Normalized protein abundance data and protein annotations for proteomic data from laboratory cultures of Ruegeria pomeroyi DSS-3 and Alteromonas macleodii MIT1002 in 2022

Website: https://www.bco-dmo.org/dataset/927507
Data Type: experimental
Version: 1
Version Date: 2024-05-14

Project
» C-CoMP Model Bacteria Physiological Studies (C-CoMP Model Bacteria)

Program
» Center for Chemical Currencies of a Microbial Planet (C-CoMP)

Contributors	Affiliation	Role
Saito, Mak A.	Woods Hole Oceanographic Institution (WHOI)	Principal Investigator
Moran, Mary Ann	University of Georgia (UGA)	Co-Principal Investigator
Cooper, Zachary Shane	University of Georgia (UGA)	Scientist
McIlvin, Matthew R.	Woods Hole Oceanographic Institution (WHOI)	Technician
Gray, Laura	Woods Hole Oceanographic Institution (WHOI)	Data Manager
York, Amber D.	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract
This dataset includes normalized protein abundance data and protein annotations for proteomic data from cultures of Ruegeria pomeroyi DSS-3 and Alteromonas macleodii MIT1002. These model marine bacteria were grown in defined culture media with either glucose, acetate, or a mix of both as carbon substrates. The data are sampled so as to capture the metabolic differences the bacteria employ when catabolizing these different substrates and when switching between them. The raw proteomics files are available on the Proteomics IDEntification Database (PRIDE) under accession PXD045824. The proteomic data accompanies the transcriptomic expression data available at BCO-DMO dataset 916134 (https://www.bco-dmo.org/dataset/916134).

Table of Contents

Coverage

Temporal Extent: 2022-09-08 - 2022-09-30

Methods & Sampling

Under well-defined laboratory conditions, we grew *R. pomeroyi* DSS-3 and *A. macleodii* MIT1002 in batch cultures on a monosaccharide (glucose) and organic acid (acetate), provided either individually or in combination, and all at the same carbon equivalent. This batch culturing approach mimicked bacterial DOC assimilation in short-lived substrate 'hot spots', such as those formed by high phytoplankton extracellular release at peak photon availability. Measurements were made of bacterial metabolite uptake, respiration, and biomass accumulation through a growth cycle. Insights into bacterial core metabolism came from gene and protein expression measured at intervals during growth. Curated genome-scale models (flux balance analysis; FBA) were used to explore the metabolic foundation of CO_2 production for insights into determinants of BGE and bCUE.

Samples were collected for proteomic analysis during the exponential growth phase from liquid cultures. Proteomic samples were pelleted by centrifugation, frozen at –80°C, and analyzed at the Bioinorganic Chemistry Laboratory at the Woods Hole Oceanographic Institution (WHOI).

Protein extracts from the biological triplicates were analyzed by liquid chromatography-mass spectrometry (LC-MS) (Michrom Advance HPLC coupled to a Thermo Scientific Fusion Orbitrap mass spectrometer with a Thermo Flex source). Each sample was concentrated onto a trap column (0.2 x 10 mm ID, 5 µm particle size, 120 Å pore size, C18 Reprosil-Gold, Dr. Maisch GmbH) and rinsed with 100 µL 0.1% formic acid, 2% acetonitrile (ACN), 97.9% water before gradient elution through a reverse phase C18 column (0.1 x 500 mm ID, 3 µm particle size, 120 Å pore size, C18 Reprosil-Gold, Dr. Maisch GmbH) at a flow rate of 250 nL/min. The chromatography consisted of a nonlinear 190 min gradient from 5% to 95% buffer B, where A was 0.1% formic acid in water and B was 0.1% formic acid in ACN (all solvents were Fisher Optima grade). The mass spectrometer was set to perform MS scans on the orbitrap (240000 resolution at 200 m/z) with a scan range of 380 m/z to 1580 m/z. MS/MS was performed on the ion trap using data-dependent settings (top speed, dynamic exclusion 15 seconds, excluding unassigned and singly charged ions, precursor mass tolerance of ±3ppm, with a maximum injection time of 50 ms).

Curated genome-scale models (flux balance analysis; FBA) were used to explore the metabolic foundation of CO_2 production for insights into determinants of bacterial growth efficiency (BGE) and bacterial carbon use efficiency (bCUE).

Data Processing Description

Search results were performed using Proteome Discoverer 2.2 (Thermo Scientific) and Scaffold 5.0 (Proteome Software Inc.).

Values for protein abundance were calculated using a normalized spectral abundance factor (NSAF). NSAF for each run was calculated as the spectral count for a peptide divided by the length of that peptide which was then divided by the sum of all individual spectral counts divided by their protein lengths. The spectral count for each peptide was multiplied by the NSAF value for that peptide for normalization which allowed us to compare protein abundances between samples (Saunders et al. 2022). Proteins were counted as present in a cell if the normalized abundance for that protein was above a cutoff of the fifth percentile of abundances, to avoid counting very low abundance or falsely assigned peptides.

Study location: This lab study took place in Athens, GA, USA within the Department of Marine Sciences at University of Georgia, U.S.A. Samples destined for proteomics analyses were shipped to the Saito Laboratory at Woods Hole Oceanographic Institution in Cape Cod, MA, U.S.A.

Organism identifiers [culture, *Genus species*, Lifesciences Identifier (LSID)]:
Ruegeria pomeroyi DSS-3, *Ruegeria pomeroyi*, urn:lsid:marinespecies.org:taxname:567965
Alteromonas macleodii MIT1002, *Alteromonas macleodii,* urn:lsid:marinespecies.org:taxname:742978

BCO-DMO Processing Description

* submitted filed "DSS3_CCoMP_CUE_norm_prot_abund_20231013.txt" and "MIT1002_CCoMP_CUE_norm_prot_abund_20231013.txt" were imported into the BCO-DMO data system for this dataset with "nan" as the missing data identifier.
** Missing data values are displayed differently based on the file format you download. They are blank in csv files, "NaN" in MatLab files, etc.

* Column names adjusted to conform to BCO-DMO naming conventions designed to support broad re-use by a variety of research tools and scripting languages. [Only numbers, letters, and underscores. Can not start with a number]

* Values "#N/A" removed from dataset.

* additional columns added "Species" and "Strain" to distinguish rows that came from "DSS3_CCoMP_CUE_norm_prot_abund_20231013.txt" and "MIT1002_CCoMP_CUE_norm_prot_abund_20231013.txt"

* columns of like type combined between the two tables (E.g. ac_mean_abund composed of "MIT1002_ac_mean_abund" rows and "DSS3_ac_mean_abund" with the strain now indicated in a separate column.

[table of contents | back to top]

Data Files

File
927507_v1_norm_prot_abund.csv (Comma Separated Values (.csv), 1.22 MB) MD5:7cc5b7b934fbea4c4304bc09f5120191 Primary data file for dataset ID 927507, version 1

[table of contents | back to top]

Related Publications

Proteome Software Inc. (n.d.) Scaffold 5. [Software]. Available from https://www.proteomesoftware.com/products/scaffold-5
Software

Saunders, J. K., McIlvin, M. R., Dupont, C. L., Kaul, D., Moran, D. M., Horner, T., Laperriere, S. M., Webb, E. A., Bosak, T., Santoro, A. E., & Saito, M. A. (2022). Microbial functional diversity across biogeochemical provinces in the central Pacific Ocean. Proceedings of the National Academy of Sciences, 119(37). https://doi.org/10.1073/pnas.2200014119
Methods

Thermo Fisher Scientific Inc. (n.d.) Proteome Discoverer (catalog number: OPTON-31105) [Software]. Available from https://www.thermofisher.com/order/catalog/product/OPTON-31105
Software

[table of contents | back to top]

Related Datasets

IsRelatedTo

Moran, M. A., Cooper, Z. S. (2023) **Metadata for transcriptomic expression data from cultures of Ruegeria pomeroyi DSS-3 and Alteromonas macleodii MIT1002 grown in defined culture media with either glucose, acetate, or a mix of both as carbon substrates.** Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 1) Version Date 2023-12-06 doi:10.26008/1912/bco-dmo.916134.1 [view at BCO-DMO]
Relationship Description: These proteomic data accompany the transcriptomic expression data "Substrate-specific metabolic responses of model marine bacteria" (https://www.bco-dmo.org/dataset/916134).

Saito, M. A., McIlvin, M. R. (2023) Dynamic metabolic efficiency of substrate utilization by copiotrophic marine bacteria. Proteomics Identifications Database (PRIDE). URL: https://www.ebi.ac.uk/pride/archive/projects/PXD045824

Parameters

Parameter	Description	Units
Species	Scientific name (Genus species) of the culture	unitless
Strain	Strain identifier of the culture	unitless
gene_callers_id	A numerical identifier unique assigned to each gene in the genome during the gene calling process for organization; not applicable outside of this specific dataset;	unitless
ac_mean_abund	Mean normalized protein abundance for culture grown with acetate	unitless
glc_mean_abund	Mean normalized protein abundance for culture grown with glucose	unitless
late_mean_abund	Mean normalized protein abundance for culture grown with acetate and glucose in late exponential phase growth	unitless
early_mean_abund	Mean normalized protein abundance for culture grown with acetate and glucose in early exponential phase growth	unitless
SPO_ID_ACCESSION	Unique IDs for genes in the genome (historical). This only applies to Ruegeria pomeroy DSS-3 in this dataset. essential for linking protein information to the genome	unitless
KOfam_ACCESSION	Unique KOfam (a customized HMM database of KEGG Orthologs (KOs)) accession ID assigned to each protein annotation	unitless
KEGG_Module	Unique KEGG module accession ID assigned to each protein annotation	unitless
COG20_FUNCTION	Function of protein as identified in the Clusters of Orthologous Groups of proteins (COGs) database	unitless
Uniprot_accession	Unique Uniprot accession ID assigned to each protein annotation, essential for linking protein information to the genome	unitless

Instruments

Dataset-specific Instrument Name	Michrom Advance HPLC
Generic Instrument Name	High-Performance Liquid Chromatograph
Dataset-specific Description	Liquid chromatography-mass spectrometry (LC-MS) (Michrom Advance HPLC coupled to a Thermo Scientific Fusion Orbitrap mass spectrometer with a Thermo Flex source)
Generic Instrument Description	A High-performance liquid chromatograph (HPLC) is a type of liquid chromatography used to separate compounds that are dissolved in solution. HPLC instruments consist of a reservoir of the mobile phase, a pump, an injector, a separation column, and a detector. Compounds are separated by high pressure pumping of the sample mixture onto a column packed with microspheres coated with the stationary phase. The different components in the mixture pass through the column at different rates due to differences in their partitioning behavior between the mobile liquid phase and the stationary phase.

Dataset-specific Instrument Name	Thermo Scientific Fusion Orbitrap mass spectrometer
Generic Instrument Name	Mass Spectrometer
Dataset-specific Description	Liquid chromatography-mass spectrometry (LC-MS) (Michrom Advance HPLC coupled to a Thermo Scientific Fusion Orbitrap mass spectrometer with a Thermo Flex source)
Generic Instrument Description	General term for instruments used to measure the mass-to-charge ratio of ions; generally used to find the composition of a sample by generating a mass spectrum representing the masses of sample components.

Project Information

C-CoMP Model Bacteria Physiological Studies (C-CoMP Model Bacteria)

The Center for Chemical Currencies of a Microbial Planet (C-CoMP) is focused on understanding how marine microorganisms determine the fate of labile carbon in the surface ocean. To study these chemical-biological dynamics across a variety of scales and in response to changing environmental conditions, the physiology of marine bacteria that drive chemical exchange must be explored in depth using a variety of microbiological, molecular biological, and integrative 'omics (e.g. proteomics, metabolomics, and genomics) methodologies. This project has been created to host data generated via these methods to investigate the physiological mechanisms underpinning the biogeochemical functions of model marine bacteria.

Program Information

Center for Chemical Currencies of a Microbial Planet (C-CoMP)

Website: https://ccomp-stc.org/

Coverage: North Atlantic, BATS, global/other

Functions carried out by microscopic inhabitants of the surface ocean affect every aspect of life on our planet, regardless of distance from the coast. Ocean phytoplankton are responsible for half of the photosynthesis on Earth, the first step in a complex system that annually withdraws 50 billion metric tons of carbon from the atmosphere to sustain their growth. Of this, 25 billion metric tons participate in a rapid cycle in which biologically reactive material is released into seawater and converted back into carbon dioxide by marine bacteria within hours to days. The chemical-microbe network at the heart of this fast cycle remains poorly constrained; consequently, its primary currencies and controls remain elusive; its sensitivities to changing ocean conditions are unknown; and its responses to future climate scenarios are not predictable. The Center for Chemical Currencies of a Microbial Planet (C-CoMP) integrates research, education and knowledge transfer activities to develop a mechanistic understanding of surface ocean carbon flux within the context of a changing ocean and through increased participation in ocean sciences. C-CoMP supports science teams that merge biology, chemistry, modeling, and informatics to close long-standing knowledge gaps in the identities and dynamics of organic molecules that serve as the currencies of elemental transfer between the ocean and atmosphere. C-CoMP fosters education, outreach, and knowledge transfer activities that engage students of all ages, broaden participation in the next generation of ocean scientists, and extend novel open-science

approaches into complementary academic and industrial communities. The Center framework is critical to this mission, uniquely facilitating an open exchange of experimental and computational science, methodological and conceptual challenges, and collaborations that establish integrated science and education partnerships. With expanded participation in ocean science research and ocean literacy across the US society, the next generation of ocean scientists will better reflect the diverse US population.

Climate-carbon feedbacks on the marine carbon reservoir are major uncertainties for future climate projections, and the trajectory and rate of ocean changes depend directly on microbial responses to temperature increases, ocean acidification, and other perturbations driven by climate change. C-CoMP research closes an urgent knowledge gap in the mechanisms driving carbon flow between ocean and atmosphere, with global implications for predictive climate models. The Center supports interdisciplinary science teams following open and reproducible science practices to address: (1) the chemical currencies of surface ocean carbon flux; (2) the structure and regulation of the chemical-microbe network that mediates this flux; and (3) sensitivity of the network and its feedbacks on climate. C-CoMP leverages emerging tools and technologies to tackle critical challenges in these themes, in synergy with existing ocean programs and consistent with NSF's Big Ideas. C-CoMP education and outreach activities seek to overcome barriers to ocean literacy and diversify participation in ocean research. The Center is developing (1) initiatives to expand ocean literacy in K-12 and the broader public, (2) ocean sciences undergraduate curricula and research opportunities that provide multiple entry points into research experiences, (3) post-baccalaureate programs to transition undergraduates into graduate education and careers in ocean science, and (4) interdisciplinary graduate student and postdoctoral programs that prepare the next generation of ocean scientists. The C-CoMP team includes education faculty who evaluate the impacts of education and outreach activities and export successful STEM initiatives to the education community. C-CoMP is revolutionizing the technologies for studying chemical transformations in microbial systems to build understanding of the outsized impact of microbes on elemental cycles. Open science, cross-disciplinary collaborations, community engagement, and inclusive practices foster strategic advances in critical science problems and STEM initiatives. C-CoMP science, education, and knowledge-transfer themes are efficiently addressed through a sustained network of scientists addressing critical research challenges while broadening the workforce that will tackle multi-disciplinary problems with academic, industrial and policy partners.

This award reflects NSF's statutory mission and has been deemed worthy of support through evaluation using the Foundation's intellectual merit and broader impacts review criteria.

The Program's Data Management Plan (DMP) is available as a PDF document.

Funding

Funding Source	Award
NSF Division of Ocean Sciences (NSF OCE)	OCE-2019589