

Depth-resolved partitioning of phytoplankton and bacterial communities in a seasonally stratified large lake

Katelyn M. Brown,¹ Mark J. Rozmarynowycz,² Tijana Glavina Del Rio,³ Lars G. Rudstam,⁴ R. Michael McKay^{1,2}

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT Prokaryotic community composition was determined through V4 16S rRNA gene sequencing during early summer stratification in Lake Ontario. Despite a distinct thermocline, bacterial communities were stable, although differences were observed in chloroplast reads attributed to eukaryotic phototrophs. Chloroplast reads contributed up to 33.0% (station 63 epilimnion) of total reads.

KEYWORDS Lake Ontario, stratification, epilimnion, hypolimnion

Warm monomictic lakes, like the Laurentian Great Lakes, alternate seasonally between isothermal mixing and thermal stratification (1). When mixed, phytoplankton occur throughout the water column, compared to stratified periods where chlorophyll *a* can be elevated below the thermocline, known as the deep chlorophyll layer (DCL) (2, 3). Since phytoplankton communities and nutrient availability differ between the surface and DCL, bacterial communities may also show patterns of partitioning. We applied 16S rRNA gene amplicon sequencing, metagenomics, and metatranscriptomics to investigate microbes in Lake Ontario's surface waters and DCL.

Depth-integrated water samples were collected from the surface and DCL at five U.S. Environmental Protection Agency (EPA) stations onboard R/V *Lake Guardian* during the 2013 Lake Ontario Cooperative Science and Monitoring Initiative (Fig. 1A, Table 1) (4). Biomass was concentrated on 0.22 µm Sterivex filters (Sigma Aldrich, St. Louis, MO) and frozen in liquid nitrogen. DNA was extracted with the PowerWater Sterivex DNA Isolation Kit (MO BIO Laboratories, Inc., Carlsbad, CA). 16S rRNA gene V4 region sequencing was completed at the Joint Genome Institute (JGI; Berkeley, CA) on an Illumina MiSeq (2 × 250 bp; Illumina, San Diego, CA) (5, 6). Primers 515F (5'-CACGGTCGKCGGCCATT-3') and a modified 806R (5, 6) were used with a HotMasterMix Amplification Kit (5PRIME, Montreal, QC). Sequences were processed with the dada2 v1.8 workflow (v1.36) (7), in R (v.4.5.1) (8), and RStudio (v.2025.05.1 + 513) (9). Chloroplast and mitochondrial reads were removed. Modifications to commands were filterAndTrim(trimLeft=c(21, 21), truncLen=c(240,240)), and mergers (minOverlap=20). Additional packages included phyloseq (v.1.52) (10), RColorBrewer (v.1.1-3) (11), ggh4x (v.0.3.1) (12), and tidyverse (v.2.0.0) (13). The map was created with sf (v.1.0-21) (14, 15), ggspatial (v.1.1.10) (16), ggplot2 (v.3.5.2) (17), and stringr (v1.5.1) (18).

Samples collected from the epilimnion from station ON 18 (May 2013) were filtered on a Sterivex and frozen in liquid nitrogen. DNA and RNA were extracted following references 19 and 20 with a DNeasy Tissue and RNeasy Mini Kit (Qiagen, Germantown, MD). DNA libraries were prepared with the KAPA-Illumina Library Creation Kit (Roche, Basel, Switzerland), and RNA libraries were rRNA depleted (Ribo-Zero rRNA Removal Kit; Illumina) and prepared with a TruSeq Stranded mRNA Kit (Illumina). Libraries were sequenced at JGI on an Illumina HiSeq 2500 (2 × 150 bp). Metagenomic assembly produced 2,170,353 contigs with an *N*₅₀ of 492,865. Reads were error-corrected (bfc v.181; kmer=21) (21) and assembled (SPAdes v.3.10.0-dev; --meta --only-assembler -k

Editor Irene L. G. Newton, Indiana University, Bloomington, Indiana, USA

Address correspondence to R. Michael McKay, Robert.Mckay@uwindsor.ca.

The authors declare no conflict of interest.

See the funding table on p. 4.

Editor's Note: A potential conflict of interest was identified during the review of this paper, and the Editor in Chief performed an additional final review of the paper.

Received 19 December 2025

Accepted 3 February 2026

Published 27 April 2026

Copyright © 2026 Brown et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

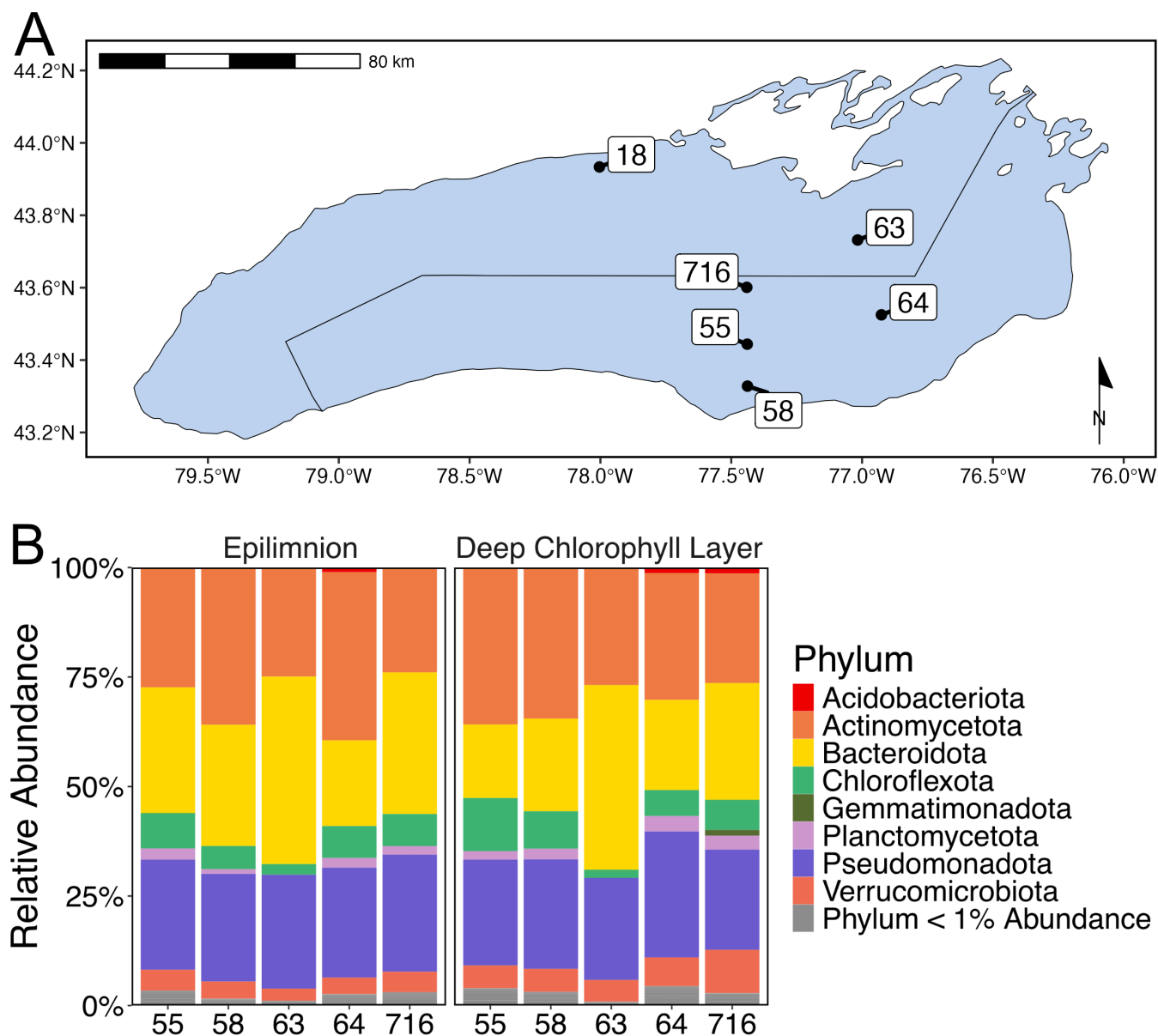


FIG 1 (A) Map of Lake Ontario with sampling stations overlaid. Five sites were targeted with amplicon sequencing, and one (station ON 18) was targeted with metagenomics and metatranscriptomics. (B) Relative abundance of prokaryotes at the phylum level for five sites with amplicon sequencing grouped by water column layer (epilimnion or deep chlorophyll layer). Taxa less than 1% abundant were aggregated at the phylum level.

21,33,55,77,99,127) (22), and filtered reads were mapped to the assembly (bwa v.0.7.15-r1142-dirty) (23). Contaminant contigs were removed (BBMap v.35.85). Default parameters were used unless otherwise noted. Metatranscriptomic reads were trimmed (ktrim=r, k=25, mink=12, tpe=t, tbo=t, qtrim=r, trimq=10, maq=10, maxns=3, minlen=50) and filtered (k=16) (BBduk v.33.95). Contaminant rRNA reads and contigs were removed by mapping to the Silva database (v.119; BBMap v.33.95; fast=t, minid=0.90, local=t) and JGI contaminants database. Reads were concatenated, merged (BBMerge v.33.95), and normalized (BBNorm v.33.95) before assembly (Rnnotator v3.0.0) (24). The metatranscriptomes contained 206,160 and 244,117 contigs with N_{50} s of 41,576 and 49,490. Bacterial phyla were consistent between the epilimnion and DCL (Fig. 1B), indicating a strong core microbiome in Lake Ontario.

TABLE 1 Location and summary of reads processing for sequenced samples^a

Station ID— layer	Latitude	Longitude	Integrated sampling depths (m)	No. of raw reads	No. of reads after initial QC	No. of reads after removing chloroplast	No. of reads after removing chloroplast/ mitochondria	Library type	Accession no.
ON 18— epilimnion	43.933506	−78.003845	2, 5	141,179,563	N/A	N/A	N/A	Metagenomic	SRR3990180
ON 18— epilimnion	43.933506	−78.003845	2, 5	116,900,025	N/A	N/A	N/A	Metatranscriptomic	SRR5249012
ON 18— epilimnion	43.933506	−78.003845	2, 5	105,151,916	N/A	N/A	N/A	Metatranscriptomic	SRR5249013
ON 55— epilimnion	43.4439	−77.4389	2, 10	91,956	72,343	67,344	66,694	Amplicon	SRR35278446
ON 55—DCL	43.4439	−77.4389	15, 25, 30	96,079	81,060	72,876	72,228	Amplicon	SRR35278448
ON 58— epilimnion	43.328	−77.43791	2, 5	100,833	82,512	61,633	60,927	Amplicon	SRR1697709
ON 58—DCL	43.328	−77.43791	10, 20, 30	112,191	92,502	82,376	81,538	Amplicon	SRR35278441
ON 63— epilimnion	43.7317	−77.0169	2, 5	71,552	59,398	39,799	38,138	Amplicon	SRR35278432
ON 63—DCL	43.7317	−77.0169	15, 25	95,412	78,255	66,380	63,490	Amplicon	SRR1697717
ON 64— epilimnion	43.52495	−76.92603	2, 8	102,643	83,229	77,812	77,067	Amplicon	SRR35278439
ON 64—DCL	43.52495	−76.92603	15, 25, 35	103,162	81,323	68,279	67,740	Amplicon	SRR1697712
ON 716— epilimnion	43.60093	−77.4406	2, 5	116,220	98,186	84,614	83,762	Amplicon	SRR35278445
ON 716—DCL	43.60093	−77.4406	10, 20, 30	99,226	85,992	76,418	75,903	Amplicon	SRR35278447

^aInitial quality control (QC) consisted of filtering, trimming, dereplication, merging, and chimera removal. DCL, deep chlorophyll layer. N/A, not applicable.

ACKNOWLEDGMENTS

We thank Derek Smith and members of the R/V Lake Guardian who assisted with sample collection in Lake Ontario.

This material is based upon work supported by the National Science Foundation under grant no. OCE-0927277 as part of the American Recovery and Reinvestment Act of 2009 (R.M.M.).

The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, was supported by the Office of Science of the U.S. Department of Energy operated under contract number DE-AC02-05CH11231 and Community Sequencing Project 723.

AUTHOR AFFILIATIONS

¹Great Lakes Institute for Environmental Research, University of Windsor, Windsor, Ontario, Canada

²Department of Biological Sciences, Bowling Green State University, Bowling Green, Ohio, USA

³Department of Energy, Joint Genome Institute, Berkeley, California, USA

⁴Department of Natural Resources and the Environment, Cornell University, Ithaca, New York, USA

AUTHOR ORCID*s*

Katelyn M. Brown  <http://orcid.org/0000-0002-8581-9490>

Tijana Glavina Del Rio  <http://orcid.org/0000-0003-1411-9814>

R. Michael McKay  <http://orcid.org/0000-0003-2723-5371>

FUNDING

Funder	Grant(s)	Author(s)
National Science Foundation	OCE-0927277	Robert Michael McKay
U.S. Department of Energy	DE-AC02-05CH11231	Tijana Glavina Del Rio
Joint Genome Institute	723	Mark J. Rozmarynowycz

AUTHOR CONTRIBUTIONS

Katelyn M. Brown, Data curation, Software, Visualization, Writing – original draft | Mark J. Rozmarynowycz, Data curation, Investigation, Methodology, Writing – review and editing | Tijana Glavina Del Rio, Methodology, Resources, Writing – review and editing | Lars G. Rudstam, Resources, Supervision, Writing – review and editing | R. Michael McKay, Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review and editing

DATA AVAILABILITY

The amplicon sequencing project has been deposited in the NCBI Sequence Read Archive under the BioProject accession no. [PRJNA255433](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA255433). The station 18 metagenome and metatranscriptomes are found on GenBank under the BioProject accessions [PRJNA329912](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA329912) and [PRJNA366186](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA366186), respectively. The accession numbers are [SRR3990180](https://www.ncbi.nlm.nih.gov/bioproject/SRR3990180) (metagenome), [SRR5249012](https://www.ncbi.nlm.nih.gov/bioproject/SRR5249012) (metatranscriptome), and [SRR5249013](https://www.ncbi.nlm.nih.gov/bioproject/SRR5249013) (metatranscriptome). Assembled metagenomic and metatranscriptomic data can also be found in the JGI Integrated Microbial Genomes and Microbiomes (IMG) database under Taxon Object IDs [3300027836](https://img.jgi.doe.gov/object/3300027836) and [3300003754](https://img.jgi.doe.gov/object/3300003754), respectively. The raw Lake Ontario sequencing data can be found in the Joint Genome Institute Genomes OnLine Database (GOLD) under the GOLD Project IDs [Gp0059905](https://gold.jgi.doe.gov/project/Gp0059905) (metagenome) and [Gp0059906](https://gold.jgi.doe.gov/project/Gp0059906) (metatranscriptomes).

REFERENCES

- Wells M, Johnson T, Robinson R, Midwood J, Shi Y, Larocque S, Eddie A, O'Malley B, Morton K, Gorsky D, Tufts B. 2025. Unique thermal mixing patterns in Lake Ontario revealed by novel year - round observations of thermal stratification. *Limnology & Oceanography* 70:3401–3416. <https://doi.org/10.1002/lno.70215>
- Scofield AE, Watkins JM, Weidel BC, Luckey FJ, Rudstam LG. 2017. The deep chlorophyll layer in Lake Ontario: extent, mechanisms of formation, and abiotic predictors. *J Great Lakes Res* 43:782–794. <https://doi.org/10.1016/j.jglr.2017.04.003>
- Bramburger AJ, Reavie ED. 2016. A comparison of phytoplankton communities of the deep chlorophyll layers and epilimnia of the Laurentian Great Lakes. *J Great Lakes Res* 42:1016–1025. <https://doi.org/10.1016/j.jglr.2016.07.004>
- Watkins JM, Weidel BC, Fisk AT, Rudstam LG. 2017. Cooperative science to inform Lake Ontario management: research from the 2013 Lake Ontario CSMI program. *J Great Lakes Res* 43:779–781. <https://doi.org/10.1016/j.jglr.2017.07.008>
- Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci USA* 108 Suppl 1:4516–4522. <https://doi.org/10.1073/pnas.1000080107>
- Tremblay J, Singh K, Fern A, Kirton ES, He S, Woyke T, Lee J, Chen F, Dangel JL, Tringe SG. 2015. Primer and platform effects on 16S rRNA tag sequencing. *Front Microbiol* 6:771. <https://doi.org/10.3389/fmicb.2015.00771>
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 13:581–583. <https://doi.org/10.1038/nmeth.3869>
- R Core Team. 2025. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Posit team. 2024. RStudio: integrated development environment for R. Posit Software, PBC, Boston, MA.
- McMurdie PJ, Holmes S. 2013. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. <https://doi.org/10.1371/journal.pone.0061217>
- Neuwirth E. 2022. RColorBrewer: ColorBrewer palettes (1.1-3)
- Brand T. 2025. Ggh4x: Hacks for “ggplot2” (0.3.1.9000)
- Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K, Ooms J, Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. 2019. Welcome to the Tidyverse. *JOSS* 4:1686. <https://doi.org/10.21105/joss.01686>
- Pebesma E. 2018. Simple features for R: standardized support for spatial vector data. *R J* 10:439. <https://doi.org/10.32614/RJ-2018-009>
- Pebesma E, Bivand R. 2023. Spatial data science: with applications in R. Chapman & Hall/CRC, New York.
- Dunnington D. 2025. Ggspsatial: spatial data framework for ggplot2
- Wickham H. 2016. Ggplot2. Springer International Publishing, Cham. <https://link.springer.com/10.1007/978-3-319-24277-4>
- Wickham H. 2023. Stringr: simple, consistent wrappers for common string operations
- Ilikchyan IN, McKay RML, Zehr JP, Dyhrman ST, Bullerjahn GS. 2009. Detection and expression of the phosphonate transporter gene *phnD* in marine and freshwater picocyanobacteria. *Environ Microbiol* 11:1314–1324. <https://doi.org/10.1111/j.1462-2920.2009.01869.x>
- Brown KM, Beall BFN, Bullerjahn GS, Finlay JC, Glavina Del Rio T, Small GE, Sterner RW, McKay RM. 2025. Depth-resolved paired metagenomes and metatranscriptomes from the Lake Erie “dead zone”. *Microbiol Resour Announc* 14:e0101325. <https://doi.org/10.1128/mra.01013-25>
- Li H. 2015. BFC: correcting Illumina sequencing errors. *Bioinformatics* 31:2885–2887. <https://doi.org/10.1093/bioinformatics/btv290>
- Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 27:824–834. <https://doi.org/10.1101/gr.213959.116>

23. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
24. Martin J, Bruno VM, Fang Z, Meng X, Blow M, Zhang T, Sherlock G, Snyder M, Wang Z. 2010. Rnnotator: an automated de novo transcriptome assembly pipeline from stranded RNA-Seq reads. *BMC Genomics* 11:663. <https://doi.org/10.1186/1471-2164-11-663>