Expression profiling of hypothetical genes in *Desulfovibrio vulgaris* leads to improved functional annotation

Dwayne A. Elias[1], Elliott C. Drury, Alyssa M. Redding, Aindrila Mukhopadhyay, Marcin Joachimiak, Huei-Che B. Yen, Matthew W. Fields[2], Terry C. Hazen, Adam P. Arkin, Jay D. Keasling, Judy D. Wall[1]

[1]University of Missouri, [2]University of Montana, Lawrence Berkeley National Laboratory, Virtual Institute for Stress and Survival (VIMSS), Berkeley, CA

Acknowledgements

This work was part of the Virtual Institute for Microbial Stress and Survival (http://VIMSS.lbl.gov) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

Abstract

Hypothetical and conserved hypothetical genes account for >30% of sequenced bacterial genomes. For the sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough, 348 of the 3534 genes are annotated as conserved hypothetical (9.7%) with 889 hypothetical genes (25.0%). Given this large genome fraction, it is plausible that some genes serve critical cellular functions. The goals of this study were to determine which genes can be expressed and to provide a more functionally based annotation. To accomplish this, expression profiles of the 1237 hypothetical and conserved hypothetical genes were obtained from transcriptomic datasets of 10 environmental stresses, complimented with iTRAQ proteomic data. Genes were divided into putatively polycistronic operons and those predicted to be monocistronic, then classified by basal expression levels and grouped according to changes in expression for one or multiple stresses. 1219 of these genes were transcribed with 265 proteins detected. There was no evidence for expression of 17 predicted genes. Except for the latter, annotation of all monocistronic genes was expanded using the above criteria and COG information. Polycistronic genes were annotated with expression information including proximity to more confidently annotated genes. Two targeted deletion mutants were used as test cases to determine the accuracy of the inferred functional annotations.