Metadata template[[1]](#footnote-1) for datasets of *L&O-Letters* articles

**Table 1.** Description of the fields needed to describe the creation of your dataset.

|  |  |
| --- | --- |
| **Title of dataset** | Sediment microbial community composition and environmental data for 14 lakes at the University of Notre Dame Environmental Research Center |
| **URL of dataset** | <https://github.com/brittnibertolet/MicrobeRegCH4> |
| **Abstract** | The objective of our study was to identify environmental conditions that structure lake sediment microbial communities and determine whether community composition explained inter-lake variation in potential methanogenesis rates. We performed a comparative analysis of microbial communities and methanogenesis rates in 14 freshwater lake sediments along gradients of pH and primary productivity. Variation in methanogen community composition and non-methanogen microbial community composition was best explained by pH and sediment organic matter content. However, these regulators of microbial community structure were not associated with differences in methanogenesis rates, highlighting a potential disconnect between microbial community composition and the functions they mediate. Instead, variation in potential lake methanogenesis rates was best explained by proxies for organic matter supplied to sediments (lake chlorophyll *a* concentration and sediment pore water total phosphorus) and the composition of the non-methanogen microbial community. Our results suggest a role for sediment bacterial community in influencing methanogenesis via the supply of growth substrates. |
| **Keywords** | methane, microbial community composition, structure-function, lake sediments, carbon cycling |
| **Lead author for the dataset** | Stuart E. Jones |
| **Title and position of lead author** | Principle Investigator |
| **Organization and address of lead author** | University of Notre Dame  264 Galvin Life Sciences, Notre Dame IN 46556 |
| **Email address of lead author** | sjones20@nd.edu |
| **Additional authors or contributors to the dataset** | Brittni L. Bertolet, William E. West, David W. Armitage |
| **Organization associated with the data** |  |
| **Funding** | Collaborative Research: Regulation of lake productivity of terrestrial dissolved organic matter, NSF DEB 1754561 (PIs: Stuart E. Jones)  Dimensions: Collaborative Research: Microbial Seed Banks: Processes and patterns of dormancy-driven biodiversity, NSF DEB 1624010 (PIs: Stuart E. Jones) |
| **License** | CCBY |
| **Geographic location – verbal description** | University of Notre Dame Environmental Research Center  Wisconsin, USA  Text description of the region of study (can include government entities as in county, province, state, country, etc) as well as finer-scaled information |
| **Geographic coverage bounding coordinates** | University of Notre Dame Environmental Research Center  (46.211411 N, -89.528792 W)  (46. 258657 N, -89.458444 W) |
| **Time frame - Begin date** | 2012-05-01 |
| **Time frame - End date** | 2012-07-30 |
| **General study design** | Field survey of 14 lakes. Each lake was sampled once for water and sediment chemistry/biology, sediment methanogenesis rates, and sediment microbial community composition. |
| **Methods description:**  **Add reference section** | Each lake was sampled once during the summer 2012. Sediments were sampled from the deepest point of the pelagic region for determination of sediment microbial community composition, sediment chemistry (including total nitrogen and total phosphorus), sediment organic matter content, and sediment methanogenesis rates. Epilimnetic water was collected for determination of water column chlorophyll a concentrations. pH was recorded at the sediment-water interface. |
| **Laboratory, field, or other analytical methods** | **Microbial community composition:**To determine bacterial and archaeal community composition, we performed 16S rRNA paired-end sequencing on sediment collected from the deepest point of each lake. Sediments were collected from the top 15 cm of the sediment surface using an Ekman dredge. The sample was homogenized and stored at -80°C until DNA extraction. A single DNA extraction was performed with 0.5 g of sediment using a MoBio PowerSoil DNA Isolation kit (Mo Bio, Carlsbad, CA, USA) following the manufacturer’s instructions (https://mobio.com/media/wysiwyg/pdfs/protocols/12888.pdf). The V4 region of the 16S rRNA gene was amplified in a 25 μl PCR reaction according to the Department of Energy Joint Genome Institute (DOE JGI) iTag sample amplification protocol with amplicon primers 515F and 805R. Amplified DNA served as template for high-throughput paired-end (2x150 bp) sequencing on an Illumina MiSeq at the DOE JGI (DOE JGI, Walnut Creek, CA, USA). Sequencing produced 3,851,982 sequence reads across the 14 lake sediment samples. Raw sequence reads are available online at DOE JGI Genome Portal under Project ID 1041357.  To quantify community membership, operational taxonomic units (OTUs) were defined at 97% similarity using the QIIME (version 1.9.1) bioinformatics pipeline for merging of paired-end reads, quality filtering, and OTU picking with the *pick\_open\_reference\_otu.py* command using the procedure applied in Caporaso et al. 2010. Representative sequences for each OTU were aligned and classified against the Greengenes database version 13.5) (McDonald et al. 2012). OTUs with a single read count were discarded. We then identified methanogen OTUs by the presence of “Methano” in taxonomic assignments, which encompasses all previously identified methanogen taxa. These OTUs were used for MCC analyses and excluded from the non-MCC analyses.  **Lake environmental conditions:**pH and dissolved oxygen (DO) were recorded at the sediment-water interface using a YSI Professional Plus Multiparameter meter (Yellow Springs Instruments, Yellow Springs, OH, USA). An integrated sample of epilimnetic lake water was collected for analysis of lake water column chlorophyll *a* (chl *a*) concentrations. Particles from 450 mL of epilimnion lake water were captured onto a 0.7 μm glass fiber filter for analysis, and analyzed using methanol extraction and fluorometry (Welschmeyer 1994). Epilimnetic chl *a* concentrations were used as proxies for supply rates of settling autochthonous carbon.  Surface sediments were also collected to determine sediment organic matter content and pore water total nitrogen (TN) and total phosphorus (TP) concentrations. Percent organic matter was determined using loss on ignition measurements of dried sediment samples (Heiri et al. 2001). To determine nutrient concentrations, 30 mL of sediment was centrifuged for 10 minutes to extract sediment pore water. Each sediment sample produced at least 10 mL of pore water, which was diluted to 20 mL and used for determination of pore water TN and TP concentrations. TN and TP concentrations were quantified as described in West et al. (2016).  **Methanogenesis potential:**During each sampling event, surface sediments were also collected to measure potential lake methanogenesis rates. Methanogenesis rates were determined using sediment incubations as described in West et al. (2016). Incubations were conducted in the laboratory in 300 mL sealed serum bottles containing 50 mL of lake sediment and 50 mL of hypolimnetic lake water. Bottles were flushed with N2 gas to maintain anoxia and stored at in-situ lake temperature in the dark for 9 days. Methanogenesis rates were then estimated by sampling the headspace three times over nine days and fitting a linear regression to the time course. CH4 concentrations were measured using gas chromatography as described in West et al. (2016).   * Caporaso et al. 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7(5): 335-336, doi:[10.1038/nmeth.f.303](https://dx.doi.org/10.1038%2Fnmeth.f.303) * Heiri et al. 2001. Loss on ignition as a method for estimating organic and carbonate content in sediments: Reproducibility and comparability of results. J. Paleolimnol. 25: 101-110, doi:10.1023/a: 1008119611481 * McDonald et al. 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. The ISME Journal 6(3): 610-618. * Welschmeyer 1994. Fluorometric analysis of chlorophyll a in the presence of chlorophyll b and pheopigments. Limnol. Oceanogr. 39(8): 1985-1992. * West et al. 2016. Productivity and depth regulate lake contributions to atmospheric methane. Limnol. Oceanogr. 61: S51-S61, doi:10.1002/lno.10247 |
| **Taxonomic species or groups** | Bacteria and Archaea |
| **Quality control** | Environmental chemistry data was subjected to a range check based on previous observations at these sites. Sequence analyses for microbial community composition included a number of standard quality filtering steps (Caporaso et al. 2010) |
| **Additional information** | *NA* |

**Table 2.** Data dictionary: description of the variables (i.e., columns) in EACH dataset.

Dataset filename: *envTable.csv*

Dataset description: *Environmental data for 14 lakes sampled, including epilimnion chlorophyll a concentrations, pH at the sediment-water interface, sediment porewater total nitrogen and total phosphorus concentration, sediment organic matter percent, and sediment methanogenesis rates.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column name** | **Description** | **Units** | **Code explanation** | **Data format** | **Missing data code** |
| LakeID | Two letter identification code for each lake | NA | BA = Bay Lake  BE = Bergner Lake  BO = Bolger Bog  BR = Brown Lake  CB = Cranberry Bog  CR = Crampton Lake  FO = Foggy Lake  HB = Hummingbird Bog  MO = Morris Lake  NG = North Gate Bog  PA = Paul Lake  PE = Peter Lake  TU = Tuesday Lake  WL = West Long Lake | Two character string | NA |
| chlA | Concentration of chlorophyll a in epilimnion | Micro grams per liter | NA | Numeric | NA |
| pH | pH at the sediment-water interface | NA | NA | Numeric | NA |
| TN | Concentration of sediment porewater total nitrogen | Micro grams per liter | NA | Numeric | NA |
| TP | Concentration of sediment porewater total phosphorus | Micro grams per liter | NA | Numeric | NA |
| OMperc | Percent organic matter of sediments | Mass percentage | NA | Numeric | NA |
| CH4prod | Rate of potential methanogenesis from sediments | Micro moles per meter squared per day | NA | Numeric | NA |

Dataset filename: *otuTable.csv*

Dataset description: *Operational taxonomic unit sequence count data for 14 lakes sampled.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column name** | **Description** | **Units** | **Code explanation** | **Data format** | **Missing data code** |
| OTU.ID | Multi-character string identifying operational taxonomic units based upon sequence homology | NA | OTU string IDs are arbitrarily assigned | Multi-character string | NA |
| BA | Count of quality filtered sequences belonging to each operational taxonomic unit (rows) for Bay Lake. | Sequences | NA | Integer | NA |
| BE | Count of quality filtered sequences belonging to each operational taxonomic unit (rows) for Bergner Lake. | Sequences | NA | Integer | NA |
| BO | Count of quality filtered sequences belonging to each operational taxonomic unit (rows) for Bolger Bog. | Sequences | NA | Integer | NA |
| BR | Count of quality filtered sequences belonging to each operational taxonomic unit (rows) for Brown Lake. | Sequences | NA | Integer | NA |
| CB | Count of quality filtered sequences belonging to each operational taxonomic unit (rows) for Cranberry Bog | Sequences | NA | Integer | NA |
| CR | Count of quality filtered sequences belonging to each operational taxonomic unit (rows) for Crampton Lake. | Sequences | NA | Integer | NA |
| FO | Count of quality filtered sequences belonging to each operational taxonomic unit (rows) for Foggy Lake. | Sequences | NA | Integer | NA |
| HB | Count of quality filtered sequences belonging to each operational taxonomic unit (rows) for Hummingbird Bog. | Sequences | NA | Integer | NA |
| MO | Count of quality filtered sequences belonging to each operational taxonomic unit (rows) for Morris Lake. | Sequences | NA | Integer | NA |
| NG | Count of quality filtered sequences belonging to each operational taxonomic unit (rows) for North Gate Bog | Sequences | NA | Integer | NA |
| PA | Count of quality filtered sequences belonging to each operational taxonomic unit (rows) for Paul Lake. | Sequences | NA | Integer | NA |
| PE | Count of quality filtered sequences belonging to each operational taxonomic unit (rows) for Peter Lake. | Sequences | NA | Integer | NA |
| TU | Count of quality filtered sequences belonging to each operational taxonomic unit (rows) for Tuesday Lake. | Sequences | NA | Integer | NA |
| WL | Count of quality filtered sequences belonging to each operational taxonomic unit (rows) for West Long Lake. | Sequences | NA | Integer | NA |
| Consensus.Lineage | Taxonomic assignment for each operational taxonomic unit, classified against the Greengenes database (version 13.5). | NA | NA | Character string | NA |

**Table 3. Data provenance**

If you used data derived from other sources, provide the information here so future users know where the data came from.

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset title** | **Dataset DOI or URL** | **Creator (name & email)** | **Contact (name & email)** |
| NA | NA | NA | NA |

**Scripts/code (software) –** *OPTIONAL*

It is recommended that you also provide your scripts along with your data, although it is not required at this time in our journal.

|  |  |  |
| --- | --- | --- |
| **File name** | **Description** | **Scripting language** |
| MicroRegCH4-StatsCode.R | Code for all statistical analyses | R |
| MicroRegCH4-OTUpicking.sh | Code for generating operational taxonomic unit sequence count data | bash |

1. *This document liberally borrows from a similar document provided by the Environmental Data Initiative* [↑](#footnote-ref-1)