Title

WHONDRS Summer 2019 Sampling Campaign: Global River Corridor Sediment FTICR-MS, Dissolved Organic Carbon, Aerobic Respiration, Elemental Composition, Grain Size, Total Nitrogen and Organic Carbon Content, Bacterial Abundance, and Stable Isotopes

Summary

The WHONDRS Summer 2019 Sampling (S19S) Campaign collected samples in 97 globally distributed river corridor systems between July and September 2019. Surficial streambed sediments were collected at three locations within each site (upstream, midstream, and downstream). Surface water was collected at the downstream site. This dataset includes a portion of the data types produced from the sediment samples and does not include any results from the surface water. Surface water data can be found at https://data.ess-dive.lbl.gov/view/doi:10.15485/1603775. Future datasets from this study will include geochemical, hydrologic, and microbial data from the surface water and sediment. The S19S campaign was designed with the science community to ask questions associated with links among core/transient metabolomes, microbial metabolism, biogeochemical function, and physical properties of watershed and river corridor systems.

This dataset contains (1) high resolution characterization of dissolved organic matter from sediment via 12 Tesla Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) through the Environmental Molecular Sciences Laboratory (EMSL; https://www.pnnl.gov/environmental-molecular-sciences-laboratory); (2) dissolved organic carbon (DOC, measured as non-purgeable organic carbon, NPOC); (3) respiration rates calculated from laboratory incubations; (4) X-ray fluorescence (XRF) elemental composition; (5) grain size distribution from the < 2 millimeter fraction of sediment; (6) total organic carbon and total nitrogen content; (7) bacterial abundance flow cytometry data; (8) sediment stable isotopes (C and N); (9) site photos; (10) field and laboratory methods metadata; (11) file-level metadata (flmd); and (12) data dictionary (dd). The FTICR-MS files are .xml, and the data package includes instructions for using Formularity (https://omics.pnl.gov/software/formularity) and an R script to process the data based on the user's specific needs. The XRF data includes .msa, .txt, and .spx files. All other data files are .csv, .pdf, .txt, .jpg, or .jpeg.

This data package was originally published in December 2020. It was updated in April 2021, May 2022, November 2022, January 2023, and February 2023. See the change history section below for detail.

Critical Details

1 – For this campaign, WHONDRS provided global collaborators with a sampling kit containing materials and protocols to collect stream surface water and surficial streambed sediments. The sampling kit had a unique ID that was part of each individual vial's label. The format for this Sample_ID was "S19S_####" in which #### is a 4 digit numeric code. Sample_IDs spanned S19S_0001 to S19S_0100. Data in this data package and other S19S data packages can be paired with a specific site based on this Sample_ID. The metadata file contains coordinates, sampling dates, and other information about each site organized by Sample_ID.

2 – Each vial had an analysis- or parameter- specific suffix appended to the Sample_ID. If replicate samples were collected for the specific suffix, the replicate indicator was differentiated with a hyphen and 1, 2, or 3 (all collected at the downstream site) for water samples or U, M, or D (upstream, midstream, downstream) for sediment samples. Replicates are biological replicates (i.e., separate samples collected into separate vials in the field).

3 – For the FTICR and NPOC samples in this data package, vials were pre-labeled with the suffix "Sed_Field_ICR" or "Sed_INC_ICR" and had three replicates (i.e., S19S_0008_Sed_Field_ICR-M is the midstream replicate for sediment FTICR/NPOC from Sample_ID S19S_0008). "Sed_Field_ICR" refers to sediment collected in the field, preserved, and analyzed. "Sed_INC_ICR" refers to sediment collected in the field, stored in the dark at 4 degrees C until ready for use, then held in the dark at 21 degrees C overnight as if it was going to be incubated to measure respiration rates, but rather than following through with the incubation, it was flash frozen to preserve it until analysis. The "Sed_INC_ICR" samples represent the start of incubation conditions. The sediment analyzed on the FTICR-MS was not actually incubated.

4 – Sample IDs for the respiration rates are structured as "S19S_0015_SED_INC-U" where U, M, or D indicates the replicate. Note that the ID is similar to the FTICR-MS sample names, but it lacks "ICR."

5 – The FTICR-MS data are reported in files named by their Sample_ID (one file per sample). Because an important methodological detail for FTICR-MS is the ion accumulation time (IAT), the IAT for each sample is appended to the associated file name. The appended characters take the form of '_p15', which indicates an IAT of 0.15 seconds, '_p2' for an IAT of 0.2 seconds, and '_p1' for an IAT of 0.1 seconds. Example file names are "S19S_0008_Sed_Field_ICR-M_p2.xml" or "S19S_0015_SED_INC_ICR-U_p2.xml." The FTICR files must be processed via additional, publicly available software; please see point 6 below. Some file names include the suffix "rr," which indicates the data were from rerunning the sample, which was needed because the initial analysis did not produce data that passed QA/QC.

6 - This data package contains peak-picked, uncalibrated, unaligned FTICR-MS data from sediment. This is done to allow merging of multiple FTICR-MS datasets. Please see included instructions in the FTICR folder for calibrating and aligning the FTICR-MS data. Please contact WHONDRS with any questions or need for help with additional analysis of FTICR-MS data. The WHONDRS team is happy to help introduce you to FTICR-MS analyses.

Overview and Data Package Navigation

1 - See the metadata folder for the sampling protocol and analysis information. The following is a general overview of the steps that most samples went through after collection: Samples were collected and stored on wet/blue ice or in a 4 degrees C refrigerator until shipped to Pacific Northwest National Laboratory (PNNL), Richland, WA. At PNNL, sediment was sub-sampled for different analyses. The XRF samples were frozen until ready for grinding, homogenization, and analysis. The "Sed_Field_ICR" samples were flash frozen and stored until ready for analysis. Prior to analysis they were thawed and water-extracted. The extracts were analyzed for NPOC. Prior to analysis on the FTICR-MS, samples were diluted as needed to achieve a consistent NPOC concentration and underwent solid phase extraction (SPE). See folder 'WHONDRS_S19S_Sediment_Metadata' for files containing sample handling details (see files with 'LabTracking' in the name). Tracking files contain methods codes used to indicate details of sample handling, including deviations from protocols.

After receipt and sub-sampling, the "SED_INC" samples were stored at 4 degrees C until ready for use, then held in the dark at 21 degrees C overnight. The FTICR/NPOC samples were then flash frozen and followed the same procedure as the "Sed_Field_ICR" samples for extraction and analysis.
For incubations, unfiltered water from the site was added to vials with sediment until there was no headspace. Before incubation, sediments were kept at 21 degrees C overnight in a temperature-controlled chamber prior to incubation and were aerated for ~10 min before adding the unfiltered water. For the incubations, the vials were placed on a shaker at 250 revolutions per minute (rpm) in the dark. Dissolved oxygen concentrations were measured non-invasively with an optical oxygen meter (Fibox 3; PreSens GmbH, Germany) every 15 minutes for 2 hours or until the dissolved oxygen

concentration reached 5 milligrams per liter (mg/L). At each time point, dissolved oxygen was measured 3 times with an interval of 5 seconds, and then an average was calculated from the three measurements. Respiration rates were calculated as the slope of the linear regression between dissolved oxygen concentration and incubation time.

2 – Information about materials provided for the field can be found in the sampling protocol. We do not have instrumentation information for measurements taken in the field with materials we did not provide (i.e., temperature, dissolved oxygen, coordinates). Information about instrumentation used in the lab can be found in the methods code csv and XRF text described in the file-level metadata (FLMD).

Citations and Acknowledgements

This research was supported by the Worldwide Hydrobiogeochemistry Observation Network for Dynamic River Systems (WHONDRS) through the U.S. Department of Energy Biological and Environmental Research (BER) Environmental System Science (ESS) program (https://ess.science.energy.gov/) through the Pacific Northwest National Laboratory (PNNL) River Corridor Science Focus Area (SFA). PNNL is operated by Battelle Memorial Institute for the U.S. Department of Energy under Contract No. DE-AC05-76RL01830.

Please acknowledge the U.S. Department of Energy (DOE) Biological and Environmental Research (BER) Environmental System Science (ESS) program — which generously provides funding to WHONDRS — in work that uses this dataset. If using FTICR-MS data, please also acknowledge that a portion of the research was performed at the Environmental Molecular Sciences Laboratory (EMSL; grid.436923.9), a DOE Office of Science BER User Facility.

Any published work that utilizes data presented in this dataset should cite this dataset with the appropriate doi number. There is no obligation to include WHONDRS members as co-authors in your work using the data.

Contact

https://whondrs.pnnl.gov
whondrs@pnnl.gov

Change History

Version 1	1 December 2020	Original data package publication
Version 2	5 April 2021	• S19S_0045_SED_INC_ICR-M and S19S_0045_SED_INC_ICR-D in NPOC and FTICR data relabeled to S19S_0045_Sed_Field. • S19S_0046_SED_INC_ICR-U_p3, S19S_0049_SED_INC_ICR-U_p15, S19S_0058_SED_INC_ICR-U_p3. • Duplicate files for S19S_0023_Sed_INC_ICR-U_p1, S19S_0042_Sed_INC_ICR-U_p2, S19S_0055_Sed_INC_ICR-U_p2 removed from data package.
Version 3	20 May 2022	• Added XRF files and methods metadata • Added data dictionary (dd.csv) and file-level metadata (flmd.csv) • Changed N/A values in DO_data_formatted.csv from "-" to "-9999" • Updated readme_WHONDRS_S19S_Sediment_v3.pdf to reflect changes

Version 4	11 November 2022	• Added grain size data and methods metadata • Added international geo-sample number (IGSN) mapping file • Updated flmd, dd, readme, and title to reflect changes • Changed “location” column in flmd to “File_Path”
Version 5	9 January 2023	• Added total organic carbon and total nitrogen data and methods metadata • Updated “WHONDRS_S19S_Sediment_GrainSize.csv” to correct errors and match other file formatting. Changed “Percent_Med_Sand”, “Percent_Tot_Sand”, and “Percent_Silt” values of sample S19S_0026_BULK-M to -9999. Changed a duplicated sample ID from S19S_0038_BULK-M to S19S_0038_BULK-U. Changed Sample_Name column to Sample_ID and added “Study_Code” column. • Updated flmd, dd, readme, and title to reflect changes
Version 6	23 January 2023	• Added bacterial abundance flow cytometry data and methods metadata • Updated flmd, dd, readme, and title to reflect changes
Version 7	23 February 2023	• Added stable isotope data and methods metadata • Updated flmd, dd, readme, and title to reflect changes