

Title: A proteomic survey of diverse gut microbes

Ernesto S Nakayasu¹, Meagan C Burnett¹, Stephen J Callister¹, **Samuel H Payne^{1*}**
(samuel.payne@pnnl.gov)

1. Biological Sciences Division, Pacific Northwest National Laboratory, Richland WA

Project Goals: This project is focused on improving algorithms and methods for mass spectrometry data analysis of metaproteomics data. Recent advances in mass spectrometry and biological separations have dramatically increased the depth of proteomic discovery. Unfortunately, traditional computational workflows are in many cases preventing researchers from realizing these benefits for microbial communities. We propose to create a new generation of computational workflows to overcome the sensitivity limitations inherent in status quo data processing schemes.

To advance our ability to annotate tandem mass spectrometry data from microbial communities, our project has been developing algorithms to match spectra from metaproteomics experiments to a library of annotated spectra. With significant improvements having been achieved in the algorithms, the next step towards a fully functioning pipeline is to greatly expand our library of annotated spectra, specifically the diversity of the microbes present in the library. Consistent with our focus on biofuels and microbial communities that degrade plant feedstock, we are specifically targeting gut microbes such as those that live in cow rumen or beetle gut. Our first release of the PNNL Biodiversity Library¹ contained over 100 bacteria and archaea from 15 phyla, but had few organisms representing this important ecological niche. Therefore this year, we have collected global proteomics data from 10 organisms in firmicute and bacterioidetes and are in the process of identifying an additional 40 organisms for future data collection. Several organisms have been analyzed with a variety of complex media conditions to understand the metabolic adaptation to different nutrients. Dramatic advances in mass spectrometry instrumentation allow us to sample the proteome more deeply with less instrument time, enabling a greater survey of biodiversity. We have begun to analyze the data by annotating proteins into KEGG, both to look at expression across pathways, but also to help identify orthologs across species.

References:

1. The Pacific Northwest National Laboratory library of bacterial and archaeal proteomic biodiversity. Payne SH, Monroe ME, Overall CC, Kiebel GR, Degan M, Gibbons BC, Fujimoto GM, Purvine SO, Adkins JN, Lipton MS, Smith RD. *Sci Data*. (2015) 2:150041. doi: 10.1038/sdata.2015.41. PMID: 26306205

This project was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Early Career Research Program.