

DATA MANAGEMENT PLAN

Three main types of data will be generated in the proposed activities: Bacterial isolates, gene sequence data and physiological data. I am committed to making all data types publicly available through peer-reviewed publications and public databases.

Bacterial Isolates: All bacterial isolates will be stored in glycerol freezer stocks and will be made available upon request. Pure cultures used for genome sequencing, growth experiments, and gene expression studies will be deposited in publicly accessible culture collections, such as the American Type Culture Collection (ATCC).

Sequence Data: Newly determined nucleotide and amino acid sequence data will be deposited in public databases and GenBank/EMBL/DDBJ accession numbers will be included in all relevant manuscripts. We have developed an in house set of bioinformatic tools and tracking software to sequence, assemble and annotate genomes. These tools, which were developed to sequence the complete genome of an uncultured microbe (Iverson et al., 2012), are now hosted on the web and will be used to sequence, assemble and annotate the complete genome of our Arctic96BD-19 isolates (GSO-PS1 and GSO-NP1).

Physiological Data: Growth experiments will be conducted to evaluate sulfur oxidation, carbon utilization on different substrates. Growth rate, cell densities, gene expression and substrate utilization data will be collected in these experiments. These data, which are critical to the proposed study and future studies, will be made available through peer-reviewed publications and included with data deposited in public databases (ATCC and GenBank).