**Description of a dataset of benthic DNA metabarcoding data from rivers and streams**

These are auxiliary data that complement the dataset associated with subproduct SSWR.401.1.2.24 and can be used to further explore and expand approaches and applications developed therein. (SSWR.401.1.2.24: Using DNA metabarcoding to characterize national scale diatom-environment relationships and to develop indicators in streams and rivers of the United States)

This dataset presents benthic periphyton 16S and rbcL DNA metabarcoding data from: 2081 EPA National Rivers and Streams Assessment samples collected in 2018 and 2019 (16S only); 124 samples collected by the Indiana Department of Environmental Management and Ohio Environmental Protection Agency in 2022 from streams in Indiana and Ohio (16S and rbcL); 648 samples collected during 2016-2019 from experimental stream mesocosms (16S and rbcL); and 128 samples collected from experimental streams and 103 samples from the East Fork of the Little Miami River Watershed in southwest Ohio in 2022 (rbcL). All samples were collected between May and October. Data are the number of gene sequence reads for amplicon sequence variants in each sample. Metadata are included in the Read me tabs of files. See Smucker et al. 2020, Pilgrim et al. 2022, Smucker et al. 2022, and Smucker et al. 2024 for detailed methods on sample processing and DNA sequencing.

**Files:**

ESF16-19\_rbcL\_ASV\_data.xlsx

These data include diatom amplicon sequence variants using rbcL DNA metabarcoding. Data are the raw number of gene reads in each sample.

ESF16-19\_16S\_ASV\_data.xlsx

These data include diatom amplicon sequence variants using 16S DNA metabarcoding. Data are the raw number of gene reads in each sample.

IN-OH2022\_rbcL\_ASV\_data.xlsx

These data include diatom amplicon sequence variants using rbcL DNA metabarcoding. Data are the raw number of gene reads in each sample.

IN-OH2022\_16S\_ASV\_data.xlsx

These data include diatom amplicon sequence variants using 16S DNA metabarcoding. Data are the raw number of gene reads in each sample.

Storage\_study\_rbcL\_ASVs.xlsx

These data include diatom amplicon sequence variants using rbcL DNA metabarcoding. Data are the raw number of gene reads in each sample.

NRSA1819\_16S\_data.xlsx

These data include diatom amplicon sequence variants using 16S DNA metabarcoding. Data are the raw number of gene reads in each sample.

**Disclaimer:**

This research dataset has been reviewed in accordance with U.S. Environmental Protection Agency (U.S. EPA), Office of Research and Development, and approved for release. Mention of brand names or vendors does not constitute an endorsement of products or services by the U.S. EPA.

**References:**

Smucker, N.J., E.M. Pilgrim, C.T. Nietch, J.A. Darling, B.R. Johnson (2020) DNA metabarcoding effectively quantifies diatom responses to nutrients in streams. *Ecological Applications* 30: e02205.

Smucker, N.J., E.M. Pilgrim, H. Wu, C.T. Nietch, J.A. Darling, M. Molina, B.R. Johnson, L.L. Yuan (2022) Characterizing temporal variability in streams supports indicator development using diatom and bacterial DNA metabarcoding. *Science of the Total Environment* 831: 154960.

Pilgrim, E.M., N.J. Smucker, H. Wu, J. Martinson, C.T. Nietch, M. Molina, J.A. Darling, B.R. Johnson (2022) Developing indicators of nutrient pollution in streams using 16S rRNA gene metabarcoding of periphyton-associated bacteria. *Water* 14: 2361.

Smucker N.J., E.M. Pilgrim, C.T. Nietch, L. Gains-Germain, C. Carpenter, J.A. Darling, L.L. Yuan, R.M. Mitchell, A.I. Pollard (2024) Using DNA metabarcoding to characterize national scale diatom-environment relationships and to develop indicators in streams and rivers of the United States. *Science of the Total Environment* 939: 173502.