

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Papers in Natural Resources

Natural Resources, School of

---

2022

## A unified dataset of colocated sewage pollution, periphyton, and benthic macroinvertebrate community and food web structure from Lake Baikal (Siberia)

Michael F. Meyer

Ted Ozersky

Kara H. Woo

Kirill Shchapov

Aaron W. E. Galloway

*See next page for additional authors*

Follow this and additional works at: <https://digitalcommons.unl.edu/natrespapers>



Part of the [Natural Resources and Conservation Commons](#), [Natural Resources Management and Policy Commons](#), and the [Other Environmental Sciences Commons](#)

---

This Article is brought to you for free and open access by the Natural Resources, School of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Papers in Natural Resources by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

---

**Authors**

Michael F. Meyer, Ted Ozersky, Kara H. Woo, Kirill Shchapov, Aaron W. E. Galloway, Julie B. Schram, Daniel D. Snow, Maxim A. Timofeyev, Dmitry Yu. Karnaukhov, Matthew R. Brousil, and Stephanie E. Hampton

---

## DATA ARTICLE

**A unified dataset of colocated sewage pollution, periphyton, and benthic macroinvertebrate community and food web structure from Lake Baikal (Siberia)**

Michael F. Meyer <sup>1,\*</sup> Ted Ozersky <sup>2</sup> Kara H. Woo <sup>3</sup> Kirill Shchapov,<sup>2</sup> Aaron W. E. Galloway <sup>4</sup> Julie B. Schram <sup>4</sup>  
Daniel D. Snow <sup>5</sup> Maxim A. Timofeyev <sup>6</sup> Dmitry Yu. Karnaukhov,<sup>6</sup> Matthew R. Brousil <sup>3</sup> Stephanie E. Hampton <sup>3</sup>

<sup>1</sup>School of the Environment, Washington State University, Pullman, Washington; <sup>2</sup>Large Lakes Observatory, University of Minnesota Duluth, Duluth, Minnesota; <sup>3</sup>Center for Environmental Research, Education, and Outreach, Washington State University, Pullman, Washington; <sup>4</sup>Oregon Institute of Marine Biology, University of Oregon, Charleston, Oregon; <sup>5</sup>School of Natural Resources, University of Nebraska-Lincoln, Lincoln, Nebraska; <sup>6</sup>Biological Research Institute, Irkutsk State University, Irkutsk, Irkutsk Oblast, Russia

**Scientific Significance Statement**

We present a unified dataset of colocated benthic littoral nutrient concentrations, sewage indicators, algal and macroinvertebrate community abundance, stable isotopes, and fatty acids from Lake Baikal (Siberia). While researchers have studied Baikal's exceptionally diverse endemic taxa for centuries, this product is the first publicly available dataset of Baikal benthic amphipod species abundance as well as amphipod fatty acid profiles in a machine-readable format with standardized metadata. Furthermore, with over 150 colocated variables, this dataset is the most extensive, publicly available description of Baikal's nearshore benthic communities and food webs. The data are highly structured and incorporate a scripted, sequential workflow, enabling the dataset to either supplement current monitoring efforts or provide data for syntheses across systems.

**URL of the Dataset and Metadata with permanent identifier:**

Environmental Data Initiative: <https://doi.org/10.6073/pasta/9554b7f19ddd4a614e854f18be978dca>

Open Science Framework: <https://doi.org/10.17605/OSF.IO/9TA8Z>

**Code URL with permanent identifier:**

Environmental Data Initiative: <https://doi.org/10.6073/pasta/9554b7f19ddd4a614e854f18be978dca>

Open Science Framework: <https://doi.org/10.17605/OSF.IO/9TA8Z>

**Measurement(s):** Chlorophyll *a*, fatty acids, pharmaceuticals and personal care products, microplastics, periphyton community abundance, benthic macroinvertebrate abundance, stable isotopes, nitrate, ammonium, total phosphorus.

**Technology Type(s):** GC/MS, LC/MS, spectrophotometry, fluorometry, microscopy.

**Temporal range:** 19–23 August 2015.

**Frequency or sampling interval:** Single snapshot in time.

**Spatial scale:** Site-based.

\*Correspondence: michael.f.meyer@wsu.edu

**Associate editor:** Dag Hessen

**Author Contribution Statement:** MFM, SEH, and TO conceptualized the project. MFM, TO, KHW, and SEH collected samples in the field. MFM, KS, JBS, DDS, TO, AWEG, and SEH processed the samples. MFM, MRB, and KHW wrote and reviewed R scripts. MFM and MRB manage the data. All authors wrote and edited the manuscript and approved the final manuscript.

**Data Availability Statement:** Data are available at the replicate level at the Environmental Data Initiative repository at <https://doi.org/10.6073/pasta/9554b7f19ddd4a614e854f18be978dca>.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

## Abstract

Sewage released from lakeside development can introduce nutrients and micropollutants that can restructure aquatic ecosystems. Lake Baikal, the world's most ancient, biodiverse, and voluminous freshwater lake, has been experiencing localized sewage pollution from lakeside settlements. Nearby increasing filamentous algal abundance suggests benthic communities are responding to localized pollution. We surveyed 40-km of Lake Baikal's southwestern shoreline from 19 to 23 August 2015 for sewage indicators, including pharmaceuticals, personal care products, and microplastics, with colocated periphyton, macroinvertebrate, stable isotope, and fatty acid samplings. The data are structured in a tidy format (a tabular arrangement familiar to limnologists) to encourage reuse. Unique identifiers corresponding to sampling locations are retained throughout all data files to facilitate interoperability among the dataset's 150+ variables. For Lake Baikal studies, these data can support continued monitoring and research efforts. For global studies of lakes, these data can help characterize sewage prevalence and ecological consequences of anthropogenic disturbance across spatial scales.

Globally, sewage pollution is a common and often concentrated source of nitrogen and phosphorus inputs that can reshape aquatic ecosystems. Sewage inputs are often associated with increased primary production (Edmondson 1970; Moore et al. 2003), which can eventually lead to nuisance algal blooms (Hall et al. 1999; Lapointe et al. 2015). Even in instances where sewage pollution is mitigated, restoring systems can be complicated and necessitate system-specific (Jeppesen et al. 2005), long-term mitigation strategies (Hall et al. 1999; Tong et al. 2020). As such, effective sewage monitoring can require merging a suite of chemical, biological, and ecological data to synthesize locations and timing of inputs with associated shifts in ecological communities (Rosenberger et al. 2008; Hampton et al. 2011).

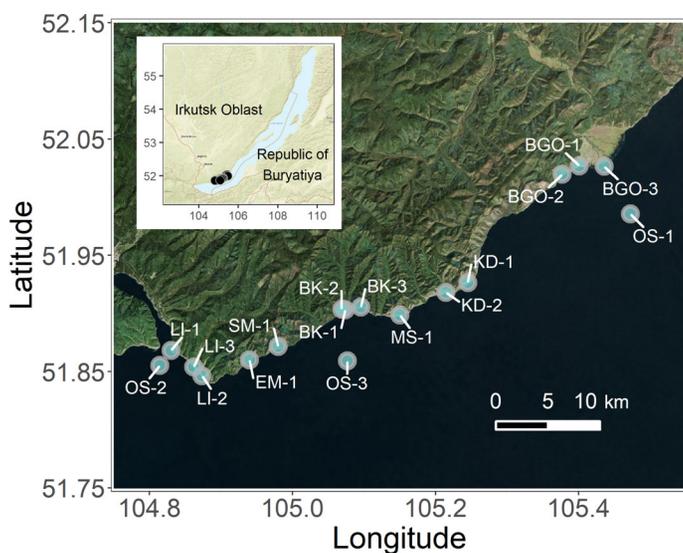
Definitively identifying sewage as the source of excess nutrients in a system can be challenging. Nutrients can originate from multiple sources, such as agriculture (Powers et al. 2016) or melting permafrost (Turetsky et al. 2000; Anisimov and Reneva 2006; Moore et al. 2009), which can obfuscate wastewater signals. Unlike nutrients, sewage-specific indicators, such as enhanced  $\delta^{15}\text{N}$  stable isotope signatures (Costanzo et al. 2001; Camilleri and Ozersky 2019), pharmaceuticals and personal care products (PPCPs) (Bendz et al. 2005; Rosi-Marshall and Royer 2012; Meyer et al. 2019), and microplastics (Barnes et al. 2009), can be highly specific to human wastewater. Accordingly, sewage-associated micropollutants have garnered global attention for their usefulness in identifying presence and quantifying magnitude of wastewater inputs. While indicators may accumulate differentially in certain taxa (Gartner et al. 2002; Green 2016; Vendel et al. 2017; Richmond et al. 2018), acutely dangerous concentrations are not common in most aquatic systems (Kolpin et al. 2002; Focazio et al. 2008; Yang et al. 2018). However, chronic exposure to microplastics and PPCPs at even minute concentrations (e.g.,  $\mu\text{g L}^{-1}$ ) can still disrupt ecological processes (Richmond et al. 2017). For example, oxazepam can increase feeding rate and decrease sociability of river perch (Brodin et al. 2013), and microplastics can release dissolved

organic carbon, thereby altering microbial communities (Romera-Castillo et al. 2018). The pervasiveness and diversity of sewage-associated micropollutants in tandem with their potency as ecologically disrupting compounds necessitates investigation within and across systems, thereby enabling synthesis of how micropollutants alter ecosystems.

When assessing biological responses to increased nutrient loading, littoral benthic algal and macroinvertebrate communities often respond most markedly, as their physical proximity to the shoreline puts them in the path of sewage pollution entering the lake (Rosenberger et al. 2008; Hampton et al. 2011). Filamentous algae, for example, can quickly increase in abundance near sewage sources (Rosenberger et al. 2008; Hampton et al. 2011). As algal communities change, food webs can also restructure. For example, change in algal communities can alter the nutritional value of primary producers or cause changes in the relative abundance of different feeding groups (e.g., increased representation of detritivores). Among the suite of food quality metrics, availability of essential fatty acids (EFAs) offers a nuanced understanding of food quality as primary producers usually maintain consistent EFA signatures (Taipale et al. 2013) and consumers acquire EFAs by grazing (Dalsgaard et al. 2003) or trophic upgrading (Sargent and Falk-Petersen 1988; Dalsgaard et al. 2003).

Together, food web structure, community composition, and sewage indicator data can be powerful tools to assess biological impacts of sewage pollution. Despite their utility, these data are not often available for many limnological systems. PPCPs, for example, have historically been less measured in lake environments (Meyer et al. 2019). In instances where data are available, efficiently merging disparate data into a single, analytically friendly format can be challenging and sometimes require complex, computationally intensive workflows (Meyer et al. 2020).

To offer a template for harmonizing sewage indicator and biological data, we present a unified data product, which contains disparate data collected from 14 littoral and 3 pelagic



**Fig. 1.** Map of all sampling locations with sites labeled with unique alphanumeric code. The entire transect included three developed sites (i.e., Listvyanka [LI], Bolshie Koty [BK], Bolshoe Goloustnoe [BGO]). Three offshore sites (OS) were also sampled to compare pelagic sewage signals to those in the littoral. Sites without adjacent lakeside development included Emelianikha Bay (EM), Maloe Kadilnoe (KD), Mys Soboliny (MS), and Sredny Mys (SM). Littoral sampling locations were all 8.90–20.75 m from shore and at a depth approximately of 0.75 m, whereas pelagic sites were approximately 2–5 km from shore and ranged in depth from 900 to 1300 m. This map was created using the R statistical environment (R Core Team 2019) and the tidyverse (Wickham et al. 2019), OpenStreetMap (Fellows and Stotz 2019), ggpubr (Kassambara 2019), cowplot (Wilke 2019), ggsp (Baquero 2019), and ggrepel (Slowikowski 2019) packages. This map was produced using data from © OpenStreetMap contributors (<https://www.openstreetmap.org/copyright>), which is licensed under the Open Data Commons Open Database License (ODbL) by the OpenStreetMap Foundation (OSMF). Base map and data from OpenStreetMap and OSMF were created using the © ESRI (inset map) and © 2021 Microsoft Corporation Earthstar Geographics SIO “bing” (zoomed-in map) tiles.

sites at Lake Baikal from 19 through 23 August 2015 (Fig. 1). Located in Siberia, Lake Baikal is the oldest, most voluminous, and deepest freshwater lake in the world (Hampton et al. 2018). Lake Baikal also has the global distinction of being the most biodiverse lake, with the highest endemism (Moore et al. 2009). The lake is experiencing rapid warming associated with climate change, including decrease in ice cover duration (Moore et al. 2009), and it exhibits offshore plankton community changes associated with warming (Hampton et al. 2008; Katz et al. 2015; Izmet'eva et al. 2016). Less is known of the change occurring in the nearshore of Lake Baikal, where not only climatic changes (Swann et al. 2020) but also human activity (Timoshkin et al. 2018) may introduce nutrients that alter the environment. Near-shore change is particularly important to understand in Lake Baikal, since the majority of the lake's biodiversity and endemic species occur in the littoral zone (Kozhova and

Izmet'eva 1998). While Lake Baikal's pelagic zone is generally ultra-oligotrophic (Yoshida et al. 2003; O'Donnell et al. 2017), littoral areas abutting lakeside settlements have recently shown distinct signs of eutrophication, such as increased filamentous green algae abundance (Timoshkin et al. 2016; Volkova et al. 2018) as well as cyanobacteria (Bondarenko et al. 2021).

As a means of identifying sewage from small, concentrated lakeside towns and the associated ecological responses, we assembled a dataset consisting of over 150 variables collected at 14 littoral and 3 pelagic sampling sites. We structured the dataset in a tidy format, where each row is a sample, each column is a variable, and each CSV file is an observable unit, where more similar variables are contained within an individual file (Wickham 2014). Independent CSV files can be merged using unique locational identifiers as relational keys, enabling future researchers to customize analyses around a particular suite of variables. As a result of the dataset's interoperability, reproducibility, and extensive variable content, it is well poised for future reuse as supporting evidence of sewage pollution in Lake Baikal. Additionally, the data's flexibility and consistent structure enable it to be merged with similar datasets, so as to synthesize biological responses to sewage across systems and scales.

To our knowledge, no raw data on Lake Baikal macroinvertebrates, periphyton, or nearshore water quality are public in a machine-readable format, for any variable (i.e., abundance, fatty acid content, stable isotopes, nutrient, and pollutant concentration), and no georeferenced data on PPCPs or microplastics appear to be publicly available for any boreal, subarctic, or arctic lakes or rivers in Siberia. Thus, the dataset fills a substantial gap for future studies, providing a window into nearshore biotic assemblages and water quality in a unique, ancient ecosystem that holds 20% of the world's liquid surface water (Moore et al. 2009).

## Data description

The final, replicate-level data products are available on the Environmental Data Initiative (EDI), where they can be freely accessed without potential barriers such as paywalls or account registrations. The final data are provided as 11 separate CSV files, each structured in a tabular format and containing a “site” column that can be used to merge tables. The repository also contains a compressed folder of R scripts (scripts.tar.gz), which were used in the main analysis of the dataset (Meyer et al. Under Review).

### site\_information.csv

This file contains metadata for each of the pelagic and littoral sampling locations. Missing data are assigned as NA.

*year*

Year sampling occurred.

*month*

Month sampling occurred.

*day*

Day of month sampling occurred.

*time*

Time sampling occurred as hours : minutes.

*site*

Unique alphanumeric identifier for a sampling location.

*lat*

Latitude of sampling location in decimal degrees.

*long*

Longitude of sampling location in decimal degrees.

*site\_description*

Researchers' description of sampling location at the time of sampling.

*distance\_to\_shore\_m*

Distance from in situ sampled location to the shoreline in meters.

*depth\_m*

Maximum depth at sampling location in meters.

*air\_temp\_celsius*

Temperature of air at sampling location in Celsius.

*surface\_temp\_celsius*

Temperature of water's surface at sampling location in Celsius.

*mid\_temp\_celsius*

Temperature of water midway (i.e.,  $\text{depth\_m}/2$ ) between surface and bottom at sampling location in Celsius.

*bottom\_temp\_celsius*

Temperature of water near sediment at sampling location in Celsius.

*comments*

Notes in the field describing sampling conditions.

*shore\_photo*

Whether or not photos of the shoreline were taken. Photos are available on the project's Open Science Framework (OSF) portal (Meyer et al. 2015).

*substrate\_photo*

Whether or not photos of the substrate were taken.

*sponges*

Whether or not sponges were present at a sampling location.

*brandtia*

Whether or not *Brandtia* spp. (endemic amphipod species) were present at a sampling location.

#### distance\_weighted\_population\_metrics.csv

This file contains inverse distance weighted (IDW), census-based human population data for each sampled location. Although the majority of sites do not have adjacent shoreline human developments, we calculated IDW population for each sampling location. IDW population is a generalized representation of the size of and proximity to a sampling location's neighboring human settlements. As these population estimates are based on census data, they reflect static populations

and do not account for seasonal population deviations from tourism. A full description of the methods used to calculate IDW population can be found in the companion manuscript Meyer et al. (Under Review).

*site*

Unique alphanumeric identifier for a sampling location.

*distance\_weighted\_population*

IDW population for a given sampling location and estimated as number of people. Because this interpolation process is a function of the size of and proximity to neighboring developed sites, values can contain decimal values.

#### nutrients.csv

This file contains nutrient concentrations for each of the associated sampling locations. Samples were collected at a depth of 0.75 m. Nutrient samples were not filtered prior to analysis, meaning that nitrogen concentrations have the potential to include intracellular nitrogen. Therefore, nitrogenous species' concentrations may be spurious. Minimal detection limits were estimated as  $0.01 \text{ mg L}^{-1}$  for nitrate,  $0.005 \text{ mg L}^{-1}$  for ammonium, and  $0.04 \text{ mg L}^{-1}$  for phosphorus.

*site*

Unique alphanumeric identifier for a sampling location.

*replicate*

Replicate for a given sampling location.

*nh4\_mg\_dm3*

Ammonium concentration in milligrams of ammonium per cubic decimeter.

*no3\_mg\_dm3*

Nitrate concentration in milligrams of nitrate per cubic decimeter.

*tp\_mg\_dm3*

Total phosphorus concentration in milligrams of phosphorus per cubic decimeter.

*tpo43\_mg\_dm3*

Total phosphate concentration as phosphate in milligrams per cubic decimeter.

#### chlorophylla.csv

This file contains chlorophyll *a* (Chl *a*) concentrations in the water column as well as fluorometric corrections for each littoral and pelagic sampling location. Minimal detection limits were estimated to be  $0.02 \text{ mg L}^{-1}$ .

*site*

Unique alphanumeric identifier for a sampling location.

*replicate*

Replicate number.

*filtered\_volume\_ml*

Lake water volume filtered in milliliters for a given replicate.

*sample\_volume\_ml*

Sample volume filtered for Chl *a* extraction.

*raw\_fluo*

Raw, uncorrected fluorometric reading for chlorophyll analysis.

*adjusted\_raw*

Corrected fluorometric reading for chlorophyll analysis.

*chl\_conc*

Chl *a* concentration in milligrams per liter.

#### ppcp.csv

This file contains PPCP concentrations in the water column at each littoral and pelagic sampling location. Detection limits were estimated to be  $0.001 \mu\text{g L}^{-1}$  based on a 500 mL sample volume.

*site*

Unique alphanumeric identifier for a sampling location.

*paraxanthine*

Concentration of paraxanthine, also known as 1,7-dimethylxanthine, in micrograms per liter. Paraxanthine is the main human metabolite of caffeine.

*acetaminophen*

Concentration of acetaminophen, also known as paracetamol, in micrograms per liter.

*amphetamine*

Concentration of amphetamine in micrograms per liter.

*caffeine*

Concentration of caffeine in micrograms per liter.

*carbamazepine*

Concentration of carbamazepine in micrograms per liter.

*cimetidine*

Concentration of cimetidine in micrograms per liter.

*cotinine*

Concentration of cotinine, which is the main human metabolite of nicotine, in micrograms per liter.

*diphenhydramine*

Concentration of diphenhydramine in micrograms per liter.

*mda*

Concentration of methylenedioxyamphetamine in micrograms per liter.

*mdma*

Concentration of methylenedioxymethamphetamine in micrograms per liter.

*methamphetamine*

Concentration of methamphetamine in micrograms per liter.

*morphine*

Concentration of morphine in micrograms per liter.

*phenazone*

Concentration of phenazone in micrograms per liter.

*sulfachloropyridazine*

Concentration of sulfachloropyridazine in micrograms per liter.

*sulfamethazine*

Concentration of *sulfamethazine* in micrograms per liter.

*sulfamethoxazole*

Concentration of sulfamethoxazole in micrograms per liter.

*thiabendazole*

Concentration of thiabendazole in micrograms per liter.

*trimethoprim*

Concentration of trimethoprim in micrograms per liter.

*collection\_year*

Year sample was collected in the field.

*collection\_month*

Month sample was collected in the field.

*collection\_day*

Day of month sample was collected in the field.

*analysis\_year*

Year sample was analyzed.

*analysis\_month*

Month sample was analyzed.

*analysis\_day*

Day of month sample was analyzed.

#### microplastics.csv

This file contains suspended microplastics counts for each of the pelagic and littoral sampling locations. Although we did not measure microplastic size, our enumeration techniques likely allowed us to reliably quantify microplastics as small as  $\sim 300 \mu\text{m}$  (Hanvey et al. 2017).

*site*

Unique alphanumeric identifier for a sampling location.

*replicate*

Replicate for a given sampling location. Replicate values of “C” indicate a control.

*fragments*

Number of microplastic fragments observed.

*fibers*

Number of microplastic fibers observed.

*beads*

Number of microplastic beads observed.

*comments*

Observer comments while enumerating microplastics.

*volume\_filtered\_ml*

Volume in milliliters for a given replicate filtered.

#### periphyton.csv

This file contains periphyton abundance data, collected from rocks at each of the sampled littoral locations. For poorly preserved samples, counts are listed as NA for each taxonomic grouping, and a note in the “comments” column is provided.

*site*

Unique alphanumeric identifier for a sampling location.

*replicate*

Replicate number for a given sampling site.

*subsamples\_counted*

Number of  $10 \mu\text{L}$  subsamples counted for a given replicate.

*diatom*

Number of diatom cells counted for a given replicate.

*spirogyra*  
Number of *Spirogyra* spp. cells counted for a given replicate.

*spirogyra\_filament*  
Number of *Spirogyra* spp. filaments counted for a given replicate.

*ulothrix*  
Number of *Ulothrix* spp. cells counted for a given replicate.

*ulothrix\_filament*  
Number of *Ulothrix* spp. filaments counted for a given replicate.

*tetrasporales*  
Number of Tetrasporales cells counted for a given replicate.

*pediastrum*  
Number of *Pediastrum* spp. cells counted for a given replicate.

*desmidiatales*  
Number of *Desmidiatales* spp. cells counted for a given replicate.

*comments*  
Notes from the observer.

#### invertebrates.csv

This file contains abundance for benthic macroinvertebrates collected at each of the 14 littoral sampling locations. Only amphipod taxa were identified to species.

*site*  
Unique alphanumeric identifier for a sampling location.

*replicate*  
Replicate for sampling location. While three replicates were collected in the field, some samples were poorly preserved, and invertebrates were not enumerated so as to prevent potential errors.

*Acroloxidae*  
Mollusk family.

*Asellidae*  
Isopod family.

*Baicaliidae*  
Mollusk family.

*Benedictidae*  
Mollusk family.

*Brandtia\_latissima*  
Endemic amphipod species. Three subspecies exist, but samples were not identified to subspecies to reduce potential errors.

*Brandtia\_parasitica\_parasitica*  
Endemic amphipod species.

*Caddisflies*  
General grouping; specimens were not identified to species.

*Cryptoropus\_inflatus*  
Endemic amphipod species.

*Cryptoropus\_pachytus*  
Endemic amphipod species.

*Cryptoropus\_rugosus*  
Endemic amphipod species.

*Eulimnogammarus\_capreolus*  
Endemic amphipod species.

*Eulimnogammarus\_cruentes*  
Endemic amphipod species.

*Eulimnogammarus\_cyaneus*  
Endemic amphipod species.

*Eulimnogammarus\_grandimanus*  
Endemic amphipod species.

*Eulimnogammarus\_juveniles*  
Endemic amphipod genus. Identification kept at genus level so as to prevent misclassification.

*Eulimnogammarus\_maackii*  
Endemic amphipod species.

*Eulimnogammarus\_marituji*  
Endemic amphipod species.

*Eulimnogammarus\_verucosus*  
Endemic amphipod species.

*Eulimnogammarus\_viridis\_viridis*  
Endemic amphipod species.

*Eulimnogammarus\_vittatus*  
Endemic amphipod species.

*Flatworms*  
Not identified beyond phylum.

*Leeches*  
Not identified beyond order, although 12 endemic species occur in Lake Baikal.

*Maackia*  
Mollusk family.

*Pallasea\_brandtia\_brandtia*  
Endemic amphipod species.

*Pallasea\_brandtii\_tenera*  
Endemic amphipod species.

*Pallasea\_cancelloides*  
Endemic amphipod species.

*Pallasea\_cancellus*  
Endemic amphipod species.

*Pallasea\_viridis*  
Endemic amphipod species.

*Planorbidae*  
Mollusk family.

*Poekilogammarus\_crassimus*  
Endemic amphipod species.

*Poekilogammarus\_ephippiatus*  
Endemic amphipod species.

*Poekilogammarus\_juveniles*  
Endemic amphipod genus. Identification kept at genus level so as to prevent misclassification.

*Poekilogammarus\_megonychus\_perpolitus*  
Endemic amphipod species.

*Poekilogammarus\_pictus*  
Endemic amphipod species.

*Valvatidae*  
Mollusk family.

stable\_isotopes.csv

This file contains carbon ( $\delta^{13}\text{C}$ ) and nitrogen ( $\delta^{15}\text{N}$ ) values for various benthic macroinvertebrate genera and periphyton collected from the 14 littoral sampling locations.

*site*

Unique alphanumeric identifier for a sampling location.

*Genus*

Genus of the analyzed organism.

*Species*

Species of the analyzed organism. When an organism was identified solely to genus, the Species value is NA.

*C13*

Carbon ( $\delta^{13}\text{C}$ ) stable isotope values in parts per thousand.

*N15*

Nitrogen ( $\delta^{15}\text{N}$ ) stable isotope values in parts per thousand.

*comments*

Quality flag column where  $\delta^{13}\text{C}$  samples were outside of the range of standards.

fatty\_acid.csv

This file contains fatty acid concentrations for various benthic macroinvertebrate genera, periphyton, and endemic *Draparnaldia* spp. benthic algae collected from the 14 littoral sampling locations.

*site*

Unique alphanumeric identifier for a sampling location.

*Genus*

Genus of the analyzed organism.

*Species*

Species of the analyzed organism. When an organism was identified solely to genus, the Species value is NA.

*c12\_0*

Concentration of 12 : 0 fatty acid as micrograms of fatty acid per milligram of tissue.

*i\_14\_0*

Concentration of i-14 : 0 fatty acid as micrograms of fatty acid per milligram of tissue.

*c14\_0*

Concentration of 14 : 0 fatty acid as micrograms of fatty acid per milligram of tissue.

*c14\_1w5*

Concentration of 14 : 1w5 fatty acid as micrograms of fatty acid per milligram of tissue.

*i\_15\_0*

Concentration of i-15 : 0 fatty acid as micrograms of fatty acid per milligram of tissue.

*a\_15\_0*

Concentration of a-15 : 0 fatty acid as micrograms of fatty acid per milligram of tissue.

*c15\_0*

Concentration of 15 : 0 fatty acid as micrograms of fatty acid per milligram of tissue.

*c15\_1w7*

Concentration of 15 : 1w7 fatty acid as micrograms of fatty acid per milligram of tissue.

*i\_16\_0*

Concentration of i-16 : 0 fatty acid as micrograms of fatty acid per milligram of tissue.

*c16\_0*

Concentration of 16 : 0 fatty acid as micrograms of fatty acid per milligram of tissue.

*c16\_1w9*

Concentration of 16 : 1w9 fatty acid as micrograms of fatty acid per milligram of tissue.

*c16\_1w8*

Concentration of 16 : 1w8 fatty acid as micrograms of fatty acid per milligram of tissue.

*c16\_1w7*

Concentration of 16 : 1w7 fatty acid as micrograms of fatty acid per milligram of tissue.

*c16\_1w6*

Concentration of 16 : 1w6 fatty acid as micrograms of fatty acid per milligram of tissue.

*c16\_1w5*

Concentration of 16 : 1w5 fatty acid as micrograms of fatty acid per milligram of tissue.

*i\_17\_0*

Concentration of i-17 : 0 fatty acid as micrograms of fatty acid per milligram of tissue.

*a\_17\_0*

Concentration of a-17 : 0 fatty acid as micrograms of fatty acid per milligram of tissue.

*c17\_0*

Concentration of 17 : 0 fatty acid as micrograms of fatty acid per milligram of tissue.

*c17\_1w7*

Concentration of 17 : 1w7 fatty acid as micrograms of fatty acid per milligram of tissue.

*c16\_2w7*

Concentration of 16 : 2w7 fatty acid as micrograms of fatty acid per milligram of tissue.

*c16\_2w6*

Concentration of 16 : 2w6 fatty acid as micrograms of fatty acid per milligram of tissue.

*c16\_2w4*

Concentration of 16 : 2w4 fatty acid as micrograms of fatty acid per milligram of tissue.

*c16\_3w6*

Concentration of 16 : 3w6 fatty acid as micrograms of fatty acid per milligram of tissue.

*c16\_3w4*

Concentration of 16 : 3w4 fatty acid as micrograms of fatty acid per milligram of tissue.

*c16\_3w3*

Concentration of 16 : 3w3 fatty acid as micrograms of fatty acid per milligram of tissue.

*c16\_4w3*  
Concentration of 16 : 4 $\omega$ 3 fatty acid as micrograms of fatty acid per milligram of tissue.

*c16\_4w1*  
Concentration of 16 : 4 $\omega$ 1 fatty acid as micrograms of fatty acid per milligram of tissue.

*c18\_0*  
Concentration of 18 : 0 fatty acid as micrograms of fatty acid per milligram of tissue.

*c18\_1w9*  
Concentration of 18 : 1 $\omega$ 9 fatty acid as micrograms of fatty acid per milligram of tissue.

*c18\_1w7*  
Concentration of 18 : 1 $\omega$ 7 fatty acid as micrograms of fatty acid per milligram of tissue.

*c18\_2w6t*  
Concentration of 18 : 2 $\omega$ 6t fatty acid as micrograms of fatty acid per milligram of tissue.

*c18\_2w6*  
Concentration of 18 : 2 $\omega$ 6 fatty acid as micrograms of fatty acid per milligram of tissue.

*c18\_3w6*  
Concentration of 18 : 3 $\omega$ 6 fatty acid as micrograms of fatty acid per milligram of tissue.

*c18\_3w3*  
Concentration of 18 : 3 $\omega$ 3 fatty acid as micrograms of fatty acid per milligram of tissue.

*c18\_4w4*  
Concentration of 18 : 4 $\omega$ 4 fatty acid as micrograms of fatty acid per milligram of tissue.

*c18\_4w3*  
Concentration of 18 : 4 $\omega$ 3 fatty acid as micrograms of fatty acid per milligram of tissue.

*c18\_5w3*  
Concentration of 18 : 5 $\omega$ 3 fatty acid as micrograms of fatty acid per milligram of tissue.

*c20\_0*  
Concentration of 20 : 0 fatty acid as micrograms of fatty acid per milligram of tissue.

*c20\_1w9*  
Concentration of 20 : 1 $\omega$ 9 fatty acid as micrograms of fatty acid per milligram of tissue.

*c20\_1w7*  
Concentration of 20 : 1 $\omega$ 7 fatty acid as micrograms of fatty acid per milligram of tissue.

*c20\_2w5\_11*  
Concentration of 20 : 2 $\omega$ 5-11 fatty acid as micrograms of fatty acid per milligram of tissue.

*c20\_2w5\_13*  
Concentration of 20 : 2 $\omega$ 5-13 fatty acid as micrograms of fatty acid per milligram of tissue.

*c20\_2w6*  
Concentration of 20 : 2 $\omega$ 6 fatty acid as micrograms of fatty acid per milligram of tissue.

*c20\_3w6*  
Concentration of 20 : 3 $\omega$ 6 fatty acid as micrograms of fatty acid per milligram of tissue.

*c20\_4w6*  
Concentration of 20 : 4 $\omega$ 6 fatty acid as micrograms of fatty acid per milligram of tissue.

*c20\_3w3*  
Concentration of 20 : 3 $\omega$ 3 fatty acid as micrograms of fatty acid per milligram of tissue.

*c20\_4w3*  
Concentration of 20 : 4 $\omega$ 3 fatty acid as micrograms of fatty acid per milligram of tissue.

*c20\_5w3*  
Concentration of 20 : 5 $\omega$ 3 fatty acid as micrograms of fatty acid per milligram of tissue.

*c22\_0*  
Concentration of 22 : 0 fatty acid as micrograms of fatty acid per milligram of tissue.

*c22\_1w9*  
Concentration of 22 : 1 $\omega$ 9 fatty acid as micrograms of fatty acid per milligram of tissue.

*c22\_1w7*  
Concentration of 22 : 1 $\omega$ 7 fatty acid as micrograms of fatty acid per milligram of tissue.

*c22\_2w6*  
Concentration of 22 : 2 $\omega$ 6 fatty acid as micrograms of fatty acid per milligram of tissue.

*c22\_4w6*  
Concentration of 22 : 4 $\omega$ 6 fatty acid as micrograms of fatty acid per milligram of tissue.

*c22\_5w6*  
Concentration of 22 : 5 $\omega$ 6 fatty acid as micrograms of fatty acid per milligram of tissue.

*c22\_3w3*  
Concentration of 22 : 3 $\omega$ 3 fatty acid as micrograms of fatty acid per milligram of tissue.

*c22\_4w3*  
Concentration of 22 : 4 $\omega$ 3 fatty acid as micrograms of fatty acid per milligram of tissue.

*c22\_5w3*  
Concentration of 22 : 5 $\omega$ 3 fatty acid as micrograms of fatty acid per milligram of tissue.

*c22\_6w3*  
Concentration of 22 : 6 $\omega$ 3 fatty acid as micrograms of fatty acid per milligram of tissue.

*c24\_0*  
Concentration of 24 : 0 fatty acid as micrograms of fatty acid per milligram of tissue.

*comments*  
Quality flag column. Two samples spilled during fatty acid extraction. These samples are flagged as such. Although concentrations are lower than other samples, proportions between fatty acids are consistent.

**total\_lipid.csv**

This file contains gravimetry data for each fatty acid sample.

*site*

Unique alphanumeric identifier for a sampling location.

*Genus*

Genus of the analyzed organism.

*Species*

Species of the analyzed organism. When organism was identified solely to genus, the Species value is NA.

*total\_lipid\_mg\_per\_g*

Total amount of lipids in a sample in milligrams of lipid per gram of tissue.

*deviation*

Samples were weighed three times and standard deviation in measurement was calculated. All values are reported in milligrams of lipid per gram of tissue.

*comments*

Quality flag column. Two samples spilled during fatty acid extraction. These samples are flagged as such.

**Methods****Site information**

The vast majority of Lake Baikal's 2100-km shoreline lacks lakeside development (Moore et al. 2009; Timoshkin et al. 2016). Our sample collection focused on a 40-km section of Lake Baikal's southwestern shoreline, which included three settlements of different sizes (Fig. 1) during a time of the year when tourism and summertime biological succession were likely at their annual peaks. Littoral locations were chosen to capture a range of sites with varying degrees of adjacent shoreline development—from “developed” (along the waterfront of human settlements) to “undeveloped” (no adjacent human settlements and complete forest cover; Fig. 1). The largest, Listvyanka, is primarily a tourist town of approximately 2000 permanent residents, although tourism can contribute significantly to the town's population with approximately 1.2 million annual visitors (Interfax-Tourism 2018). The other two settlements are the villages Bolshie Koty and Bolshoe Goloustnoe, which have approximately 80 and 600 permanent residents, respectively. Bolshie Koty is home to two field research stations and several small tourist accommodations. Bolshoe Goloustnoe has several hotels and tourist camps.

To assess disturbance gradients and ecological responses from littoral-to-pelagic zones and laterally along the shoreline, our transect consisted of 17 sampling sites that were meant to characterize differences along these gradients. Pelagic sites were located 2 to 5 km offshore from each of the developed sites in water depths of 900–1300 m (Fig. 1; Table 1). All littoral sites were sampled at approximately the same depth (max depth of ~ 1.25 m) at a distance of 8.90–20.75 m from shore (Table 1), which allowed us to collect samples without the need for SCUBA but precluded us from

**Table 1.** Locational information for each of the 17 sampling stations. “OS” refers to pelagic locations (i.e., “offshore”), whereas other site abbreviations refer to littoral sampling locations.

Site	Latitude	Longitude	Depth (m)	Distance to shore (m)
BK-1	51.90316	105.074	0.7	10
BK-2	51.90365	105.069	0.9	17.5
BK-3	51.90536	105.0957	0.8	10
BGO-1	52.02693	105.401	0.9	18
BGO-2	52.0197	105.3771	1.1	14
BGO-3	52.02649	105.4358	0.7	21
OS-1	51.98559	105.4724	900	NA
KD-1	51.92646	105.245	0.8	20.75
KD-2	51.91807	105.2146	0.9	14.5
MS-1	51.89863	105.1502	0.6	10.5
SM-1	51.87152	104.9801	0.9	11.5
LI-1	51.86825	104.8304	0.6	8.9
LI-2	51.84626	104.8736	0.8	9.4
LI-3	51.85407	104.8622	0.7	9.25
EM-1	51.86005	104.94	0.7	15.5
OS-2	51.8553	104.8148	1300	NA
OS-3	51.85911	105.0769	1400	5000

sampling deeper littoral environments. Due to this constraint, only littoral sites contain macroinvertebrate and algal samples. Otherwise, data are available for both littoral and pelagic sites. At each site, air temperature was measured with a mercury thermometer, and photographs were taken of the substrate and the shoreline. Visual inspection of substrate photographs suggested that littoral sites' substrate was consistent among sites and generally was characterized by large, oblate rocks and gravel.

**IDW population calculation for each sampling location**

We recognized that sewage indicator concentrations at each sampling location may be related to a sampling location's spatial position relative to both the size and proximity of neighboring developed sites. Therefore, we created the IDW population metric to compress, into a single metric, information about human population size, density, and location along the shoreline as well as distance between developed sites and sampling locations.

Our workflow for calculating IDW population required five steps. First, we traced polygons of each lakeside development's perimeter and line geometries of each development's shorelines from satellite imagery for each developed site in Google Earth. Polygons were traced for the entire area of visible development. Similarly, shoreline traces only reflected shoreline length for which there was visible development. Second,

polygon and line geometries were downloaded from Google Earth as a .kml file. Third, the .kml file was imported into the R statistical environment (R Core Team 2019), where using the *sf* package (Pebesma 2018) we calculated shoreline length, polygon area, and centroid location for each developed site. Fourth, we joined point locations of each sampling site with the spatial polygons to calculate the distance from each sampling location to each developed site's centroid. Fifth, we calculated IDW population for each sampling location, using formula (1)

$$I_j = \frac{P_{LI} \times L_{LI}}{D_{j,LI}^2} + \frac{P_{BK} \times L_{BK}}{D_{j,BK}^2} + \frac{P_{BGO} \times L_{BGO}}{D_{j,BGO}^2} \quad (1)$$

where  $I$  is the IDW population at sampling location  $j$ ,  $P$  is the population at each of the three developed sites Listvyanka (LI), Bolshie Koty (BK), Bolshoe Goloustnoe (BGO),  $A$  is the area of a developed site in  $\text{km}^2$ ,  $L$  is the shoreline length at a developed site in km, and  $D$  is the distance from developed site  $j$  to each developed site's centroid in km. As these population estimates are based on census data, they reflect current, static populations and do not account for seasonal population swings from tourism.

### Nutrients

Water samples for nutrient analyses were collected in 150 mL glass jars that had been washed with phosphate-free soap and rinsed three times with water from the sampling location. Samples were collected at a depth of approximately 0.75 m in duplicates and immediately frozen at  $-20^\circ\text{C}$  until processing at the A. P. Vinogradov Institute of Geochemistry (Siberian Branch of the Russian Academy of Sciences, Irkutsk). Samples were not filtered prior to freezing, meaning that nitrogen and ammonium concentrations may include intracellular nitrogen and overestimate dissolved nitrogenous forms in the water column.

For ammonium (RD:52.24.383-2018 2018) and nitrate (RD:52.24.380-2017 2018) concentrations, samples were analyzed with a spectrophotometer (SF-26). GSO 7258-96 and 7259-96 standards of  $1 \text{ g L}^{-1}$  stock concentration were used to calibrate nitrate and ammonium measurements, respectively. When nitrate and ammonium analyses could be performed within 24 h after thawing, samples were kept at  $2-8^\circ\text{C}$  without addition of preservative agents. When nitrate analyses were performed between 24 and 48 h after thawing, samples were kept at  $3-5^\circ\text{C}$  and chloroform was added as a preservative at a ratio of 2–4 mL per 1 L of sample volume. When ammonium analyses were performed within 24–96 h after thawing, samples were kept at  $3-5^\circ\text{C}$  and  $\sim 10\%$  sulfuric acid solution was added as a preservative. Phosphorus concentration was measured with a spectrophotometer (SF-46) following the addition of persulfate (GOST:18309-2014 2016). When possible, samples were analyzed within 3 h of thawing. When analyses could not be performed within 3 h, samples

were kept at  $3-5^\circ\text{C}$  and chloroform was added as a preservative at a ratio of 2–4 mL per 1 L of sample volume. Minimal detection limits were estimated as  $0.01 \text{ mg L}^{-1}$  for nitrate,  $0.005 \text{ mg L}^{-1}$  for ammonium, and  $0.04 \text{ mg L}^{-1}$  for phosphorus. Concentrations are reported in  $\text{mg L}^{-1}$  of each analyte.

For comparable methods in English, we recommend data users consult International Standards Organization (ISO) (1984) and ISO (2004) as analogs. Copies of the Russian-language methods are included in the OSP portal within the directory “Nearshore\_sampling/methods.”

### Chlorophyll *a*

Water samples were collected in 1.5 liters plastic bottles from a depth of approximately 0.75 m. Although we did not note the plastic bottles' materials within the field, all bottles for Chl *a* measurement were cleaned, beverage bottles and likely made of polyethylene terephthalate. Within 12 h of collection, three subsamples (up to 150 mL each) were filtered through 25-mm diameter,  $0.2 \mu\text{m}$  pore size nitrocellulose filters. Filters were then placed in a 35 mm petri dish, which was wrapped with aluminum foil to prevent light exposure, and frozen in the dark until processing.

Chlorophyll samples were processed in a manner similar to that of Welschmeyer (1994). Nitrocellulose filters were ground in 10 mL of 90% HPLC-grade acetone, in which chlorophyll extraction was allowed to proceed overnight. Chlorophyll extract was then analyzed using a Turner Designs 10-AU fluorometer (Turner Design) using an excitation wavelength of 436 nm and emission of 680 nm. 10-AU Secondary Solid Standard (P/N 10-AU-904) was used to calibrate fluorometer prior to samples being processed. Blank samples registered a raw fluorescence of approximately 0.1 FL units. Concentrations were calculated using formula 2

$$\text{Chlorophyll concentration} = (\text{extract reading} - \text{blank reading}) \times \frac{\text{mL of extract}}{\text{mL of filtered sample}} \quad (2)$$

Detection limits are estimated to be approximately  $0.02 \text{ mg L}^{-1}$ . Concentrations are reported as  $\text{mg L}^{-1}$ .

### Pharmaceuticals and personal care products

Water samples for PPCP analysis were collected in 250 mL amber glass bottles that were rinsed with either methanol or acetone and then three times with sample water prior to collections. Following collection, samples were refrigerated and kept in the dark until solid phase extraction (SPE).

Within 12 h of collection, samples were filtered directly from the amber glass bottle using an in-line Teflon filter holder with glass microfiber GMF ( $1.0 \mu\text{m}$  pore size, WhatmanGrad 934-AH) in tandem with a SPE cartridge (200 mg HLB, Waters Corporation) connected to a 1-liter vacuum flask. Lab personnel wore gloves and face masks to

minimize contamination. Prior to filtration, SPE cartridges were primed with at least 5 mL of either methanol or acetone and then washed with at least 5 mL of sample water. Rate of extraction was maintained at approximately 1 drop per second. Extraction proceeded until water could no longer pass through the SPE cartridge or until all collected water was filtered. Cartridges were stored in Whirlpaks at  $-20^{\circ}\text{C}$  until analysis for 18 PPCP residues using liquid chromatography tandem mass spectrometry (LC–MS–MS) following methods of Lee et al. (2016) and D’Alessio et al. (2018) with labeled internal standards ( $^{13}\text{C}_3$ -caffeine, methamphetamine-d8, MDMA-d8, morphine-d3, and  $^{13}\text{C}_6$ -sulfamethazine). Detection limits are estimated to be  $0.001\ \mu\text{g L}^{-1}$  based on a 500 mL sample volume. Concentrations are reported in  $\mu\text{g L}^{-1}$ .

### Microplastics

At each location, samples were collected at a depth of approximately 0.75 m in triplicate using 1.5 liters clear plastic bottles that were washed thoroughly with sample water before each collection. Samples were collected by hand for each littoral site and with a metal bucket from aboard the ship for pelagic sites.

For processing, each sample was vacuum filtered on to a 47-mm diameter GF/F filter. During filtration, aluminum foil was used to cover the filtration funnel to prevent contamination from airborne microplastic particles. After filtration, filters were dried under vacuum pressure and then stored in 50-mm petri dishes. Following filtration of all three replicates, the filtrate was collected and then re-filtered through a GF/F filter as a control for contamination from the plastic vacuum funnel or potentially airborne microplastics.

Microplastic counting involved visual inspection of the entire GF/F in a similar manner to methods described in Hanvey et al. (2017). Visual enumeration was conducted under a stereo microscope with  $\sim 100\times$  magnification, and microplastics were classified into one of three categories: fibers, fragments, or beads. For all categories, plastics were defined as observed objects with apparent artificial colors, so as to not enumerate plastics potentially contributed from the sampling bottle itself. Fibers were defined as smooth, long plastics with consistent diameters. Fragments were defined as plastics with irregularly sharp or jagged edges. Beads were defined as spherical plastics. Although we did not measure microplastic size, this technique likely allowed us to reliably quantify microplastics as small as  $\sim 300\ \mu\text{m}$  (Hanvey et al. 2017). During enumeration, GF/Fs remained covered in the petri dish to minimize potential for contamination from the air.

It is worth noting that since the time of our field sampling, evidence has accumulated that our methods likely dramatically underestimated microplastic abundance (Wang and Wang 2018; Brandon et al. 2020). Recent investigations of microplastics in Lake Baikal near Bolshie Koty (BK) used analogous methods and measured similar microplastic

concentrations (Karnaukhov et al. 2020). Future studies aiming to use these data for comparison or supplementing potential data gaps should consider the minimum microplastic size that could be reliably detected by the method, so as to ensure data are comparable across methods.

### Periphyton collection and abundance estimates

At each littoral site, we haphazardly selected three rocks representative of local substrate. A plastic stencil was used to define a surface area of each rock from which we scraped a standardized  $14.5\ \text{cm}^2$  patch of periphyton. Samples were preserved with Lugol’s solution and stored in plastic scintillation vials. Additional periphyton was collected in composite from each site for fatty acid and stable isotope analysis.

Periphyton taxonomic identification and enumeration was performed by subsampling  $10\ \mu\text{L}$  aliquots from each preserved sample, containing approximately 10–15 mL of preserved periphyton. For all  $10\ \mu\text{L}$  aliquots, cells, filaments, and colonies were counted, for the entire subsample, until at least 300 cells were identified for a given sampling replicate. If the first aliquot contained less than 300 cells, we counted additional subsamples until we reached at least 300 cells in total. In instances when 300 cells were counted before finishing a subsample, we still counted the entire aliquot. Taxa were classified into broad categories consistent with Baikal algal taxonomy (Izhboldina 2007), using coarse groupings to capture general patterns in relative algal abundance. As a result, algal groups consisted of diatoms, *Ulothrix* spp., *Spirogyra* spp., and the green algal Order Tetrasporales.

Separate periphyton samples for stable isotope and fatty acid analyses were also collected. Instead of preserving samples in Lugol’s solution, these samples were immediately frozen at  $-20^{\circ}\text{C}$  at the field station. The samples were later transferred to the lab in the United States via a Dewar flask with dry ice.

### Benthic macroinvertebrate collection and abundance estimates

Three kick-net samples were collected for assessment of benthic community composition and abundance. Using a D-net, we collected macroinvertebrates by flipping over 1–3 rocks, and then sweeping five times in a left-to-right motion across approximately 1 m. After the series of sweeps, the catch was rinsed into a plastic bucket. For each replicate, bucket contents were concentrated using a  $64\text{-}\mu\text{m}$  mesh and placed in glass jars with 40% ethanol (vodka; the only preservative available to us at the time) for preservation and refrigerated at  $4^{\circ}\text{C}$  aboard the research vessel. The 40% ethanol preservative was replaced with  $\sim 80\%$  ethanol upon return to the lab within 24–48 h, and samples were stored at  $\sim 4^{\circ}\text{C}$ .

Invertebrate taxonomic identification and enumeration were performed under a stereo microscope. All adult amphipods were identified to species according to Takhteev and Didorenko (2015), whereas juveniles were identified to genus. Mollusks were identified to the family level according to

Sitnikova (2012). Leeches were enumerated at the subclass level, but were likely all from the family Glossiphoniidae based on size, depth of sampling locations, and invertebrate communities sampled (Kaygorodova 2012). Like mollusks, caddisflies were also enumerated at the order level, although Baikal does contain over 14 species of caddisfly (Valuyskiy et al. 2020). Flatworms were enumerated at the phylum level. All isopods enumerated were from the family Asellidae. Aside from having limited time available to spend with Baikal taxonomists during our field campaign, our choice of taxonomic resolution ultimately was a result of relative abundance for each taxonomic group, where amphipods were the most abundant taxa and flatworms were among the least abundant taxa across all sites. All samples contained oligochaetes and polychaetes, but due to poor preservation, these taxa were not counted. Six samples of the 42 collected were not well-preserved and were excluded from further analyses, in order to reduce errors in identification. KD-1 and LI-1 were the only sites with 1 sample counted. BK-2 and KD-2 each had two samples counted.

Separate collections were conducted for invertebrate fatty acid and stable isotope analyses. Invertebrates were collected using a D-net and by hand. Organisms collected by hand included amphipod species that were observed from the community composition D-net collections but not readily observed in the stable isotope and fatty acid D-net collections. Collected organisms were live-sorted, identified to species, and then frozen at  $-20^{\circ}\text{C}$  at the field station. The samples were later transferred to the lab in the United States via a Dewar flask with dry ice.

Due to some samples warming in transit, we only processed samples that were completely frozen upon arrival to the United States. Given the potential for fatty acids to highlight more subtle, multivariate ecological responses along our transect than stable isotopes, we prioritized both periphyton and macroinvertebrate fatty acid analyses over stable isotope analyses. As such, there is an imbalance across species' abundance, stable isotope, and fatty acid data. Dominant taxa, such as *Eulimnogammarus verucosus* and *Eulimnogammarus vittatus*, though have paired data throughout the transect, whereas less common taxa, such as *Brandtia* spp., only have abundance estimates. Table 2 summarizes data available for each variable and taxonomic group.

### Stable isotope analysis

Following freeze-drying, measurements of periphyton and macroinvertebrate  $\delta^{15}\text{N}$  and  $\delta^{13}\text{C}$  values were performed on an elemental analyzer-isotope ratio mass spectrometer (EA-IRMS; Finnigan DELTAplus XP, Thermo Scientific) at the Large Lakes Observatory, University of Minnesota Duluth. Stable isotope values were calibrated against certified reference materials including L-glutamic acid (NIST SRM 8574), low organic soil and sorghum flour (standards B-2153 and B-2159

from Elemental Micro-analysis Ltd.) and in-house standards (acetanilide and caffeine).

### Fatty acid analysis

Following freeze-drying, samples were transferred to 10 mL glass centrifuge vials, and 2 mL of 100% chloroform was added to each under nitrogen gas. Samples were allowed to sit in chloroform overnight at  $-80^{\circ}\text{C}$ . Fatty acid extractions generally involved three phases: (1) 100% chloroform extraction, (2) chloroform-methanol extraction, and (3) fatty acid methylation. Fatty acid extraction methods were adapted from Schram et al. (2018).

After overnight chloroform extraction, samples underwent a chloroform-methanol extraction three times. To each sample, we added 1 mL cooled 100% methanol, 1 mL chloroform : methanol solution (2 : 1), and 0.8 mL 0.9% NaCl solution. Samples were inverted three times and sonicated on ice for 10 min. Next, samples were vortexed for 1 min, and centrifuged for 5 min (3000 rpm) at  $4^{\circ}\text{C}$ . Using a double pipette technique, the lower organic layer was removed and kept under nitrogen. After the third extraction, samples were evaporated under nitrogen flow, and resuspended in 1.5 mL chloroform and stored at  $-20^{\circ}\text{C}$  overnight.

Once resuspended in chloroform, 1 mL of chloroform extract was transferred to a glass centrifuge tube with a glass syringe as well as an internal standard of 4  $\mu\text{L}$  of 19-carbon fatty acid. Samples were then evaporated under nitrogen, and then 1 mL of toluene and 2 mL of 1% sulfuric acid-methanol was added. The vial was closed under nitrogen gas and then incubated in  $50^{\circ}\text{C}$  water bath for 16 h. After incubation, samples were removed from the bath, allowed to reach room temperature and stored on ice. Next, we performed a potassium carbonate-hexane extraction twice. To each sample, we added 2 mL of 2% potassium bicarbonate and 5 mL of 100% hexane, inverting the capped vial so as to mix the solution. Samples were centrifuged for 3 min (1500 rpm) at  $4^{\circ}\text{C}$ . The upper hexane layer was then removed and placed in a vial to evaporate under nitrogen flow. Once almost evaporated, 1 mL of 100% hexane was added and stored in a glass amber autosampler vial for GC/MS quantification. GC/MS quantification was performed with a Shimadzu QP2020 GC/MS following Schram et al. (2018). As part of our peak quantification protocol, we quantified and identified every lipid compound that showed up in the chromatogram. Each sample contained peaks that were associated with known fatty acids, and among the 59 fatty acids contained in our dataset, few fatty acids were completely absent from a sample. Consequently, it is difficult for us to definitively ascribe a minimal detection limit to this analysis, but based on standards used, we estimate that this procedure had a minimal detection limit of  $1\text{ ng mL}^{-1}$ .

Following methylation, remaining extracts were assessed for total lipid masses. Remaining sample extracts ( $\sim 0.5\text{ mL}$ ) were allowed to evaporate to dryness under a fume hood overnight. Dried samples were then left in a weight room to

**Table 2.** Summary table of algal and macroinvertebrate data within the dataset. Although fatty acids contain data on *Hyalella* spp., these specimens were likely misidentified in the field before processing. For consistency and detailing the breadth of fatty acid profiles among Baikal's littoral amphipods, we have included them in the dataset, but caution should be taken when considering these fatty acids explicitly as those representative of *Hyalella* spp.

Variable	Course taxonomic grouping	Finest taxonomic group in dataset
Abundance estimates	Amphipoda	<i>Brandtia latissima</i> subsp. (Dorogostaiskii 1930; Dybowsky 1874)
		<i>Brandtia parasitica parasitica</i> (Dybowsky 1874)
		<i>Cryptoropus inflatus</i> (Dybowsky 1874)
		<i>Cryptoropus pachytus</i> (Dybowsky 1874)
		<i>Cryptoropus rugosus</i> (Dybowsky 1874)
		<i>Eulimnogammarus capreolus</i> (Dybowsky 1874)
		<i>Eulimnogammarus cruentus</i> (Dorogostaiskii 1930)
		<i>Eulimnogammarus cyaneus</i> (Dybowsky 1874)
		<i>Eulimnogammarus grandimanus</i> (Bazikalova 1945)
		<i>Eulimnogammarus maacki</i> (Gerstfeldt 1858)
		<i>Eulimnogammarus marituji</i> (Bazikalova 1945)
		<i>Eulimnogammarus verucossus</i> (Gerstfeldt 1858)
		<i>Eulimnogammarus viridis viridis</i> (Dybowsky 1874)
		<i>Eulimnogammarus vittatus</i> (Dybowsky 1874)
		<i>Pallasea brandtia brandita</i> (Dybowsky 1874)
		<i>Pallasea brandtii tenera</i> (Sovinskii 1930)
		<i>Pallasea cancelloides</i> (Gerstfeldt 1858)
		<i>Pallasea cancellus</i> (Pallas 1776)
		<i>Pallasea viridis</i> (Garjajev 1901)
		<i>Poekilogammarus crassimus</i> (Sovinskii 1915)
<i>Poekilogammarus ephippiatus</i> (Dybowsky 1874)		
<i>Poekilogammarus megonychus perpolitus</i> (Takhteev 2002)		
<i>Poekilogammarus pictus</i> (Dybowsky 1874)		
Molluska		Acroloxiidae
		Baicaliidae
		Benedictidae
		Maackia
		Planorbidae
		Valvatidae
Other macroinvertebrates		Asellidae
		Caddisflies
		Hirudinea
		Planaria
Benthic algae		Diatom
		<i>Ulothrix</i> spp.
		<i>Spirogyra</i> spp.
		Tetrasporales
Stable isotopes	Amphipoda	<i>Eulimnogammarus cyaneus</i> (Dybowsky 1874)
		<i>Eulimnogammarus verucossus</i> (Gerstfeldt 1858)
		<i>Eulimnogammarus vittatus</i> (Dybowsky 1874)
		<i>Pallasea cancellus</i> (Pallas 1776)
	Benthic algae	Periphyton

(Continues)

**Table 2.** Continued

Variable	Course taxonomic grouping	Finest taxonomic group in dataset
Fatty acids	Amphipoda	<i>Eulimnogammarus cyaneus</i> (Dybowsky 1874) <i>Eulimnogammarus verucossus</i> (Gerstfeldt 1858) <i>Eulimnogammarus vittatus</i> (Dybowsky 1874) <i>Hyalella</i> spp. <i>Pallasea cancellus</i> (Pallas 1776)
	Molluska	Processed in composite and not identified to family
	Benthic algae	Periphyton <i>Draparnaldia</i> spp.

acclimatize for 30–60 min and then massed within the scintillation vials. To calculate an average lipid mass, samples were massed three times, so as also to assess deviation in measurements. Lipid gravimetry is reported as the mg of lipids per g of dry-weight tissue.

### Technical validation

The dataset had three main validation procedures: taxonomic, analytical, and reproducible.

For taxonomic validation, all phylogenetic groupings were based off most recent identification keys. Amphipods were identified according to Takhteev & Didorenko (2015). Mollusks were identified according to Sitnikova (2012). Algal taxa were identified according to Izhboldina (2007). For consistency, all taxa were identified by one person (Michael F. Meyer), who was trained by experts in Baikal algal and macroinvertebrate taxonomy.

For analytical validation, internal standards were used for all mass-spectroscopy analyses. PPCP analyses involved labeled internal standards ( $^{13}\text{C}_3$ -caffeine, methamphetamine-d8, MDMA-d8, morphine-d3, and  $^{13}\text{C}_6$ -sulfamethazine). Stable isotope values were calibrated against certified reference materials including L-glutamic acid (NIST SRM 8574), low organic soil and sorghum flour (standards B-2153 and B-2159 from Elemental Micro-analysis Ltd.) and in-house standards (acetanilide and caffeine). Replicate analyses of external standards showed a mean standard deviation of 0.06‰ and 0.09‰, for  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$ , respectively. Finally, fatty acid estimations used an internal 19 : 0 standard to assess oxidation of fatty acids during extraction, methylation, and quantification.

For data reproducibility, data aggregation and harmonization procedures were conducted in the R statistical environment (R Core Team 2019), using the tidyverse (Wickham et al. 2019) packages. As part of the data aggregation, an initial cleaning script (00\_disaggregated\_data\_cleaning.R) removed incorrect spellings, erroneous data values, and inconsistent column names from raw data. This step created the

standardized CSV files detailed above, which are available on the EDI repository (this study). Raw data files are available on the project's OSF portal (Meyer et al. 2015) but are not included in the EDI repository to prevent confusion or incorrect usage. Data hosted on EDI are at the replicate-level but can be aggregated to the sampling-site-level using script "01\_data\_cleaning.R." In addition to aggregation scripts, six R scripts used for analyses in Meyer et al. (Under Review) are also available on the EDI repository within the compressed entity "scripts.tar.gz." All R code for data aggregation was written by one person (Michael F. Meyer) and then independently reviewed by two others (Matthew R. Brouil and Kara H. Woo) to confirm that code performed as intended, was well documented, and annotations were complete.

### A commitment to FAIR and TRUST principles

Throughout the dataset's development, we strove to incorporate both FAIR (Findable, Accessible, Interoperable, and Reproducible) and TRUST (Transparency, Responsibility, User Focus, Sustainability, and Technology) principles where applicable.

With respect to FAIR principles (Wilkinson et al. 2016), the data are openly accessible in a standardized, replicate-level format on the EDI portal. The 11 CSV files contained within the dataset are entirely interoperable using the "site" column, enabling all variables to efficiently be merged together. Finally, all analytical and some data wrangling scripts are available on the EDI portal in a compressed format, such that future users can reproduce data manipulation and analyses described in Meyer et al. (Under Review).

With respect to TRUST principles (Lin et al. 2020), we strove to document additional metadata and data-cleaning practices in a public OSF repository (Meyer et al. 2015). These steps are not necessarily critical to the core EDI dataset, but provide increased transparency for future users wishing recreate the dataset de novo. All "raw" data are provided in the OSF portal, including an initial cleaning script (00\_disaggregated\_data\_cleaning.R) to remove incorrect spellings,

erroneous data values, and inconsistent column names. This repository also includes photographs of both field notes as well as photographs of shoreline and substrate from sampling locations. To empower and expedite future reuse, all directories are accompanied with documentation that details directory contents, and all associated scripts are documented and annotated. While many of the files are redundant from the EDI repository, the OSF repository is meant to supplement the EDI repository, so as to enable sustainable, user-focused transparency of how data were collected and cleaned from their raw formats.

### Data use and recommendations for reuse

Recognizing the potential for continued low-level, sewage pollution at Lake Baikal (Timoshkin et al. 2016, 2018; Volkova et al. 2018) and lakes worldwide (Yang et al. 2018; Meyer et al. 2019), the final dataset can be applied to a suite of research questions pertaining to ecological responses to human disturbance. We highlight two main areas for immediate application.

First, the final data products can be harmonized with other littoral sampling efforts throughout Lake Baikal, so as to enhance spatial coverage and data diversity. Since 2010, Lake Baikal has experienced increasing filamentous algal abundance, especially near larger lakeside developments (Kravtsova et al. 2014; Timoshkin et al. 2016, 2018; Volkova et al. 2018). Recent benthic algal surveys throughout Lake Baikal's entirety, but especially near our sampling locations, have suggested that cosmopolitan filamentous algae, such as *Spirogyra* spp., tend to be more abundant near larger lakeside developments (Timoshkin et al. 2016; Volkova et al. 2018). For example, Listvyanka is a small town located at the beginning of the Angara River, Lake Baikal's only surface outflow. While Listvyanka's permanent population is approximately 2000 persons, the town is a growing tourism hub, and hosts over 1.2 million tourists per year (Interfax-Tourism 2018). Surveys conducted near Listvyanka have suggested increased *Spirogyra* spp. abundance is associated with wastewater release (Timoshkin et al. 2016). Although wastewater inputs are likely low and are diluted to negligible concentrations offshore (Meyer et al. Under Review), combining monitoring efforts across spatial and temporal scales is necessary to evaluate the spatial and temporal extent of wastewater entering Lake Baikal. As such, our data could complement previous, current, and future monitoring efforts, where observations may be missing.

Second, the final data products are useful to expanding freshwater PPCP, microplastic, and associated biological responses across large spatial scales. Recent syntheses of the PPCP literature have reported that studies involving lakes are less abundant relative to those focused on lotic systems (Meyer et al. 2019). Likewise, microplastic studies have noted

that freshwater environments are less represented in the literature relative to marine ecosystems (Horton et al. 2017). For both PPCPs and microplastics, toxic responses to even minute concentrations can be uncertain and differ between ecosystem types (e.g., Rosi-Marshall et al. 2013 for lotic and Shaw et al. 2015 for lentic). As a result of PPCPs and microplastics garnering increasing attention worldwide, sampling of PPCPs and microplastics with colocated biological data across multiple spatial and temporal scales would be necessary to synthesize biotic responses to micropollutants across systems. Although our data constitute a limited sample number of PPCP and microplastic data that exist globally, our final data products are highly structured and flexible for merging with similar datasets. Additionally, our dataset's sequential harmonization workflow could be adopted by similar monitoring efforts, thereby facilitating data interoperability. Through integration with similar monitoring efforts, our dataset can contribute to global synthesis of emerging contaminant consequences, especially in a region of the world that is often not easily accessible to many researchers.

### References

- Anisimov, O., and S. Reneva. 2006. Permafrost and changing climate: The Russian perspective. *Ambio* **35**: 169–175.
- Baquero, O. S. 2019. ggsm: North symbols and scale bars for maps created with “ggplot2” or “ggmap”.
- Barnes, D. K. A., F. Galgani, R. C. Thompson, and M. Barlaz. 2009. Accumulation and fragmentation of plastic debris in global environments. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364**: 1985–1998. doi:10.1098/rstb.2008.0205
- Bendz, D., N. A. Paxéus, T. R. Ginn, and F. J. Loge. 2005. Occurrence and fate of pharmaceutically active compounds in the environment, a case study: Höje River in Sweden. *J. Hazard. Mater.* **122**: 195–204. doi:10.1016/j.jhazmat.2005.03.012
- Bondarenko, N. A., and others. 2021. *Dolichospermum lemmermannii* (Nostocales) bloom in world's deepest Lake Baikal (East Siberia): Abundance, toxicity and factors influencing growth. *Limnol. Freshw. Biol.* **1**: 1101–1110. doi:10.31951/2658-3518-2021-A-1-1101
- Brandon, J. A., A. Freibott, and L. M. Sala. 2020. Patterns of suspended and salp-ingested microplastic debris in the North Pacific investigated with epifluorescence microscopy. *Limnol. Oceanogr. Lett.* **5**: 46–53. doi:10.1002/lo2.10127
- Brodin, T., J. Fick, M. Jonsson, and J. Klaminder. 2013. Dilute concentrations of a psychiatric drug alter behavior of fish from natural populations. *Science* **339**: 814–815. doi:10.1126/science.1226850
- Camilleri, A. C., and T. Ozersky. 2019. Large variation in periphyton  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  values in the upper Great Lakes: Correlates and implications. *J. Great Lakes Res.* **45**: 986–990. doi:10.1016/j.jglr.2019.06.003

- Costanzo, S. D., M. J. O'Donohue, W. C. Dennison, N. R. Loneragan, and M. Thomas. 2001. A new approach for detecting and mapping sewage impacts. *Mar. Pollut. Bull.* **42**: 149–156. doi:[10.1016/S0025-326X\(00\)00125-9](https://doi.org/10.1016/S0025-326X(00)00125-9)
- D'Alessio, M., S. Onanong, D. D. Snow, and C. Ray. 2018. Occurrence and removal of pharmaceutical compounds and steroids at four wastewater treatment plants in Hawai'i and their environmental fate. *Sci. Total Environ.* **631–632**: 1360–1370. doi:[10.1016/j.scitotenv.2018.03.100](https://doi.org/10.1016/j.scitotenv.2018.03.100)
- Dalsgaard, J., M. S. John, G. Kattner, D. Müller-Navarra, and W. Hagen. 2003. Fatty acid trophic markers in the pelagic marine environment, p. 225–340. *In Advances in marine biology*. Elsevier.
- Edmondson, W. T. 1970. Phosphorus, nitrogen, and algae in Lake Washington after diversion of sewage. *Science* **169**: 690–691.
- Fellows, I., and J. P. Stotz. 2019. OpenStreetMap: Access to open street map raster images.
- Focazio, M. J., D. W. Kolpin, K. K. Barnes, E. T. Furlong, M. T. Meyer, S. D. Zaugg, L. B. Barber, and M. E. Thurman. 2008. A national reconnaissance for pharmaceuticals and other organic wastewater contaminants in the (United States - II) untreated drinking water sources. *Sci. Total Environ.* **402**: 201–216. doi:[10.1016/j.scitotenv.2008.02.021](https://doi.org/10.1016/j.scitotenv.2008.02.021)
- Gartner, A., P. Lavery, and A. J. Smit. 2002. Use of  $\delta\text{N-15}$  signatures of different functional forms of macroalgae and filter-feeders to reveal temporal and spatial patterns in sewage dispersal. *Mar. Ecol. Prog. Ser.* **235**: 63–73. doi:[10.3354/meps235063](https://doi.org/10.3354/meps235063)
- GOST:18309-2014. 2016. Methods for determination of phosphorus-containing matters (with corrections) (Методы определения фосфорсодержащих веществ).
- Green, D. S. 2016. Effects of microplastics on European flat oysters, *Ostrea edulis* and their associated benthic communities. *Environ. Pollut.* **216**: 95–103. doi:[10.1016/j.envpol.2016.05.043](https://doi.org/10.1016/j.envpol.2016.05.043)
- Hall, R. I., P. R. Leavitt, R. Quinlan, A. S. Dixit, and J. P. Smol. 1999. Effects of agriculture, urbanization, and climate on water quality in the northern Great Plains. *Limnol. Oceanogr.* **44**: 739–756. doi:[10.4319/lo.1999.44.3\\_part\\_2.0739](https://doi.org/10.4319/lo.1999.44.3_part_2.0739)
- Hampton, S. E., S. C. Fradkin, P. R. Leavitt, and E. E. Rosenberger. 2011. Disproportionate importance of near-shore habitat for the food web of a deep oligotrophic lake. *Mar. Freshw. Res.* **62**: 350. doi:[10.1071/MF10229](https://doi.org/10.1071/MF10229)
- Hampton, S. E., L. R. Izmet'eva, M. V. Moore, S. L. Katz, B. Dennis, and E. A. Silow. 2008. Sixty years of environmental change in the world's largest freshwater Lake - Lake Baikal, Siberia. *Glob. Chang. Biol.* **14**: 1947–1958. doi:[10.1111/j.1365-2486.2008.01616.x](https://doi.org/10.1111/j.1365-2486.2008.01616.x)
- Hampton, S. E., and others. 2018. Recent ecological change in ancient lakes. *Limnol. Oceanogr.* **63**: 2277–2304. doi:[10.1002/lno.10938](https://doi.org/10.1002/lno.10938)
- Hanvey, J. S., P. J. Lewis, J. L. Lavers, N. D. Crosbie, K. Pozo, and B. O. Clarke. 2017. A review of analytical techniques for quantifying microplastics in sediments. *Anal. Methods* **9**: 1369–1383. doi:[10.1039/C6AY02707E](https://doi.org/10.1039/C6AY02707E)
- Horton, A. A., A. Walton, D. J. Spurgeon, E. Lahive, and C. Svendsen. 2017. Microplastics in freshwater and terrestrial environments: Evaluating the current understanding to identify the knowledge gaps and future research priorities. *Sci. Total Environ.* **586**: 127–141. doi:[10.1016/j.scitotenv.2017.01.190](https://doi.org/10.1016/j.scitotenv.2017.01.190)
- Interfax-Tourism. 2018. Байкал с января по август 2018 года посетили 1,2 миллиона туристов (1.2 million tourists visited Baikal from January through August 2018). Interfax-Tourism, October 25.
- International Standards Organization (ISO). 1984. ISO 6777: 1984(en) Water quality—determination of nitrite—molecular absorption spectrometric method. ISO 6777. ISO 6777 ISO.
- International Standards Organization (ISO). 2004. ISO 6878: 2004(en) Water quality—determination of phosphorus—ammonium molybdate spectrometric method. ISO 6878. ISO 6878 ISO.
- Izboldina, L. A. 2007. *Guide and key to benthic and periphyton algae of Lake Baikal (meio- and macrophytes) with brief notes on their ecology*. Nauka-Centre.
- Izmet'eva, L. R., and others. 2016. Lake-wide physical and biological trends associated with warming in Lake Baikal. *J. Great Lakes Res.* **42**: 6–17. doi:[10.1016/j.jglr.2015.11.006](https://doi.org/10.1016/j.jglr.2015.11.006)
- Jeppesen, E., and others. 2005. Lake responses to reduced nutrient loading – an analysis of contemporary long-term data from 35 case studies. *Freshw. Biol.* **50**: 1747–1771. doi:[10.1111/j.1365-2427.2005.01415.x](https://doi.org/10.1111/j.1365-2427.2005.01415.x)
- Karnaukhov, D., S. Birietskaya, E. Dolinskaya, M. Teplykh, N. Silenko, Y. Ermolaeva, and E. Silow. 2020. Pollution by macro- and microplastic of large lacustrine ecosystems in Eastern Asia. *Poll. Res.* **2**: 353–355.
- Kassambara, A. 2019. ggpubr: “ggplot2” based publication ready plots.
- Katz, S. L., L. R. Izmet'eva, S. E. Hampton, T. Ozersky, K. Shchapov, M. V. Moore, S. V. Shimaraeva, and E. A. Silow. 2015. The “Melosira years” of Lake Baikal: Winter environmental conditions at ice onset predict under-ice algal blooms in spring. *Limnol. Oceanogr.* **60**: 1950–1964. doi:[10.1002/lno.10143](https://doi.org/10.1002/lno.10143)
- Kaygorodova, I. 2012. An illustrated checklist of leech species from Lake Baikal (Eastern Siberia, Russia). *Dataset Papers in Biology* **2013**: e261521. doi:[10.7167/2013/261521](https://doi.org/10.7167/2013/261521).
- Kolpin, D. W., E. T. Furlong, M. T. Meyer, E. M. Thurman, S. D. Zaugg, L. B. Barber, and H. T. Buxton. 2002. Pharmaceuticals, hormones, and other organic wastewater contaminants in U.S. Streams, 1999–2000: A national reconnaissance. *Environ. Sci. Technol.* **36**: 1202–1211. doi:[10.1021/es011055j](https://doi.org/10.1021/es011055j)
- Kozhova, O. M., and L. R. Izmet'eva. 1998. *Lake Baikal: Evolution and biodiversity*. Backhuys Publishers.

- Kravtsova, L. S., and others. 2014. Nearshore benthic blooms of filamentous green algae in Lake Baikal. *J. Great Lakes Res.* **40**: 441–448. doi:[10.1016/j.jglr.2014.02.019](https://doi.org/10.1016/j.jglr.2014.02.019)
- Lapointe, B. E., L. W. Herren, D. D. Debortoli, and M. A. Vogel. 2015. Evidence of sewage-driven eutrophication and harmful algal blooms in Florida's Indian River Lagoon. *Harmful Algae* **43**: 82–102. doi:[10.1016/j.hal.2015.01.004](https://doi.org/10.1016/j.hal.2015.01.004)
- Lee, S. S., A. M. Paspalof, D. D. Snow, E. K. Richmond, E. J. Rosi-Marshall, and J. J. Kelly. 2016. Occurrence and potential biological effects of amphetamine on stream communities. *Environ. Sci. Technol.* **50**: 9727–9735. doi:[10.1021/acs.est.6b03717](https://doi.org/10.1021/acs.est.6b03717)
- Lin, D., and others. 2020. The TRUST principles for digital repositories. *Sci. Data* **7**: 144. doi:[10.1038/s41597-020-0486-7](https://doi.org/10.1038/s41597-020-0486-7)
- Meyer, M., T. Ozersky, K. Woo, A. W. E. Galloway, M. R. Brousil, and S. Hampton. 2015. Baikal food webs. doi:[10.17605/OSF.IO/9TA8Z](https://doi.org/10.17605/OSF.IO/9TA8Z)
- Meyer, M. F., S. G. Labou, A. N. Cramer, M. R. Brousil, and B. T. Luff. 2020. The global lake area, climate, and population dataset. *Sci. Data* **7**: 174. doi:[10.1038/s41597-020-0517-4](https://doi.org/10.1038/s41597-020-0517-4)
- Meyer, M. F., S. M. Powers, and S. E. Hampton. 2019. An evidence synthesis of pharmaceuticals and personal care products (PPCPs) in the environment: Imbalances among compounds, sewage treatment techniques, and ecosystem types. *Environ. Sci. Technol.* **53**: 12961–12973. doi:[10.1021/acs.est.9b02966](https://doi.org/10.1021/acs.est.9b02966)
- Meyer, M. F., and others. Under Review. Effects of spatially heterogeneous lakeside development on nearshore biotic communities in a large, deep, oligotrophic lake (Lake Baikal, Siberia).
- Moore, J. W., D. E. Schindler, M. D. Scheuerell, D. Smith, and J. Frodge. 2003. Lake eutrophication at the urban fringe, Seattle region, USA. *AMBIO J. Hum. Environ.* **32**: 13–18.
- Moore, M. V., S. E. Hampton, L. R. Izmet'eva, E. A. Silow, E. V. Peshkova, and B. K. Pavlov. 2009. Climate change and the world's "Sacred Sea"—Lake Baikal, Siberia. *Bioscience* **59**: 405–417. doi:[10.1525/bio.2009.59.5.8](https://doi.org/10.1525/bio.2009.59.5.8)
- O'Donnell, D. R., P. Wilburn, E. A. Silow, L. Y. Yampolsky, and E. Litchman. 2017. Nitrogen and phosphorus colimitation of phytoplankton in Lake Baikal: Insights from a spatial survey and nutrient enrichment experiments. *Limnol. Oceanogr.* **62**: 1383–1392. doi:[10.1002/lno.10505](https://doi.org/10.1002/lno.10505)
- Pebesma, E. 2018. Simple features for R: Standardized support for spatial vector data. *R J.* **10**: 439–446. doi:[10.32614/RJ-2018-009](https://doi.org/10.32614/RJ-2018-009)
- Powers, S. M., and others. 2016. Long-term accumulation and transport of anthropogenic phosphorus in three river basins. *Nat. Geosci.* **9**: 353–356. doi:[10.1038/ngeo2693](https://doi.org/10.1038/ngeo2693)
- R Core Team. 2019. R: A language and environment for statistical computing. RD:52.24.380-2017. 2018. Nitrate concentration in waters: Photometric methods with Giress reagent following standardization in a cadmium reducer (Массовая концентрация нитратного азота в водах: Методика измерений фотометрическим методом с реактивом Грисса после восстановления в камиевом редукторе). RD:52.24.383-2018. 2018. Working Document: Concentration of aqueous ammonium: Method for measuring with a photometer using indophenol blue (Руководящий Документ: Массовая концентрация аммонийного азота в водах: Методика измерений фотометрическим методом в виде индофенолового сингео). RD:52.24.383-2018. RD:52.24.383-2018.
- Richmond, E. K., M. R. Grace, J. J. Kelly, A. J. Reisinger, E. J. Rosi, and D. M. Walters. 2017. Pharmaceuticals and personal care products (PPCPs) are ecological disrupting compounds (EcoDC). *Elem. Sci. Anth.* **5**: 66. doi:[10.1525/elementa.252](https://doi.org/10.1525/elementa.252)
- Richmond, E. K., E. J. Rosi, D. M. Walters, J. Fick, S. K. Hamilton, T. Brodin, A. Sundelin, and M. R. Grace. 2018. A diverse suite of pharmaceuticals contaminates stream and riparian food webs. *Nat. Commun.* **9**: 4491. doi:[10.1038/s41467-018-06822-w](https://doi.org/10.1038/s41467-018-06822-w)
- Romera-Castillo, C., M. Pinto, T. M. Langer, X. A. Álvarez-Salgado, and G. J. Herndl. 2018. Dissolved organic carbon leaching from plastics stimulates microbial activity in the ocean. *Nat. Commun.* **9**: 1–7. doi:[10.1038/s41467-018-03798-5](https://doi.org/10.1038/s41467-018-03798-5)
- Rosenberger, E. E., S. E. Hampton, S. C. Fradkin, and B. P. Kennedy. 2008. Effects of shoreline development on the nearshore environment in large deep oligotrophic lakes. *Freshw. Biol.* **53**: 1673–1691. doi:[10.1111/j.1365-2427.2008.01990.x](https://doi.org/10.1111/j.1365-2427.2008.01990.x)
- Rosi-Marshall, E. J., D. W. Kincaid, H. A. Bechtold, T. V. Royer, M. Rojas, and J. J. Kelly. 2013. Pharmaceuticals suppress algal growth and microbial respiration and alter bacterial communities in stream biofilms. *Ecol. Appl.* **23**: 583–593. doi:[10.1890/12-0491.1](https://doi.org/10.1890/12-0491.1)
- Rosi-Marshall, E. J., and T. V. Royer. 2012. Pharmaceutical compounds and ecosystem function: An emerging research challenge for aquatic ecologists. *Ecosystems* **15**: 867–880. doi:[10.1007/s10021-012-9553-z](https://doi.org/10.1007/s10021-012-9553-z)
- Sargent, J. R., and S. Falk-Petersen. 1988. The lipid biochemistry of calanoid copepods. *Hydrobiologia* **167–168**: 101–114. doi:[10.1007/BF00026297](https://doi.org/10.1007/BF00026297)
- Schram, J. B., J. N. Kobelt, M. N. Dethier, and A. W. E. Galloway. 2018. Trophic transfer of macroalgal fatty acids in two urchin species: Digestion, egestion, and tissue building. *Front. Ecol. Evol.* **6**: 1–13. doi:[10.3389/fevo.2018.00083](https://doi.org/10.3389/fevo.2018.00083)
- Shaw, L., C. Phung, and M. Grace. 2015. Pharmaceuticals and personal care products alter growth and function in lentic biofilms. *Environ. Chem.* **12**: 301. doi:[10.1071/EN14141](https://doi.org/10.1071/EN14141)
- Sitnikova, T. Ya. 2012. *Key of the Gastropod Molluscs in the Bay of Bolshie Koty (South-West shoreline of Lake Baikal)*, Irkutsk State University.
- Slowikowski, K. 2019. ggrepel: Automatically Position Non-Overlapping Text Labels with "ggplot2".

- Swann, G. E. A., and others. 2020. Changing nutrient cycling in Lake Baikal, the world's oldest lake. *PNAS* **117**: 27211–27217. doi:[10.1073/pnas.2013181117](https://doi.org/10.1073/pnas.2013181117)
- Taipale, S., U. Strandberg, E. Peltomaa, A. W. E. Galloway, A. Ojala, and M. T. Brett. 2013. Fatty acid composition as biomarkers of freshwater microalgae: Analysis of 37 strains of microalgae in 22 genera and in seven classes. *Aquat. Microb. Ecol.* **71**: 165–178. doi:[10.3354/ame01671](https://doi.org/10.3354/ame01671)
- Takhteev, V. V., and D. I. Didorenko. 2015. *Fauna and ecology of amphipods of Lake Baikal: A Training manual*, V.B. Sochava Institute of Geography SB RAS.
- Timoshkin, O. A., and others. 2016. Rapid ecological change in the coastal zone of Lake Baikal (East Siberia): Is the site of the world's greatest freshwater biodiversity in danger? *J. Great Lakes Res.* **42**: 487–497. doi:[10.1016/j.jglr.2016.02.011](https://doi.org/10.1016/j.jglr.2016.02.011)
- Timoshkin, O. A., and others. 2018. Groundwater contamination by sewage causes benthic algal outbreaks in the littoral zone of Lake Baikal (East Siberia). *J. Great Lakes Res.* **44**: 230–244. doi:[10.1016/j.jglr.2018.01.008](https://doi.org/10.1016/j.jglr.2018.01.008)
- Tong, Y., and others. 2020. Improvement in municipal wastewater treatment alters lake nitrogen to phosphorus ratios in populated regions. *Proc. Natl. Acad. Sci. USA* **117**: 11566–11572. doi:[10.1073/pnas.1920759117](https://doi.org/10.1073/pnas.1920759117)
- Turetsky, M. R., R. K. Wieder, C. J. Williams, and D. H. Vitt. 2000. Organic matter accumulation, peat chemistry, and permafrost melting in peatlands of boreal Alberta. *Écoscience* **7**: 115–122. doi:[10.1080/11956860.2000.11682608](https://doi.org/10.1080/11956860.2000.11682608)
- Valuyskiy, M. Y., S. I. Melnitsky, and V. D. Ivanov. 2020. Structure and evolution of the antennal sensory surface in endemic caddisfly tribes baicalinini and Thamastini (Trichoptera: Apataniidae) from Lake Baikal. *J. Evol. Biochem Phys* **56**: 318–332. doi:[10.1134/S0022093020040031](https://doi.org/10.1134/S0022093020040031)
- Vendel, A. L., F. Bessa, V. E. N. Alves, A. L. A. Amorim, J. Patrício, and A. R. T. Palma. 2017. Widespread microplastic ingestion by fish assemblages in tropical estuaries subjected to anthropogenic pressures. *Mar. Pollut. Bull.* **117**: 448–455. doi:[10.1016/j.marpolbul.2017.01.081](https://doi.org/10.1016/j.marpolbul.2017.01.081)
- Volkova, E. A., N. A. Bondarenko, and O. A. Timoshkin. 2018. Morphotaxonomy, distribution and abundance of *Spirogyra* (Zygnematophyceae, Charophyta) in Lake Baikal, East Siberia. *Phycologia* **57**: 298–308. doi:[10.2216/17-69.1](https://doi.org/10.2216/17-69.1)
- Wang, W., and J. Wang. 2018. Investigation of microplastics in aquatic environments: An overview of the methods used, from field sampling to laboratory analysis. *TrAC Trends Anal. Chem.* **108**: 195–202. doi:[10.1016/j.trac.2018.08.026](https://doi.org/10.1016/j.trac.2018.08.026)
- Welschmeyer, N. A. 1994. Fluorometric analysis of chlorophyll *a* in the presence of chlorophyll *b* and pheopigments. *Limnol. Oceanogr.* **39**: 1985–1992. doi:[10.4319/lo.1994.39.8.1985](https://doi.org/10.4319/lo.1994.39.8.1985)
- Wickham, H. 2014. Tidy data. *J. Stat. Soft.* **59**: 1–23. doi:[10.18637/jss.v059.i10](https://doi.org/10.18637/jss.v059.i10)
- Wickham, H., and others. 2019. Welcome to the tidyverse. *J. Open Source Softw.* **4**: 1686. doi:[10.21105/joss.01686](https://doi.org/10.21105/joss.01686)
- Wilke, C. O. 2019. cowplot: Streamlined Plot Theme and Plot Annotations for “ggplot2”.
- Wilkinson, M. D., and others. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**: 160018. doi:[10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)
- Yang, Y., W. Song, H. Lin, W. Wang, L. Du, and W. Xing. 2018. Antibiotics and antibiotic resistance genes in global lakes: A review and meta-analysis. *Environ. Int.* **116**: 60–73. doi:[10.1016/j.envint.2018.04.011](https://doi.org/10.1016/j.envint.2018.04.011)
- Yoshida, T., and others. 2003. Seasonal dynamics of primary production in the pelagic zone of southern Lake Baikal. *Limnology* **4**: 53–62. doi:[10.1007/s10201-002-0089-3](https://doi.org/10.1007/s10201-002-0089-3)

#### Acknowledgments

We would like to thank the faculty, students, staff, and mariners of the Irkutsk State University's Biological Research Institute Biostation for their expert field, taxonomic, and laboratory support; Marianne Moore and Bart De Stasio for helpful advice; the researchers and students of the Siberian Branch of the Russian Academy of Sciences Limnological Institute for expert taxonomic and logistical assistance; Oleg A. Timoshkin, Tatiana Ya. Sitnikova, Irina V. Mekhanikova, Nina A. Bondorenko, Ekaterina Volkova, and Vadim V. Takhteev for offering insights and taxonomic training throughout the development of this project. We would also like to thank Dag O. Hessen and an anonymous reviewer for helping us improve the clarity of our data article as well as the reproducibility and transparency of our data products. Funding was provided by the National Science Foundation (NSF-DEB-1136637) to SEH, a Fulbright Fellowship to MFM, a NSF Graduate Research Fellowship to MFM (NSF-DGE-1347973), and the Russian Ministry of Science and Education (N FZZE-2020-0026; N FZZE-2020-0023). This work serves as one chapter of MFM's doctoral dissertation in Environmental and Natural Resource Sciences at Washington State University.

Submitted 07 January 2021

Revised 01 July 2021

Accepted 28 September 2021