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

ESPP2 Environmental Stress Pathway Project

Dwayne A. Elias, Elliott C. Drury, Alyssa M. Redding, Aindrila Mukhopadyay, Huei-Che B. Yen, Kat H. Huang, Terry C. Hazen, Adam P. Arkin, Judy D. Wall

DOE GENOMICS:GTL ACCELERATING DISCOVERY FOR ENERGY AND ENVIRONMENT

VIMSS Virtual Institute for Microbial Stress and Survival | Massachusetts Institute of Technology | Missouri | University of Washington | Montana State University | Southern Illinois University Carbondale | Oklahoma | Sandia National Laboratories | Oak Ridge National Laboratory | Berkeley | http://vimss.lbl.gov/

ABSTRACT

Hypothetical (HP) and conserved hypothetical (CHP) proteins consistently make up >30% of sequenced bacterial genomes. It is likely that many of these proteins serve significant functions ranging from regulation to presently unknown steps in carbon or electron flux.

Expression profiles for the expected 1237 HP and CHP in *D. vulgaris* were obtained from VIMSS/ESPP transcriptomic and MS-based iTRAQ proteomic datasets of controlled cultures subjected to 10 environmental stresses, comparison with a megaplasmid lacking strain, and two Fur-regulon mutants. The genes were divided into two groups; those in polycistronic operons and those that are monocistronic.

We are presently able to confirm the expression at the mRNA level for 1219 genes and for 265 genes at the mRNA and protein level, while there was no evidence for either mRNA or protein detectable for 17 genes. We constructed and tested two deletion mutants; an operon of DVU0303-0304 that responded to multiple stresses and the orphan DVUA0095 that only increased transcription with chromate stress. In both cases the mutants responded as predicted, validating the microarray results and the functional annotations.

EXPRESSION OF HP AND CHP GENES

The sequenced genome of *Dv. vulgaris* suggests that there are 348 CHP and 889 HP. In order to attempt to assign putative functions to these proteins, we first probed all available microarray and proteomics data to eliminate the genes that had never shown expression on either level.

Table 1: Hypothetical and conserved hypothetical proteins with evidence of expression

Current Annotation	Operon	Number of Possible Genes	Transcript Identified	Protein Identified
Hypothetical	Polycistronic	346	343	72
Conserved Hypothetical	Polycistronic	211	201	70
Hypothetical	Monocistronic	543	541	69
Conserved Hypothetical	Monocistronic	137	134	53

Once those hypothetical and conserved hypothetical genes for which expression was not detected were eliminated, those showing expression were grouped according to their microarray expression profiles.

Table 2: Definitions for Gene Grouping based on Microarray Expression Profiles

Category	Definition
No Expression	No record of binding of the RNA to the oligonucleotide of the microarray in any control or stress experiments.
Normal Expression	Differential expression was not detected and was not in the top 1/8 of genes expressed.
High Expression	Differential expression was not detected but was in the top 1/8 of genes expressed.
Differential Expression to One Stress	mRNA level showed a $\log^2$ R value >= 1.2 change relative to the experimental control in greater than 20% of the time points in only one stress condition.
Differential Expression to Multiple Stresses	mRNA level showed a $\log^2$ R value >= 1.2 change relative to the experimental control in greater than 20% of the time points in more than one stress condition.

EXPRESSION PROFILING OF HP GENES

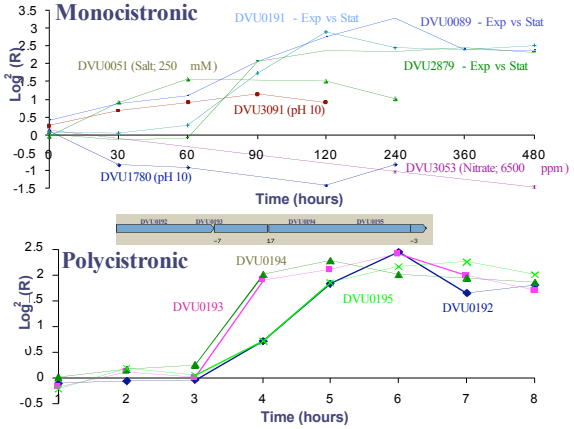

Figure 1: Microarray expression profiling of (top) monocistronic HP with various stresses and (bottom) polycistronic HP from exponential growth vs stationary phase allows for a putative functional assignment using the profile and gene association. Up-regulation of a predicted 4 gene operon DVU0192 (adenine specific DNA methyltransferase) and DVU0194 (terminase) are well conserved while DVU0193 and DVU0195 were annotated as hypothetical proteins.

STRESS RESPONSE DISTRIBUTION OF HP/CHP GENES

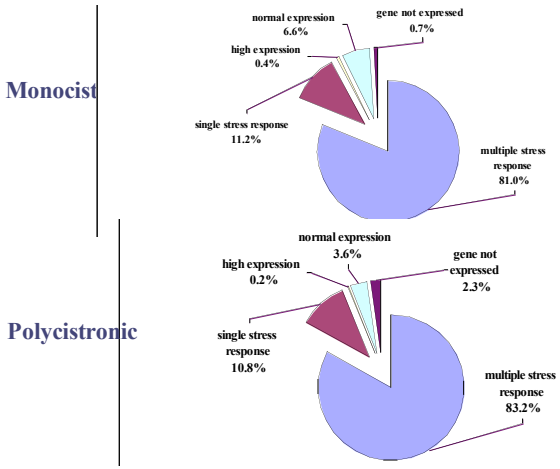

Figure 2: Pie charts showing the stress response distribution of all (top) 680 monocistronic HP and CHP genes and (bottom) 557 HP and CHP genes predicted to be in polycistronic operons. In each case the genes displaying a single stress response accounted for ~10% of the genes while >80% were differentially expressed in more than one stress.

USE OF TARGETED MUTANTS TO ASSIGN FUNCTIONS

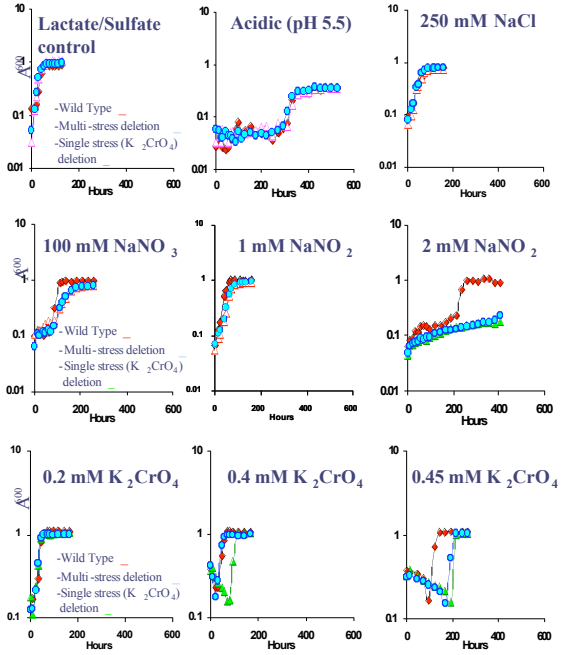

Figure 4: Growth curves of the two targeted deletion mutants (genes DVU0303-0304 that were differentially expressed in 13 and 9 stresses, respectively (●), a deletion of DVUA0095 that was only up-regulated in the presence of chromium (▲) and wild type *D. vulgaris* (♦) under various stress conditions. The phenotype of the deletions under selected stresses compared to their observed expression patterns gives confidence to the datasets as well as the new functional assignments.

FUTURE PLANS

1) Test functional assignments via deletion of single stress genes, followed by multiple stress genes.

2) Find genes in other organisms that are closely related to single stress HP and CHP genes in *D. vulgaris* (≥98% amino acid sequence similarity) and delete/impair them to determine if revised annotations are valid in other organisms.

ACKNOWLEDGEMENT

ESPP2 (MDCASE) is part of the Virtual Institute for Microbial Stress and Survival (VIMSS) 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.