

## Large Structured Non-coding RNAs Revealed by Bacterial Metagenome Analysis

Zasha Weinberg<sup>†,‡</sup>, Jonathan Perreault<sup>‡</sup>, Michelle M. Meyer<sup>‡</sup> and Ronald R. Breaker<sup>†, ‡, §,\*</sup>

<sup>†</sup>Howard Hughes Medical Institute, <sup>‡</sup>Department of Molecular, Cellular and Developmental Biology, <sup>§</sup>Department of Molecular Biophysics and Biochemistry, Yale University, Box 208103, New Haven, CT 06520-8103, USA.

Estimates of the total number of bacteria species suggest that existing DNA sequence databases carry only a tiny fraction of the total amount of novel DNA sequence space present in this division of life. Indeed, environmental DNA samples have been shown to encode numerous examples of previously unknown proteins and RNAs. Bioinformatics searches of genomic DNA from known bacteria commonly identify novel non-coding RNAs such as riboswitch candidates<sup>1</sup>. In rare instances, RNAs that exhibit extraordinary sequence and structural conservation across a wide range of bacteria are encountered. Given that large structured RNAs are known to carry out complex biochemical functions such as protein synthesis and RNA processing reactions, identifying additional RNAs of great size and intricate structure are likely to reveal additional complex functions that can be achieved by RNA.

We applied an updated computational pipeline<sup>2</sup> to discover<sup>3</sup> non-coding RNAs that rival the known large ribozymes in size and structural complexity or that are among the most abundant RNAs in bacteria that encode them. These RNAs would have been difficult or impossible to detect without examining environmental DNA sequences, suggesting that numerous large and highly-structured RNAs remain to be discovered in the vastness of unexplored bacterial sequence space.

1. Weinberg, Z. *et al.* Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res.* **35**, 4809-4819.
2. Tseng, H. H., Weinberg, Z., Gore, J., Breaker, R. R. & Ruzzo, W. L. Finding non-coding RNAs through genome-scale clustering. *J. Bioinform. Comput. Biol.* **7**, 373-88 (2009).
3. Weinberg, Z., Perreault, J., Meyer, M. M. & Breaker, R. R. Large structured non-coding RNAs revealed by bacterial metagenome analysis. (manuscript in preparation).