10.1042/BST20180171</a>
Publisher	Portland Press
Journal	Biochemical Society Transactions
Rights	Attribution-NonCommercial-ShareAlike 4.0 International
Download date	04/08/2022 17:39:05
Item License	<a href="http://creativecommons.org/licenses/by-nc-sa/4.0/">http://creativecommons.org/licenses/by-nc-sa/4.0/</a>
Link to Item	<a href="http://hdl.handle.net/10033/621797">http://hdl.handle.net/10033/621797</a>

# Transcriptional noise and exaptation as sources for bacterial sRNAs

Bethany R Jose<sup>1,2,\*</sup>, Paul P Gardner<sup>1,2</sup>, Lars Barquist<sup>3,4,\*</sup>

keywords: sRNAs, ncRNA, bacteria, selection, transcriptional noise, de-novo evolution, exaptation

1. Biomolecular Interactions Centre, School of Biological Sciences, University of Canterbury, Christchurch, New Zealand.
2. Department of Biochemistry, Otago University, Dunedin, New Zealand.
3. Helmholtz Institute for RNA-based Infection Research (HIRI), Helmholtz-Centre for Infection Research (HZI), Würzburg, Germany.
4. Faculty of Medicine, University of Würzburg, Würzburg, Germany.

\*Correspondence to: [bethany.jose@postgrad.otago.ac.nz](mailto:bethany.jose@postgrad.otago.ac.nz),  
[lars.barquist@helmholtz-hiri.de](mailto:lars.barquist@helmholtz-hiri.de)

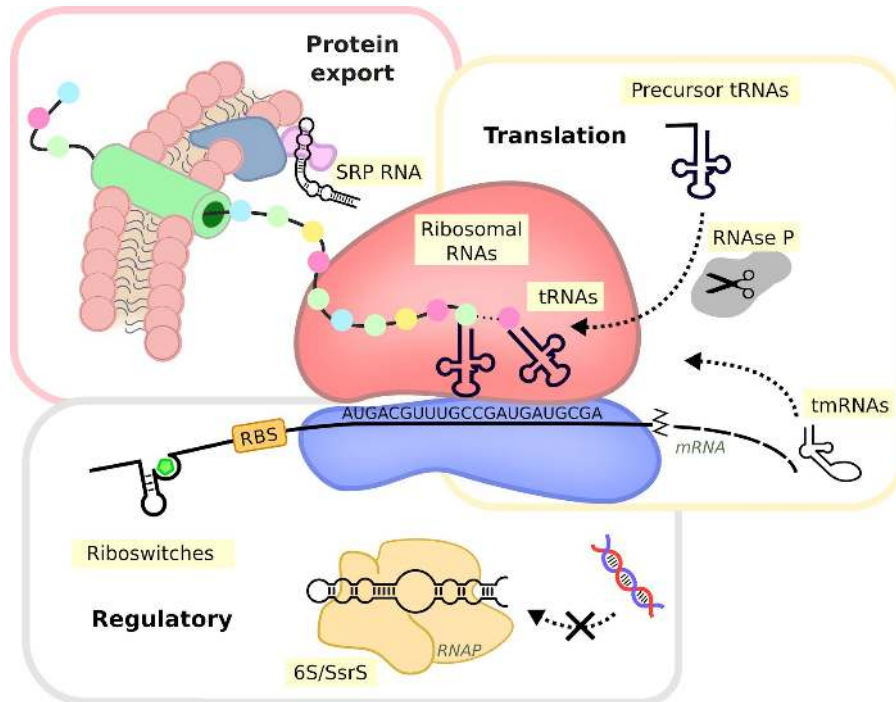
## Abstract

Understanding how new genes originate and integrate into cellular networks is key to understanding evolution. Bacteria present unique opportunities for both the natural history and experimental study of gene origins, due to their large effective population sizes, rapid generation times, and ease of genetic manipulation. Bacterial small non-coding RNAs (sRNAs) in particular, many of which operate through a simple antisense regulatory logic, may serve as tractable models for exploring processes of gene origin and adaptation. Understanding how and on what timescales these regulatory molecules arise has important implications for understanding the evolution of bacterial regulatory networks, in particular for the design of comparative studies of sRNA function. Here we introduce relevant concepts from evolutionary biology and review recent work that has begun to shed light on the timescales and processes through which non-functional transcriptional noise is co-opted to provide regulatory functions. We explore possible scenarios for sRNA origin, focusing on the co-option, or exaptation, of existing genomic structures which may provide protected spaces for sRNA evolution.

## Introduction

Far from being relics of a lost RNA world (1), most bacterial non-coding RNAs (ncRNAs) appear to be relatively recent innovations, with only a handful that are involved in translation, transcription, and translocation exhibiting deep conservation (2) (**Figure 1**). One class of bacterial ncRNA that appears to be exceptionally dynamic over evolutionary time are the small non-coding RNAs (sRNAs) - functionally heterogeneous transcripts of around 50-500

nucleotides, the majority of which appear to base-pair with mRNA targets to affect translation and/or mRNA stability (3). Many of these functions are mediated by RNA-binding proteins, such as Hfq (4), though some operate by sequestering RNA-binding proteins, such as the global regulator CsrA (5) or RNA polymerase (6).

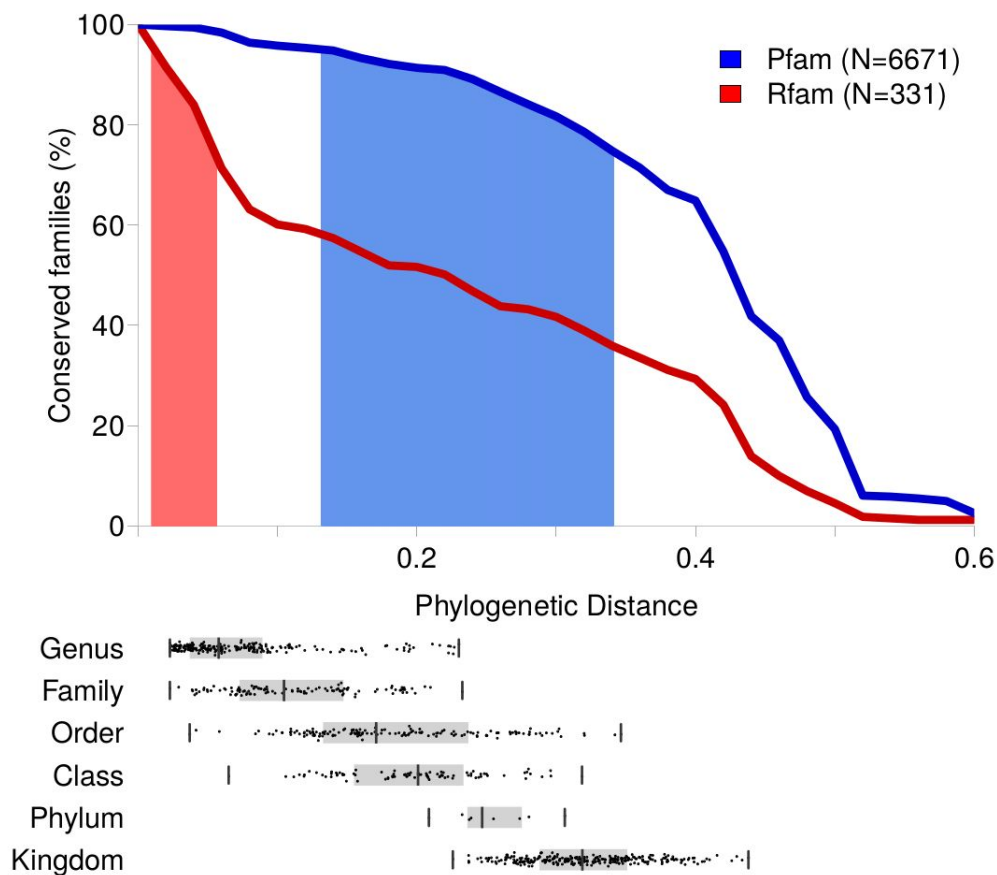


**Figure 1: broadly conserved ncRNAs in bacteria.** These ncRNAs are involved in core cellular processes: Translation (ribosomal RNAs, tRNAs, RNase P, and tmRNAs), regulation of transcription (the 6S RNA, aka SsrS) and translation (TPP riboswitch, aka Thi-element), and protein export (the Signal Recognition Particle (SRP) RNA, aka 4.5S). Conserved ncRNAs were identified by phylogenetic distribution of ncRNA annotations in the Rfam database (Hoeppner et al. 2012).

The development of RNA-seq-based techniques to identify transcripts has led to an explosion of uncharacterised putative sRNAs (7), with whole-genome screens regularly reporting hundreds of novel sRNAs per genome (8–11). While functional roles for sRNAs are well established in the Proteobacteria *Escherichia coli* and *Salmonella enterica*, the continuing discovery of sRNAs in other phyla, such as Cyanobacteria (12) and Firmicutes (13), indicate that sRNA-based regulation is prevalent across the bacterial phylogeny.

While sRNAs appear to be a widespread feature of bacterial transcriptomes, individual sRNA loci themselves are generally not. BLAST-based analyses of *Escherichia* (14), *Salmonella* (10,15), *Listeria* (16) and *Campylobacter* (17) sRNAs indicate that the majority are order or genus-specific, with some specific to serovars or even strains. It could be argued this result is an artifact of the sensitivity of homology search tools, as the lower bound for local alignment of nucleotide sequences is around ~50-60% sequence identity (18), and sRNAs may be relatively robust to mutation given that their function is primarily attributed to ‘seed regions’

of complementarity to their targets that may be as short as 6-8nt (19,20). However, phylogeny-wide analysis of ncRNA conservation with more sensitive methods, including both profile hidden Markov models (HMMs) and curated covariance models (CMs) from the Rfam database (21), suggests a similar overall pattern of narrow conservation, with a sharp drop in the number of conserved ncRNAs at the genus to family level (22) (**Figure 2**). It is important to note that this analysis will overestimate conservation, since Rfam families are overwhelmingly constructed from conserved sequences. This lead us previously to propose a narrow ‘Goldilocks Zone’ where comparative transcriptomics studies would be most useful for identifying conserved transcripts (22).



**Figure 2: A comparison of conservation of RNA and protein families.** Rfam and Pfam families were identified in 2,562 bacterial genomes using the Infernal and HMMER packages, respectively. The maximum phylogenetic distance spanned by each family was estimated using pairwise phylogenetic distances estimated from 16S rRNA sequences. Top: The percentage of Rfam (N=331) or Pfam (N=6,671) families that are conserved over phylogenetic distances ranging from 0 (closely related) to 0.6 (very divergent). The shaded regions under the curves indicate the “Goldilocks” phylogenetic ranges where between 95 and 75% of RNA (light red) or protein (light blue) families are conserved. Bottom: The distribution of phylogenetic distances between randomly sampled pairs of genomes at 6

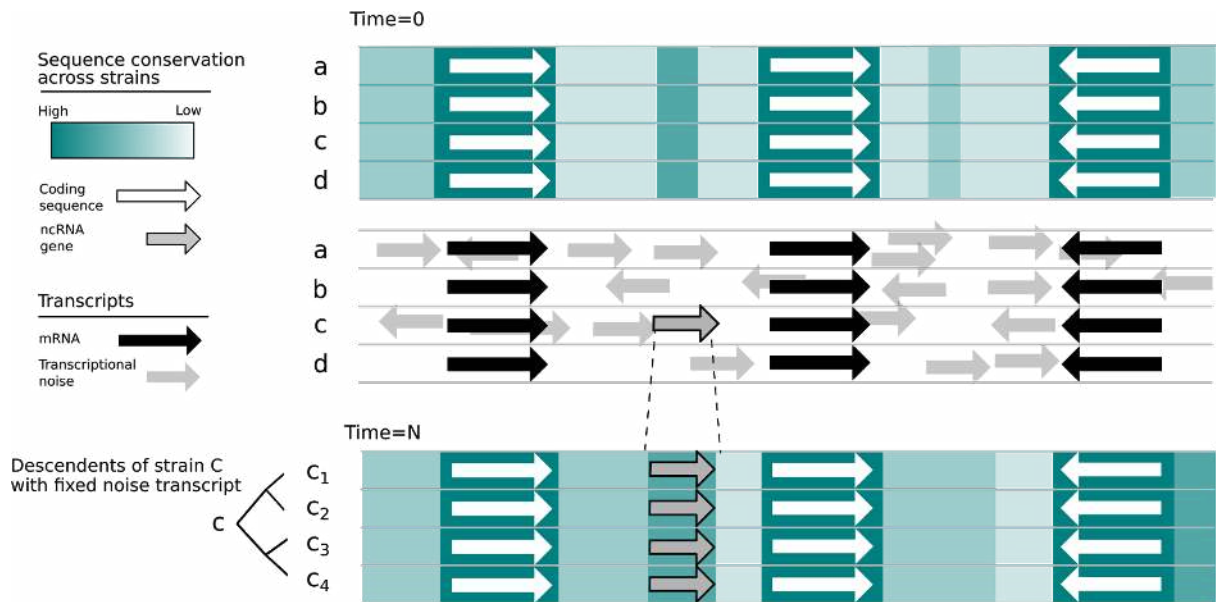
taxonomic ranks: genus, family, order, class, phylum and kingdom. At each taxonomic level, pairs were chosen from e.g. the same genus but different species, and so on up the taxonomic hierarchy until pairs of species from the same kingdom (bacteria) and different phyla are shown. Figure adapted from (Lindgreen et al. 2014).

The near universal presence of a large cohort of non-coding transcripts paired with their apparent narrow conservation raises a number of questions: where do these transcripts come from? How are they turned over? How many serve biological functions in the cell, and how many are just transcriptional noise? Two excellent recent reviews have addressed how specific features of sRNA evolution may have evolved (23,24). Here we aim to complement this previous work by investigating evolutionary dynamics underlying the generation and removal of transcriptional noise, and how this could serve as the raw stuff from which functional sRNAs emerge.

### **Transcriptional noise**

What do we mean by transcriptional noise? The term has been used to discuss ‘noisy’ interactions between the transcriptional machinery and DNA, in at least four distinct senses (25). In this article, we use it in the sense of non-functional transcripts that accumulate over evolutionary time due the interaction of permissive transcriptional machinery with random mutations in the genome (26) — however, it should be noted that even the concept of “function” is subject to multiple, competing definitions (27).

To clarify the sense in which we use the term transcriptional noise here, the Random Genome Project, a thought experiment proposed by Sean Eddy (28), is illustrative. Given the low information content of many protein binding sites, we would expect that any sufficiently long random DNA sequence will reproducibly interact with DNA-binding proteins, such as transcription factors. This will lead to the generation of discrete non-coding transcripts indistinguishable from native transcripts, including for instance condition-specific expression or affinity for RNA-binding proteins. By extension, random mutation will similarly lead to the generation of new, discrete transcripts over time as the accumulation of small changes lead to the generation of promoter-like sequences. Some proportion of these nascent transcripts will fix in the population depending on the effectiveness of selection and the dynamics of the host genome (**Figure 3**), discussed below. Importantly in the absence of perfectly efficient selection, a beneficial function is not a necessary condition for fixation.



**Figure 3: Fixation of transcriptional noise.** At Time=0: The top panel shows hypothetical genome sequences for four related bacterial strains (a, b, c, d) with regions of varying conservation along the sequence. The middle panel shows transcripts produced by each strain - dark colours indicate coding sequences, light colours indicate non-coding sequences. Noise transcripts are common, but not consistently expressed across strains and typically have low expression levels. At Time=N (bottom panel): In descendants of strain ‘C’ ( $C_1$ ,  $C_2$ ,  $C_3$ ,  $C_4$ ,  $C_5$ ) a noise transcript from the ancestral ‘C’ has been fixed due to drift, a selective sweep or population bottleneck.

For the purposes of this review, we define transcriptional noise as transcription which does not have a fitness benefit to the organism in its native environment(s) sufficient to drive fixation, providing a null model for observed transcripts (29). While testing this condition in multicellular eukaryotic cells may prove difficult, the development of high-throughput approaches to phenotyping has made testing this condition theoretically tractable in most bacteria (30,31), limited mainly by our understanding of and ability to model bacterial ecology. In particular, technologies like transposon insertion sequencing (32), CRISPRi (33), and robotics platforms (34) are increasingly making it easier to test the fitness effects of genes in a range of physiologically relevant conditions. Applying these technologies in concert with experimental evolution (35) may allow for the calibration of the selection coefficients necessary to maintain a gene in a given population, at least under laboratory conditions.

### Estimating the rate of de-novo sRNA evolution

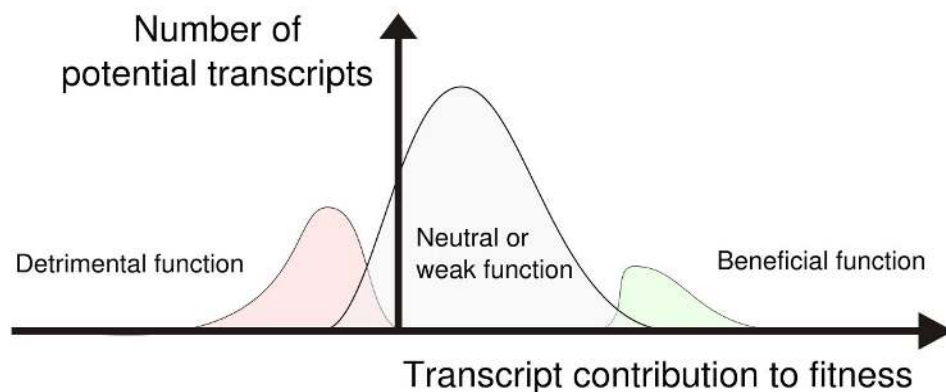
So, how likely is the *de novo* generation of a functional sRNA? We may never know for certain, but attempting to establish the parameters necessary to estimate this probability can help to clarify our expectations and direct investigation. In this back-of-the-envelope spirit, we have tried to lay out the key parameters for sRNA generation in Figure 4A, inspired by a

similar Drake-like equation for the *de novo* generation of protein-coding genes (36). This comprises three primary components: the transcript birth rate and the transcript death rate, which together establish the reservoir of transcriptional noise available for selection to act on, and finally the probability of acquiring a beneficial function. The fitness effects of any particular nascent transcript on the host organism will represent a draw from the distribution of potential effects (Figure 4B).

A.

$$\begin{aligned}
 P_{\text{(sRNA gene)}} &= P_{\text{(transcript birth)}} \times P_{\text{(no transcript death)}} \times P_{\text{(beneficial function)}} \\
 &= [P_{\text{(promoter)}} \times P_{\text{(terminator)}} \times P_{\text{(stability)}}] \\
 &\quad \times [\{1 - P_{\text{(neutral removal)}}\} \times \{1 - P_{\text{(detrimental function)}}\}] \\
 &\quad \times P_{\text{(beneficial function)}}
 \end{aligned}$$

B.



**Figure 4A: Estimating the probability of *de novo* sRNA formation with a Drake-like equation.** Details of each term are discussed in detail in the text. Briefly, the probability of sRNA formation has been decomposed into three relatively independent factors, the probabilities of transcript birth, transcript death and of acquiring a beneficial function. These probabilities can be further decomposed, into e.g. the probabilities of promoter and terminator acquisition, as well as acquiring stabilisation factors for the transcript birth probability.

**4B Theoretical distribution of fitness for potential transcripts of detrimental (red), neutral (grey) or beneficial (green) function.** While  $P[\text{beneficial function}]$  and  $P[\text{detrimental}$

function] are not independent, they represent extremes of the distribution of potential transcript fitness, separated by a large number of transcripts with a neutral or 'weak' function. (Based on Figure 1, (37))

Transcript birth requires three basic components: promotion, termination, and transcript stabilization. A number of lines of evidence suggest the generation of spontaneous promoters is common. Many bacterial promoter sequences are degenerate, and transcription factor binding sites may arise from just a few point mutations (38). This result has recently been confirmed by experimental evolution in *E. coli* (39), showing that random sequences can rapidly acquire promoter activity. These theoretical and laboratory results are supported by observations that antisense promoters are common (40,41) but rarely conserved (42,43), suggesting a continuous turn-over in these sites.

It is less clear how transcriptional termination might evolve. Classical Rho-independent terminators are complex, requiring both a stable stem loop and poly-uridine stretch (44). It is not obvious how they might frequently arise, though the frequent occurrence of inverted repeat structures in intergenic sequences suggests it may be possible (45,46). These repeats, through their provision of secondary structure, might also serve as a reservoir of important stability determinants by protecting nascent transcripts from exonuclease digestion (47). The acquisition of secondary structures and/or associations with protective proteins may be especially important for the survival of non-coding transcripts, which unlike mRNAs do not benefit from protection against degradation by the presence of translating ribosomes (48).

Alternatives to de novo evolution of Rho-independent terminators include the deposition of terminators by mobile elements (49), or through genomic rearrangements, either of which could provide a space in which an sRNA could evolve. A recent study has indicated the generation of a genus-specific non-coding transcript through just such a rearrangement in *Salmonella* (50). Additionally, Rho-dependent termination appears to occur in a number of intergenic transcripts (51), although as the features promoting Rho recruitment are less well understood, the evolutionary implications of this are unclear.

Once expressed, a nascent transcript has to survive long enough to acquire function. Pervasive transcription from non-coding regions is now an accepted feature of eukaryotic genomes (52), and it appears selective constraints in eukaryotes are weak enough to allow for nearly every possible non-coding transcript to be sampled over evolutionary time (53). However, relative to eukaryotes, bacteria have extremely compact genomes with gene contents averaging ~88% (54), reducing the sequence space available for the generation of intergenic transcriptional noise. This is driven by two factors: differences in mutational biases between bacteria and eukaryotes, and differences in the selective pressures felt by each.

In contrast to eukaryotes, there is a strong bias towards deletion mutations in bacteria in the absence of selection (55,56). This has important consequences for genome dynamics and the



maintenance of non-coding sequence. In populations with small effective population sizes where selection is less effective, including many important human pathogens (57), genetic drift leads to reduced genome sizes (58). In these bacteria, transcript removal will be dominated by the neutral removal rate. Somewhat counterintuitively, this can also lead to a higher proportion of non-coding sequence in these genomes, due to the accumulation of pseudogenes and proliferation of selfish elements (59).

On the other hand, some environmental bacteria have large population sizes, increasing the effectiveness of selection in these populations compared to eukaryotes, and their smaller cell size increases their sensitivity to the energetic costs of gene maintenance relative to eukaryotes (60). Large population sizes may also make nascent transcripts more vulnerable to clonal interference (61), where neutral or weakly beneficial mutations are rapidly removed through competition with other, more beneficial mutations that arise independently. As it is unlikely a novel transcript will be sufficient to confer a competitive advantage organism without time for selection to act, the rate of acquisition of beneficial mutation in the population as a whole will impact the rate of transcript fixation.

Additional fitness costs may be imposed by the biochemical activity of the noise transcript itself, for instance through stochastic interactions with other transcripts (62) — these may be increased by the relative lack of compartmentalization in bacterial cells. This selective pressure results in a reduced non-coding:coding DNA ratio (59), in extreme cases leading to another form of genome reduction driven by streamlining (63) as observed in the SAR11 clade of marine bacteria (64).

The final term in our Drake-like equation is acquisition of a beneficial function. Biochemical activity, in the form of affinity for other transcripts or RNA-binding proteins, would likely exist in any random transcript. However, such interactions are unlikely to have a strong effect on fitness prior to selection (Fig 4B). The observation that for many sRNAs, at least one of their target sites either pre-dates or originated at the same time as the sRNA (14) certainly suggests acquisition of interactions can be rapid; however, the majority of random interactions are likely detrimental to organismal fitness. As has been suggested for eukaryotic microRNAs (65), bacterial noise transcripts that survive initial purification by selection are likely to have weak initial interactions with target genes, and low-level, condition-specific expression. Initial analysis of expression of young non-coding transcripts in *Escherichia coli* and *Salmonella enterica* suggest this is in fact the case (66), and it has similarly been estimated that most antisense transcripts are expressed at levels too low to have major impacts on fitness (67). This line of argument also suggests affinity for RNA-binding proteins should follow a similar pattern. The ability to increase the effectiveness of interactions through acquisition of protein chaperone affinity (see **box 1: RNA-binding proteins**) may provide a route to rapidly tune the activity of a nascent sRNA.

The terms in our sRNA Drake-like equation are of course all contingent on the specifics of each bacterial genome. Some examples include that genome-wide A/T content may make promoters and terminators more or less likely to occur (67), or that large effective population sizes may make selection more effective in some bacteria than others (59), or that the presence of RNA-binding proteins may make sRNA-based regulation more effective in some lineages than others (14). These terms can also be locally manipulated, by taking advantage of existing genomic features through the process of exaptation.

### **Box 1: RNA-binding proteins**

Many sRNAs require RNA-binding proteins, such as Hfq, CsrA, and ProQ, to function (68). For instance, association with Hfq can greatly increase the rate of formation and the stability of sRNA-mRNA duplexes (69,70), and is necessary for the function of many *E. coli* and *Salmonella* trans-acting sRNAs. The requirement for Hfq appears to vary by genome, GC content, and sRNA composition, although studies are limited to a small number of model organisms with functionally characterised sRNAs (71). Other sRNAs, such as the Csr/Rsm family, appear to operate primarily through titration of RNA-binding proteins themselves (5).

RNA-binding protein interaction sites appear to be relatively simple. Hfq has multiple binding surfaces with different sequence specificities (72), primarily consisting of A- or U-rich tracts. The CsrA binding site is similarly small, requiring just a GGA sequence within a hairpin loop (5), though other factors can influence binding affinity (73). The size and degeneracy of these motifs may provide a quick route to acquiring chaperone affinity, and many may require only small modifications of existing structures. For example, mutagenesis and CLIP-seq studies have shown that many Rho-independent terminators associate with Hfq (74,75).

### **Exaptation in the evolution of sRNAs**

Exaptation is the process through which existing features, sometimes referred to as pre-adaptations, are co-opted to provide a new function not previously selected for (76). In the simplest case, processes such as duplication or horizontal transfer allow for transplantation of transcribed sequences, removing the need for de novo evolution of promoters, terminators, or stability determinants and hence increasing the probability of sRNA generation. Importantly, both duplications and horizontal insertions provide non-deleterious sequences that are already adapted to the host, presenting a low-risk route to acquiring new ncRNAs.

Recognizable sRNA duplications appear fairly commonly (77). Notable examples including the Qrr sRNAs involved in quorum sensing in *Vibrio* species (78), the PrrF sRNAs involved in regulating iron homeostasis in *Pseudomonas* (79), and the OmrA/B sRNAs in *Escherichia coli* (80). Such duplications could allow duplicated sRNAs to diverge and gain new functions through a process of subfunctionalization and neofunctionalization (81). While the functional

divergence of duplicated sRNAs remains understudied, there are some tantalizing indications of regulon drift between duplicated sRNAs (77).

Many bacterial genomes contain horizontally-acquired genomic islands (GIs) (82), often referred to as pathogenicity islands in virulent bacteria, which are frequent sources of novel sRNAs (83–85). The existence of such sRNAs is perhaps unsurprising, as many common mobile elements utilize ncRNA regulation to control their replication and integration (86), or to ensure their maintenance, such as the sRNA antitoxins in Type I toxin-antitoxin systems (87). One recent example of an apparent exaptation event is the art200 antisense RNA derived from *tnpA* transposase loci that has been suggested to regulate *Salmonella* Typhimurium virulence by targeting invasion genes (88), and may affect gastrointestinal colonization of mice (89).

### Protected spaces for sRNA evolution

Beyond the direct duplication or import of transcriptional units, reusing the existing genome architecture provides another way to tweak the parameters of our Drake-like equation so as to increase the rate of sRNA formation. In particular, mRNA operon structures harbor expressed untranslated regions (UTRs) that are protected from large-scale deletions by the selective pressures maintaining the associated coding sequences, eliminating the need for de novo evolution of transcription and lowering the probability of transcript removal. Recent work has suggested ncRNAs may be generated from within coding sequences themselves (90), though no specific function has been proposed for these fragments as of yet.

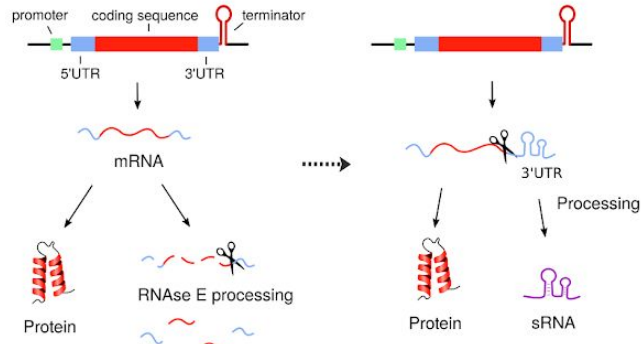
3' UTRs at the end of mRNA transcripts appear to be a particularly fertile ground for sRNA formation, with several instances well characterized (91). The presence of rho-independent terminators at many gene ends provides a natural termination point, stable secondary structure, and possibly a precursor to an Hfq binding site (**Box 1**). The natural degradation of mRNA transcripts may by chance liberate such a stable fragment which is then free to acquire a secondary function. Indeed, such a stepwise evolution has been proposed for the *cpxQ* 3' UTR derived sRNA (92) (**Figure 5A**). The expression of such sRNAs would be tied to the expression of their parent transcript, possibly limiting the functional space the sRNA could explore. A natural extension then would be the evolution of independent cryptic promoters that could unlink expression of the sRNA from the parent transcript. This is the case for the sRNA MicL expressed from the 3' UTR of the *cutC* locus (**Figure 5B**), which has been shown to be responsible for the copper sensitivity phenotype previously attributed to the coding sequence (93). Intriguingly, MicL is transcribed as a 308 nucleotide sequence that is then processed to its active 80 nucleotide length. In fact, many well characterized intergenic sRNAs, for instance ArcZ (94), are similarly processed. Are these processing sites genomic relics of a past 3' UTR origin?

Many 5' UTRs contain transcriptional attenuators, including riboswitches, RNA thermometers, peptide leaders, and RNA-binding protein motifs (95). As a side-effect of attenuation some of these generate stable, or at least abundant, short transcripts that are then free to explore new functional spaces (**Figure 5C**), though this would be constrained by the need to conserve the original attenuator function. Functionally characterised examples of such dual-purpose cis/trans regulators have been found derived from attenuators involved in virulence, such as the riboswitch-derived SreA in *Listeria monocytogenes* (96), and the ATP- and charged tRNA-sensitive leader sequence of the *Salmonella mgtCBR* operon (97).

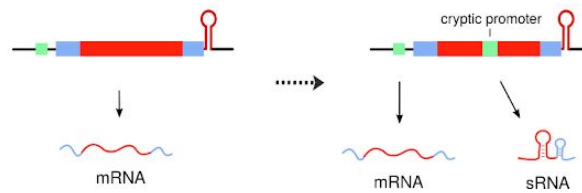
Beyond coding sequences, core ncRNAs such as ribosomal and tRNAs are often arranged in operons that are processed to generate functional molecules, producing various intermediates and byproducts along the way. The tRNAs have been proposed as promising candidates in particular (98), as a number of tRNAs and their intermediates have been suggested to interact with Hfq (99,100). Spacers excised from pre-tRNA transcripts during tRNA processing have been indeed found to act as sRNA sponges (**Figure 5D**) (101). Given the consistency of the tRNA complement across related species (102) it is unclear how often such transcripts might make the transition to independent sRNAs through duplication or decay of tRNA function, though the presence of enigmatic expressed repeats in the vicinity of some tRNA loci (103) suggests it may be possible. A recent report has additionally suggested a role for Hfq in ribosomal RNA biogenesis (104), in particular providing evidence for Hfq binding of spacer regions flanking the mature 16S sequence, again raising the question of whether these sequences could acquire secondary functions.

## Evolution within 3' UTRs

### A. Exaptation of processing product (Chao & Vogel 2016)

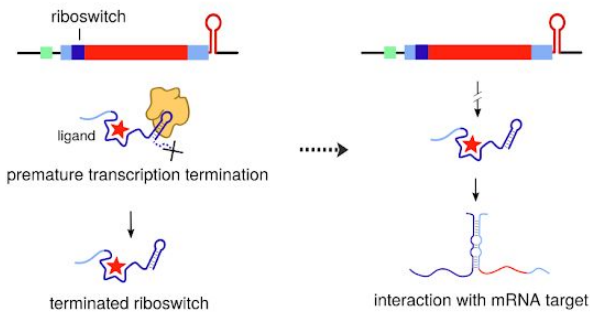


### B. Cryptic promoter formation (Guo *et al.* 2014)

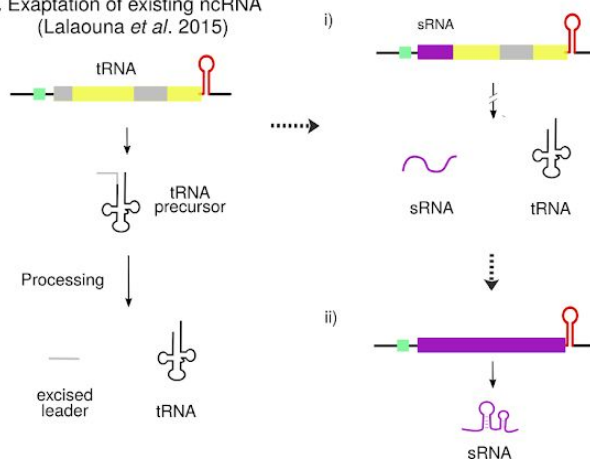


## Evolution within 5' UTRs

### C. Exaptation of cis-regulatory element (Loh *et al.* 2009)



### D. Exaptation of existing ncRNA (Lalaouna *et al.* 2015)



**Figure 5 : Routes for sRNA evolution within “protected spaces”** - Functional genomic elements sheltered from deletion by selection on neighbouring genes, that may increase the probability of sRNA evolution. These include known examples from within 3' UTRs, either derived from processing products (A), or cryptic promoters (B); from existing cis-regulatory elements in 5' UTRs (C); or from existing ncRNAs, such as tRNAs (D).

## Future Directions

Here we have attempted to sketch the outline of a quantitative understanding of sRNA birth. Understanding the parameters of this equation will take significant work. One promising approach is through molecular archeology, attempting to infer the sequence of events that have led to the present sRNA repertoire through the examination of increasingly plentiful bacterial genome sequences. Already, this approach has provided the first putative examples of sRNA creation through recombination (50). Scaling this approach to the entire sRNA complement of a bacterial clade will be a challenge, with complications we have not touched upon in this review — for instance, the selective pressures we have examined are not constant but change as a bacterial population is exposed to different environmental circumstances. As many bacterial sRNAs appear to be involved in stress responses, these variable selective pressures may be particularly important to consider. Perhaps by combining this molecular archeology approach with new techniques for high-throughput characterization of sRNA function (7), we can begin to understand how transcriptional noise is domesticated to provide regulatory functions in the bacterial cell.

A second approach is to look forward, rather than back. Can we manipulate the terms of the sRNA formation equation to make it more likely that we observe the emergence of a functional bacterial sRNA under laboratory conditions? Could engineering the protected spaces we have reviewed here, coupled with the application of experimental evolution (35,105), provide a means for us to directly and reproducibly drive the *de novo* birth of a gene? It is unclear how feasible this might be, but then aspirations do not need to be attainable to be useful.

## Perspectives

- Short non-coding transcripts are omnipresent in bacterial transcriptomes; however, their conservation tends to be limited.
- This implies that these transcripts are generated and removed regularly over the course of evolution.
- This dynamic pool of nonfunctional noise transcripts may serve as one reservoir on which selection can act to produce functional sRNAs; the specific rates of noise

transcript birth and death are likely depend on the characteristics of the bacterial population under study.

- Non-coding regions of the genome protected from removal by their association with functional sequences, for example the UTRs of coding sequences or processing intermediaries of tRNAs, may provide a more stable reservoir for sRNA evolution through exaptation.
- Understanding the parameters governing the process of functional sRNA generation will require the application of both backward-looking (molecular archaeology) and forward-looking (synthetic biology and experimental evolution) approaches.

## Acknowledgements

The authors would like to thank Alex Westermann and Erik Holmqvist and the anonymous reviewers for insightful comments on the draft manuscript.

## Funding

BRJ and PPG are funded by the Bio-Protection Research Centre, PPG receives funding from the New Zealand Ministry of Business Innovation and Employment (MBIE) Smart Ideas (UOOX1709) and the Marsden Fund (17-UOO-050).

## References

1. Eddy SR. Non-coding RNA genes and the modern RNA world. *Nat Rev Genet.* 2001 Dec 1;2(12):919–29.
2. Hoepfner MP, Gardner PP, Poole AM. Comparative analysis of RNA families reveals distinct repertoires for each domain of life. *PLoS Comput Biol.* 2012 Nov 1;8(11):e1002752.
3. Gorski SA, Vogel J, Doudna JA. RNA-based recognition and targeting: sowing the seeds of specificity. *Nat Rev Mol Cell Biol.* 2017 Apr;18(4):215–28.
4. Vogel J, Luisi BF. Hfq and its constellation of RNA. *Nat Rev Microbiol.* 2011 Aug;9(8):578–89.
5. Romeo T, Babitzke P. Global Regulation by CsrA and Its RNA Antagonists. *Microbiol Spectr* [Internet]. 2018 Mar;6(2). Available from: <http://dx.doi.org/10.1128/microbiolspec.RWR-0009-2017>
6. Wassarman KM. 6S RNA: a regulator of transcription. *Mol Microbiol.* 2007 Sep;65(6):1425–31.
7. Barquist L, Vogel J. Accelerating Discovery and Functional Analysis of Small RNAs with New Technologies. *Annu Rev Genet.* 2015 Nov 23;49:367–94.

8. Miotto P, Forti F, Ambrosi A, Pellin D, Veiga DF, Balazsi G, et al. Genome-wide discovery of small RNAs in *Mycobacterium tuberculosis*. *PLoS One*. 2012 Dec 19;7(12):e51950.
9. Gómez-Lozano M, Marvig RL, Molina-Santiago C, Tribelli PM, Ramos J-L, Molin S. Diversity of small RNAs expressed in *Pseudomonas* species. *Environ Microbiol Rep*. 2015 Apr;7(2):227–36.
10. Kröger C, Dillon SC, Cameron ADS, Papenfort K, Sivasankaran SK, Hokamp K, et al. The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. *Proc Natl Acad Sci U S A*. 2012 Apr 25;109(20):E1277–86.
11. Rau MH, Bojanovič K, Nielsen AT, Long KS. Differential expression of small RNAs under chemical stress and fed-batch fermentation in *E. coli*. *BMC Genomics*. 2015 Dec 10;16:1051.
12. Klähn S, Schaal C, Georg J, Baumgartner D, Knippen G, Hagemann M, et al. The sRNA NsiR4 is involved in nitrogen assimilation control in cyanobacteria by targeting glutamine synthetase inactivating factor IF7. *Proc Natl Acad Sci U S A*. 2015 Nov 10;112(45):E6243–52.
13. Durand S, Braun F, Helfer A-C, Romby P, Condon C. sRNA-mediated activation of gene expression by inhibition of 5'-3' exonucleolytic mRNA degradation. *Elife* [Internet]. 2017 Apr 24;6. Available from: <http://dx.doi.org/10.7554/eLife.23602>
14. Peer A, Margalit H. Evolutionary patterns of *Escherichia coli* small RNAs and their regulatory interactions. *RNA*. 2014 Jul;20(7):994–1003.
15. Kröger C, Colgan A, Srikumar S, Händler K, Sivasankaran SK, Hammarlöf DL, et al. An infection-relevant transcriptomic compendium for *Salmonella enterica* Serovar Typhimurium. *Cell Host Microbe*. 2013 Dec 11;14(6):683–95.
16. Cerutti F, Mallet L, Painset A, Hoede C, Moisan A, Bécavin C, et al. Unraveling the evolution and coevolution of small regulatory RNAs and coding genes in *Listeria*. *BMC Genomics*. 2017 Nov 16;18(1):882.
17. Dugar G, Herbig A, Förstner KU, Heidrich N, Reinhardt R, Nieselt K, et al. High-resolution transcriptome maps reveal strain-specific regulatory features of multiple *Campylobacter jejuni* isolates. *PLoS Genet*. 2013 May;9(5):e1003495.
18. Gardner PP, Wilm A, Washietl S. A benchmark of multiple sequence alignment programs upon structural RNAs. *Nucleic Acids Res*. 2005 Apr 28;33(8):2433–9.
19. Papenfort K, Bouvier M, Mika F, Sharma CM, Vogel J. Evidence for an autonomous 5' target recognition domain in an Hfq-associated small RNA. *Proceedings of the National Academy of Sciences*. 2010 Nov 23;107(47):20435–40.
20. Richter AS, Backofen R. Accessibility and conservation: general features of bacterial small RNA-mRNA interactions? *RNA Biol*. 2012 Jul;9(7):954–65.
21. Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, et al. Rfam



- 12.0: updates to the RNA families database. *Nucleic Acids Res.* 2015 Jan;43(Database issue):D130–7.
22. Lindgreen S, Umu SU, Lai AS-W, Eldai H, Liu W, McGimpsey S, et al. Robust identification of noncoding RNA from transcriptomes requires phylogenetically-informed sampling. *PLoS Comput Biol.* 2014 Oct;10(10):e1003907.
  23. Updegrove TB, Shabalina SA, Storz G. How do base-pairing small RNAs evolve? *FEMS Microbiol Rev.* 2015;fuv014.
  24. Dutcher HA, Raghavan R. Origin, Evolution, and Loss of Bacterial Small RNAs. *Microbiol Spectr* [Internet]. 2018 Apr;6(2). Available from: <http://dx.doi.org/10.1128/microbiolspec.RWR-0004-2017>
  25. Raser JM, O’Shea EK. Noise in gene expression: origins, consequences, and control. *Science.* 2005 Sep 23;309(5743):2010–3.
  26. Struhl K. Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat Struct Mol Biol.* 2007 Feb;14(2):103–5.
  27. Doolittle WF. We simply cannot go on being so vague about “function.” *Genome Biol.* 2018 Dec 18;19(1):223.
  28. Eddy SR. The ENCODE project: missteps overshadowing a success. *Curr Biol.* 2013 Apr 8;23(7):R259–61.
  29. Koonin EV. Splendor and misery of adaptation, or the importance of neutral null for understanding evolution. *BMC Biol.* 2016 Dec 23;14(1):114.
  30. Gray AN, Koo B-M, Shiver AL, Peters JM, Osadnik H, Gross CA. High-throughput bacterial functional genomics in the sequencing era. *Curr Opin Microbiol.* 2015 Aug 31;27:86–95.
  31. Brochado AR, Typas A. High-throughput approaches to understanding gene function and mapping network architecture in bacteria. *Curr Opin Microbiol.* 2013 Apr;16(2):199–206.
  32. Chao MC, Abel S, Davis BM, Waldor MK. The design and analysis of transposon insertion sequencing experiments. *Nat Rev Microbiol.* 2016 Jan 19;14(2):119–28.
  33. Peters JM, Silvis MR, Zhao D, Hawkins JS, Gross CA, Qi LS. Bacterial CRISPR: accomplishments and prospects. *Curr Opin Microbiol.* 2015 Oct;27:121–6.
  34. Kritikos G, Banzhaf M, Herrera-Dominguez L, Koumoutsi A, Wartel M, Zietek M, et al. A tool named Iris for versatile high-throughput phenotyping in microorganisms. *Nat Microbiol.* 2017 Feb 17;2:17014.
  35. Cooper VS. Experimental Evolution as a High-Throughput Screen for Genetic Adaptations. *mSphere.* 2018 Jun 27;3(3):e00121–18.
  36. Weisman CM, Eddy SR. Gene Evolution: Getting Something from Nothing. *Curr Biol.*

2017 Jul 10;27(13):R661–3.

37. Martin G, Roques L. The Nonstationary Dynamics of Fitness Distributions: Asexual Model with Epistasis and Standing Variation. *Genetics*. 2016 Dec;204(4):1541–58.
38. Stone JR, Wray GA. Rapid evolution of cis-regulatory sequences via local point mutations. *Mol Biol Evol*. 2001 Sep;18(9):1764–70.
39. Yona AH, Alm EJ, Gore J. Random sequences rapidly evolve into de novo promoters. *Nat Commun*. 2018 Apr 18;9(1):1530.
40. Dornenburg JE, Devita AM, Palumbo MJ, Wade JT. Widespread antisense transcription in *Escherichia coli*. *MBio* [Internet]. 2010 May 18;1(1). Available from: <http://dx.doi.org/10.1128/mBio.00024-10>
41. Thomason MK, Bischler T, Eisenbart SK, Förstner KU, Zhang A, Herbig A, et al. Global Transcriptional Start Site Mapping Using Differential RNA Sequencing Reveals Novel Antisense RNAs in *Escherichia coli*. *J Bacteriol*. 2015 Jan 1;197(1):18–28.
42. Raghavan R, Sloan DB, Ochman H. Antisense transcription is pervasive but rarely conserved in enteric bacteria. *MBio* [Internet]. 2012 Aug 7;3(4). Available from: <http://dx.doi.org/10.1128/mBio.00156-12>
43. Shao W, Price MN, Deutschbauer AM, Romine MF, Arkin AP. Conservation of transcription start sites within genes across a bacterial genus. *MBio*. 2014 Jul 1;5(4):e01398–14.
44. Ray-Soni A, Bellecourt MJ, Landick R. Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annu Rev Biochem*. 2016 Jun 2;85:319–47.
45. Ladoukakis ED, Eyre-Walker A. The excess of small inverted repeats in prokaryotes. *J Mol Evol*. 2008 Sep;67(3):291–300.
46. Lillo F, Basile S, Mantegna RN. Comparative genomics study of inverted repeats in bacteria. *Bioinformatics*. 2002 Jul;18(7):971–9.
47. Dar D, Sorek R. High-resolution RNA 3'-ends mapping of bacterial Rho-dependent transcripts. *Nucleic Acids Res*. 2018 Jul 27;46(13):6797–805.
48. Edri S, Tuller T. Quantifying the effect of ribosomal density on mRNA stability. *PLoS One*. 2014 Jul 14;9(7):e102308.
49. Naville M, Gautheret D. Premature terminator analysis sheds light on a hidden world of bacterial transcriptional attenuation. *Genome Biol*. 2010 Sep 29;11(9):R97.
50. Raghavan R, Kacharia FR, Millar JA, Sislak CD, Ochman H. Genome rearrangements can make and break small RNA genes. *Genome Biol Evol* [Internet]. 2015 Jan 18; Available from: <http://dx.doi.org/10.1093/gbe/evv009>
51. Peters JM, Mooney RA, Kuan PF, Rowland JL, Keles S, Landick R. Rho directs widespread termination of intragenic and stable RNA transcription. *Proc Natl Acad Sci*

- U S A. 2009 Sep 8;106(36):15406–11.
52. Deveson IW, Hardwick SA, Mercer TR, Mattick JS. The Dimensions, Dynamics, and Relevance of the Mammalian Noncoding Transcriptome. *Trends Genet.* 2017 Jul;33(7):464–78.
  53. Neme R, Tautz D. Fast turnover of genome transcription across evolutionary time exposes entire non-coding DNA to de novo gene emergence. *Elife.* 2016 Feb 2;5:e09977.
  54. Mira A, Ochman H, Moran NA. Deletional bias and the evolution of bacterial genomes. *Trends Genet.* 2001 Oct;17(10):589–96.
  55. Kuo C-H, Ochman H. Deletional bias across the three domains of life. *Genome Biol Evol.* 2009 Jun 27;1:145–52.
  56. Nilsson AI, Koskiniemi S, Eriksson S, Kugelberg E, Hinton JCD, Andersson DI. Bacterial genome size reduction by experimental evolution. *Proc Natl Acad Sci U S A.* 2005 Aug 23;102(34):12112–6.
  57. Weinert LA, Welch JJ. Why Might Bacterial Pathogens Have Small Genomes? *Trends Ecol Evol.* 2017 Dec;32(12):936–47.
  58. Kuo C-H, Moran NA, Ochman H. The consequences of genetic drift for bacterial genome complexity. *Genome Res.* 2009 Aug;19(8):1450–4.
  59. Novichkov PS, Wolf YI, Dubchak I, Koonin EV. Trends in prokaryotic evolution revealed by comparison of closely related bacterial and archaeal genomes. *J Bacteriol.* 2009 Jan;191(1):65–73.
  60. Lynch M, Marinov GK. The bioenergetic costs of a gene. *Proc Natl Acad Sci U S A.* 2015 Dec 22;112(51):15690–5.
  61. Fogle CA, Nagle JL, Desai MM. Clonal interference, multiple mutations and adaptation in large asexual populations. *Genetics.* 2008 Dec;180(4):2163–73.
  62. Umu SU, Poole AM, Dobson RC, Gardner PP. Avoidance of stochastic RNA interactions can be harnessed to control protein expression levels in bacteria and archaea. *Elife* [Internet]. 2016;5. Available from: <http://dx.doi.org/10.7554/eLife.13479>
  63. Batut B, Knibbe C, Marais G, Daubin V. Reductive genome evolution at both ends of the bacterial population size spectrum. *Nat Rev Microbiol.* 2014 Dec;12(12):841–50.
  64. Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, et al. Genome streamlining in a cosmopolitan oceanic bacterium. *Science.* 2005 Aug 19;309(5738):1242–5.
  65. Chen K, Rajewsky N. The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet.* 2007 Feb;8(2):93–103.
  66. 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 Apr;84(4):204–13.
67. Lloréns-Rico V, Cano J, Kamminga T, Gil R, Latorre A, Chen W-H, et al. Bacterial antisense RNAs are mainly the product of transcriptional noise. *Sci Adv.* 2016 Mar;2(3):e1501363.
  68. Holmqvist E, Vogel J. RNA-binding proteins in bacteria. *Nat Rev Microbiol* [Internet]. 2018 Jul 11; Available from: <http://dx.doi.org/10.1038/s41579-018-0049-5>
  69. Fender A, Elf J, Hampel K, Zimmermann B, Wagner EGH. RNAs actively cycle on the Sm-like protein Hfq. *Genes Dev.* 2010 Dec 1;24(23):2621–6.
  70. Panja S, Santiago-Frangos A, Schu DJ, Gottesman S, Woodson SA. Acidic Residues in the Hfq Chaperone Increase the Selectivity of sRNA Binding and Annealing. *J Mol Biol.* 2015 Nov 6;427(22):3491–500.
  71. Jouselin A, Metzinger L, Felden B. On the facultative requirement of the bacterial RNA chaperone, Hfq. *Trends Microbiol.* 2009 Sep;17(9):399–405.
  72. Santiago-Frangos A, Woodson SA. Hfq chaperone brings speed dating to bacterial sRNA. *Wiley Interdiscip Rev RNA.* 2018 Jul;9(4):e1475.
  73. Duss O, Michel E, Diarra dit Konté N, Schubert M, Allain FH-T. Molecular basis for the wide range of affinity found in Csr/Rsm protein-RNA recognition. *Nucleic Acids Res.* 2014 Apr;42(8):5332–46.
  74. Otaka H, Ishikawa H, Morita T, Aiba H. PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action. *Proc Natl Acad Sci U S A.* 2011 Aug 9;108(32):13059–64.
  75. Holmqvist E, Wright PR, Li L, Bischler T, Barquist L, Reinhardt R, et al. Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo. *EMBO J.* 2016 May 2;35(9):991–1011.
  76. Gould SJ, Vrba ES. Exaptation—a Missing Term in the Science of Form. *Paleobiology.* 1982;8(1):4–15.
  77. 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 Oct 28;4:151.
  78. Pappenfort K, Bassler BL. Quorum sensing signal-response systems in Gram-negative bacteria. *Nat Rev Microbiol.* 2016 Aug 11;14(9):576–88.
  79. Wilderman PJ, Sowa NA, FitzGerald DJ, FitzGerald PC, Gottesman S, Ochsner UA, et al. Identification of tandem duplicate regulatory small RNAs in *Pseudomonas aeruginosa* involved in iron homeostasis. *Proc Natl Acad Sci U S A.* 2004 Jun 29;101(26):9792–7.
  80. Holmqvist E, Reimegård J, Sterk M, Grantcharova N, Römling U, Wagner EGH. Two

- antisense RNAs target the transcriptional regulator CsgD to inhibit curli synthesis. *EMBO J.* 2010 Jun 2;29(11):1840–50.
81. Rastogi S, Liberles DA. Subfunctionalization of duplicated genes as a transition state to neofunctionalization. *BMC Evol Biol.* 2005 Apr 14;5:28.
  82. Dobrindt U, Hochhut B, Hentschel U, Hacker J. Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol.* 2004 May;2(5):414–24.
  83. Pichon C, Felden B. Small RNA genes expressed from *Staphylococcus aureus* genomic and pathogenicity islands with specific expression among pathogenic strains. *Proc Natl Acad Sci U S A.* 2005 Oct 4;102(40):14249–54.
  84. Padalon-Brauch G, Hershberg R, Elgrably-Weiss M, Baruch K, Rosenshine I, Margalit H, et al. Small RNAs encoded within genetic islands of *Salmonella typhimurium* show host-induced expression and role in virulence. *Nucleic Acids Res.* 2008 Apr;36(6):1913–27.
  85. Tree JJ, Granneman S, McAteer SP, Tollervey D, Gally DL. Identification of bacteriophage-encoded anti-sRNAs in pathogenic *Escherichia coli*. *Mol Cell.* 2014 Jul 17;55(2):199–213.
  86. Wagner EG, Simons RW. Antisense RNA control in bacteria, phages, and plasmids. *Annu Rev Microbiol.* 1994;48:713–42.
  87. Brantl S. Bacterial type I toxin-antitoxin systems. *RNA Biol.* 2012 Dec;9(12):1488–90.
  88. Ellis MJ, Trussler RS, Charles O, Haniford DB. A transposon-derived small RNA regulates gene expression in *Salmonella Typhimurium*. *Nucleic Acids Res [Internet].* 2017 Feb 21; Available from: <https://academic.oup.com/nar/article/3038232/A>
  89. Ellis MJ, Carfrae LA, Macnair CR, Trussler RS, Brown ED, Haniford DB. Silent but deadly: IS200 promotes pathogenicity in *Salmonella Typhimurium*. *RNA Biol.* 2018 Feb 1;15(2):176–81.
  90. Dar D, Sorek R. Bacterial Noncoding RNAs Excised from within Protein-Coding Transcripts. *MBio [Internet].* 2018 Sep 25;9(5). Available from: <http://dx.doi.org/10.1128/mBio.01730-18>
  91. Miyakoshi M, Chao Y, Vogel J. Regulatory small RNAs from the 3' regions of bacterial mRNAs. *Curr Opin Microbiol.* 2015 Apr;24(0):132–9.
  92. Chao Y, Vogel J. A 3' UTR-Derived Small RNA Provides the Regulatory Noncoding Arm of the Inner Membrane Stress Response. *Mol Cell [Internet].* 2016; Available from: <http://www.sciencedirect.com/science/article/pii/S1097276515010059>
  93. Guo MS, Updegrove TB, Gogol EB, Shabalina SA, Gross CA, Storz G. MicL, a new  $\sigma$ E-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. *Genes Dev.* 2014 Jul 15;28(14):1620–34.
  94. Papenfort K, Said N, Welsink T, Lucchini S, Hinton JCD, Vogel J. Specific and

- pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small RNA. *Mol Microbiol.* 2009 Oct;74(1):139–58.
95. Naville M, Gautheret D. Transcription attenuation in bacteria: theme and variations. *Brief Funct Genomics.* 2010 Mar;9(2):178–89.
  96. Loh E, Dussurget O, Gripenland J, Vaitkevicius K, Tiensuu T, Mandin P, et al. A trans-acting riboswitch controls expression of the virulence regulator PrfA in *Listeria monocytogenes*. *Cell.* 2009 Nov 13;139(4):770–9.
  97. Choi E, Han Y, Cho Y-J, Nam D, Lee E-J. A trans-acting leader RNA from a *Salmonella* virulence gene. *Proc Natl Acad Sci U S A.* 2017 Sep 19;114(38):10232–7.
  98. Gottesman S, Storz G. Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harb Perspect Biol* [Internet]. 2011 Dec;3(12). Available from: <http://dx.doi.org/10.1101/cshperspect.a003798>
  99. Lee T, Feig AL. The RNA binding protein Hfq interacts specifically with tRNAs. *RNA.* 2008 Mar;14(3):514–23.
  100. Zhang A, Wassarman KM, Rosenow C, Tjaden BC, Storz G, Gottesman S. Global analysis of small RNA and mRNA targets of Hfq. *Mol Microbiol.* 2003 Nov;50(4):1111–24.
  101. Lalaouna D, Carrier M-C, Semsey S, Brouard J-S, Wang J, Wade JT, et al. A 3' External Transcribed Spacer in a tRNA Transcript Acts as a Sponge for Small RNAs to Prevent Transcriptional Noise. *Mol Cell.* 2015 May 7;58(3):393–405.
  102. Marck C, Grosjean H. tRNomics: analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. *RNA.* 2002 Oct;8(10):1189–232.
  103. Bösl M, Kersten H. A novel RNA product of the *tyrT* operon of *Escherichia coli*. *Nucleic Acids Res.* 1991 Nov 11;19(21):5863–70.
  104. Andrade JM, Dos Santos RF, Chelysheva I, Ignatova Z, Arraiano CM. The RNA-binding protein Hfq is important for ribosome biogenesis and affects translation fidelity. *EMBO J* [Internet]. 2018 Jun 1;37(11). Available from: <http://dx.doi.org/10.15252/embj.201797631>
  105. Cvijović I, Nguyen Ba AN, Desai MM. Experimental Studies of Evolutionary Dynamics in Microbes. *Trends Genet.* 2018 Sep;34(9):693–703.