



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

Journal:	Limnology and Oceanography Letters
Manuscript ID	LOL2-21-0001.R1
Wiley - Manuscript type:	Data Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Meyer, Michael; Washington State University, School of the Environment Ozersky, Ted; University of Minnesota Duluth, Large Lakes Observatory Woo, Kara; Washington State University, Center for Environmental Research, Education, and Outreach Shchapov, Kirill; University of Minnesota Duluth, Large Lakes Observatory Galloway, Aaron; University of Oregon Oregon Institute of Marine Biology Schram, Julie; University of Oregon Oregon Institute of Marine Biology Snow, Daniel; University of Nebraska-Lincoln Timofeev, Maxim; Research Institute of Biology Irkutsk State University Karnaukhov, Dmitrii; Research Institute for Development Brousil, Matthew; Washington State University, Center for Environmental Research, Education, and Outreach Hampton, Stephanie; Washington State University, Center for Environmental Research, Education and Outreach
Search Terms:	pharmaceuticals, microplastics, fatty acids, stable isotopes, amphipod, spirogyra
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 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.

SCHOLARONE™ Manuscripts

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 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 co-located 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.



1	A unified dataset of co-located sewage pollution, periphyton, and benthic macroinvertebrate
2	community and food web structure from Lake Baikal (Siberia)
3	
4	Michael F. Meyer ¹ *
5	Ted Ozersky ²
6	Kara H. Woo ³
7	Kirill Shchapov ²
8	Aaron W. E. Galloway ⁴
9	Julie B. Schram ⁴
10	Daniel D. Snow ⁵
11	Maxim A. Timofeyev ⁶
12	Dmitry Yu. Karnaukhov ⁶
13	Matthew R. Brousil ³
14	Stephanie E. Hampton ³
15	
16	^{1.} School of the Environment, Washington State University, Pullman, WA, USA
17	^{2.} Large Lakes Observatory, University of Minnesota - Duluth, Duluth, MN, USA
18	^{3.} Center for Environmental Research, Education, and Outreach, Washington State University,
19	Pullman, WA, USA
20	^{4.} Oregon Institute of Marine Biology, University of Oregon, Charleston, OR, USA
21	^{5.} School of Natural Resources, University of Nebraska-Lincoln, Lincoln, NE, USA
22	^{6.} Biological Research Institute, Irkutsk State University, Irkutsk, Irkutsk Oblast, Russia
23	
24	*corresponding author: michael.f.meyer@wsu.edu
25	
26	Author Contribution Statement
27	Conceptualized the project: MFM, SEH, TO
28	Collected samples in the field: MFM, TO, KHW, SEH
29	Processed samples: MFM, KS, JBS, DDS, TO, AWEG, SEH
30	Wrote and Reviewed R scripts: MFM, MRB, KHW
31	Data management: MFM, MRB
32	Wrote and edited the manuscript: All authors
33	Approved the final manuscript: All authors
34	
35	Grant sponsor information:
36	Funding was provided by the National Science Foundation (NSF-DEB-1136637) to S.E.H., a Fulbright
37	Fellowship to M.F.M., a NSF Graduate Research Fellowship to M.F.M. (NSF-DGE-1347973), and the
38	Russian Ministry of Science and Education (N FZZE-2020-0026; N FZZE-2020-0023).
39	
40	Key Words: pharmaceuticals, microplastics, fatty acids, stables isotopes, amphipod, mollusk,
41	diatom, spirogyra
42	
43	URL of the Dataset and Metadata with permanent identifier:
44	• Environmental Data Initiative:
45	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:
- 50 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,
- 53 Microplastics, Periphyton community abundance, benthic macroinvertebrate abundance, Stable
- Isotopes, nitrate, ammonium, total phosphorus
- 55 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

79 in 80 pr 81 al 82 mr 83 20

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

```
180
       Year sampling occurred.
181
182
       month
183
       Month sampling occurred.
184
185
       day
186
       Day of month sampling occurred.
187
188
189
       Time sampling occurred as Hours:Minutes.
190
191
192
       Unique alphanumeric identifier for a sampling location.
193
194
195
       Latitude of sampling location in decimal degrees.
196
197
198
       Longitude of sampling location in decimal degrees.
199
200
       site description
201
       Researchers' description of sampling location at the time of sampling.
202
203
       distance to shore m
204
       Distance from in situ sampled location to the shoreline in meters.
205
206
       depth m
207
       Maximum depth at sampling location in meters.
208
209
       air temp celsius
210
       Temperature of air at sampling location in Celsius.
211
212
       surface temp celsius
213
       Temperature of water's surface at sampling location in Celsius.
214
215
       mid temp celsius
216
       Temperature of water midway (i.e., depth m/2) between surface and bottom at sampling location in
217
       Celsius.
218
219
       bottom temp celsius
220
       Temperature of water near sediment at sampling location in Celsius.
221
222
       comments
223
       Notes in the field describing sampling conditions.
224
225
       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).
- 228
- 229 substrate photo
- Whether or not photos of the substrate were taken.
- 231
- 232 sponges
- 233 Whether or not sponges were present at a sampling location.
- 234
- 235 brandtia
- Whether or not *Brandtia* spp. (endemic amphipod species) were present at a sampling location.
- 237 238
- distance weighted population metrics.csv
- 239
- 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).

nutrients.csv

- 248
- 249 *site*
- 250 Unique alphanumeric identifier for a sampling location.
- 251
- 252 distance weighted population
- 253 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.
- 256
- 257
- 258259
 - 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
- were collected at a depth of 0.75 m. Nutrient samples were not filtered prior to analysis, meanin that nitrogen concentrations have the potential to include intracellular nitrogen. Therefore,
- 262 nitrogenous species' concentrations may be spurious. Minimal detection limits were estimated as
- 263 0.01 mg/L for nitrate, 0.005 mg/L for ammonium, and 0.04 mg/L for phosphorus.
- 264
- 265 *site*
- 266 Unique alphanumeric identifier for a sampling location.
- 267
- 268 replicate
- 269 Replicate for a given sampling location.
- 270
- 271 nh4 mg dm3

272 Ammonium concentration in milligrams of ammonium per cubic decimeter. 273 274 no3 mg dm3 275 Nitrate concentration in milligrams of nitrate per cubic decimeter 276 277 tp mg dm3 278 Total phosphorus concentration in milligrams of phosphorus per cubic decimeter. 279 280 tpo43 mg dm3 281 Total phosphate concentration as phosphate in milligrams per cubic decimeter. 282 283 chlorophylla.csv 284 285 This file contains chlorophyll a concentrations in the water column as well as fluorometric 286 corrections for each littoral and pelagic sampling location. Minimal detection limits were estimated 287 to be 0.02 mg/L. 288 289 290 Unique alphanumeric identifier for a sampling location. 291 292 replicate 293 Replicate number. 294 295 filtered volume ml 296 Lake water volume filtered in milliliters for a given replicate. 297 298 sample volume ml 299 Sample volume filtered for chlorophyll a extraction. 300 Raw, uncorrected fluorometric reading for chlorophyll analysis. 301 302 303 304 adjusted raw 305 Corrected fluorometric reading for chlorophyll analysis. 306 307 chl conc Chlorophyll a concentration in milligrams per liter. 308 309 310 ppcp.csv 311 312 This file contains Pharmaceutical and Personal Care Product (PPCP) concentrations in the water 313 column at each littoral and pelagic sampling location. Detection limits are estimated to be 0.001 314 µg/L based on a 500 mL sample volume. 315 316 site 317 Unique alphanumeric identifier for a sampling location.

318	
319	paraxanthine
320	Concentration of paraxanthine, also known as 1,7-dimethylxanthine, in micrograms per liter.
321	Paraxanthine is the main human metabolite of caffeine.
322	
323	acetaminophen
324	Concentration of acetaminophen, also known as paracetamol, in micrograms per liter.
325	
326	amphetamine
327	Concentration of amphetamine in micrograms per liter.
328	
329	caffeine
330	Concentration of caffeine in micrograms per liter.
331	
332	carbamazepine
333	Concentration of carbamazepine in micrograms per liter.
334	
335	cimetidine
336	Concentration of cimetidine in micrograms per liter.
337	
338	cotinine
339	Concentration of cotinine, which is the main human metabolite of nicotine, in micrograms per liter.
340	
341	diphenhydramine
342	Concentration of diphenhydramine in micrograms per liter.
343	
344	mda
345	Concentration of methylenedioxyamphetamine in micrograms per liter.
346	
347	mdma
348	Concentration of methylenedioxymethamphetamine in micrograms per liter.
349	
350	methamphetamine
351	Concentration of methamphetamine in micrograms per liter.
352	
353	morphine
354	Concentration of morphine in micrograms per liter.
355	
356	phenazone
357	Concentration of phenazone in micrograms per liter.
358	
359	sulfachloropyridazine
360	Concentration of sulfachloropyridazine in micrograms per liter.
361	
362	sulfamethazine
363	Concentration of <i>sulfamethazine</i> 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). 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. 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 Number of diatom cells counted for a given replicate. 435 436 437 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

456 457	Number of Desmidales spp. cells counted for a given replicate.
458	comments
459 460	Notes from the observer.
461 462	<u>invertebrates.csv</u>
463 464 465	This file contains abundance for benthic macroinvertebrates collected at each of the 14 littoral sampling locations. Only amphipod taxa were identified to species.
466	site
467 468	Unique alphanumeric identifier for a sampling location.
469	replicate
470	Replicate for sampling location. While three replicates were collected in the field, some samples
471 472	were poorly preserved, and invertebrates were not enumerated so as to prevent potential errors.
473	Acroloxidae
474	Mollusk family.
475	
476	Asellidae
477 478	Isopod family.
479	Baicaliidae
480	Mollusk family.
481	
482	Benedictidae
483 484	Mollusk family.
485	Brandtia latissima
486	Endemic amphipod species. Three subspecies exist, but samples were not identified to subspecies to
487	reduce potential errors.
488	
489	Brandtia parasitica parasitica
490	Endemic amphipod species.
491	
492	Caddisflies
493 494	General grouping; specimens were not identified to species.
495	Cryptoropus inflatus
496	Endemic amphipod species.
497	
498	Cryptoropus pachytus
499 500	Endemic amphipod species.
501	Cryptoropus_rugosus

502	Endomio amplino de masica
502 503	Endemic amphipod species.
	Eulimnogammamus aanvaalus
504	Eulimnogammarus_capreolus
505	Endemic amphipod species.
506	
507	Eulimnogammarus_cruentes
508	Endemic amphipod species.
509	
510	Eulimnogammarus_cyaneus
511	Endemic amphipod species.
512	
513	Eulimnogammarus_grandimanus
514	Endemic amphipod species.
515	
516	Eulimnogammarus_juveniles
517	Endemic amphipod genus. Identification kept at genus level so as to prevent misclassification.
518	
519	Eulimnogammarus maackii
520	Endemic amphipod species.
521	
522	Eulimnogammarus marituji
523	Endemic amphipod species.
524	
525	Eulimnogammarus verucossus
526	Endemic amphipod species.
527	
528	Eulimnogammarus viridis viridis
529	Endemic amphipod species.
530	
531	Eulimnogammarus vittatus
532	Endemic amphipod species.
533	
534	Flatworms
535	Not identified beyond phylum.
536	1 tot lacitifica degona phytam.
537	Leeches
538	Not identified beyond order, although 12 endemic species occur in Lake Baikal.
539	Two identified beyond order, attribugh 12 endernie species occur in Lake Baikar.
540	Maackia
541	Mollusk family.
542	Worldsk failing.
	Dallagga hugudtia hugudtia
543	Pallasea_brandtia_brandtia
544 545	Endemic amphipod species.
545 546	Dallar on handit toward
546 547	Pallasea_brandtii_tenera
547	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 Poekilogammarus ephippiatus 564 Endemic amphipod species. 565 566 567 Poekilogammarus juveniles Endemic amphipod genus. Identification kept at genus level so as to prevent misclassification. 568 569 570 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 This file contains carbon (δ^{13} C) and nitrogen (δ^{15} N) values for various benthic macroinvertebrate 581 582 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

594 C13 595 Carbon (δ^{13} C) stable isotope values in parts per thousand. 596 597 N15 598 Nitrogen (δ^{15} N) stable isotope values in parts per thousand. 599 600 comments 601 Quality flag column where δ^{13} C samples were outside of the range of standards. 602 603 fatty acid.csv 604 605 This file contains fatty acid concentrations for various benthic macroinvertebrate genera, periphyton, and endemic *Draparnaldia* spp. benthic algae collected from the 14 littoral sampling 606 607 locations. 608 609 site 610 Unique alphanumeric identifier for a sampling location. 611 612 Genus Genus of the analyzed organism. 613 614 615 Species 616 Species of the analyzed organism. When an organism was identified solely to genus, the Species value is NA. 617 618 619 c12 0 Concentration of 12:0 fatty acid as micrograms of fatty acid per milligram of tissue. 620 621 622 i 14 0 Concentration of i-14:0 fatty acid as micrograms of fatty acid per milligram of tissue. 623 624 625 c14 0 Concentration of 14:0 fatty acid as micrograms of fatty acid per milligram of tissue. 626 627 628 c14 1w5 629 Concentration of 14:1ω5 fatty acid as micrograms of fatty acid per milligram of tissue. 630 631 i 15 0 632 Concentration of i-15:0 fatty acid as micrograms of fatty acid per milligram of tissue. 633 634 a 15 0 Concentration of a-15:0 fatty acid as micrograms of fatty acid per milligram of tissue. 635 636 637 c15 0

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

638

```
640 c15 1w7
```

Concentration of 15:1\omega7 fatty acid as micrograms of fatty acid per milligram of tissue.

642 643 *i* 16 0

648

651

654

657

660

663

666

669

672

675

678

681

684

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

645 646 *c16* 0

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

649 *c16 1w9*

650 Concentration of 16:1ω9 fatty acid as micrograms of fatty acid per milligram of tissue.

652 *c16 1w8*

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

655 *c16 1w7*

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

658 *c16 1w6*

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

661 *c16 1w5*

662 Concentration of 16:1ω5 fatty acid as micrograms of fatty acid per milligram of tissue.

664 *i* 17 0

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

667 *a* 17 0

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

670 *c17 0*

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

673 *c17 1w7*

674 Concentration of 17:1ω7 fatty acid as micrograms of fatty acid per milligram of tissue.

676 c16 2w7

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

679 *c16 2w6*

680 Concentration of 16:2\omega\text{6} fatty acid as micrograms of fatty acid per milligram of tissue.

682 c16 2w4

Concentration of 16:2\omega4 fatty acid as micrograms of fatty acid per milligram of tissue.

685 c16 3w6

- 686 Concentration of 16:3ω6 fatty acid as micrograms of fatty acid per milligram of tissue.
- *c16_3w4*
- 689 Concentration of 16:3ω4 fatty acid as micrograms of fatty acid per milligram of tissue.
- *c16_3w3*
- 692 Concentration of 16:3ω3 fatty acid as micrograms of fatty acid per milligram of tissue.
- *c16 4w3*
- 695 Concentration of 16:4\omega3 fatty acid as micrograms of fatty acid per milligram of tissue.

- *c16 4w1*
- 698 Concentration of 16:4ω1 fatty acid as micrograms of fatty acid per milligram of tissue.

- *c18* 0
- 701 Concentration of 18:0 fatty acid as micrograms of fatty acid per milligram of tissue.

- 703 c18 1w9
- 704 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*
- 710 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*
- 716 Concentration of 18:3ω6 fatty acid as micrograms of fatty acid per milligram of tissue.

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

- *c18 4w4*
- 722 Concentration of 18:4\omega4 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*
- 728 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*
- 734 Concentration of 20:1ω9 fatty acid as micrograms of fatty acid per milligram of tissue.
- 736 c20_1w7
 737 Concentration of 20:1ω7 fatty acid as micrograms of fatty acid per milligram of tissue.
- *c20 2w5 11*
- 740 Concentration of 20:2ω5-11 fatty acid as micrograms of fatty acid per milligram of tissue.
- *c20 2w5 13*
- 743 Concentration of 20:2ω5-13 fatty acid as micrograms of fatty acid per milligram of tissue.
- *c20 2w6*

- 746 Concentration of 20:2ω6 fatty acid as micrograms of fatty acid per milligram of tissue.
- *c20 3w6*
- 749 Concentration of 20:3ω6 fatty acid as micrograms of fatty acid per milligram of tissue.
- *c20 4w6*
- 752 Concentration of 20:4ω6 fatty acid as micrograms of fatty acid per milligram of tissue.
 753
- *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*
- 761 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*
- 767 Concentration of 22:1ω9 fatty acid as micrograms of fatty acid per milligram of tissue. 768
- *c22 1w7*
- Concentration of 22:1ω7 fatty acid as micrograms of fatty acid per milligram of tissue.
- *c22 2w6*
- Concentration of 22:2\omega6 fatty acid as micrograms of fatty acid per milligram of tissue.
- *c22 4w6*

776 Concentration of $22:4\omega 6$ fatty acid as micrograms of fatty acid per milligram of tissue. 777

- 778 *c22 5w6*
- 779 Concentration of 22:5\omega6 fatty acid as micrograms of fatty acid per milligram of tissue.
- 780
- 781 *c22 3w3*
- 782 Concentration of 22:3\omega3 fatty acid as micrograms of fatty acid per milligram of tissue.
- 783
- 784 *c22 4w3*
- 785 Concentration of 22:4ω3 fatty acid as micrograms of fatty acid per milligram of tissue.
- 786 787
- 787 *c22 5w3*

c22 6w3

- 788 Concentration of 22:5ω3 fatty acid as micrograms of fatty acid per milligram of tissue.
 789
- 790

792

795

- 791 Concentration of 22:6ω3 fatty acid as micrograms of fatty acid per milligram of tissue.
- 793 *c24* 0
- 794 Concentration of 24:0 fatty acid as micrograms of fatty acid per milligram of tissue.
- 796 comments
- 797 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
- 799 consistent.
- 800
- 801 total lipid.csv
- 802

804

- This file contains gravimetry data for each fatty acid sample.
- 805 *site*
- 806 Unique alphanumeric identifier for a sampling location.
- 808 Genus
- 809 Genus of the analyzed organism.
- 810811 *Species*
- Species of the analyzed organism. When organism was identified solely to genus, the Species value
- 813 is NA.
- 814
- 815 total lipid mg per g
- Total amount of lipids in a sample in milligrams of lipid per gram of tissue.
- 817
- 818 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.
- 821
- 822 *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.

869 870

871

872

873 874

875

876

877

878

879

880

882

883 884

885

886

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,

881

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

where I is the IDW population at sampling location i, 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 km², L is the shoreline length at a developed site in km, and D is the distance from developed site *i* 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.

887 888 889

Nutrients

890 891

892

893

894

895

896

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.

897 898 899

900

901

902

903

904 905

906

907

908

909

910

911

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.

914

915 For comparable methods in English, we recommend data users consult International Standards 916 Organization (ISO) (1984) and ISO (2004) as analogs. Copies of the Russian-language methods are 917 included in the Open Science Framework portal within the directory "Nearshore sampling/methods". 918

919

Chlorophyll a

920 921 922

923

924

925

926

927

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.

928 929

932

Chlorophyll samples were processed in a manner similar to that of Welschmever (1994).

930 Nitrocellulose filters were ground in 10 mL of 90% HPLC-grade acetone, in which chlorophyll 931

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

933 nm and emission of 680 nm. 10-AU Secondary Solid Standard (P/N 10-AU-904) was used to 934

calibrate fluorometer prior to samples being processed. Blank samples registered a raw fluorescence

935 of approximately 0.1 FL units. Concentrations were calculated using formula 2 (2)

Chlorophyll concentration = (extract reading – blank reading) * $\frac{mL \text{ of extract}}{mL \text{ of filtered sample}}$. 936 937

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

938 939

Pharmaceuticals and Personal Care Products (PPCPs)

volume. Concentrations are reported in µg/L.

940 941

942

943

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).

944 945

946 Within 12 h of collection, samples were filtered directly from the amber glass bottle using an in-line 947 Teflon filter holder with glass microfiber GMF (1.0 µm pore size, WhatmanGrad 934-AH) in 948 tandem with a solid phase extraction (SPE) cartridge (200 mg HLB, Waters Corporation, Milford, 949 MA) connected to a 1-liter vacuum flask. Lab personnel wore gloves and face masks to minimize 950 contamination. Prior to filtration, SPE cartridges were primed with at least 5 mL of either methanol 951 or acetone and then washed with at least 5 mL of sample water. Rate of extraction was maintained 952 at approximately 1 drop per second. Extraction proceeded until water could no longer pass through 953 the SPE cartridge or until all collected water was filtered. Cartridges were stored in Whirlpacks at -954 20°C until analysis for 18 PPCP residues using liquid chromatography tandem mass spectrometry 955 (LC-MS-MS) following methods of Lee et al. (2016) and D'Alessio et al (2018) with labeled 956 internal standards (¹³C₃-caffeine, methamphetamine-d8, MDMAd8, morphine-d3, and ¹³C₆-957 sulfamethazine). Detection limits are estimated to be 0.001 µg/L based on a 500 mL sample

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 $\sim \! 100 x$ 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 $\sim \! 300 \, \mu m$ (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 cm² 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 $10 \,\mu L$ aliquots from each preserved sample, containing approximately 10-15 mL of preserved periphyton. For all $10 \,\mu L$ 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°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- μ 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°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.

1100 1101 1102

1103

1104

1105

1106

1098

1099

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 dryweight tissue.

1107 1108 1109

Technical Validation

1110 1111 1112

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

1113 1114

1115

1116 1117 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.

1118 1119 1120

1121 1122

1123

1124

1125

1126

1127

For analytical validation, internal standards were used for all mass-spectroscopy analyses. PPCP analyses involved labeled internal standards (13C3-caffeine, methamphetamine-d8, MDMAd8, morphine-d3, and ¹³C₆-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 δ^{13} C and δ^{15} N, respectively. Finally, fatty acid estimations used an internal 19:0 standard to assess oxidation of fatty acids during extraction, methylation, and quantification.

1128 1129

1130 For data reproducibility, data aggregation and harmonization procedures were conducted in the R 1131 statistical environment (R Core Team 2019), using the tidyverse (Wickham et al. 2019) packages. 1132 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. 1133 1134 This step created the standardized CSV files detailed above, which are available on the EDI 1135 repository (Meyer et al. 2021). Raw data files are available on the project's Open Science 1136 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 1137 1138 the sampling-site-level using script "01 data cleaning.R". In addition to aggregation scripts, six R 1139 scripts used for analyses in Meyer et al. (*Under Review*) are also available on the EDI repository 1140 within the compressed entity "scripts.tar.gz". All R code for data aggregation was written by one 1141 person (Michael F. Meyer) and then independently reviewed by two others (Matthew R. Brousil 1142 and Kara H. Woo) to confirm that code performed as intended, was well documented, and annotations were complete.

1144 1145

A commitment to FAIR and TRUST principles

1146 1147

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.

1149 1150 1151

1152

1153

1154

1155

1148

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).

1156 1157 1158

1159

1160

1161 1162

1163

1164

1165 1166

1167

1168

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.

1169 1170 1171

Data Use and Recommendations for Reuse

1173 1174

1175

1172

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.

1176 1177

1178 First, the final data products can be harmonized with other littoral sampling efforts throughout Lake 1179 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 1180 1181 (Kravtsova et al. 2014; Timoshkin et al. 2016, 2018; Volkova et al. 2018). Recent benthic algal 1182 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 1183 1184 near larger lakeside developments (Timoshkin et al. 2016; Volkova et al. 2018). For example, 1185 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 1186 1187 growing tourism hub, and hosts over 1.2 million tourists per year (Interfax-Tourism 2018). Surveys 1188 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.

1193 1194 1195

1196

1197

1198

1199

1200

1201

1202

1203

1204

1205

1206

1207 1208

1209

1210

1190

1191

1192

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.

1214 1215	References
1216	Anisimov, O., and S. Reneva. 2006. Permafrost and changing climate: The Russian perspective.
1217	Ambio 35 : 169–175.
1218	Baquero, O. S. 2019. ggsn: North symbols and scale bars for maps created with "ggplot2" or
1219	"ggmap,."
1220	Barnes, D. K. A., F. Galgani, R. C. Thompson, and M. Barlaz. 2009. Accumulation and
1221	fragmentation of plastic debris in global environments. Philos Trans R Soc Lond B Biol Sci
1222	364 : 1985–1998. doi:10.1098/rstb.2008.0205
1223	Bendz, D., N. A. Paxéus, T. R. Ginn, and F. J. Loge. 2005. Occurrence and fate of pharmaceutically
1224	active compounds in the environment, a case study: Höje River in Sweden. Journal of
1225	Hazardous Materials 122 : 195–204. doi:10.1016/j.jhazmat.2005.03.012
1226	Bondarenko, N. A., I. V. Tomberg, A. A. Shirokaya, and others. 2021. Dolichospermum
1227	lemmermannii (Nostocales) bloom in world's deepest Lake Baikal (East Siberia):
1228	abundance, toxicity and factors influencing growth. Limnology and Freshwater Biology 1:
1229	1101–1110. doi:10.31951/2658-3518-2021-A-1-1101
1230	Brandon, J. A., A. Freibott, and L. M. Sala. 2020. Patterns of suspended and salp-ingested
1231	microplastic debris in the North Pacific investigated with epifluorescence microscopy.
1232	Limnology and Oceanography Letters 5: 46–53. doi:10.1002/lol2.10127
1233	Brodin, T., J. Fick, M. Jonsson, and J. Klaminder. 2013. Dilute concentrations of a psychiatric drug
1234	alter behavior of fish from natural populations. Science 339: 814-815.
1235	doi:10.1126/science.1226850

1236	Camilleri, A. C., and T. Ozersky. 2019. Large variation in periphyton δ 13C and δ 15N 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. EcolProg. 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 U.S. Streams, 1999–2000: A national reconnaissance. Environmental Science & 1310 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. <i>Under Review</i> . 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

1431	Yang, Y., W. Song, H. Lin, W. Wang, L. Du, and W. Xing. 2018. Antibiotics and antibiotic
1432	resistance genes in global lakes: A review and meta-analysis. Environment International
1433	116: 60–73. doi:10.1016/j.envint.2018.04.011
1434	Yoshida, T., T. Sekino, M. Genkai-Kato, and others. 2003. Seasonal dynamics of primary
1435	production in the pelagic zone of southern Lake Baikal. Limnology 4: 53-62.
1436	doi:10.1007/s10201-002-0089-3
1437 1438 1439	

Acknowledgements

State University.

1440 1441 1442

1443

1444

1445

1446

1447

1448

1449

1450

1451

1452

1453

1454

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 1. (N. FZZE-2 vironmental . 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

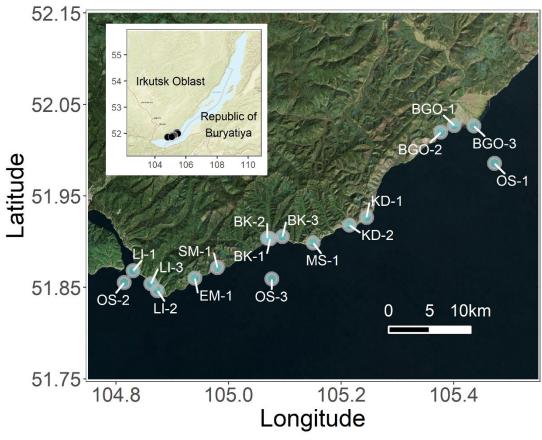


Figure 1: Map of all sampling locations with sites labeled with unique alphanumeric code. The entire transect included three developed sites (i.e., Listvyanka (LI), Bolshie Koty (BK), Bolshoe Goloustnoe (BGO)). Three offshore sites (OS) were also sampled to compare pelagic sewage signals to those in 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
		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)
	O ₄	Eulimnogammarus cyaneus (Dybowsky 1874)
		Eulimnogammarus grandimanus (Bazikalova 1945)
		Eulimnogammarus maacki (Gerstfeldt 1858)
		Eulimnogammarus marituji (Bazikalova 1945)
	Amphipoda	Eulimnogammarus verucossus (Gerstfeldt 1858)
		Eulimnogammarus viridis viridis (Dybowsky 1874)
		Eulimnogammarus vittatus (Dybowsky 1874)
Abundance Estimates		Pallasea brandtia brandita (Dybowsky 1874)
Abundance Estimates		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	
		Diatom	
	Benthic Algae	Ulothrix spp.	
	Bentine Aigae	Spirogyra spp.	
		Tetrasporales	
		Eulimnogammarus cyaneus (Dybowsky 1874)	
	Amphipoda	Eulimnogammarus verucossus (Gerstfeldt 1858)	
Stable Isotopes	Ampinipoda	Eulimnogammarus vittatus (Dybowsky 1874)	
		Pallasea cancellus (Pallas 1776)	
	Benthic Algae	Periphyton	
		Eulimnogammarus cyaneus (Dybowsky 1874)	
		Eulimnogammarus verucossus (Gerstfeldt 1858)	
	Amphipoda	Eulimnogammarus vittatus (Dybowsky 1874)	
Fatty Acids		Hyalella spp.	
ratty Acius		Pallasea cancellus (Pallas 1776)	
	Molluska	Processed in composite and not identified to family.	
	Donthic Algae	Periphyton	
	Benthic Algae	Draparnaldia spp.	

1 2	A unified dataset of co-located sewage pollution, periphyton, and benthic macroinvertebrate community and food web structure from Lake Baikal (Siberia)
3	
4	Michael F. Meyer ^{1*}
5	Ted Ozersky ²
6	Kara H. Woo ³
7	Kirill Shchapov ²
8	Aaron W. E. Galloway ⁴
9	Julie B. Schram ⁴
10	Daniel D. Snow ⁵
11	Maxim A. Timofeyev ⁶
12	Dmitry Yu. Karnaukhov ⁶
13	Matthew R. Brousil ³
14	Stephanie E. Hampton ³
15	
16	^{1.} School of the Environment, Washington State University, Pullman, WA, USA
17	² Large Lakes Observatory, University of Minnesota - Duluth, Duluth, MN, USA
18	^{3.} Center for Environmental Research, Education, and Outreach, Washington State University,
19	Pullman, WA, USA
20	^{4.} Oregon Institute of Marine Biology, University of Oregon, Charleston, OR, USA
21	^{5.} School of Natural Resources, University of Nebraska-Lincoln, Lincoln, NE, USA
22	⁶ Biological Research Institute, Irkutsk State University, Irkutsk, Irkutsk Oblast, Russia
23	
24	*corresponding author: michael.f.meyer@wsu.edu
25	
26	Author Contribution Statement
27	Conceptualized the project: MFM, SEH, TO
28	Collected samples in the field: MFM, TO, KHW, SEH
29	Processed samples: MFM, KS, JBS, DDS, TO, AWEG, SEH
30	Wrote and Reviewed R scripts: MFM, MRB, KHW
31	Data management: MFM, MRB
32	Wrote and edited the manuscript: All authors
33	Approved the final manuscript: All authors
34	
35	Grant sponsor information:
36	Funding was provided by the National Science Foundation (NSF-DEB-1136637) to S.E.H., a Fulbright
37	Fellowship to M.F.M., a NSF Graduate Research Fellowship to M.F.M. (NSF-DGE-1347973), and the
38	Russian Ministry of Science and Education (N FZZE-2020-0026; N FZZE-2020-0023).
39	
40	Key Words: pharmaceuticals, microplastics, fatty acids, stables isotopes, amphipod, mollusk,
41	diatom, spirogyra
42	
43	URL of the Dataset and Metadata with permanent identifier:
44	Environmental Data Initiative: doi:10.6073/pasta/76f43144015ec795679bac508efa044b
45	Open Science Framework: https://doi.org/10.17605/OSF.IO/9TA8Z

Code URL with permanent identifier:

- Environmental Data Initiative: doi:10.6073/pasta/76f43144015ec795679bac508efa044b
- 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
- 52 Isotopes, nitrate, ammonium, total phosphorus
- 53 Technology Type(s): GC/MS, LC/MS, Spectrophotometry, Spectrophotometry, Fluorometry,
- 54 <u>Microscopy</u>
- **Temporal range:** 19 23 August 2015
- 56 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 this-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 samplingsdata. 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. The data are structured in a tidy format (a tabular arrangement familiar to limnologists) to encourage reuse. 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 require relatively complex, computationally intensive workflows (Meyer et al. 2020a).

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 blooms (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. 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).

chlorophylla.csv

This file contains chlorophyll a concentrations <u>in the water column</u> as well as fluorometric corrections for each littoral and pelagic sampling location.

179 sii

Unique alphanumeric identifier for a sampling location.

- 182 replicate
- 183 Replicate number.

- 185 filtered volume ml
- Lake water volume filtered in milliliters for a given replicate.

187

- 188 sample volume ml
- 189 Sample volume filtered for chlorophyll a extraction.

190

- 191 raw fluo
- Raw, uncorrected fluorometric reading for chlorophyll analysis.

193

- 194 adjusted raw
- 195 Corrected fluorometric reading for chlorophyll analysis.

196 197

- 97 chl_conc
- 198 Chlorophyll a concentration in milligrams per liter.

199 200

distance weighted population metrics.csv

201 202

203

204

205

206

207

This file contains <u>inverse distance weighted</u>, <u>census-based</u> 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. <u>As these population estimates are based on census data, they reflect static populations and do not account for seasonal population deviations from tourism.</u> A full description of the methods used to calculate IDW population can be found in the companion manuscript Meyer et al. (Under Review).

208209

- 210211
- 212 Unique alphanumeric identifier for a sampling location.

213

214 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.

218

219 fatty acid.csv

site

220

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.

224

- 225 sit
- 226 Unique alphanumeric identifier for a sampling location.

- 228 Genus
- Genus of the analyzed organism.
- 230
- 231 Species
- Species of the analyzed organism. When an oorganism was identified solely to genus, the Species
- value is NA.
- 234
- 235 *c12 0*
- 236 Concentration of 12:0 fatty acid as micrograms of fatty acid per milligram of tissue.
- 237
- 238 *i* 14 0
- 239 Concentration of i-14:0 fatty acid as micrograms of fatty acid per milligram of tissue.
- 240
- 241 *c14* 0
- 242 Concentration of 14:0 fatty acid as micrograms of fatty acid per milligram of tissue.
- 243
- 244 *c14_4w5*

i 15 0

- 245 Concentration of 14:4n-5 fatty acid as micrograms of fatty acid per milligram of tissue.
- 246247
- 248 Concentration of i-15:0 fatty acid as micrograms of fatty acid per milligram of tissue.
- 249
- 250 *a_15_0*
- 251 Concentration of a-15:0 fatty acid as micrograms of fatty acid per milligram of tissue.
- 252 253 *c15* 0
- 254 Concentration of 15:0 fatty acid as micrograms of fatty acid per milligram of tissue.
- 255 256 *c15 lw7*
- 257 Concentration of 15:1ω7 fatty acid as micrograms of fatty acid per milligram of tissue.
- 258 259 *i* 16 0
- 260 Concentration of i-16:0 fatty acid as micrograms of fatty acid per milligram of tissue.
- 261 262 *c16 0*
- Concentration of 16:0 fatty acid as micrograms of fatty acid per milligram of tissue.
- 265 *c16 1w9*
- Concentration of 16:1ω9 fatty acid as micrograms of fatty acid per milligram of tissue.
- 268 *c16_1w8*
- 269 Concentration of 16:1ω8 fatty acid as micrograms of fatty acid per milligram of tissue.
- 271 *c16 lw7*
- 272 Concentration of 16:1ω7 fatty acid as micrograms of fatty acid per milligram of tissue.
- 273

```
274 c16 1w6
```

- 275 Concentration of 16:1ω6 fatty acid as micrograms of fatty acid per milligram of tissue.
- *c16 1w5*
- 278 Concentration of 16:1ω5 fatty acid as micrograms of fatty acid per milligram of tissue.
- *i* 17 0
- 281 Concentration of i-17:0 fatty acid as micrograms of fatty acid per milligram of tissue.
- *a 17 0*

- 284 Concentration of a-17:0 fatty acid as micrograms of fatty acid per milligram of tissue.
- *c17 0*
- 287 Concentration of 17:0 fatty acid as micrograms of fatty acid per milligram of tissue.
- *c17 1w7*
- 290 Concentration of 17:1n-7 fatty acid as micrograms of fatty acid per milligram of tissue.
- *c16 2w7*
- 293 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*
- 299 Concentration of 16:2\omega4 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*
- 305 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*
- 314 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*

320 Concentration of 18:1ω9 fatty acid as micrograms of fatty acid per milligram of tissue.

- *c18_1w7*
- Concentration of 18:1\omega7 fatty acid as micrograms of fatty acid per milligram of tissue.

- *c18_2w6t*
- 326 Concentration of 18:2\omega\text{6t} fatty acid as micrograms of fatty acid per milligram of tissue.

- *c18 2w6*
- 329 Concentration of 18:2ω6 fatty acid as micrograms of fatty acid per milligram of tissue.

- *c18 3w6*
- 332 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*
- 338 Concentration of 18:4\omega4 fatty acid as micrograms of fatty acid per milligram of tissue.

- *c18 4w3*
- Concentration of 18:4\omega3 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*
- 350 Concentration of 20:1ω9 fatty acid as micrograms of fatty acid per milligram of tissue.

- *c20 1w7*
- 353 Concentration of 20:1ω7 fatty acid as micrograms of fatty acid per milligram of tissue.

- *c20 2w5 11*
- 356 Concentration of 20:2-5-11 fatty acid as micrograms of fatty acid per milligram of tissue.

- *c20 2w5 13*
- Concentration of 20:2-5-13 fatty acid as micrograms of fatty acid per milligram of tissue.

- *c20 2w6*
- 362 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.

366 367 c20 4w6 Concentration of 20:4\omega\text{6} fatty acid as micrograms of fatty acid per milligram of tissue. 368 369 370 c20 3w3 371 Concentration of 20:3ω3 fatty acid as micrograms of fatty acid per milligram of tissue. 372 373 c20 4w3 374 Concentration of 20:4\omega3 fatty acid as micrograms of fatty acid per milligram of tissue. 375 376 c20 5w3 377 Concentration of 20:5ω3 fatty acid as micrograms of fatty acid per milligram of tissue. 378 379 c22 0 380 Concentration of 22:0 fatty acid as micrograms of fatty acid per milligram of tissue. 381 382 c22 1w9 383 Concentration of 22:1\omega9 fatty acid as micrograms of fatty acid per milligram of tissue. 384 385 *c22 1w7* Concentration of 22:1ω7 fatty acid as micrograms of fatty acid per milligram of tissue. 386 387 388 c22 2w6 389 Concentration of 22:2\omega6 fatty acid as micrograms of fatty acid per milligram of tissue. 390 391 c22 4w6 Concentration of 22:4\omega\text{6} fatty acid as micrograms of fatty acid per milligram of tissue. 392 393 394 c22 5w6 395 Concentration of 22:5\omega6 fatty acid as micrograms of fatty acid per milligram of tissue. 396 397 *c22 3w3* 398 Concentration of 22:3\omega3 fatty acid as micrograms of fatty acid per milligram of tissue. 399 400 c22 4w3 401 Concentration of 22:4\omega3 fatty acid as micrograms of fatty acid per milligram of tissue. 402 403 *c22 5w3* Concentration of 22:5\omega3 fatty acid as micrograms of fatty acid per milligram of tissue. 404 405 406 *c22 6w3* Concentration of 22:6\omega3 fatty acid as micrograms of fatty acid per milligram of tissue. 407 408 409 $c24 \ 0$ 410 Concentration of 24:0 fatty acid as micrograms of fatty acid per milligram of tissue.

412	comments
413	Quality flag column. Two samples spilled during fatty acid extraction. These samples are flagged as
414	such. Although concentrations are lower than other samples, proportions between fatty acids are
415	consistent.
416	
417	<u>invertebrates.csv</u>
418	
419	This file contains abundance for benthic macroinvertebrates collected at each of the 14 littoral
420	sampling locations. Only amphipod taxa were identified to species. Mollusks and isopods were
421	identified to genus.
422	
423	site
424	Unique alphanumeric identifier for a sampling location.
425	
426	replicate
427	Replicate for sampling location. While three replicates were collected in the field, some samples
428	were poorly preserved, and invertebrates were not enumerated so as to prevent potential errors.
429	
430	Acroloxidae
431	Mollusk genus <u>family</u>
432	
433	Asellidae
434	Endemic iIsopod familygenus
435	
436	Baicaliidae
437	Mollusk <u>familygenus</u> , most of which are endemic
438	
439	Benedictidae
440	Mollusk <u>family</u> genus, most of which are endemic
441	
442	
443	Brandtia_latissima
444	Endemic amphipod species. Three subspecies exist, but samples were not identified to subspecies to
445	reduce potential errors.
446	D. Lie vie vie
447	Brandtia_parasitica_parasitica
448	Endemic amphipod species
449	
450	Caddisflies
451	General grouping; were not identified to species.
452	Commence in Autom
453	Cryptoropus_inflatus
454	Endemic amphipod species
455	Competance of a charter
456	Cryptoropus_pachytus
457	Endemic amphipod species

458	
459	Cryptoropus rugosus
460	Endemic amphipod species
461	
462	Eulimnogammarus capreolus
463	Endemic amphipod species
464	
465	Eulimnogammarus cruentes
466	Endemic amphipod species
467	
468	Eulimnogammarus cyaneus
469	Endemic amphipod species
470	Enderme umpmpod species
471	Eulimnogammarus grandimanus
472	Endemic amphipod species
473	
474	Eulimnogammarus juveniles
475	Endemic amphipod genus. Identification kept at genus level so as to prevent misclassification.
476	
477	Eulimnogammarus maackii
478	Endemic amphipod species
479	
480	Eulimnogammarus marituji
481	Endemic amphipod species
482	
483	Eulimnogammarus verucossus
484	Endemic amphipod species
485	
486	Eulimnogammarus_viridis_viridis
487	Endemic amphipod species
488	Eulimnogammarus_viridis_viridis Endemic amphipod species
489	Eulimnogammarus_vittatus
490	Endemic amphipod species
491	
492	Flatworms
493	Not identified beyond orderphylum.
494	
495	Leeches
496	Not identified beyond order, although 12 endemic species do exist.
497	
498	Maackia
499	Mollusk <u>familygenus</u> , most of which are endemic
500	
501	Pallasea_brandtia_brandtia
502	Endemic amphipod species
503	

504	Pallasea brandtii tenera
505	Endemic amphipod species
506	
507	Pallasea cancelloides
508	Endemic amphipod species
509	
510	Pallasea cancellus
511	Endemic amphipod species
512	
513	Pallasea viridis
514	Endemic amphipod species
515	
516	Planorbidae
517	Mollusk family genus, most of which are endemic
518	
519	Poekilogammarus crassimus
520	Endemic amphipod species
521	
522	Poekilogammarus ephippiatus
523	Endemic amphipod species
524	
525	Poekilogammarus juveniles
526	Endemic amphipod genus. <u>Identification kept at genus level so as to prevent</u>
527	misclassification. Identifying to species introduced risk of misclassification.
528	
529	Poekilogammarus megonychus perpolitus
530	Endemic amphipod species
531	
532	Poekilogammarus_pictus
533	Endemic amphipod species
534	
535	Valvatidae
536	Mollusk genus, most of which are endemicfamily
537	
538	site_informationmetadata.csv
539	
540	This file contains metadata for each of the pelagic and littoral sampling locations. Missing data are
541	assigned as NA.
542	
543	year
544	Year sampling occurred.
545	
546	month
547	Month sampling occurred.
548	
549	day

```
550
       Day of month sampling occurred.
551
552
       time
553
       Time sampling occurred as Hours:Minutes.
554
555
       site
556
       Unique alphanumeric identifier for a sampling location.
557
558
559
       Latitude of sampling location in decimal degrees.
560
561
562
       Longitude of sampling location in decimal degrees.
563
564
       site description
565
       Researchers' description of sampling location at the time of sampling.
566
567
       distance to shore m
       Distance from in situ sampled location to the shoreline in meters.
568
569
570
       depth m
571
       Maximum dDepth at in situ sampling location in meters.
572
573
       air temp celsius
574
       Temperature of air at sampling location in Celsius.
575
576
       surface temp celsius
577
       Temperature of water's surface at sampling location in Celsius.
578
579
       mid temp celsius
580
       Temperature of water midway (i.e., depth m/2) between surface and bottom at sampling location in
581
       Celsius.
582
583
       bottom temp celsius
584
       Temperature of water near sediment at sampling location in Celsius.
585
586
       comments
587
       Notes in the field describing sampling conditions.
588
589
       shore photo
590
       Whether or not photos of the shoreline were taken. Photos are available on the project's Open
591
       Science Framework page portal (Meyer et al. 2015).
592
593
       substrate photo
594
       Whether or not photos of the substrate were taken.
595
```

sponges Whether or not sponges were present at a sampling location. brandtia Whether or not *Brandtia spp.* (endemic amphipod <u>species</u>) was were present at a sampling location. microplastics.csv This file contains suspended microplastics counts for each of the pelagic and littoral sampling locations. site Unique alphanumeric identifier for a sampling location. replicate Replicate for a given sampling location. Replicate values of "C" indicate a control. Number of microplastic fragments observed. Number of microplastic fibers observed. Number of microplastic beads observed. comments Observer comments while enumerating microplastics volume filtered ml Volume in milliliters for a given replicate filtered. 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. site Unique alphanumeric identifier for a sampling location. replicate Replicate for a given sampling location. nh4 mg dm3

642 643	Ammonium concentration in milligrams of ammonium per cubic decimeter.
644	no3 mg dm3
645	Nitrate concentration in milligrams of nitrate per cubic decimeter
646	8 r r
647	tp mg dm3
648	Total phosphorus concentration in milligrams of phosphorus per cubic decimeter.
649	- com prospersor construction
650	tpo43 mg dm3
651	Total phosphate concentration as phosphate in milligrams per cubic decimeter.
652	Surface Control of the Control of th
653	periphyton.csv
654	
655	This file contains periphyton abundance data, collected from rocks at for each of the sampled
656	littoral locations. For poorly preserved samples, counts are listed as NA for each taxonomic
657	grouping, and a note in the "comments" column is provided.
658	8-0 sp-0-8, man in accordance to the contract of participation.
659	site
660	Unique alphanumeric identifier for a sampling location.
661	
662	replicate
663	Replicate number for a given sampling site.
664	represent name of for a gryon sampling site.
665	subsamples counted
666	Number of $\frac{1}{10}$ microliter subsamples counted for a given replicate.
667	The state of the s
668	diatom
669	Number of diatom cells counted for a given replicate.
670	
671	spirogyra
672	Number of <i>Spirogyra spp.</i> cells counted for a given replicate.
673	
674	spirogyra filament
675	Number of <i>Spirogyra spp</i> . filaments counted for a given replicate.
676	
677	ulothrix
678	Number of <i>Ulothrix spp.</i> cells counted for a given replicate.
679	
680	ulothrix filament
681	Number of <i>Ulothrix spp</i> . filaments counted for a given replicate.
682	
683	tetrasporales
684	Number of Tetrasporales-spp. cells counted for a given replicate
685	
686	pediastrum
687	Number of <i>Pediastrum spp.</i> cells counted for a given replicate.

688 689 desmidales 690 Number of *Desmidales* spp. cells counted for a given replicate. 691 692 comments 693 Notes from the observer. 694 695 ppcp.csv 696 697 This file contains Pharmaceutical and Personal Care Product (PPCP) concentrations in the water 698 column for at each littoral and pelagic sampling location. Detection limits are estimated to be 0.001 699 µg/L based on a 500 mL sample volume. 700 701 site 702 Unique alphanumeric identifier for a sampling location. 703 704 paraxanthine 705 Concentration of paraxanthine, also known as 1,7-dimethylxanthine, in micrograms per liter. 706 Paraxanthine is the main human metabolite of caffeine 707 708 acetaminophen 709 Concentration of acetaminophen, also known as paracetamol, in micrograms per liter. 710 711 amphetamine 712 Concentration of amphetamine in micrograms per liter. 713 714 715 Concentration of caffeine in micrograms per liter. 716 717 carbamazepine 718 Concentration of carbamazepine in micrograms per liter. 719 720 cimetidine 721 Concentration of cimetidine in micrograms per liter. 722 723 724 Concentration of cotinine, which is the main human metabolite of nicotine, in micrograms per liter. 725 726 diphenhydramine 727 Concentration of diphenhydramine in micrograms per liter. 728 729 mda 730 Concentration of methylenedioxyamphetamine in micrograms per liter. 731 732 mdma 733 Concentration of methylenedioxymethamphetamine in micrograms per liter.

734	
735	methamphetamine
736	Concentration of methamphetamine in micrograms per liter.
737	
738	morphine
739	Concentration of morphine in micrograms per liter.
740	
741	phenazone
742	Concentration of phenazone in micrograms per liter.
743	
744	sulfachloropyridazine
745	Concentration of sulfachloropyridazine in micrograms per liter.
746	
747	sulfamethazine
748	Concentration of <i>sulfamethazine</i> in micrograms per liter.
749	
750	sulfamethoxazole
751	Concentration of sulfamethoxazole in micrograms per liter.
752	
753	thiabendazole
754	Concentration of thiabendazole in micrograms per liter.
755	
756	trimethoprim
757	Concentration of trimethoprim in micrograms per liter.
758	
759	collection year
760	Year sample was collected in the field.
761	
762	collection month
763	Month sample was collected in the field.
764	•
765	collection day
766	Day of month sample was collected in the field.
767	
768	analysis year
769	Year sample was analyzed.
770	
771	analysis_month
772	Month sample was analyzed.
773	
774	analysis_day
775	Day of month sample was analyzed.
776	
777	stable_isotopes.csv
778	

This file contains carbon (δ^{13} C) and nitrogen (δ^{15} N) values for various benthic macroinvertebrate 779 780 genera and periphyton collected from the 14 littoral sampling locations. 781 782 *C13* 783 Carbon (δ^{13} C) stable isotope values in parts per thousand. 784 785 N15 786 Nitrogen (δ^{15} N) stable isotope values in parts per thousand. 787 788 site Unique alphanumeric identifier for a sampling location. 789 790 791 Genus 792 Genus of the analyzed organism. 793 794 Species Species of the analyzed organism. When organism was identified solely to genus, the Species value 795 796 is NA. 797 798 comments Quality flag column where δ^{13} C samples were outside of the range of standards. 799 800 801 total lipid.csv 802 803 This file contains gravimetry data for each fatty acid sample. 804 805 806 Unique alphanumeric identifier for a sampling location. 807 Genus 808 809 Genus of the analyzed organism. 810 811 Species Species of the analyzed organism. When organism was identified solely to genus, the Species value 812 813 is NA. 814 815 total lipid mg per g 816 Total amount of lipids in a sample in milligrams of lipid per gram of tissue. 817 818 deviation 819 Samples were weighed three times and standard deviation in measurement was calculated. All values are reported in milligrams of lipid per gram of tissue. 820 821 822

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

823

824

such.

Data Availability

Data are available at the replicate level at the Environmental Data Initiative (doi.org/10.6073/pasta/76f43144015ec795679bac508efa044b).

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 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 that can extend to depths of 100 m (Takhteev and Didorenko 2015). 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).

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

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 km², 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 potentially include intracellular nitrogen and overestimate dissolved nitrogenous forms in the water column.

For each water sample, nitrate, ammonium, and total phosphorus concentrations were measured. For ammonium (RD:52.24.383-2018 2018)(GOST:33045-2014 2016a)(2016a) and nitrate (RD:52.24.380-2017 2018)(2017) concentrations, samples were analyzed with a spectrophotometer (SF-26). following the addition of Nessler's reagent and disulfuric acid respectively. 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. Total Pphosphorus concentration was measured with a spectrophotometer (SF-46) following the addition of persulfate (GOST:18309-2014 2016)(GOST:18309-2014 2016b)(2016b). 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.

9<mark>17</mark>

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 users looking for comparable methods in English, we recommend data users consult International Standards Organization (ISO) (1984) and ISO (2004) as analogs.

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) Parsons and Strickland (1963) and Lorenzen (1967). Nitrocellulose filters were ground in 10 mL of 90% HPLC-grade acetone, in which chlorophyll extraction was allowed to proceed overnight. Samples were then centrifuged for 15-20 minutes. After centrifugation, absorbance of the eChlorophyll 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 nmmeasured in a spectrophotometer at 630, 645, 665, and 750 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 the formula: Chlorophyll concentration = (extract reading – blank reading) * $\frac{mL \text{ of } \text{ extract}}{mL \text{ of } \text{ filtered sample}}$; where A is the absorbance value of a particular wavelength, V₁ is the volume of the filtered water, and V₂ is the volume of extract. 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 $\sim \! 100 x$ 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 $\sim \! 300 \, \mu m$ (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 abundance

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 cm² 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.

1006 Periphyton taxonomic identification and enumeration was performed by subsampling 10 uL 1007 aliquots from each preserved sample, containing approximately 10-15 mL of preserved periphyton. 1008 For all 10 µL aliquots, cells, filaments, and colonies were counted, for the entire subsample, until at 1009 least 300 cells were identified for a given sampling replicate. If the first aliquot contained less than 1010 300 cells, we counted additional subsamples until we reached at least 300 cells in total. In instances 1011 when 300 cells were counted before finishing a subsample, we still counted the entire aliquot. Taxa 1012 were classified into broad categories consistent with Baikal algal taxonomy (Izhboldina 2007), 1013 using coarse groupings to capture general patterns in relative algal abundance. As a result, algal 1014 groups consisted of diatoms, *Ulothrix*, *Spirogyra*, and the green algal Order Tetrasporales.

1015 1016

1017

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.

10<mark>18</mark> 1019 1020

Benthic macroinvertebrate abundance collection and abundance estimates

1021 1022

1023

1024

1025

1026

1027

1028

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.

1029 1030 1031

1032

1033

1034

1035

1036

1037

1038

1039

1040

1041

1042

1043

1044

1045

1046

1047

1048

1049

1050

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 (Valuvskiy 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. Invertebrate taxonomic identification and enumeration were performed under a stereo microscope. All invertebrates were identified to species with the exception of juveniles (Takhteev and Didorenko (2015) for amphipods; Sitnikova (2012) for mollusks; Table 2). 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 more sparse 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, Mmeasurements 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.

11<mark>2</mark>1

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 acidmethanol 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 to assess deviation in measurements. Lipid gravimetry is reported as the mg of lipids per g of dryweight 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.
- 1145 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. 2020b). 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
- 1150 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.

1157 1158

A commitment to FAIR and TRUST principles

1159

- 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.

1163 1164

- 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 10 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*).

1170

- 1171 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.

1183 1184

Data Use and Recommendations for Reuse

1185

- 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.

1190 1191

1192

1193

1194

1195

1196

1197

1198

1199

1200

1201

1202

1203

1204

1205

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 (Kraytsova 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.

1206 1207 1208

1209

1210

1211

1212

1213

1214 1215

1216

1217

1218 1219

1220

1221

1222

1223

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.

1224 1225 1226

1227 1228	References
1229	Anisimov, O., and S. Reneva. 2006. Permafrost and Changing Climate: The Russian Perspective.
1230	<u>Ambio 35: 169–175.</u>
1231	Barnes, D. K. A., F. Galgani, R. C. Thompson, and M. Barlaz. 2009. Accumulation and
1232	fragmentation of plastic debris in global environments. Philos Trans R Soc Lond B Biol Sci
1233	364: 1985–1998. doi:10.1098/rstb.2008.0205
1234	Bendz, D., N. A. Paxéus, T. R. Ginn, and F. J. Loge. 2005. Occurrence and fate of pharmaceutically
1235	active compounds in the environment, a case study: Höje River in Sweden. Journal of
1236	<u>Hazardous Materials</u> 122 : 195–204. doi:10.1016/j.jhazmat.2005.03.012
1237	Bondarenko, N. A., I. V. Tomberg, A. A. Shirokaya, and others. 2021. Dolichospermum
1238	lemmermannii (Nostocales) bloom in world's deepest Lake Baikal (East Siberia):
1239	abundance, toxicity and factors influencing growth. Limnology and Freshwater Biology 1:
1240	1101–1110. doi:10.31951/2658-3518-2021-A-1-1101
1241	Brandon, J. A., A. Freibott, and L. M. Sala. 2020. Patterns of suspended and salp-ingested
1242	microplastic debris in the North Pacific investigated with epifluorescence microscopy.
1243	Limnology and Oceanography Letters 5: 46–53. doi:10.1002/lol2.10127
1244	Brodin, T., J. Fick, M. Jonsson, and J. Klaminder. 2013. Dilute Concentrations of a Psychiatric
1245	Drug Alter Behavior of Fish from Natural Populations. Science 339: 814–815.
1246	doi:10.1126/science.1226850
1247	Camilleri, A. C., and T. Ozersky. 2019. Large variation in periphyton δ13C and δ15N values in the
1248	upper Great Lakes: Correlates and implications. Journal of Great Lakes Research 45: 986–
1249	990. doi:10.1016/j.jglr.2019.06.003

1250	Costanzo, S. D., M. J. O'Donohue, W. C. Dennison, N. R. Loneragan, and M. Thomas. 2001. A
1251	New Approach for Detecting and Mapping Sewage Impacts. Marine Pollution Bulletin 42:
1252	149–156. doi:10.1016/S0025-326X(00)00125-9
1253	D'Alessio, M., S. Onanong, D. D. Snow, and C. Ray. 2018. Occurrence and removal of
1254	pharmaceutical compounds and steroids at four wastewater treatment plants in Hawai'i and
1255	their environmental fate. Science of The Total Environment 631–632: 1360–1370.
1256	doi:10.1016/j.scitotenv.2018.03.100
1257	Dalsgaard, J., M. St. John, G. Kattner, D. Müller-Navarra, and W. Hagen. 2003. Fatty acid trophic
1258	markers in the pelagic marine environment, p. 225–340. In Advances in Marine Biology.
1259	Elsevier.
1260	Edmondson, W. T. 1970. Phosphorus, Nitrogen, and Algae in Lake Washington after Diversion of
1261	Sewage. Science 169: 690–691.
1262	Fellows, I., and using the Jm. library by J. P. Stotz. 2019. OpenStreetMap: Access to Open Street
1263	Map Raster Images,.
1264	Focazio, M. J., D. W. Kolpin, K. K. Barnes, E. T. Furlong, M. T. Meyer, S. D. Zaugg, L. B. Barber,
1265	and M. E. Thurman. 2008. A national reconnaissance for pharmaceuticals and other organic
1266	wastewater contaminants in the United States - II) Untreated drinking water sources.
1267	SCIENCE OF THE TOTAL ENVIRONMENT 402: 201–216.
1268	doi:10.1016/j.scitotenv.2008.02.021
1269	Gartner, A., P. Lavery, and A. J. Smit. 2002. Use of delta N-15 signatures of different functional
1270	forms of macroalgae and filter-feeders to reveal temporal and spatial patterns in sewage
1271	dispersal. Mar. EcolProg. Ser. 235: 63–73. doi:10.3354/meps235063

1272	GOST:18309-2014. 2016. Methods for determination of phosphorus-containing matters (with
1273	corrections) (Методы определения фосфорсодержащих веществ).
1274	Green, D. S. 2016. Effects of microplastics on European flat oysters, Ostrea edulis and their
1275	associated benthic communities. Environmental Pollution 216: 95–103.
1276	doi:10.1016/j.envpol.2016.05.043
1277	Hall, R. I., P. R. Leavitt, R. Quinlan, A. S. Dixit, and J. P. Smol. 1999. Effects of agriculture,
1278	urbanization, and climate on water quality in the northern Great Plains. Limnology and
1279	Oceanography 44: 739–756. doi:10.4319/lo.1999.44.3_part_2.0739
1280	Hampton, S. E., S. C. Fradkin, P. R. Leavitt, and E. E. Rosenberger. 2011. Disproportionate
1281	importance of nearshore habitat for the food web of a deep oligotrophic lake. Marine and
1282	Freshwater Research 62: 350. doi:10.1071/MF10229
1283	Hampton, S. E., L. R. Izmest'Eva, M. V. Moore, S. L. Katz, B. Dennis, and E. A. Silow. 2008.
1284	Sixty years of environmental change in the world's largest freshwater lake - Lake Baikal,
1285	Siberia. Global Change Biology 14: 1947–1958. doi:10.1111/j.1365-2486.2008.01616.x
1286	Hampton, S. E., S. McGowan, T. Ozersky, and others. 2018. Recent ecological change in ancient
1287	lakes. Limnology and Oceanography 63: 2277–2304. doi:10.1002/lno.10938
1288	Hanvey, J. S., P. J. Lewis, J. L. Lavers, N. D. Crosbie, K. Pozo, and B. O. Clarke. 2017. A review
1289	of analytical techniques for quantifying microplastics in sediments. Anal. Methods 9: 1369-
1290	1383. doi:10.1039/C6AY02707E
1291	Horton, A. A., A. Walton, D. J. Spurgeon, E. Lahive, and C. Svendsen. 2017. Microplastics in
1292	freshwater and terrestrial environments: Evaluating the current understanding to identify the
1293	knowledge gaps and future research priorities. Science of The Total Environment 586: 127-
1294	141. doi:10.1016/j.scitotenv.2017.01.190

1295	Interfax-Tourism. 2018. Байкал с января по август 2018 года посетили 1,2 миллиона туристов
1296	(1.2 million tourists vistied Baikal from January through August 2018). Interfax-Tourism,
1297	October 25
1298	<u>International Standards Organization (ISO). 1984. ISO 6777:1984(en) Water quality</u>
1299	<u>Determination of nitrite</u> — <u>Molecular absorption spectrometric method. ISO 6777. ISO</u>
1300	<u>6777 ISO.</u>
1301	International Standards Organization (ISO). 2004. ISO 6878:2004(en) Water quality —
1302	<u>Determination of phosphorus — Ammonium molybdate spectrometric method. ISO 6878.</u>
1303	<u>ISO 6878 ISO.</u>
1304	Izhboldina, L. A. 2007. Guide and Key to Benthic and Periphyton Algae of Lake Baikal (meio- and
1305	macrophytes) with Brief Notes on Their Ecology, Nauka-Centre.
1306	Izmest'eva, L. R., M. V. Moore, S. E. Hampton, and others. 2016. Lake-wide physical and
1307	biological trends associated with warming in Lake Baikal. Journal of Great Lakes Research
1308	42: 6–17. doi:10.1016/j.jglr.2015.11.006
1309	Jeppesen, E., M. Søndergaard, J. P. Jensen, and others. 2005. Lake responses to reduced nutrient
1310	<u>loading – an analysis of contemporary long-term data from 35 case studies. Freshwater</u>
1311	Biology 50 : 1747–1771. doi:10.1111/j.1365-2427.2005.01415.x
1312	Karnaukhov, D., S. Biritskaya, E. Dolinskaya, M. Teplykh, N. Silenko, Y. Ermolaeva, and E.
1313	Silow. 2020. POLLUTION BY MACRO- AND MICROPLASTIC OF LARGE
1314	LACUSTRINE ECOSYSTEMS IN EASTERN ASIA. Pollution Research 2: 353–355.
1315	Kassambara, A. 2019. ggpubr: "ggplot2" Based Publication Ready Plots,.
1316	Katz, S. L., L. R. Izmest'eva, S. E. Hampton, T. Ozersky, K. Shchapov, M. V. Moore, S. V.
1317	Shimaraeva, and E. A. Silow. 2015. The "Melosira years" of Lake Baikal: Winter
1	

1318	environmental conditions at ice onset predict under-ice algal blooms in spring. Limnology
1319	and Oceanography 60: 1950–1964. doi:10.1002/lno.10143
1320	Kolpin, D. W., E. T. Furlong, M. T. Meyer, E. M. Thurman, S. D. Zaugg, L. B. Barber, and H. T.
1321	Buxton. 2002. Pharmaceuticals, Hormones, and Other Organic Wastewater Contaminants in
1322	U.S. Streams, 1999–2000: A National Reconnaissance. Environmental Science &
1323	<u>Technology</u> 36 : 1202–1211. doi:10.1021/es011055j
1324	Kozhova, O. M., and L. R. Izmest'eva. 1998. Lake Baikal: Evolution and Biodiversity, Backhuys
1325	Publishers.
1326	Kravtsova, L. S., L. A. Izhboldina, I. V. Khanaev, and others. 2014. Nearshore benthic blooms of
1327	filamentous green algae in Lake Baikal. Journal of Great Lakes Research 40: 441-448.
1328	doi:10.1016/j.jglr.2014.02.019
1329	Lapointe, B. E., L. W. Herren, D. D. Debortoli, and M. A. Vogel. 2015. Evidence of sewage-driven
1330	eutrophication and harmful algal blooms in Florida's Indian River Lagoon. Harmful Algae
1331	43: 82–102. doi:10.1016/j.hal.2015.01.004
1332	Lee, S. S., A. M. Paspalof, D. D. Snow, E. K. Richmond, E. J. Rosi-Marshall, and J. J. Kelly. 2016.
1333	Occurrence and Potential Biological Effects of Amphetamine on Stream Communities.
1334	Environmental Science & Technology 50: 9727–9735. doi:10.1021/acs.est.6b03717
1335	Lin, D., J. Crabtree, I. Dillo, and others. 2020. The TRUST Principles for digital repositories.
1336	Scientific Data 7: 144. doi:10.1038/s41597-020-0486-7
1337	Meyer, M. F., S. G. Labou, A. N. Cramer, M. R. Brousil, and B. T. Luff. 2020a. The global lake
1338	area, climate, and population dataset. Scientific Data 7: 174. doi:10.1038/s41597-020-0517-
1339	<u>4</u>

1340	Meyer, M. F., T. Ozersky, K. H. Woo, and others. 2020b. A unified dataset of co-located sewage
1341	pollution, periphyton, and benthic macroinvertebrate community and food web structure
1342	from Lake Baikal (Siberia).doi:10.6073/PASTA/76F43144015EC795679BAC508EFA044E
1343	Meyer, M. F., T. Ozersky, K. H. Woo, and others. Effects of spatially heterogeneous lakeside
1344	development on nearshore biotic communities in a large, deep, oligotrophic lake (Lake
1345	Baikal, Siberia).
1346	Meyer, M. F., S. M. Powers, and S. E. Hampton. 2019. An Evidence Synthesis of Pharmaceuticals
1347	and Personal Care Products (PPCPs) in the Environment: Imbalances among Compounds,
1348	Sewage Treatment Techniques, and Ecosystem Types. Environ. Sci. Technol. 53: 12961-
1349	12973. doi:10.1021/acs.est.9b02966
1350	Meyer, M., T. Ozersky, K. Woo, A. W. E. Galloway, M. R. Brousil, and S. Hampton. 2015. Baikal
1351	Food Webs.doi:10.17605/OSF.IO/9TA8Z
1352	Moore, J. W., D. E. Schindler, M. D. Scheuerell, D. Smith, and J. Frodge. 2003. Lake
1353	eutrophication at the urban fringe, Seattle region, USA. AMBIO: A Journal of the Human
1354	Environment 32: 13–18.
1355	Moore, M. V., S. E. Hampton, L. R. Izmest'eva, E. A. Silow, E. V. Peshkova, and B. K. Pavlov.
1356	2009. Climate Change and the World's "Sacred Sea"-Lake Baikal, Siberia. Bioscience 59:
1357	405-417. doi:10.1525/bio.2009.59.5.8
1358	O'Donnell, D. R., P. Wilburn, E. A. Silow, L. Y. Yampolsky, and E. Litchman. 2017. Nitrogen and
1359	phosphorus colimitation of phytoplankton in Lake Baikal: Insights from a spatial survey and
1360	nutrient enrichment experiments. Limnology and Oceanography 62: 1383–1392.
1361	doi:10.1002/lno.10505

1362	Pebesma, E. 2018. Simple Features for R: Standardized Support for Spatial Vector Data. The R
1363	Journal 10: 439–446. doi:10.32614/RJ-2018-009
1364	Powers, S. M., T. W. Bruulsema, T. P. Burt, and others. 2016. Long-term accumulation and
1365	transport of anthropogenic phosphorus in three river basins. Nature Geoscience 9: 353–356.
1366	doi:10.1038/ngeo2693
1367	R Core Team. 2019. R: A Language and Environment for Statistical Computing,.
1368	RD:52.24.380-2017. 2018. Nitrate concentration in waters: Photometric methods with Giress
1369	reagent following stabilization in a cadmium reducer (Массовая концентрация нитратного
1370	азота в водах: Методика измерений фотометрическим методом с реактивом Грисса
1371	после восстановления в камиевом редукторе).
1372	RD:52.24.383-2018. 2018. Working Document: Concentration of aqueous ammonium: Method for
1373	measuring with a photometer using indophenol blue (Руководящий Документ: Массовая
1374	концентрация аммонийного азота в водах: Методика измерений фотометрическим
1375	методом в виде индофенолового сингео). RD:52.24.383-2018. RD:52.24.383-2018.
1376	Richmond, E. K., M. R. Grace, J. J. Kelly, A. J. Reisinger, E. J. Rosi, and D. M. Walters. 2017.
1377	Pharmaceuticals and personal care products (PPCPs) are ecological disrupting compounds
1378	(EcoDC). Elem Sci Anth 5: 66. doi:10.1525/elementa.252
1379	Richmond, E. K., E. J. Rosi, D. M. Walters, J. Fick, S. K. Hamilton, T. Brodin, A. Sundelin, and M.
1380	R. Grace. 2018. A diverse suite of pharmaceuticals contaminates stream and riparian food
1381	webs. Nature Communications 9: 4491. doi:10.1038/s41467-018-06822-w
1382	Romera-Castillo, C., M. Pinto, T. M. Langer, X. A. Álvarez-Salgado, and G. J. Herndl. 2018.
1383	Dissolved organic carbon leaching from plastics stimulates microbial activity in the ocean.
1384	Nat Commun 9: 1–7. doi:10.1038/s41467-018-03798-5

1385	Rosenberger, E. E., S. E. Hampton, S. C. Fradkin, and B. P. Kennedy. 2008. Effects of shoreline
1386	development on the nearshore environment in large deep oligotrophic lakes. Freshwater
1387	Biology 53 : 1673–1691. doi:10.1111/j.1365-2427.2008.01990.x
1388	Rosi-Marshall, E. J., D. W. Kincaid, H. A. Bechtold, T. V. Royer, M. Rojas, and J. J. Kelly. 2013.
1389	Pharmaceuticals suppress algal growth and microbial respiration and alter bacterial
1390	communities in stream biofilms. Ecological Applications 23: 583–593. doi:10.1890/12-
1391	<u>0491.1</u>
1392	Rosi-Marshall, E. J., and T. V. Royer. 2012. Pharmaceutical Compounds and Ecosystem Function:
1393	An Emerging Research Challenge for Aquatic Ecologists. Ecosystems 15: 867–880.
1394	doi:10.1007/s10021-012-9553-z
1395	Sargent, J. R., and S. Falk-Petersen. 1988. The lipid biochemistry of calanoid copepods.
1396	Hydrobiologia 167–168: 101–114. doi:10.1007/BF00026297
1397	Schram, J. B., J. N. Kobelt, M. N. Dethier, and A. W. E. Galloway. 2018. Trophic Transfer of
1398	Macroalgal Fatty Acids in Two Urchin Species: Digestion, Egestion, and Tissue Building.
1399	Front. Ecol. Evol. 6. doi:10.3389/fevo.2018.00083
1400	Shaw, L., C. Phung, and M. Grace. 2015. Pharmaceuticals and personal care products alter growth
1401	and function in lentic biofilms. Environmental Chemistry 12: 301. doi:10.1071/EN14141
1402	Slowikowski, K. 2019. ggrepel: Automatically Position Non-Overlapping Text Labels with
1403	<u>"ggplot2,."</u>
1404	Swann, G. E. A., V. N. Panizzo, S. Piccolroaz, and others. 2020. Changing nutrient cycling in Lake
1405	Baikal, the world's oldest lake. PNAS 117: 27211–27217. doi:10.1073/pnas.2013181117
1406	Taipale, S., U. Strandberg, E. Peltomaa, A. W. E. Galloway, A. Ojala, and M. T. Brett. 2013. Fatty
1407	acid composition as biomarkers of freshwater microalgae: analysis of 37 strains of
1	

1408	microalgae in 22 genera and in seven classes. Aquatic Microbial Ecology 71: 165–178.
1409	doi:10.3354/ame01671
1410	Timoshkin, O. A., M. V. Moore, N. N. Kulikova, and others. 2018. Groundwater contamination by
1411	sewage causes benthic algal outbreaks in the littoral zone of Lake Baikal (East Siberia).
1412	Journal of Great Lakes Research. doi:10.1016/j.jglr.2018.01.008
1413	Timoshkin, O. A., D. P. Samsonov, M. Yamamuro, and others. 2016. Rapid ecological change in
1414	the coastal zone of Lake Baikal (East Siberia): Is the site of the world's greatest freshwater
1415	biodiversity in danger? Journal of Great Lakes Research 42: 487–497.
1416	doi:10.1016/j.jglr.2016.02.011
1417	Tong, Y., M. Wang, J. Peñuelas, and others. 2020. Improvement in municipal wastewater treatment
1418	alters lake nitrogen to phosphorus ratios in populated regions. Proc Natl Acad Sci USA 117:
1419	11566–11572. doi:10.1073/pnas.1920759117
1420	Turetsky, M. R., R. K. Wieder, C. J. Williams, and D. H. Vitt. 2000. Organic matter accumulation,
1421	peat chemistry, and permafrost melting in peatlands of boreal Alberta. Écoscience 7: 115-
1422	122. doi:10.1080/11956860.2000.11682608
1423	Vendel, A. L., F. Bessa, V. E. N. Alves, A. L. A. Amorim, J. Patrício, and A. R. T. Palma. 2017.
1424	Widespread microplastic ingestion by fish assemblages in tropical estuaries subjected to
1425	anthropogenic pressures. Marine Pollution Bulletin 117: 448–455.
1426	doi:10.1016/j.marpolbul.2017.01.081
1427	Volkova, E. A., N. A. Bondarenko, and O. A. Timoshkin. 2018. Morphotaxonomy, distribution and
1428	abundance of Spirogyra (Zygnematophyceae, Charophyta) in Lake Baikal, East Siberia.
1429	Phycologia 57: 298–308. doi:10.2216/17-69.1

1430	Wang, W., and J. Wang. 2018. Investigation of microplastics in aquatic environments: An overview
1431	of the methods used, from field sampling to laboratory analysis. TrAC Trends in Analytical
1432	Chemistry 108: 195–202. doi:10.1016/j.trac.2018.08.026
1433	Welschmeyer, N. A. 1994. Fluorometric analysis of chlorophyll a in the presence of chlorophyll b
1434	and pheopigments. Limnol. Oceanogr. 39: 1985–1992. doi:10.4319/lo.1994.39.8.1985
1435	Wickham, H. 2014. Tidy Data. Journal of Statistical Software 59 : 1–23. doi:10.18637/jss.v059.i10
1436	Wickham, H., M. Averick, J. Bryan, and others. 2019. Welcome to the tidyverse. Journal of Open
1437	Source Software 4: 1686. doi:10.21105/joss.01686
1438	Wilke, C. O. 2019. cowplot: Streamlined Plot Theme and Plot Annotations for "ggplot2,."
1439	Wilkinson, M. D., M. Dumontier, Ij. J. Aalbersberg, and others. 2016. The FAIR Guiding
1440	Principles for scientific data management and stewardship. Sci Data 3.
1441	doi:10.1038/sdata.2016.18
1442	Yang, Y., W. Song, H. Lin, W. Wang, L. Du, and W. Xing. 2018. Antibiotics and antibiotic
1443	resistance genes in global lakes: A review and meta-analysis. Environment International
1444	116: 60–73. doi:10.1016/j.envint.2018.04.011
1445	Yoshida, T., T. Sekino, M. Genkai-Kato, and others. 2003. Seasonal dynamics of primary
1446	production in the pelagic zone of southern Lake Baikal. Limnology 4: 53–62.
1447	doi:10.1007/s10201-002-0089-3
1448	Anisimov, O., and S. Reneva. 2006. Permafrost and Changing Climate: The Russian Perspective.
1449	Ambio 35: 169–175.
1450	Barnes, D. K. A., F. Galgani, R. C. Thompson, and M. Barlaz. 2009. Accumulation and
1451	fragmentation of plastic debris in global environments. Philos Trans R Soc Lond B Biol Sci
1452	364: 1985–1998. doi:10.1098/rstb.2008.0205

1453	Bendz, D., N. A. Paxéus, T. R. Ginn, and F. J. Loge. 2005. Occurrence and fate of pharmaceutically
1454	active compounds in the environment, a case study: Höje River in Sweden. Journal of
1455	Hazardous Materials 122: 195–204. doi:10.1016/j.jhazmat.2005.03.012
1456	Bondarenko, N. A., I. V. Tomberg, A. A. Shirokaya, and others. 2021. Dolichospermum
1457	lemmermannii (Nostocales) bloom in world's deepest Lake Baikal (East Siberia):
1458	abundance, toxicity and factors influencing growth. Limnology and Freshwater Biology 1:
1459	1101–1110. doi:10.31951/2658-3518-2021-A-1-1101
1460	Brandon, J. A., A. Freibott, and L. M. Sala. 2020. Patterns of suspended and salp-ingested
1461	microplastic debris in the North Pacific investigated with epifluorescence microscopy.
1462	Limnology and Oceanography Letters 5: 46–53. doi:10.1002/lol2.10127
1463	Brodin, T., J. Fick, M. Jonsson, and J. Klaminder. 2013. Dilute Concentrations of a Psychiatric
1464	Drug Alter Behavior of Fish from Natural Populations. Science 339: 814-815.
1465	doi:10.1126/science.1226850
1465 1466	doi:10.1126/science.1226850 Camilleri, A. C., and T. Ozersky. 2019. Large variation in periphyton δ13C and δ15N values in the
1466	Camilleri, A. C., and T. Ozersky. 2019. Large variation in periphyton δ13C and δ15N values in the
1466 1467	Camilleri, A. C., and T. Ozersky. 2019. Large variation in periphyton δ13C and δ15N values in the upper Great Lakes: Correlates and implications. Journal of Great Lakes Research 45: 986
1466 1467 1468	Camilleri, A. C., and T. Ozersky. 2019. Large variation in periphyton δ13C and δ15N values in the upper Great Lakes: Correlates and implications. Journal of Great Lakes Research 45: 986–990. doi:10.1016/j.jglr.2019.06.003
1466 1467 1468 1469	Camilleri, A. C., and T. Ozersky. 2019. Large variation in periphyton δ13C and δ15N values in the upper Great Lakes: Correlates and implications. Journal of Great Lakes Research 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
1466 1467 1468 1469 1470	Camilleri, A. C., and T. Ozersky. 2019. Large variation in periphyton δ13C and δ15N values in the upper Great Lakes: Correlates and implications. Journal of Great Lakes Research 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. Marine Pollution Bulletin 42:
1466 1467 1468 1469 1470	Camilleri, A. C., and T. Ozersky. 2019. Large variation in periphyton δ13C and δ15N values in the upper Great Lakes: Correlates and implications. Journal of Great Lakes Research 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. Marine Pollution Bulletin 42: 149–156. doi:10.1016/S0025-326X(00)00125-9
1466 1467 1468 1469 1470 1471	Camilleri, A. C., and T. Ozersky. 2019. Large variation in periphyton 813C and 815N values in the upper Great Lakes: Correlates and implications. Journal of Great Lakes Research 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. Marine Pollution Bulletin 42: 149–156. doi:10.1016/S0025-326X(00)00125-9 D'Alessio, M., S. Onanong, D. D. Snow, and C. Ray. 2018. Occurrence and removal of
1466 1467 1468 1469 1470 1471 1472	Camilleri, A. C., and T. Ozersky. 2019. Large variation in periphyton 813C and 815N values in the upper Great Lakes: Correlates and implications. Journal of Great Lakes Research 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. Marine Pollution Bulletin 42: 149–156. doi: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

1476	Dalsgaard, J., M. St. John, G. Kattner, D. Müller-Navarra, and W. Hagen. 2003. Fatty acid trophic
1477	markers in the pelagic marine environment, p. 225–340. <i>In</i> Advances in Marine Biology.
1478	Elsevier.
1479	Edmondson, W. T. 1970. Phosphorus, Nitrogen, and Algae in Lake Washington after Diversion of
1480	Sewage. Science 169: 690–691.
1481	Fellows, I., and using the Jm. library by J. P. Stotz. 2019. OpenStreetMap: Access to Open Street
1482	Map Raster Images,.
1483	Focazio, M. J., D. W. Kolpin, K. K. Barnes, E. T. Furlong, M. T. Meyer, S. D. Zaugg, L. B. Barber
1484	and M. E. Thurman. 2008. A national reconnaissance for pharmaceuticals and other organic
1485	wastewater contaminants in the United States - II) Untreated drinking water sources.
1486	SCIENCE OF THE TOTAL ENVIRONMENT 402: 201–216.
1487	doi:10.1016/j.scitotenv.2008.02.021
1488	Gartner, A., P. Lavery, and A. J. Smit. 2002. Use of delta N-15 signatures of different functional
1489	forms of macroalgae and filter-feeders to reveal temporal and spatial patterns in sewage
1490	dispersal. Mar. EcolProg. Ser. 235: 63-73. doi:10.3354/meps235063
1491	GOST:18309-2014. 2016b. Methods for determination of phosphorus containing matters (with
1492	corrections) (Методы определения фосфорсодержащих веществ).
1493	GOST:33045-2014. 2016a. Methods for determination of nitrogen-containing matters (with
1494	corrections) (Методы определения азотсодержащих веществ (с Поправками)).
1495	GOST:33045-2014. GOST:33045-2014 Intergovernmental committe for standardization,
1496	regulation, and metrology.

1497	Green, D. S. 2016. Effects of microplastics on European flat oysters, Ostrea edulis and their					
1498	associated benthic communities. Environmental Pollution 216: 95–103.					
1499	doi:10.1016/j.envpol.2016.05.043					
1500	Hall, R. I., P. R. Leavitt, R. Quinlan, A. S. Dixit, and J. P. Smol. 1999. Effects of agriculture,					
1501	urbanization, and climate on water quality in the northern Great Plains. Limnology and					
1502	Oceanography 44: 739-756. doi:10.4319/lo.1999.44.3_part_2.0739					
1503	Hampton, S. E., S. C. Fradkin, P. R. Leavitt, and E. E. Rosenberger. 2011. Disproportionate					
1504	importance of nearshore habitat for the food web of a deep oligotrophic lake. Marine and					
1505	Freshwater Research 62: 350. doi:10.1071/MF10229					
1506	Hampton, S. E., L. R. Izmest'Eva, M. V. Moore, S. L. Katz, B. Dennis, and E. A. Silow. 2008.					
1507	Sixty years of environmental change in the world's largest freshwater lake - Lake Baikal,					
1508	Siberia. Global Change Biology 14: 1947–1958. doi:10.1111/j.1365-2486.2008.01616.x					
1509	Hampton, S. E., S. McGowan, T. Ozersky, and others. 2018. Recent ecological change in ancient					
1510	lakes. Limnology and Oceanography 63: 2277 2304. doi:10.1002/lno.10938					
1511	Hanvey, J. S., P. J. Lewis, J. L. Lavers, N. D. Crosbie, K. Pozo, and B. O. Clarke. 2017. A review					
1512	of analytical techniques for quantifying microplastics in sediments. Anal. Methods 9: 1369					
1513	1383. doi:10.1039/C6AY02707E					
1514	Horton, A. A., A. Walton, D. J. Spurgeon, E. Lahive, and C. Svendsen. 2017. Microplastics in					
1515	freshwater and terrestrial environments: Evaluating the current understanding to identify the					
1516	knowledge gaps and future research priorities. Science of The Total Environment 586: 127-					
1517	141. doi:10.1016/j.scitotenv.2017.01.190					

1518	Interfax-Tourism. 2018. Байкал с января по август 2018 года посетили 1,2 миллиона туристов
1519	(1.2 million tourists vistied Baikal from January through August 2018). Interfax-Tourism,
1520	October 25
1521	International Standards Organization (ISO). 1984. ISO 6777:1984(en) Water quality—
1522	Determination of nitrite — Molecular absorption spectrometric method. ISO 6777. ISO
1523	6777 ISO.
1524	International Standards Organization (ISO). 2004. ISO 6878:2004(en) Water quality—
1525	Determination of phosphorus — Ammonium molybdate spectrometric method. ISO 6878.
1526	ISO 6878 ISO.
1527	Izhboldina, L. A. 2007. Guide and Key to Benthic and Periphyton Algae of Lake Baikal (meio- and
1528	macrophytes) with Brief Notes on Their Ecology, Nauka-Centre.
1529	Izmest'eva, L. R., M. V. Moore, S. E. Hampton, and others. 2016. Lake-wide physical and
1530	biological trends associated with warming in Lake Baikal. Journal of Great Lakes Research
1531	42 : 6 17. doi:10.1016/j.jglr.2015.11.006
1532	Jeppesen, E., M. Søndergaard, J. P. Jensen, and others. 2005. Lake responses to reduced nutrient
1533	loading an analysis of contemporary long-term data from 35 case studies. Freshwater
1534	Biology 50 : 1747–1771. doi:10.1111/j.1365-2427.2005.01415.x
1535	Karnaukhov, D., S. Biritskaya, E. Dolinskaya, M. Teplykh, N. Silenko, Y. Ermolaeva, and E.
1536	Silow. 2020. POLLUTION BY MACRO- AND MICROPLASTIC OF LARGE
1537	LACUSTRINE ECOSYSTEMS IN EASTERN ASIA. Pollution Research 2: 353–355.
1538	Kassambara, A. 2019. ggpubr: "ggplot2" Based Publication Ready Plots,.
1539	Katz, S. L., L. R. Izmest'eva, S. E. Hampton, T. Ozersky, K. Shchapov, M. V. Moore, S. V.
1540	Shimaraeva, and E. A. Silow. 2015. The "Melosira years" of Lake Baikal: Winter
1	

1541	environmental conditions at ice onset predict under-ice algal blooms in spring. Limnology
1542	and Oceanography 60: 1950–1964. doi:10.1002/lno.10143
1543	Kolpin, D. W., E. T. Furlong, M. T. Meyer, E. M. Thurman, S. D. Zaugg, L. B. Barber, and H. T.
1544	Buxton. 2002. Pharmaceuticals, Hormones, and Other Organic Wastewater Contaminants in
1545	U.S. Streams, 1999–2000: A National Reconnaissance. Environmental Science &
1546	Technology 36: 1202–1211. doi:10.1021/es011055j
1547	Kozhova, O. M., and L. R. Izmest'eva. 1998. Lake Baikal: Evolution and Biodiversity, Backhuys
1548	Publishers.
1549	Kravtsova, L. S., L. A. Izhboldina, I. V. Khanaev, and others. 2014. Nearshore benthic blooms of
1550	filamentous green algae in Lake Baikal. Journal of Great Lakes Research 40: 441-448.
1551	doi:10.1016/j.jglr.2014.02.019
1552	Lapointe, B. E., L. W. Herren, D. D. Debortoli, and M. A. Vogel. 2015. Evidence of sewage-driven
1553	eutrophication and harmful algal blooms in Florida's Indian River Lagoon. Harmful Algae
1554	43: 82 102. doi:10.1016/j.hal.2015.01.004
1555	Lee, S. S., A. M. Paspalof, D. D. Snow, E. K. Richmond, E. J. Rosi-Marshall, and J. J. Kelly. 2016.
1556	Occurrence and Potential Biological Effects of Amphetamine on Stream Communities.
1557	Environmental Science & Technology 50: 9727-9735. doi:10.1021/acs.est.6b03717
1558	Lin, D., J. Crabtree, I. Dillo, and others. 2020. The TRUST Principles for digital repositories.
1559	Scientific Data 7: 144. doi:10.1038/s41597-020-0486-7
1560	Meyer, M. F., S. G. Labou, A. N. Cramer, M. R. Brousil, and B. T. Luff. 2020a. The global lake
1561	area, climate, and population dataset. Scientific Data 7: 174. doi:10.1038/s41597-020-0517-
1562	4

1563	Meyer, M. F., T. Ozersky, K. H. Woo, and others. 2020b. A unified dataset of co-located sewage
1564	pollution, periphyton, and benthic macroinvertebrate community and food web structure
1565	from Lake Baikal (Siberia).doi:10.6073/PASTA/76F43144015EC795679BAC508EFA044B
1566	Meyer, M. F., T. Ozersky, K. H. Woo, and others. Effects of spatially heterogeneous lakeside
1567	development on nearshore biotic communities in a large, deep, oligotrophic lake (Lake
1568	Baikal, Siberia).
1569	Meyer, M. F., S. M. Powers, and S. E. Hampton. 2019. An Evidence Synthesis of Pharmaceuticals
1570	and Personal Care Products (PPCPs) in the Environment: Imbalances among Compounds,
1571	Sewage Treatment Techniques, and Ecosystem Types. Environ. Sci. Technol. 53: 12961
1572	12973. doi:10.1021/acs.est.9b02966
1573	Meyer, M., T. Ozersky, K. Woo, A. W. E. Galloway, M. R. Brousil, and S. Hampton. 2015. Baikal
1574	Food Webs.doi:10.17605/OSF.IO/9TA8Z
1575	Moore, J. W., D. E. Schindler, M. D. Scheuerell, D. Smith, and J. Frodge. 2003. Lake
1576	eutrophication at the urban fringe, Seattle region, USA. AMBIO: A Journal of the Human
1577	Environment 32: 13–18.
1578	Moore, M. V., S. E. Hampton, L. R. Izmest'eva, E. A. Silow, E. V. Peshkova, and B. K. Pavlov.
1579	2009. Climate Change and the World's "Sacred Sea" Lake Baikal, Siberia. Bioscience 59:
1580	405-417. doi:10.1525/bio.2009.59.5.8
1581	O'Donnell, D. R., P. Wilburn, E. A. Silow, L. Y. Yampolsky, and E. Litchman. 2017. Nitrogen and
1582	phosphorus colimitation of phytoplankton in Lake Baikal: Insights from a spatial survey and
1583	nutrient enrichment experiments. Limnology and Oceanography 62: 1383–1392.
1584	doi:10.1002/lno.10505

1585	Pebesma, E. 2018. Simple Features for R: Standardized Support for Spatial Vector Data. The R
1586	Journal 10: 439-446. doi:10.32614/RJ-2018-009
1587	Powers, S. M., T. W. Bruulsema, T. P. Burt, and others. 2016. Long-term accumulation and
1588	transport of anthropogenic phosphorus in three river basins. Nature Geoscience 9: 353–356.
1589	doi:10.1038/ngeo2693
1590	R Core Team. 2019. R: A Language and Environment for Statistical Computing,.
1591	RD:52.24.380-2017. 2018. Nitrate concentration in waters: Photometric methods with Giress
1592	reagent following stabilization in a cadmium reducer (Массовая концентрация нитратного
1593	азота в водах: Методика измерений фотометрическим методом с реактивом Грисса
1594	после восстановления в камиевом редукторе).
1595	Richmond, E. K., M. R. Grace, J. J. Kelly, A. J. Reisinger, E. J. Rosi, and D. M. Walters. 2017.
1596	Pharmaceuticals and personal care products (PPCPs) are ecological disrupting compounds
1597	(EcoDC). Elem Sci Anth 5: 66. doi:10.1525/elementa.252
1598	Richmond, E. K., E. J. Rosi, D. M. Walters, J. Fick, S. K. Hamilton, T. Brodin, A. Sundelin, and M.
1599	R. Grace. 2018. A diverse suite of pharmaceuticals contaminates stream and riparian food
1600	webs. Nature Communications 9: 4491. doi:10.1038/s41467-018-06822-w
1601	Romera-Castillo, C., M. Pinto, T. M. Langer, X. A. Álvarez-Salgado, and G. J. Herndl. 2018.
1602	Dissolved organic carbon leaching from plastics stimulates microbial activity in the ocean.
1603	Nat Commun 9: 1 7. doi:10.1038/s41467-018-03798-5
1604	Rosenberger, E. E., S. E. Hampton, S. C. Fradkin, and B. P. Kennedy. 2008. Effects of shoreline
1605	development on the nearshore environment in large deep oligotrophic lakes. Freshwater
1606	Biology 53 : 1673–1691. doi:10.1111/j.1365-2427.2008.01990.x
1	

1607	Rosi-Marshall, E. J., D. W. Kincaid, H. A. Bechtold, T. V. Royer, M. Rojas, and J. J. Kelly. 2013.
1608	Pharmaceuticals suppress algal growth and microbial respiration and alter bacterial
1609	communities in stream biofilms. Ecological Applications 23: 583–593. doi:10.1890/12-
1610	0491.1
1611	Rosi-Marshall, E. J., and T. V. Royer. 2012. Pharmaceutical Compounds and Ecosystem Function:
1612	An Emerging Research Challenge for Aquatic Ecologists. Ecosystems 15: 867–880.
1613	doi:10.1007/s10021-012-9553-z
1614	Sargent, J. R., and S. Falk-Petersen. 1988. The lipid biochemistry of calanoid copepods.
1615	Hydrobiologia 167–168: 101–114. doi:10.1007/BF00026297
1616	Schram, J. B., J. N. Kobelt, M. N. Dethier, and A. W. E. Galloway. 2018. Trophic Transfer of
1617	Macroalgal Fatty Acids in Two Urchin Species: Digestion, Egestion, and Tissue Building.
1618	Front. Ecol. Evol. 6. doi:10.3389/fevo.2018.00083
1619	Shaw, L., C. Phung, and M. Grace. 2015. Pharmaceuticals and personal care products alter growth
1620	and function in lentic biofilms. Environmental Chemistry 12: 301. doi:10.1071/EN14141
1621	Slowikowski, K. 2019. ggrepel: Automatically Position Non-Overlapping Text Labels with
1622	"ggplot2,."
1623	Swann, G. E. A., V. N. Panizzo, S. Piccolroaz, and others. 2020. Changing nutrient cycling in Lake
1624	Baikal, the world's oldest lake. PNAS 117: 27211-27217. doi:10.1073/pnas.2013181117
1625	Taipale, S., U. Strandberg, E. Peltomaa, A. W. E. Galloway, A. Ojala, and M. T. Brett. 2013. Fatty
1626	acid composition as biomarkers of freshwater microalgae: analysis of 37 strains of
1627	microalgae in 22 genera and in seven classes. Aquatic Microbial Ecology 71: 165–178.
1628	doi:10.3354/ame01671

1629	Timoshkin, O. A., M. V. Moore, N. N. Kulikova, and others. 2018. Groundwater contamination by
1630	sewage causes benthic algal outbreaks in the littoral zone of Lake Baikal (East Siberia).
1631	Journal of Great Lakes Research. doi:10.1016/j.jglr.2018.01.008
1632	Timoshkin, O. A., D. P. Samsonov, M. Yamamuro, and others. 2016. Rapid ecological change in
1633	the coastal zone of Lake Baikal (East Siberia): Is the site of the world's greatest freshwater
1634	biodiversity in danger? Journal of Great Lakes Research 42: 487–497.
1635	doi:10.1016/j.jglr.2016.02.011
1636	Tong, Y., M. Wang, J. Peñuelas, and others. 2020. Improvement in municipal wastewater treatment
1637	alters lake nitrogen to phosphorus ratios in populated regions. Proc Natl Acad Sci USA 117:
1638	11566–11572. doi:10.1073/pnas.1920759117
1639	Turetsky, M. R., R. K. Wieder, C. J. Williams, and D. H. Vitt. 2000. Organic matter accumulation,
1640	peat chemistry, and permafrost melting in peatlands of boreal Alberta. Écoscience 7: 115-
1641	122. doi:10.1080/11956860.2000.11682608
1642	Vendel, A. L., F. Bessa, V. E. N. Alves, A. L. A. Amorim, J. Patrício, and A. R. T. Palma. 2017.
1643	Widespread microplastic ingestion by fish assemblages in tropical estuaries subjected to
1644	anthropogenic pressures. Marine Pollution Bulletin 117: 448-455.
1645	doi:10.1016/j.marpolbul.2017.01.081
1646	Volkova, E. A., N. A. Bondarenko, and O. A. Timoshkin. 2018. Morphotaxonomy, distribution and
1647	abundance of Spirogyra (Zygnematophyceae, Charophyta) in Lake Baikal, East Siberia.
1648	Phycologia 57: 298–308. doi:10.2216/17-69.1
1649	Wang, W., and J. Wang. 2018. Investigation of microplastics in aquatic environments: An overview
1650	of the methods used, from field sampling to laboratory analysis. TrAC Trends in Analytical
1651	Chemistry 108: 195–202. doi:10.1016/j.trac.2018.08.026

1652	Welschmeyer, N. A. 1994. Fluorometric analysis of chlorophyll a in the presence of chlorophyll b
1653	and pheopigments. Limnol. Oceanogr. 39: 1985–1992. doi:10.4319/lo.1994.39.8.1985
1654	Wickham, H. 2014. Tidy Data. Journal of Statistical Software 59: 1–23. doi:10.18637/jss.v059.i10
1655	Wickham, H., M. Averick, J. Bryan, and others. 2019. Welcome to the tidyverse. Journal of Open
1656	Source Software 4: 1686. doi:10.21105/joss.01686
1657	Wilke, C. O. 2019. cowplot: Streamlined Plot Theme and Plot Annotations for "ggplot2,."
1658	Wilkinson, M. D., M. Dumontier, Ij. J. Aalbersberg, and others. 2016. The FAIR Guiding
1659	Principles for scientific data management and stewardship. Sci Data 3.
1660	doi:10.1038/sdata.2016.18
1661	Yang, Y., W. Song, H. Lin, W. Wang, L. Du, and W. Xing. 2018. Antibiotics and antibiotic
1662	resistance genes in global lakes: A review and meta-analysis. Environment International
1663	116: 60-73. doi:10.1016/j.envint.2018.04.011
1664	Yoshida, T., T. Sekino, M. Genkai-Kato, and others. 2003. Seasonal dynamics of primary
1665	production in the pelagic zone of southern Lake Baikal. Limnology 4: 53–62.
1666	doi:10.1007/s10201-002-0089-3
1667	doi:10.1007/s10201-002-0089-3
1668 1669	

Acknowledgements

1670 1671

1672 We would like to thank the faculty, students, staff, and mariners of the Irkutsk State University's 1673 Biological Research Institute Biostation for their expert field, taxonomic, and laboratory support; 1674 Marianne Moore and Bart De Stasio for helpful advice; the researchers and students of the Siberian 1675 Branch of the Russian Academy of Sciences Limnological Institute for expert taxonomic and 1676 logistical assistance; Oleg A. Timoshkin, Tatiana Ya. Sitnikova, Irina V. Mekhanikova, Nina A. 1677 Bondorenko, Ekaterina Volkova, and Vadim V. Takhteev for offering insights and taxonomic training throughout the development of this project. Funding was provided by the National Science 1678 1679 Foundation (NSF-DEB-1136637) to S.E.H., a Fulbright Fellowship to M.F.M., a NSF Graduate 1680

Research Fellowship to M.F.M. (NSF-DGE-1347973), and Russian Ministry of Science and

ZE-2. ral and N 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

1683 University.

1684

1681 1682



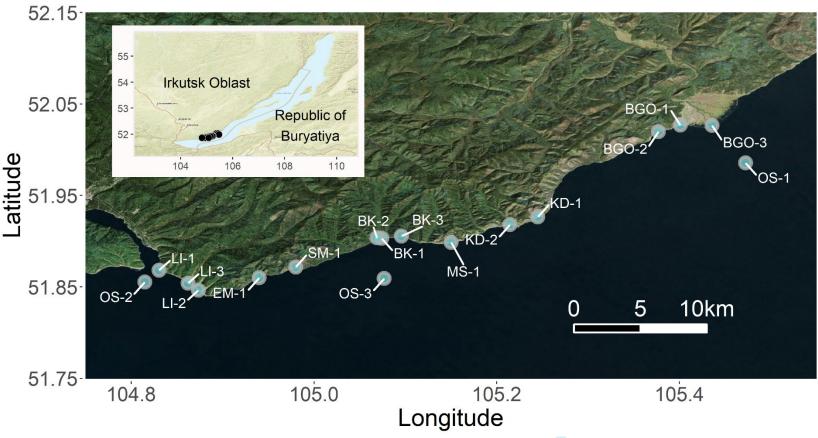


Figure 1: Map of all sampling locations with sites labeled with unique alphanumeric code. The entire transect included three developed sites (i.e., Listvyanka (LI), Bolshie Koty (BK), Bolshoe Goloustnoe (BGO)). Three offshore sites (OS) were also sampled to compare pelagic sewage signals to those in 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), and ggrepel (Slowikowski 2019) packages.

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	
		Brandtia latissima <u>subspp. intermida</u> (Dorogostaiskii 1930 <u>; Dybowsky 1874</u>)	
		Brandtia latissima lata (Dybowsky 1874)	
		Brandtia latissima latior (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)	
	Amphipoda	Eulimnogammarus marituji (Bazikalova 1945) Eulimnogammarus verucossus (Gerstfeldt 1858) Eulimnogammarus viridis viridis (Dybowsky 1874)	
bundance Estimates			
Abundance Estimates		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)	
		Acroloxidae	
		Baicaliidae	
	Molluska	Benedictidate	
		Maackia	
		Planorbidae	

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

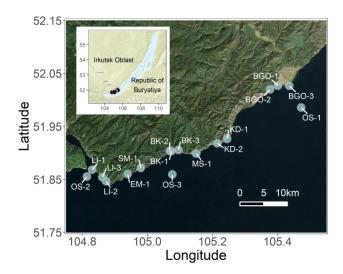


Figure 1: Map of all sampling locations with sites labeled with unique alphanumeric code. The entire transect included three developed sites (i.e., Listvyanka (LI), Bolshie Koty (BK), Bolshoe Goloustnoe (BGO)). Three offshore sites (OS) were also sampled to compare pelagic sewage signals to those in 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
	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)
Abundance Estimates		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 Managein and hunter	Asellidae
		Caddisflies
	Other Macroinvertebrates	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.		
Draparnaldia spp.				