Researchers Utilize Proteomics to Reveal a Core Proteome

SJ Callister,[(a)] LA McCue,[(a)] JE Turse,[(a)] ME Monroe,[(a)] KJ Auberry,[(b)] RD Smith,[(a)] JN Adkins,[(a)] and MS Lipton[(a)]

(a) Pacific Northwest National Laboratory, Richland, Washington

(b) EMSL, Richland, Washington

Researchers leveraged proteomics technology at EMSL to reveal the existence of a core proteome among 17 diverse bacteria. While many researchers make genomic comparisons among different bacteria and often report a set of common genes or core genome (Koonin 2003), the expression of this core genome as a core proteome is not generally verified. The verification of these genes as proteins has important implications to defining a set of basal proteins important to bacterial life that could aid in the construction of synthetic life-like systems, or synthetic biology. Ultimately, this research demonstrated the ability to use proteomic data in a comparative manner outside of conventional norms.

Enabled by a proteome database that encompassed ~967,000 experimentally determined unique peptides linked to specific protein information and publicly available genome sequences, the observation of proteins predicted from genomic comparisons among 17 environmental and pathogenic bacteria was investigated (Callister et al. 2008). Bacteria selected for this investigation included the metabolically diverse organisms *Rhodobacter sphaeroides*, *Shewanella oneidensis*, and *Synechocystis sp.* PCC6803, as well pathogens such as *Yersina pestis* and *Salmonella typhimurium*. Facilitated by successful collaborations that have made samples available for proteomic analysis, this investigation represents the gathering and evaluation of proteomic measurements made over the past six years.

Genomic comparisons among the 17 bacteria predicted the existence of a core genome composed of 144 genes (Figure 1). Proteins from 74% of these genes were observed within the database, with each protein identified by two unique peptides. The large percentage of the core genome being observed surprised researchers because of the somewhat diverse make-up of the bacteria selected, and the different number and variety of conditions used to culture the organisms. This observation has led to the hypothesis that although the number of genes making up a core genome may expand or contract depending on the number and diversity of organisms included, the percent of these genes being expressed as proteins in nature will be high.

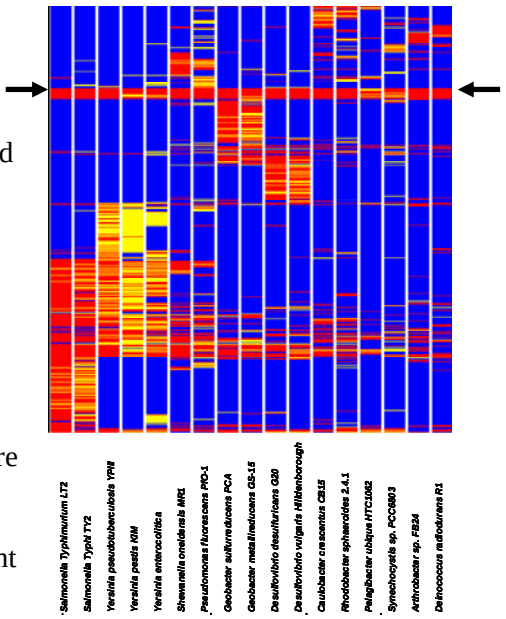

Figure 1. *The core proteome is revealed for 17 bacteria. Genomic comparisons identified genes common to two or more bacteria (orange) resulting in a core genome of 144 genes. Proteomic measurements were used to verify the existence of these genes as proteins (red) resulting in the identification of a core proteome (arrow).*

A functional analysis revealed that a majority of core proteome proteins (~55%) have functions related to protein synthesis, not surprising as the ability of a bacterial cell to make proteins for cell maintenance and growth is a vital function. However, what was surprising to researchers was the observation of

proteins (~7%) having very little or no functional characterization; for example, the observation of the iojap-like protein. This gene was predicted in our study as homologous across all 17 bacteria and is also found outside the bacterial domain as well. The product of this gene is a small protein so the number of detectable peptides is also potentially small relative to some of the larger ribosomal proteins we observed. Yet, little is known about this protein's function in the bacterial cell (Galperin and Koonin 2004). The observation of these relatively uncharacterized proteins emphasizes the need for a better understanding of basal bacterial functions.

Citations:

Callister SJ, LA McCue, JE Turse, ME Monroe, KJ Auberry, RD Smith, JN Adkins, and MS Lipton. 2008. "Comparative Bacterial Proteomics: Analysis of the Core Genome Concept." *PLoS ONE* 3(2):e1542.

Galperin MY, and EV Koonin. 2004. "'Conserved Hypothetical' Proteins: Prioritization of Targets for Experimental Study." *Nucleic Acids Research* 32(18):5452-5463.

Koonin EV. 2003. "Comparative Genomics, Minimal Gene-Sets and the Last Universal Common Ancestor." *Nature Reviews Microbiology* 1(2):127-136.