

Title

Data and scripts associated with the manuscript “Organic Molecules are Deterministically Assembled in River Sediments”

Summary

This data package is associated with the publication “Organic Molecules are Deterministically Assembled in River Sediments” submitted to Scientific Reports (Stegen et al., 2024).

The study applies community ecology methods to dissolved organic matter (DOM) chemistry from variably inundated riverbed sediments to uncover principles governing DOM composition at a reach-scale. This data package documents the workflow used to process and generate the main findings in the manuscript. The R scripts reference the raw, unprocessed Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) data from another data package, available on ESS-DIVE at <https://data.ess-dive.lbl.gov/view/doi:10.15485/1834208>. The scripts then process the raw FTICR-MS data and generate the findings and figures presented in the associated manuscript. In brief, this study demonstrates that DOM assemblages in variably inundated sediments are primarily governed by deterministic variable selection, including sediment moisture effecting the degree of deterministic assembly. See the manuscript for more details pertaining to interpretation and implications of the findings.

This data package is associated with the GitHub repository found at https://github.com/WHONDRS-Hub/ECA_2020_Sed.

Brief Overview of Methods

Sediment samples were collected along variably inundated river regions in a transect of up to 10 locations spanning upstream to downstream. Some sediment was weighed and dried in a series of steps for sediment moisture measurements. The remaining sediment was freeze-dried and extracted with water prior to analysis for dissolved organic carbon (DOC; measured as non-purgeable organic carbon; NPOC), total dissolved nitrogen (TN), and Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS). The dissolved organic matter (DOM) characteristics served as inputs to the ecological null modeling, which was used to quantify the relative influences of deterministic and stochastic processes over the DOM assemblages. This data package contains the scripts, figures, and processed FTICR-MS data. See the methods section of the manuscript for more details.

Critical Details

For those who wish to re-run any of the scripts, once FTICR data are processed, you can run the scripts in the following order:

- **01_fticr_blanks_github.r:** This file cleans the FTICR data so it is free of contamination. The outputs are stored in the repository, so it does not need to be rerun.
- **02_Generate_MCD_github.R:** This file generates molecular characteristics dendrograms (MCDs) for each field site. The outputs are stored in the repository, so it does not need to be rerun.
- **03_FTICR_bNTI_create-nulls_github.R:** This file generates null model runs. The outputs from this file are not stored in the repository due to a large number of files.

- **04_FTICR_bNTI_merge-nulls_github.R:** This file compiles null model runs. The outputs are stored in the repository, so it does not need to be rerun.
- **05_ECA_2020_ICR_Null_Figures_Stats_github.r:** This file conducts analyses and generates figures. Figures in the paper are generated by this script.
 - The files ECA2_FTICR_BetaDisp.csv and VGC_texture.csv need to be in the repository for a piece of the code to run, but the contents are not used for any analysis associated with the manuscript.
 - The file merged_weights.csv contains moisture content data in the format used by the scripts, which deviates from the format used on the ESS-DIVE data package. It is read in by the scripts below.

Data Package Structure

This data package is comprised of 6 scripts and 7 folders. The file-level metadata file (file ending in “flmd.csv”) lists all files contained in this data package and descriptions for each. The data dictionary (file ending in “dd.csv”) describes all tabular data columns and their respective definitions and units.

The FTICR_Processing_Scripts produce the outputs found in the “Processed_Data” folder. The remaining scripts (located in the parent directory) produce the outputs found in the following four folders: (1) “MCD_Dendrograms”, “MCD_Randomizations”, “MCD_bNTI_Outcomes”, and “OM_Null_Modeling”. The fifth script additionally takes the three comma-separated values (CSV) files found in the parent directory as input (“VGC_texture.csv”, “merged_weights.csv”, and “ECA2_FTICR_BetaDisp.csv”). The outputs of each of the five scripts serve as the input to the following script, with the final outputs stored in the folder “OM_Null_Modeling”.

Citations and Acknowledgements

Cite this data package with the appropriate DOI. Cite the associated manuscript in any work that that uses analyses or conclusions presented in the manuscript. To cite the paper:

James C. Stegen, Vanessa A. Garayburu-Caruso, Robert E. Danczak, Rosey Chu, Amy E. Goldman, Sophia McKeever, Lupita Renteria, Jason Toyoda, and WHONDRS Consortium. *Organic Molecules are Deterministically Assembled in River Sediments*. 2024. Scientific Reports.

This research was supported by the U.S. Department of Energy (DOE) Office of Science, Early Career Program. PNNL is operated by Battelle Memorial Institute for the DOE under Contract No. DE-AC05-76RL01830.

FTICR-MS data were generated at the Environmental Molecular Sciences Laboratory, a DOE BER User Facility (EMSL; <https://ror.org/04rc0xn13>), under the EMSL User Proposal 51399.

Citations:

- Garayburu-Caruso V A ; Goldman A E ; Chu R ; Danczak R E ; McCall M L ; McKeever S A ; Renteria L ; Tolic N ; Torgeson J M ; Toyoda J G ; Stegen J C ; WHONDRS Consortium T (2021): FTICR, NPOC, TN, and Moisture of Variably Inundated Sediment across 48 North American Rivers. Early Career Research Program: Watershed Perturbation-Response Traits Derived Through Ecological Theory - Worldwide Hydrobiogeochemistry Observation Network for Dynamic River Systems (WHONDRS),

ESS-DIVE repository. Dataset. doi:10.15485/1834208 accessed via <https://data.ess-dive.lbl.gov/datasets/doi:10.15485/1834208> in June 2021.

Contact

James Stegen, james.stegen@pnnl.gov

Change History

Data Package Version	Changes
Version 1 <i>May 2024</i>	Original data package publication