Scientific Significance Statement

We present a unified dataset of co-located benthic littoral nutrient concentrations, sewage indicators, algal and macroinvertebrate community abundance, stable isotopes, and fatty acids from Lake Baikal (Siberia). While researchers have studied Baikalâ€™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 co-located variables, this dataset is the most extensive, publicly available description of Baikalâ€™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.

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

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

[1.] School of the Environment, Washington State University, Pullman, WA, USA
[2.] Large Lakes Observatory, University of Minnesota - Duluth, Duluth, MN, USA
[3.] Center for Environmental Research, Education, and Outreach, Washington State University, Pullman, WA, USA
[4.] Oregon Institute of Marine Biology, University of Oregon, Charleston, OR, USA
[5.] School of Natural Resources, University of Nebraska-Lincoln, Lincoln, NE, USA
[6.] Biological Research Institute, Irkutsk State University, Irkutsk, Irkutsk Oblast, Russia

*corresponding author: michael.f.meyer@wsu.edu

Author Contribution Statement

Conceptualized the project: MFM, SEH, TO
Collected samples in the field: MFM, TO, KHW, SEH
Processed samples: MFM, KS, JBS, DDS, TO, AWEG, SEH
Wrote and Reviewed R scripts: MFM, MRB, KHW
Data management: MFM, MRB
Wrote and edited the manuscript: All authors
Approved the final manuscript: All authors

Grant sponsor information:
Funding was provided by the National Science Foundation (NSF-DEB-1136637) to S.E.H., a Fulbright Fellowship to M.F.M., a NSF Graduate Research Fellowship to M.F.M. (NSF-DGE-1347973), and the Russian Ministry of Science and Education (N FZZE-2020-0026; N FZZE-2020-0023).

Key Words: pharmaceuticals, microplastics, fatty acids, stables isotopes, amphipod, mollusk, diatom, spirogyra

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

Abstract (150 of 150 words)

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 19-23 August 2015 for sewage indicators, including pharmaceuticals, personal care products, and microplastics, with co-located 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.

Background and Motivation

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 $\delta^{15}N$ 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., µg/L) 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 sites at Lake Baikal from 19 through 23 August 2015 (Figure 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; Izmest'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. Nearshore 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 Izmest'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 pharmaceuticals and personal care products 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 (Meyer et al. 2021). 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., 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 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, census-based human population data for each sampled location. Although the majority of sites do not have adjacent shoreline human developments, we calculated inverse distance weighted (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
Inverse distance weighted 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 mg/L for nitrate, 0.005 mg/L for ammonium, and 0.04 mg/L 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 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 mg/L.

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 chlorophyll a extraction.

raw_fluo
Raw, uncorrected fluorometric reading for chlorophyll analysis.

adjusted_raw
Corrected fluorometric reading for chlorophyll analysis.

chl_conc
Chlorophyll a concentration in milligrams per liter.

ppcp.csv

This file contains Pharmaceutical and Personal Care Product (PPCP) concentrations in the water column at each littoral and pelagic sampling location. Detection limits were estimated to be 0.001 µg/L 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 ~300 μm (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.

410 *beads*
411 Number of microplastic beads observed.
412
413 *comments*
414 Observer comments while enumerating microplastics.
415
416 *volume_filtered_ml*
417 Volume in milliliters for a given replicate filtered.
418
419 periphyton.csv
420
421 This file contains periphyton abundance data, collected from rocks at each of the sampled littoral
422 locations. For poorly preserved samples, counts are listed as NA for each taxonomic grouping, and
423 a note in the "comments" column is provided.
424
425 *site*
426 Unique alphanumeric identifier for a sampling location.
427
428 *replicate*
429 Replicate number for a given sampling site.
430
431 *subsamples_counted*
432 Number of 10 microliter subsamples counted for a given replicate.
433
434 *diatom*
435 Number of diatom cells counted for a given replicate.
436
437 *spirogyra*
438 Number of *Spirogyra* spp. cells counted for a given replicate.
439
440 *spirogyra_filament*
441 Number of *Spirogyra* spp. filaments counted for a given replicate.
442
443 *ulothrix*
444 Number of *Ulothrix* spp. cells counted for a given replicate.
445
446 *ulothrix_filament*
447 Number of *Ulothrix* spp. filaments counted for a given replicate.
448
449 *tetrasporales*
450 Number of Tetrasporales cells counted for a given replicate.
451
452 *pediastrum*
453 Number of *Pediastrum* spp. cells counted for a given replicate.
454
455 *desmidales*

Number of *Desmidales* 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_verucossus
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.

548
549 *Pallasea_cancelloides*
550 Endemic amphipod species.
551
552 *Pallasea_cancellus*
553 Endemic amphipod species.
554
555 *Pallasea_viridis*
556 Endemic amphipod species.
557
558 *Planorbidae*
559 Mollusk family.
560
561 *Poekilogammarus_crassimus*
562 Endemic amphipod species.
563
564 *Poekilogammarus_ephippiatus*
565 Endemic amphipod species.
566
567 *Poekilogammarus_juveniles*
568 Endemic amphipod genus. Identification kept at genus level so as to prevent misclassification.
569
570 *Poekilogammarus_megonychus_perpolitus*
571 Endemic amphipod species.
572
573 *Poekilogammarus_pictus*
574 Endemic amphipod species.
575
576 *Valvatidae*
577 Mollusk family.
578
579 stable_isotopes.csv
580
581 This file contains carbon ($\delta^{13}C$) and nitrogen ($\delta^{15}N$) values for various benthic macroinvertebrate
582 genera and periphyton collected from the 14 littoral sampling locations.
583
584 *site*
585 Unique alphanumeric identifier for a sampling location.
586
587 *Genus*
588 Genus of the analyzed organism.
589
590 *Species*
591 Species of the analyzed organism. When an organism was identified solely to genus, the Species
592 value is NA.
593

C13
Carbon ($\delta^{13}C$) stable isotope values in parts per thousand.

N15
Nitrogen ($\delta^{15}N$) stable isotope values in parts per thousand.

comments
Quality flag column where $\delta^{13}C$ 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 14:1ω5 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 15:1ω7 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 16:1ω9 fatty acid as micrograms of fatty acid per milligram of tissue.

c16_1w8
Concentration of 16:1ω8 fatty acid as micrograms of fatty acid per milligram of tissue.

c16_1w7
Concentration of 16:1ω7 fatty acid as micrograms of fatty acid per milligram of tissue.

c16_1w6
Concentration of 16:1ω6 fatty acid as micrograms of fatty acid per milligram of tissue.

c16_1w5
Concentration of 16:1ω5 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 17:1ω7 fatty acid as micrograms of fatty acid per milligram of tissue.

c16_2w7
Concentration of 16:2ω7 fatty acid as micrograms of fatty acid per milligram of tissue.

c16_2w6
Concentration of 16:2ω6 fatty acid as micrograms of fatty acid per milligram of tissue.

c16_2w4
Concentration of 16:2ω4 fatty acid as micrograms of fatty acid per milligram of tissue.

c16_3w6

Concentration of 16:3ω6 fatty acid as micrograms of fatty acid per milligram of tissue.

c16_3w4
Concentration of 16:3ω4 fatty acid as micrograms of fatty acid per milligram of tissue.

c16_3w3
Concentration of 16:3ω3 fatty acid as micrograms of fatty acid per milligram of tissue.

c16_4w3
Concentration of 16:4ω3 fatty acid as micrograms of fatty acid per milligram of tissue.

c16_4w1
Concentration of 16:4ω1 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 18:1ω9 fatty acid as micrograms of fatty acid per milligram of tissue.

c18_1w7
Concentration of 18:1ω7 fatty acid as micrograms of fatty acid per milligram of tissue.

c18_2w6t
Concentration of 18:2ω6t fatty acid as micrograms of fatty acid per milligram of tissue.

c18_2w6
Concentration of 18:2ω6 fatty acid as micrograms of fatty acid per milligram of tissue.

c18_3w6
Concentration of 18:3ω6 fatty acid as micrograms of fatty acid per milligram of tissue.

c18_3w3
Concentration of 18:3ω3 fatty acid as micrograms of fatty acid per milligram of tissue.

c18_4w4
Concentration of 18:4ω4 fatty acid as micrograms of fatty acid per milligram of tissue.

c18_4w3
Concentration of 18:4ω3 fatty acid as micrograms of fatty acid per milligram of tissue.

c18_5w3
Concentration of 18:5ω3 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 20:1ω9 fatty acid as micrograms of fatty acid per milligram of tissue.

c20_1w7
Concentration of 20:1ω7 fatty acid as micrograms of fatty acid per milligram of tissue.

c20_2w5_11
Concentration of 20:2ω5-11 fatty acid as micrograms of fatty acid per milligram of tissue.

c20_2w5_13
Concentration of 20:2ω5-13 fatty acid as micrograms of fatty acid per milligram of tissue.

c20_2w6
Concentration of 20:2ω6 fatty acid as micrograms of fatty acid per milligram of tissue.

c20_3w6
Concentration of 20:3ω6 fatty acid as micrograms of fatty acid per milligram of tissue.

c20_4w6
Concentration of 20:4ω6 fatty acid as micrograms of fatty acid per milligram of tissue.

c20_3w3
Concentration of 20:3ω3 fatty acid as micrograms of fatty acid per milligram of tissue.

c20_4w3
Concentration of 20:4ω3 fatty acid as micrograms of fatty acid per milligram of tissue.

c20_5w3
Concentration of 20:5ω3 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 22:1ω9 fatty acid as micrograms of fatty acid per milligram of tissue.

c22_1w7
Concentration of 22:1ω7 fatty acid as micrograms of fatty acid per milligram of tissue.

c22_2w6
Concentration of 22:2ω6 fatty acid as micrograms of fatty acid per milligram of tissue.

c22_4w6
Concentration of 22:4ω6 fatty acid as micrograms of fatty acid per milligram of tissue.

c22_5w6
Concentration of 22:5ω6 fatty acid as micrograms of fatty acid per milligram of tissue.

c22_3w3
Concentration of 22:3ω3 fatty acid as micrograms of fatty acid per milligram of tissue.

c22_4w3
Concentration of 22:4ω3 fatty acid as micrograms of fatty acid per milligram of tissue.

c22_5w3
Concentration of 22:5ω3 fatty acid as micrograms of fatty acid per milligram of tissue.

c22_6w3
Concentration of 22:6ω3 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.

Data Availability

Data are available at the replicate level at the Environmental Data Initiative (doi.org/10.6073/pasta/9554b7f19ddd4a614e854f18be978dca).

Methods

Site Information

The vast majority of Lake Baikal's 2,100-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 (Figure 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; Figure 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 to 1300 m (Figure 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 to 20.75 m from shore (Table 1), which allowed us to collect samples without the need for SCUBA but precluded us from 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.

Inverse distance weighted (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 inverse distance weighted (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)

$$(1)\quad Ij = \frac{\frac{P_{LI}}{A_{LI}} * L_{LI}}{D_{j,LI}} + \frac{\frac{P_{BK}}{A_{BK}} * L_{BK}}{D_{j,BK}} + \frac{\frac{P_{BGO}}{A_{BGO}} * L_{BGO}}{D_{j,BGO}}$$

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 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°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 g/L 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°C without addition of preservative agents. When nitrate analyses were performed between 24-48 h after thawing, samples were kept at 3-5°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°C and ~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 three hours of thawing. When analyses could not be performed within three hours, samples were kept at 3-5°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 mg/L for nitrate, 0.005 mg/L for ammonium, and 0.04 mg/L for phosphorus. Concentrations are reported in mg/L 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 Open Science Framework portal within the directory "Nearshore_sampling/methods".

Chlorophyll a

Water samples were collected in 1.5 L 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 chlorophyll 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 µm 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, Sunnyvale, CA) 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 (2)

$$Chlorophyll\ concentration = (extract\ reading - blank\ reading) * \frac{mL\ of\ extract}{mL\ of\ filtered\ sample}.$$

Detection limits are estimated to be approximately 0.02 mg/L. Concentrations are reported as mg/L.

Pharmaceuticals and Personal Care Products (PPCPs)

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 µm pore size, WhatmanGrad 934-AH) in tandem with a solid phase extraction (SPE) cartridge (200 mg HLB, Waters Corporation, Milford, MA) 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 Whirlpacks at -20°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 ($^{13}C_3$-caffeine, methamphetamine-d8, MDMAd8, morphine-d3, and $^{13}C_6$-sulfamethazine). Detection limits are estimated to be 0.001 µg/L based on a 500 mL sample volume. Concentrations are reported in µg/L.

Microplastics

At each location, samples were collected at a depth of approximately 0.75 m in triplicate using 1.5 L 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 ~100x 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 ~300 µm (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 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 µL aliquots from each preserved sample, containing approximately 10-15 mL of preserved periphyton. For all 10 µL 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°C at the field station. The samples were later transferred to the lab in the U.S. 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-μm 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°C aboard the research vessel. The 40% ethanol preservative was replaced with ~80% ethanol upon return to the lab within 24 to 48 hours, and samples were stored at ~4°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°C at the field station. The samples were later transferred to the lab in the U.S. 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 *E. veruccosus* and *E. 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 $\delta^{15}N$ and $\delta^{13}C$ 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., Okehampton, UK) 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°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 minutes. Next, samples were vortexed for 1 minute, and centrifuged for 5 minutes (3,000 rpm) at 4°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°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 µL 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°C water bath for 16 hours. 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 minutes (1,500 rpm) at 4°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 ng/mL.

Following methylation, remaining extracts were assessed for total lipid masses. Remaining sample extracts (~0.5 mL) were allow to evaporate to dryness under a fume hood overnight. Dried samples were then left in a weigh room to acclimatize for 30-60 mins 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 ($^{13}C_3$-caffeine, methamphetamine-d8, MDMAd8, morphine-d3, and $^{13}C_6$-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., Okehampton, UK) and in-house standards (acetanilide and caffeine). Replicate analyses of external standards showed a mean standard deviation of 0.06 ‰ and 0.09 ‰, for $\delta^{13}C$ and $\delta^{15}N$, 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 (Meyer et al. 2021). Raw data files are available on the project's Open Science Framework 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. Brousil 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 Open Science Framework (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 2,000 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 are 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 co-located 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. ggsn: 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. Journal of Hazardous Materials **122**: 195–204. doi:10.1016/j.jhazmat.2005.03.012

Bondarenko, N. A., I. V. Tomberg, A. A. Shirokaya, and others. 2021. *Dolichospermum lemmermannii* (Nostocales) bloom in world's deepest Lake Baikal (East Siberia): abundance, toxicity and factors influencing growth. Limnology and Freshwater Biology **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. Limnology and Oceanography Letters **5**: 46–53. doi:10.1002/lol2.10127

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

1236 Camilleri, A. C., and T. Ozersky. 2019. Large variation in periphyton δ13C and δ15N values in the
1237 upper Great Lakes: Correlates and implications. Journal of Great Lakes Research **45**: 986–
1238 990. doi:10.1016/j.jglr.2019.06.003
1239 Costanzo, S. D., M. J. O'Donohue, W. C. Dennison, N. R. Loneragan, and M. Thomas. 2001. A
1240 new approach for detecting and mapping sewage impacts. Marine Pollution Bulletin **42**:
1241 149–156. doi:10.1016/S0025-326X(00)00125-9
1242 D'Alessio, M., S. Onanong, D. D. Snow, and C. Ray. 2018. Occurrence and removal of
1243 pharmaceutical compounds and steroids at four wastewater treatment plants in Hawai'i and
1244 their environmental fate. Science of The Total Environment **631–632**: 1360–1370.
1245 doi:10.1016/j.scitotenv.2018.03.100
1246 Dalsgaard, J., M. St. John, G. Kattner, D. Müller-Navarra, and W. Hagen. 2003. Fatty acid trophic
1247 markers in the pelagic marine environment, p. 225–340. Advances in Marine Biology.
1248 Elsevier.
1249 Edmondson, W. T. 1970. Phosphorus, nitrogen, and algae in Lake Washington after diversion of
1250 sewage. Science **169**: 690–691.
1251 Fellows, I., and using the Jm. library by J. P. Stotz. 2019. OpenStreetMap: Access to Open Street
1252 Map Raster Images.
1253 Focazio, M. J., D. W. Kolpin, K. K. Barnes, E. T. Furlong, M. T. Meyer, S. D. Zaugg, L. B. Barber,
1254 and M. E. Thurman. 2008. A national reconnaissance for pharmaceuticals and other organic
1255 wastewater contaminants in the United States - II) Untreated drinking water sources.
1256 Science of the Total Environment **402**: 201–216. doi:10.1016/j.scitotenv.2008.02.021

1257 Gartner, A., P. Lavery, and A. J. Smit. 2002. Use of δN-15 signatures of different functional forms
1258 of macroalgae and filter-feeders to reveal temporal and spatial patterns in sewage dispersal.
1259 Mar. Ecol.-Prog. Ser. **235**: 63–73. doi:10.3354/meps235063
1260 GOST:18309-2014. 2016. Methods for determination of phosphorus-containing matters (with
1261 corrections) (Методы определения фосфорсодержащих веществ).
1262 Green, D. S. 2016. Effects of microplastics on European flat oysters, *Ostrea edulis* and their
1263 associated benthic communities. Environmental Pollution **216**: 95–103.
1264 doi:10.1016/j.envpol.2016.05.043
1265 Hall, R. I., P. R. Leavitt, R. Quinlan, A. S. Dixit, and J. P. Smol. 1999. Effects of agriculture,
1266 urbanization, and climate on water quality in the northern Great Plains. Limnology and
1267 Oceanography **44**: 739–756. doi:10.4319/lo.1999.44.3_part_2.0739
1268 Hampton, S. E., S. C. Fradkin, P. R. Leavitt, and E. E. Rosenberger. 2011. Disproportionate
1269 importance of nearshore habitat for the food web of a deep oligotrophic lake. Marine and
1270 Freshwater Research **62**: 350. doi:10.1071/MF10229
1271 Hampton, S. E., L. R. Izmest'Eva, M. V. Moore, S. L. Katz, B. Dennis, and E. A. Silow. 2008.
1272 Sixty years of environmental change in the world's largest freshwater lake - Lake Baikal,
1273 Siberia. Global Change Biology **14**: 1947–1958. doi:10.1111/j.1365-2486.2008.01616.x
1274 Hampton, S. E., S. McGowan, T. Ozersky, and others. 2018. Recent ecological change in ancient
1275 lakes. Limnology and Oceanography **63**: 2277–2304. doi:10.1002/lno.10938
1276 Hanvey, J. S., P. J. Lewis, J. L. Lavers, N. D. Crosbie, K. Pozo, and B. O. Clarke. 2017. A review
1277 of analytical techniques for quantifying microplastics in sediments. Anal. Methods **9**: 1369–
1278 1383. doi:10.1039/C6AY02707E

1279 Horton, A. A., A. Walton, D. J. Spurgeon, E. Lahive, and C. Svendsen. 2017. Microplastics in
1280 freshwater and terrestrial environments: Evaluating the current understanding to identify the
1281 knowledge gaps and future research priorities. Science of The Total Environment **586**: 127–
1282 141. doi:10.1016/j.scitotenv.2017.01.190
1283 Interfax-Tourism. 2018. Байкал с января по август 2018 года посетили 1,2 миллиона туристов
1284 (1.2 million tourists vistied Baikal from January through August 2018). Interfax-Tourism,
1285 October 25
1286 International Standards Organization (ISO). 1984. ISO 6777:1984(en) Water quality —
1287 Determination of nitrite — Molecular absorption spectrometric method. ISO 6777. ISO
1288 6777 ISO.
1289 International Standards Organization (ISO). 2004. ISO 6878:2004(en) Water quality —
1290 Determination of phosphorus — Ammonium molybdate spectrometric method. ISO 6878.
1291 ISO 6878 ISO.
1292 Izhboldina, L. A. 2007. Guide and Key to Benthic and Periphyton Algae of Lake Baikal (meio- and
1293 macrophytes) with Brief Notes on Their Ecology, Nauka-Centre.
1294 Izmest'eva, L. R., M. V. Moore, S. E. Hampton, and others. 2016. Lake-wide physical and
1295 biological trends associated with warming in Lake Baikal. Journal of Great Lakes Research
1296 **42**: 6–17. doi:10.1016/j.jglr.2015.11.006
1297 Jeppesen, E., M. Søndergaard, J. P. Jensen, and others. 2005. Lake responses to reduced nutrient
1298 loading – an analysis of contemporary long-term data from 35 case studies. Freshwater
1299 Biology **50**: 1747–1771. doi:10.1111/j.1365-2427.2005.01415.x

1300 Karnaukhov, D., S. Biritskaya, E. Dolinskaya, M. Teplykh, N. Silenko, Y. Ermolaeva, and E.
1301 Silow. 2020. Pollution by macro- and microplastic of large lacustrine ecosystems in Eastern
1302 Asia. Pollution Research **2**: 353–355.
1303 Kassambara, A. 2019. ggpubr: "ggplot2" Based Publication Ready Plots.
1304 Katz, S. L., L. R. Izmest'eva, S. E. Hampton, T. Ozersky, K. Shchapov, M. V. Moore, S. V.
1305 Shimaraeva, and E. A. Silow. 2015. The "Melosira years" of Lake Baikal: Winter
1306 environmental conditions at ice onset predict under-ice algal blooms in spring. Limnology
1307 and Oceanography **60**: 1950–1964. doi:10.1002/lno.10143
1308 Kolpin, D. W., E. T. Furlong, M. T. Meyer, E. M. Thurman, S. D. Zaugg, L. B. Barber, and H. T.
1309 Buxton. 2002. Pharmaceuticals, hormones, and other organic wastewater contaminants in
1310 U.S. Streams, 1999−2000: A national reconnaissance. Environmental Science &
1311 Technology **36**: 1202–1211. doi:10.1021/es011055j
1312 Kozhova, O. M., and L. R. Izmest'eva. 1998. Lake Baikal: Evolution and Biodiversity, Backhuys
1313 Publishers.
1314 Kravtsova, L. S., L. A. Izhboldina, I. V. Khanaev, and others. 2014. Nearshore benthic blooms of
1315 filamentous green algae in Lake Baikal. Journal of Great Lakes Research **40**: 441–448.
1316 doi:10.1016/j.jglr.2014.02.019
1317 Lapointe, B. E., L. W. Herren, D. D. Debortoli, and M. A. Vogel. 2015. Evidence of sewage-driven
1318 eutrophication and harmful algal blooms in Florida's Indian River Lagoon. Harmful Algae
1319 **43**: 82–102. doi:10.1016/j.hal.2015.01.004
1320 Lee, S. S., A. M. Paspalof, D. D. Snow, E. K. Richmond, E. J. Rosi-Marshall, and J. J. Kelly. 2016.
1321 Occurrence and potential biological effects of amphetamine on stream communities.
1322 Environmental Science & Technology **50**: 9727–9735. doi:10.1021/acs.est.6b03717

1323 Lin, D., J. Crabtree, I. Dillo, and others. 2020. The TRUST Principles for digital repositories.
1324 Scientific Data **7**: 144. doi:10.1038/s41597-020-0486-7
1325 Meyer, M. F., S. G. Labou, A. N. Cramer, M. R. Brousil, and B. T. Luff. 2020. The global lake
1326 area, climate, and population dataset. Scientific Data **7**: 174. doi:10.1038/s41597-020-0517-
1327 4
1328 Meyer, M. F., T. Ozersky, K. H. Woo, and others. 2021. A unified dataset of co-located sewage
1329 pollution, periphyton, and benthic macroinvertebrate community and food web structure
1330 from Lake Baikal (Siberia).
1331 doi:https://doi.org/10.6073/pasta/9554b7f19ddd4a614e854f18be978dca
1332 Meyer, M. F., T. Ozersky, K. H. Woo, and others. *Under Review.* Effects of spatially heterogeneous
1333 lakeside development on nearshore biotic communities in a large, deep, oligotrophic lake
1334 (Lake Baikal, Siberia).
1335 Meyer, M. F., S. M. Powers, and S. E. Hampton. 2019. An evidence synthesis of pharmaceuticals
1336 and personal care products (PPCPs) in the environment: Imbalances among compounds,
1337 sewage treatment techniques, and ecosystem types. Environ. Sci. Technol. **53**: 12961–
1338 12973. doi:10.1021/acs.est.9b02966
1339 Meyer, M., T. Ozersky, K. Woo, A. W. E. Galloway, M. R. Brousil, and S. Hampton. 2015. Baikal
1340 Food Webs. doi:10.17605/OSF.IO/9TA8Z
1341 Moore, J. W., D. E. Schindler, M. D. Scheuerell, D. Smith, and J. Frodge. 2003. Lake
1342 eutrophication at the urban fringe, Seattle region, USA. AMBIO: A Journal of the Human
1343 Environment **32**: 13–18.

1344 Moore, M. V., S. E. Hampton, L. R. Izmest'eva, E. A. Silow, E. V. Peshkova, and B. K. Pavlov.
1345 2009. Climate change and the world's "Sacred Sea"- Lake Baikal, Siberia. Bioscience **59**:
1346 405–417. doi:10.1525/bio.2009.59.5.8
1347 O'Donnell, D. R., P. Wilburn, E. A. Silow, L. Y. Yampolsky, and E. Litchman. 2017. Nitrogen and
1348 phosphorus colimitation of phytoplankton in Lake Baikal: Insights from a spatial survey and
1349 nutrient enrichment experiments. Limnology and Oceanography **62**: 1383–1392.
1350 doi:10.1002/lno.10505
1351 Pebesma, E. 2018. Simple Features for R: Standardized support for spatial vector data. The R
1352 Journal **10**: 439–446. doi:10.32614/RJ-2018-009
1353 Powers, S. M., T. W. Bruulsema, T. P. Burt, and others. 2016. Long-term accumulation and
1354 transport of anthropogenic phosphorus in three river basins. Nature Geoscience **9**: 353–356.
1355 doi:10.1038/ngeo2693
1356 R Core Team. 2019. R: A Language and Environment for Statistical Computing,.
1357 RD:52.24.380-2017. 2018. Nitrate concentration in waters: Photometric methods with Giress
1358 reagent following stabilization in a cadmium reducer (Массовая концентрация нитратного
1359 азота в водах: Методика измерений фотометрическим методом с реактивом Грисса
1360 после восстановления в камиевом редукторе).
1361 RD:52.24.383-2018. 2018. Working Document: Concentration of aqueous ammonium: Method for
1362 measuring with a photometer using indophenol blue (Руководящий Документ: Массовая
1363 концентрация аммонийного азота в водах: Методика измерений фотометрическим
1364 методом в виде индофенолового сингео). RD:52.24.383-2018. RD:52.24.383-2018.

1365 Richmond, E. K., M. R. Grace, J. J. Kelly, A. J. Reisinger, E. J. Rosi, and D. M. Walters. 2017.
1366 Pharmaceuticals and personal care products (PPCPs) are ecological disrupting compounds
1367 (EcoDC). Elem Sci Anth **5**: 66. doi:10.1525/elementa.252
1368 Richmond, E. K., E. J. Rosi, D. M. Walters, J. Fick, S. K. Hamilton, T. Brodin, A. Sundelin, and M.
1369 R. Grace. 2018. A diverse suite of pharmaceuticals contaminates stream and riparian food
1370 webs. Nature Communications **9**: 4491. doi:10.1038/s41467-018-06822-w
1371 Romera-Castillo, C., M. Pinto, T. M. Langer, X. A. Álvarez-Salgado, and G. J. Herndl. 2018.
1372 Dissolved organic carbon leaching from plastics stimulates microbial activity in the ocean.
1373 Nat Commun **9**: 1–7. doi:10.1038/s41467-018-03798-5
1374 Rosenberger, E. E., S. E. Hampton, S. C. Fradkin, and B. P. Kennedy. 2008. Effects of shoreline
1375 development on the nearshore environment in large deep oligotrophic lakes. Freshwater
1376 Biology **53**: 1673–1691. doi:10.1111/j.1365-2427.2008.01990.x
1377 Rosi-Marshall, E. J., D. W. Kincaid, H. A. Bechtold, T. V. Royer, M. Rojas, and J. J. Kelly. 2013.
1378 Pharmaceuticals suppress algal growth and microbial respiration and alter bacterial
1379 communities in stream biofilms. Ecological Applications **23**: 583–593. doi:10.1890/12-
1380 0491.1
1381 Rosi-Marshall, E. J., and T. V. Royer. 2012. Pharmaceutical compounds and ecosystem function:
1382 an emerging research challenge for aquatic ecologists. Ecosystems **15**: 867–880.
1383 doi:10.1007/s10021-012-9553-z
1384 Sargent, J. R., and S. Falk-Petersen. 1988. The lipid biochemistry of calanoid copepods.
1385 Hydrobiologia **167–168**: 101–114. doi:10.1007/BF00026297

1386 Schram, J. B., J. N. Kobelt, M. N. Dethier, and A. W. E. Galloway. 2018. Trophic transfer of
1387 macroalgal fatty acids in two urchin species: Digestion, egestion, and tissue building. Front.
1388 Ecol. Evol. **6**. doi:10.3389/fevo.2018.00083
1389 Shaw, L., C. Phung, and M. Grace. 2015. Pharmaceuticals and personal care products alter growth
1390 and function in lentic biofilms. Environmental Chemistry **12**: 301. doi:10.1071/EN14141
1391 Slowikowski, K. 2019. ggrepel: Automatically Position Non-Overlapping Text Labels with
1392 "ggplot2,."
1393 Swann, G. E. A., V. N. Panizzo, S. Piccolroaz, and others. 2020. Changing nutrient cycling in Lake
1394 Baikal, the world's oldest lake. PNAS **117**: 27211–27217. doi:10.1073/pnas.2013181117
1395 Taipale, S., U. Strandberg, E. Peltomaa, A. W. E. Galloway, A. Ojala, and M. T. Brett. 2013. Fatty
1396 acid composition as biomarkers of freshwater microalgae: analysis of 37 strains of
1397 microalgae in 22 genera and in seven classes. Aquatic Microbial Ecology **71**: 165–178.
1398 doi:10.3354/ame01671
1399 Timoshkin, O. A., M. V. Moore, N. N. Kulikova, and others. 2018. Groundwater contamination by
1400 sewage causes benthic algal outbreaks in the littoral zone of Lake Baikal (East Siberia).
1401 Journal of Great Lakes Research. doi:10.1016/j.jglr.2018.01.008
1402 Timoshkin, O. A., D. P. Samsonov, M. Yamamuro, and others. 2016. Rapid ecological change in
1403 the coastal zone of Lake Baikal (East Siberia): Is the site of the world's greatest freshwater
1404 biodiversity in danger? Journal of Great Lakes Research **42**: 487–497.
1405 doi:10.1016/j.jglr.2016.02.011
1406 Tong, Y., M. Wang, J. Peñuelas, and others. 2020. Improvement in municipal wastewater treatment
1407 alters lake nitrogen to phosphorus ratios in populated regions. Proc Natl Acad Sci USA **117**:
1408 11566–11572. doi:10.1073/pnas.1920759117

1409 Turetsky, M. R., R. K. Wieder, C. J. Williams, and D. H. Vitt. 2000. Organic matter accumulation,
1410 peat chemistry, and permafrost melting in peatlands of boreal Alberta. Écoscience **7**: 115–
1411 122. doi:10.1080/11956860.2000.11682608
1412 Vendel, A. L., F. Bessa, V. E. N. Alves, A. L. A. Amorim, J. Patrício, and A. R. T. Palma. 2017.
1413 Widespread microplastic ingestion by fish assemblages in tropical estuaries subjected to
1414 anthropogenic pressures. Marine Pollution Bulletin **117**: 448–455.
1415 doi:10.1016/j.marpolbul.2017.01.081
1416 Volkova, E. A., N. A. Bondarenko, and O. A. Timoshkin. 2018. Morphotaxonomy, distribution and
1417 abundance of *Spirogyra* (Zygnematophyceae, Charophyta) in Lake Baikal, East Siberia.
1418 Phycologia **57**: 298–308. doi:10.2216/17-69.1
1419 Wang, W., and J. Wang. 2018. Investigation of microplastics in aquatic environments: An overview
1420 of the methods used, from field sampling to laboratory analysis. TrAC Trends in Analytical
1421 Chemistry **108**: 195–202. doi:10.1016/j.trac.2018.08.026
1422 Welschmeyer, N. A. 1994. Fluorometric analysis of chlorophyll a in the presence of chlorophyll b
1423 and pheopigments. Limnol. Oceanogr. **39**: 1985–1992. doi:10.4319/lo.1994.39.8.1985
1424 Wickham, H. 2014. Tidy Data. Journal of Statistical Software **59**: 1–23. doi:10.18637/jss.v059.i10
1425 Wickham, H., M. Averick, J. Bryan, and others. 2019. Welcome to the tidyverse. Journal of Open
1426 Source Software **4**: 1686. doi:10.21105/joss.01686
1427 Wilke, C. O. 2019. cowplot: Streamlined Plot Theme and Plot Annotations for "ggplot2."
1428 Wilkinson, M. D., M. Dumontier, Ij. J. Aalbersberg, and others. 2016. The FAIR Guiding
1429 Principles for scientific data management and stewardship. Sci Data **3**.
1430 doi: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. Environment International **116**: 60–73. doi:10.1016/j.envint.2018.04.011

Yoshida, T., T. Sekino, M. Genkai-Kato, 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

Acknowledgements

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 S.E.H., a Fulbright Fellowship to M.F.M., a NSF Graduate Research Fellowship to M.F.M. (NSF-DGE-1347973), and Russian Ministry of Science and Education (N FZZE-2020-0026; N FZZE-2020-0023). This work serves as one chapter of M.F.M.'s doctoral dissertation in Environmental and Natural Resource Sciences at Washington State University.

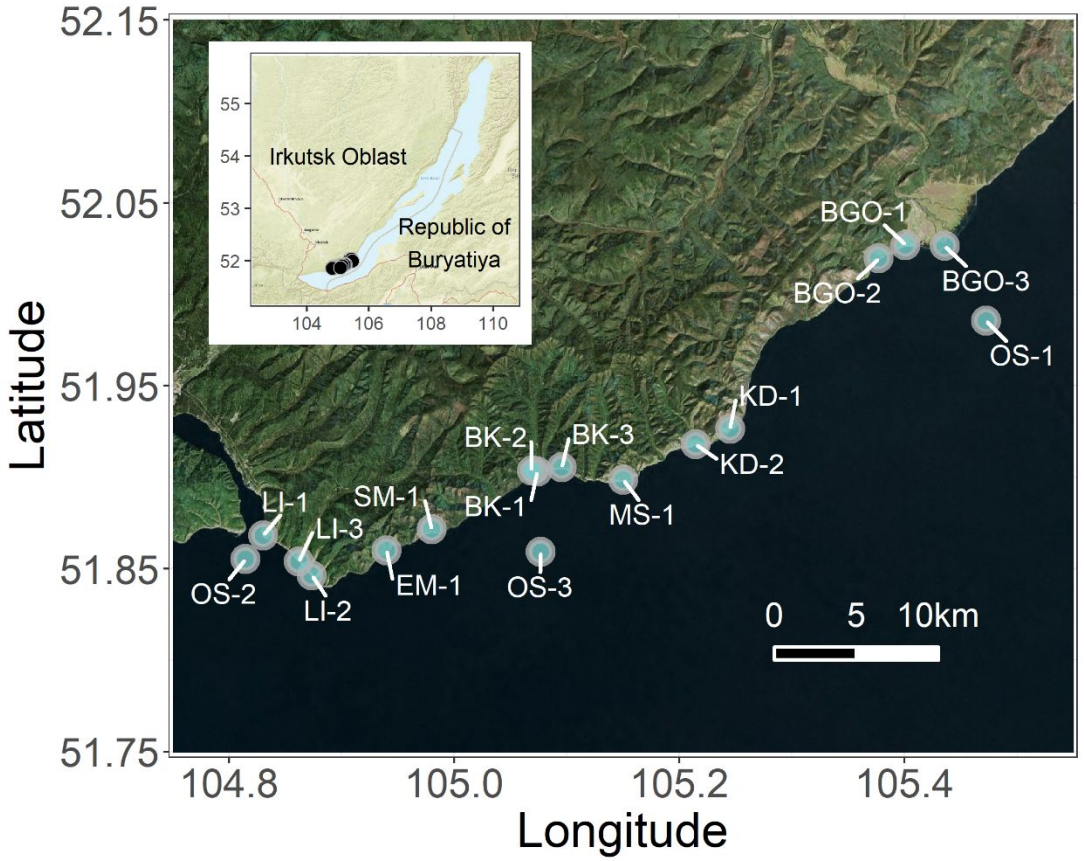

Figure 1: Map of all sampling locations with sites labeled with unique alphanumeric code. The entire transect included three developed sites (i.e., Listvyanka (LI), Bolshie Koty (BK), Bolshoe Goloustnoe (BGO)). Three offshore sites (OS) were also sampled to compare pelagic sewage signals to those in the littoral. Sites without adjacent lakeside development included Emelyanikha Bay (EM), Maloe Kadilnoe (KD), Mys Soboliny (MS), 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), OpenStreeetMap (Fellows and Stotz 2019), ggpubr (Kassambara 2019), cowplot (Wilke 2019), ggsn (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.

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

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.

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	*Brandtia latissima* subspp. (Dorogostaiskii 1930; Dybowsky 1874) *Brandtia parasitica parasitica* (Dybowsky 1874) *Cryptoropus inflatus* (Dybowsky 1874) *Cryptoropus pachytus* (Dybowsky 1874) *Cryptoropus rugosus* (Dybowsky 1874) *Eulimnogammarus capreolus* (Dybowsky 1874) *Eulimnogammarus cruentes* (Dorogostaiskii 1930) *Eulimnogammarus cyaneus* (Dybowsky 1874) *Eulimnogammarus grandimanus* (Bazikalova 1945) *Eulimnogammarus maacki* (Gerstfeldt 1858) *Eulimnogammarus marituji* (Bazikalova 1945) *Eulimnogammarus verucossus* (Gerstfeldt 1858) *Eulimnogammarus viridis viridis* (Dybowsky 1874) *Eulimnogammarus vittatus* (Dybowsky 1874) *Pallasea brandtia brandita* (Dybowsky 1874) *Pallasea brandtii tenera* (Sovinskii 1930) *Pallasea cancelloides* (Gerstfeldt 1858) *Pallasea cancellus* (Pallas 1776) *Pallasea viridis* (Garjajev 1901) *Poekilogammarus crassimus* (Sovinskii 1915) *Poekilogammarus ephippiatus* (Dybowsky 1874) *Poekilogammarus megonychus perpolitus* (Takhteev 2002) *Poekilogammarus pictus* (Dybowsky 1874)
	Molluska	Acroloxidae Baicaliidae Benedictidate Maackia Planorbidae Valvatidae
	Other Macroinvertebrates	Asellidae

		Caddisflies Hirudinea Planaria
	Benthic Algae	Diatom *Ulothrix* spp. *Spirogyra* spp. Tetrasporales
Stable Isotopes	Amphipoda	*Eulimnogammarus cyaneus* (Dybowsky 1874) *Eulimnogammarus verucossus* (Gerstfeldt 1858) *Eulimnogammarus vittatus* (Dybowsky 1874) *Pallasea cancellus* (Pallas 1776)
	Benthic Algae	Periphyton
Fatty Acids	Amphipoda	*Eulimnogammarus cyaneus* (Dybowsky 1874) *Eulimnogammarus verucossus* (Gerstfeldt 1858) *Eulimnogammarus vittatus* (Dybowsky 1874) *Hyalella* spp. *Pallasea cancellus* (Pallas 1776)
	Molluska	Processed in composite and not identified to family.
	Benthic Algae	Periphyton *Draparnaldia* spp.

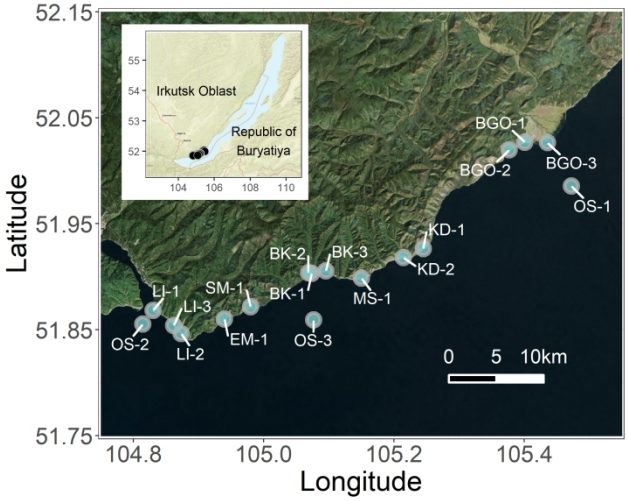

Figure 1: Map of all sampling locations with sites labeled with unique alphanumeric code. The entire transect included three developed sites (i.e., Listvyanka (LI), Bolshie Koty (BK), Bolshoe Goloustnoe (BGO)). Three offshore sites (OS) were also sampled to compare pelagic sewage signals to those in the littoral. Sites without adjacent lakeside development included Emelyanikha Bay (EM), Maloe Kadilnoe (KD), Mys Soboliny (MS), 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), OpenStreeetMap (Fellows and Stotz 2019), ggpubr (Kassambara 2019), cowplot (Wilke 2019), ggsn (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.

774x387mm (118 x 118 DPI)

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

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.

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	*Brandtia latissima* subspp. (Dorogostaiskii 1930; Dybowsky 1874) *Brandtia parasitica parasitica* (Dybowsky 1874) *Cryptoropus inflatus* (Dybowsky 1874) *Cryptoropus pachytus* (Dybowsky 1874) *Cryptoropus rugosus* (Dybowsky 1874) *Eulimnogammarus capreolus* (Dybowsky 1874) *Eulimnogammarus cruentes* (Dorogostaiskii 1930) *Eulimnogammarus cyaneus* (Dybowsky 1874) *Eulimnogammarus grandimanus* (Bazikalova 1945) *Eulimnogammarus maacki* (Gerstfeldt 1858) *Eulimnogammarus marituji* (Bazikalova 1945) *Eulimnogammarus verucossus* (Gerstfeldt 1858) *Eulimnogammarus viridis viridis* (Dybowsky 1874) *Eulimnogammarus vittatus* (Dybowsky 1874) *Pallasea brandtia brandita* (Dybowsky 1874) *Pallasea brandtii tenera* (Sovinskii 1930) *Pallasea cancelloides* (Gerstfeldt 1858) *Pallasea cancellus* (Pallas 1776) *Pallasea viridis* (Garjajev 1901)

		Poekilogammarus crassimus (Sovinskii 1915) *Poekilogammarus ephippiatus* (Dybowsky 1874) *Poekilogammarus megonychus perpolitus* (Takhteev 2002) *Poekilogammarus pictus* (Dybowsky 1874)
	Molluska	Acroloxidae Baicaliidae Benedictidate Maackia Planorbidae Valvatidae
	Other Macroinvertebrates	Asellidae Caddisflies Hirudinea Planaria
	Benthic Algae	Diatom *Ulothrix* spp. *Spirogyra* spp. Tetrasporales
Stable Isotopes	Amphipoda	*Eulimnogammarus cyaneus* (Dybowsky 1874) *Eulimnogammarus verucossus* (Gerstfeldt 1858) *Eulimnogammarus vittatus* (Dybowsky 1874) *Pallasea cancellus* (Pallas 1776)
	Benthic Algae	Periphyton
Fatty Acids	Amphipoda	*Eulimnogammarus cyaneus* (Dybowsky 1874)

		Eulimnogammarus verucossus (Gerstfeldt 1858) *Eulimnogammarus vittatus* (Dybowsky 1874) *Hyalella* spp. *Pallasea cancellus* (Pallas 1776)
	Molluska	Processed in composite and not identified to family.
	Benthic Algae	Periphyton *Draparnaldia* spp.