Supplemental Text File S1

Classification of the predicted sRNAs. Experimentally-tested sRNAs were computationally classified into Rfam sRNA families (Supplemental Tables S3 and S4) using GraPPLE, which uses the information available from graph properties and the predictive power of support vector machines to classify RNA sequences (1). This computational classification showed a large diversity among the predicted sRNAs. New members were identified even for some well studied sRNA families. For example, sCLK_3642, sCAC1132 and sCAC3825 were identified as new riboswitches in *C. botulinum A3 Loch Maree* and *C. acetobutylicum*. GraPPLE gives a probability for each sRNA family that it classifies the sequence into. If we use a probability of 0.5 as a confident classification threshold, 22 out of 51 experimentally tested (by Q-RT-PCR and/or Northern analysis) sequences could not be classified into any known family with a probability above this threshold. Among these 22 sequences, 10 were validated by either Q-RT-PCR and/or Northern analysis. Because Grapple results are sensitive to the input sequences, some of these unclassifiable sequences were probably due to the inaccurate predictions of sRNA length or boundaries.

Reference for this supplemental file

1. **Childs, L., Z. Nikoloski, P. May, and D. Walther.** 2009. Identification and classification of ncRNA molecules using graph properties. Nucleic Acids Res **37:**e66.