Karen M. Wassarman and Patricia J. Kiley

Published Ahead of Print 23 October 2009.

Updated information and services can be found at:
http://jb.asm.org/content/192/1/26

These include:

**REFERENCES**
This article cites 16 articles, 4 of which can be accessed free at:
http://jb.asm.org/content/192/1/26#ref-list-1

**CONTENT ALERTS**
Receive: RSS Feeds, eTOCs, free email alerts (when new articles cite this article), more»

Information about commercial reprint orders: http://journals.asm.org/site/misc/reprints.xhtml
To subscribe to another ASM Journal go to: http://journals.asm.org/site/subscriptions/
Global Approaches for Finding Small RNA and Small Open Reading Frame Functions

Karen M. Wassarman and Patricia J. Kiley

Department of Bacteriology and Department of Biomolecular Chemistry
University of Wisconsin—Madison, Madison, Wisconsin 53706

Sequencing of the first Escherichia coli (K-12) genome revealed that there were ~4,300 open reading frames (ORFs) expected to encode proteins and many stable RNA genes that encode tRNA and rRNA (3). Surprisingly, at that time only 43% of the ORFs had been previously described, and 38% could not be assigned even predicted functions. These observations suggested that there was a wealth of information still awaiting discovery even in this extremely well-studied model organism, which spurred great interest and focused studies on the ORFs of unknown function. Over the past decade it also has become increasingly apparent that there is an abundance of additional genes that play important roles in cellular physiology; these genes encode small RNAs (sRNAs) and small proteins (small ORFs [sORFs], defined as ≤50 amino acids) that were overlooked in the initial annotations due to their small size and/or lack of open reading frames (8, 16). Most of the sRNA and sORF genes are currently of unknown function; however, sRNAs and sORFs of known function often have regulatory roles, frequently in signal transduction pathways and in coordinating regulatory networks (6, 8, 16). Therefore, there is great interest in these gene classes and the identification of their functions, and two papers in this issue of the Journal of Bacteriology, by Hobbs et al. (9) and Hemm et al. (7), present new genome-based approaches to search for functions of sRNAs and sORFs.

Although directed genetic and genomic approaches (1, 2) have been quite successful in identifying functions for protein-encoding genes and in reducing the number of ORFs of unknown function (10), these approaches have been less profitable for identifying sRNA and sORF gene functions. Loss of function of regulatory sRNA-encoding genes typically leads to more subtle phenotypes than commonly observed when regulatory proteins are mutated, perhaps due to the observation that sRNA regulation is usually more modulatory in nature than that of their protein counterparts that commonly direct large-scale changes (16). As a consequence, phenotypes associated with mutations in sRNA genes can be difficult to recognize without additional information pointing toward a specific time, event, or condition to explore. In addition, sRNA and sORF genes are significantly smaller targets for traditional genetic mutation, and sRNA genes are not subject to frameshift or nonsense mutations (since they do not code for proteins), making it more difficult to generate loss-of-function alleles by general mutation. Therefore, it is perhaps not surprising that sRNA and sORF genes have not been identified readily in classical genetic screens or selections. More global approaches, such as mRNA expression profiling and proteomic studies, also have been extremely powerful for characterization of many ORFs and have led to identification of many cellular functions. However, once again such approaches have not been as successful for studies of sRNAs and sORFs, due in part to the fact that most commercially available microarrays do not include representation of these genes, as they have only been recently identified, and that general protocols used for proteomic studies, including mass spectrometry or two-dimensional gel electrophoresis, are not well suited for studying small proteins.

The two highlighted papers from Gisela Storz’s group, the reports of Hobbs et al. (9) and Hemm et al. (7), tackle the problem of characterization and identification of function for sRNAs and sORFs, respectively, and present their progress in large-scale directed studies especially suited to the study of these genes. Importantly, the approaches described in these papers are also generally applicable to functional studies of any unknown gene in E. coli or other microorganisms.

In the first paper by Hobbs et al. (9), the authors demonstrate the power of analyzing DNA bar-coded mutants in mixed population studies by uncovering mutant phenotypes associated with several small RNAs and small proteins. For these experiments, strains were generated in which genes of interest were individually replaced by bar codes that were originally designed for construction of mutant libraries in yeast (5, 11). The bar code contains two 20-mer sequences (UP and DN) that are unique for each mutation generated. Flanking the specific UP and DN sequences are three additional sequence elements (Fig. 1, common sequences com1, -2, and -3) that are identical for all bar codes. These “common” sequences serve as priming sites for PCR amplification of the UP or DN bar codes from each strain (Fig. 1, arrows 1a and 1b for UP and 2a and 2b for DN). Hence, isolation of genomic DNA from a mixed population of mutant cells allows the amplification of bar codes from each strain (Fig. 1, arrows 1a and 1b for UP and 2a and 2b for DN). Thus, isolation of genomic DNA from a mixed population of mutant cells allows the amplification of the UP and DN barcodes using just two sets of primers (Fig. 1).

The relative abundance of each bar code within the PCR products should be representative of the relative abundance of each strain within the population. In this way, each mutant can be independently tracked, even in a culture containing many different genotypes, by hybridization of amplified bar code DNA to a commercially available DNA array designed to detect the bar codes. Note that each bar code has two sequence elements (UP and DN) that can be analyzed independently.

The bar-coded mutant approach clearly facilitates analysis of a large number of mutants at once, providing quantitative representation of each strain present. However, in addition, it
is the competition for growth within a mixed population that is likely to be critical for enhancing detection of subtle mutant phenotypes commonly associated with loss of sRNAs. For example, competitive growth experiments were required to reveal altered phenotypes for cells lacking 6S RNA in previous work (13). Thus, the use of bar-coded mutants dramatically expands the potential to search for phenotypes, even when individual mutant cells are a minority population in a complex mixture of cells, demonstrating an important advance in genetic approaches for identifying sRNA and small protein function.

Tracking individual DNA bar codes in a mixed cell population allowed the authors to specifically screen their library for strains that showed either increased or decreased resistance to their test conditions: acid or cell envelope stress. Previous studies showed that several sRNAs regulate synthesis of outer membrane proteins (15) and 70% of the small proteins are membrane localized (8), providing a rationale that the selected stress conditions to assess when these small proteins might be functional. In addition, for several of these proteins, the mRNA transcripts were also studied and transcription factors were identified that mediate the observed regulation. Overall, the authors found that 21 of 51 proteins tested were induced under at least one condition tested (heat shock, oxidative stress, zinc limitation, oxygen limitation, acid or envelope stress, or changes in carbon source). Remarkably, the levels of over half of these proteins were increased during heat shock, suggesting that these small proteins may play a specific role in the response to increased temperatures.

In summary, the two highlighted papers from this issue demonstrate the successful use of two approaches that can be applied globally to learn more about the functions of sRNAs and small proteins. Extension of these approaches to additional growth or stress conditions should provide an even richer data set to fully comprehend the physiological function of these exciting gene products.

REFERENCES
The views expressed in this Commentary do not necessarily reflect the views of the journal or of ASM.