AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology

Chloroflexi CL500-11 Populations That Predominate Deep-Lake Hypolimnion Bacterioplankton Rely on Nitrogen-Rich Dissolved Organic Matter Metabolism and C_1 Compound Oxidation

Vincent J. Denef,[a] Ryan S. Mueller,[b] Edna Chiang,[a] James R. Liebig,[c] Henry A. Vanderploeg[c]

Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan, USA[a]; Department of Microbiology, Oregon State University, Corvallis, Oregon, USA[b]; NOAA Great Lakes Environmental Research Laboratory, Ann Arbor, Michigan, USA[c]

The *Chloroflexi* CL500-11 clade contributes a large proportion of the bacterial biomass in the oxygenated hypolimnia of deep lakes worldwide, including the world's largest freshwater system, the Laurentian Great Lakes. Traits that allow CL500-11 to thrive and its biogeochemical role in these environments are currently unknown. Here, we found that a CL500-11 population was present mostly in offshore waters along a transect in ultraoligotrophic Lake Michigan (a Laurentian Great Lake). It occurred throughout the water column in spring and only in the hypolimnion during summer stratification, contributing up to 18.1% of all cells. Genome reconstruction from metagenomic data suggested an aerobic, motile, heterotrophic lifestyle, with additional energy being gained through carboxidovory and methylovory. Comparisons to other available streamlined freshwater genomes revealed that the CL500-11 genome contained a disproportionate number of cell wall/capsule biosynthesis genes and the most diverse spectrum of genes involved in the uptake of dissolved organic matter (DOM) substrates, particularly peptides. *In situ* expression patterns indicated the importance of DOM uptake and protein/peptide turnover, as well as type I and type II carbon monoxide dehydrogenase and flagellar motility. Its location in the water column influenced its gene expression patterns the most. We observed increased bacteriorhodopsin gene expression and a response to oxidative stress in surface waters compared to its response in deep waters. While CL500-11 carries multiple adaptations to an oligotrophic lifestyle, its investment in motility, its large cell size, and its distribution in both oligotrophic and mesotrophic lakes indicate its ability to thrive under conditions where resources are more plentiful. Our data indicate that CL500-11 plays an important role in nitrogen-rich DOM mineralization in the extensive deep-lake hypolimnion habitat.

Freshwater lakes are disproportionally active sites of carbon cycling relative to the surface area that they cover due to strong linkages to the surrounding land, from which they receive inorganic nutrients as well as organic carbon (1, 2). Of the estimated 1.9 Pg of terrestrial organic carbon that freshwater systems process per year, nearly half is respired by bacteria (3–5). When soil dissolved organic carbon outgassing is included, net freshwater carbon emissions are of the same order of magnitude as net oceanic uptake (2). While photochemical mineralization of organic carbon can predominate in lake habitats with high levels of photosynthetically active radiation (6), bacterial contributions to dissolved organic matter (DOM) mineralization are important as well (7).

Nevertheless, linkages between the metabolism of organic carbon and specific populations remain limited, particularly in the less-studied hypolimnia of lakes, even for ubiquitous and highly abundant taxa, due to challenges with the isolation of representatives of these taxa (8). In recent years, the use of culture-independent methods has provided insights into the metabolic potential of some key heterotrophic freshwater lineages, such as LD12, the freshwater clade that is a sister to marine clade SAR11 (*Alphaproteobacteria* [9]), acI (*Actinobacteria* [10, 11]), and *Polynucleobacter* (*Betaproteobacteria* [12, 13]). A common feature of these ubiquitous and abundant heterotrophic freshwater taxa is a streamlined genome whose sizes range from 1 to 2.5 Mbp and which contains a variety of genes encoding the ability to take up and metabolize DOM.

We currently have no information regarding the geochemical role of the CL500-11 lineage of the *Chloroflexi*. *Chloroflexi* are rarely observed in freshwater pelagic zones when oxygen is plentiful (8), but CL500-11-like populations are a notable exception and are emerging as a taxon restricted to lakes that are deep enough to maintain low temperatures (<10°C) in the hypolimnion after stratification (14). This group was first observed in Crater Lake, OR, where it is abundant throughout the year, contributing up to 50% of all cells in the deep (15, 16). Since its initial discovery, a similar predominance has been observed in deep lakes around the world, including in Western Europe, East Asia, and, most recently, two of the largest freshwater lakes in the world, Lake Superior and Lake Huron (14, 17). It would be worthwhile to know more about the role of CL500-11 populations in deep-lake hypolimnia, as (i) a large proportion of all surface freshwater is contained in the hypolimnia of deep lakes, and (ii) CL500-11 pop-

Received 14 September 2015 **Accepted** 14 December 2015

Accepted manuscript posted online 18 December 2015

Citation Denef VJ, Mueller RS, Chiang E, Liebig JR, Vanderploeg HA. 2016. *Chloroflexi* CL500-11 populations that predominate deep-lake hypolimnion bacterioplankton rely on nitrogen-rich dissolved organic matter metabolism and C_1 compound oxidation. Appl Environ Microbiol 82:1423–1432. doi:10.1128/AEM.03014-15.

Editor: A. M. Spormann, Stanford University

Address correspondence to Vincent J. Denef, vdenef@umich.edu.

This is GLERL contribution number 1794.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.03014-15.

ulations contribute a large proportion of all bacterial biomass in these habitats through a combination of numerical dominance and large cell size.

In this study, we analyzed the occurrence of a CL500-11-lineage population along a near-shore to offshore transect in Lake Michigan based on sequencing of the V4 region of the 16S rRNA gene as well as catalyzed reporter deposition fluorescent *in situ* hybridization (CARD-FISH). To understand its success and role in the deep-lake habitat, we reconstructed a nearly complete genomic data set from Lake Michigan metagenomic data and identified functions in the CL500-11 genome that were overrepresented relative to their representation in the genomes of other ubiquitous freshwater heterotrophs with streamlined genomes. To gain further insights into the *in situ* functioning of this organism, we generated metatranscriptomic data and identified highly expressed functions, as well as differential expression patterns, based on season and location in the water column.

MATERIALS AND METHODS

Physical and geochemical analyses. A plankton survey system (PSS) was continuously lowered and raised at ~0.25 m s^{-1} in a sinusoidal path from 1 to 2 m beneath the lake surface to 2 to 4 m above the bottom as the R/V *Laurentian* moved at ~1.8 m s^{-1} while logging data every 0.5 s (18). The PSS contained sensors mounted on a V fin to measure the chlorophyll *a* concentration (Wet Labs ECO fluorometer, Sea-Bird Scientific), the amount of photosynthetically active radiation (PAR; determined with a 4-pi sensor; model QSP2300; Biospherical Instruments Inc.), the dissolved oxygen concentration (SBE43; Sea-Bird Scientific), turbidity (Wet Labs ECO NTU; Sea-Bird Scientific), and temperature. The fluorometer output (in volts) was converted to derived chlorophyll *a* concentrations by regression between the fluorometer output and laboratory chlorophyll *a* measurements (18). Replicate samples were analyzed for the concentrations of dissolved organic carbon (DOC), particulate organic carbon (POC), particulate organic nitrogen (PON), total phosphorus (TP), particulate phosphorus (PP), and soluble reactive phosphorus (SRP) according to NOAA Great Lakes Environmental Research Laboratory (GLERL) standard operating procedures (19). Total dissolved phosphorus (TDP) values were calculated by subtracting the PP value from the TP value.

Bacterial samples. Water samples originated from 5 m below the surface and 2 to 5 m above the lake floor at the near-shore station (43°11′17″N, 86°20′38″W, where samples were obtained on 23 April, 15 July, and 23 September 2013) and the offshore station (43°11′59″N, 86°34′11″ W, where samples were obtained on 23 April, 16 July, and 24 September 2013) along the NOAA Lake Michigan-Muskegon, MI, transect (Fig. 1). Water was collected using a 30-liter Niskin bottle and prefiltered through 210- and 20-μm-mesh-size Nitex meshes into 10-liter carboys. The carboys, funnels, and meshes were cleaned with bleach, rinsed with Milli-Q water until no bleach odor remained, and rinsed twice with sample water. Prefiltered water was sequentially filtered onto 3.0-μm-pore-size polycarbonate filters and 0.22-μm-pore-size polyethersulfone filter membranes (142 mm; Millipore) using a Masterflex I/P peristaltic pump (Cole Parmer) between settings 11 and 13. Filters were folded with the bacterial biomass facing inwards and submersed into the RNAlater reagent (Ambion). Samples were stored at −20°C onboard and transferred to a −80°C freezer within 48 h of sampling. Sample filtering was limited to 10 min, and all samples were stored in RNAlater within 20 min of sampling.

DNA/RNA extraction. Duplicate nucleic acid extractions from the same 142-mm filter membrane were performed for each of the field samples using a modified AllPrep DNA/RNA/microRNA universal kit protocol (Qiagen) (20). A part of each of the RNA fractions was converted to cDNA using a ProtoScript II first-strand cDNA synthesis kit (New England BioLabs).

16S rRNA gene sequencing and analysis. DNA and cDNA were submitted to the Joint Genome Institute (JGI) for 16S rRNA gene amplicon sequencing, targeting the V4 region of the 16S rRNA gene (515F/806R universal primers) (21). Pooled libraries were sequenced on an Illumina MiSeq sequencer, using v2 chemistry and two 250-cycle (500 cycles) paired-end reads. RTA software (v1.17.28) and MCS software (v2.2.0) were used to generate data. A random subset of 40,000 read pairs was used for each of the field data samples to reduce the computational time. We used mothur software (v.1.34.3) to generate an operational taxonomic unit (OTU; 97% sequence similarity) table and rarefy data at a subsampling level that allowed inclusion of all samples (n = 4,500; certain samples contained >85% chloroplast sequences). We used the MiSeq standard operating protocol, accessed on 17 December 2014, using the SILVA (release 119) database for alignment and classification (22, 23). The OTU table was imported into Excel software, and OTU count data for the OTUs classified as members of the *Anaerolineaceae* were extracted for the generation of Fig. 2. All data are available on the Joint Genome Institute's genome data portal (http://genome.jgi.doe.gov/; project identifiers, 1041195 and 1041198).

Metagenomic sequencing and analyses. The Joint Genome Institute generated metagenomic data from field samples (the 0.22- to 3-μm fraction only) collected at the offshore station in spring (at depths of 5 m [surface water; Integrated Microbial Genomes {IMG} Genomes Online Database {GOLD} accession number Ga0007769] and 108 m [deep water; IMG GOLD accession number Ga0007770]), summer (at depths of 108 m [IMG GOLD accession number Ga0007777] and 35 m [the depth with the maximum chlorophyll concentration; IMG GOLD accession number Ga0007778]), and fall (at a depth of 108 m only [low chlorophyll was present throughout the water column, so no sampling at 35 m was performed in the fall; IMG GOLD accession number Ga0007786]). Paired-end 150-bp reads were generated (HiSeq 2000 sequencing system) from ~200-bp-fragment-size Nextera (summer and fall) or TruSeq (spring) libraries (Illumina). Metagenomic data from individual samples were assembled by use of the JGI pipeline. In brief, the BBDuk tool (filterk = 27, trimk = 27) was used to remove Illumina adapters, Illumina artifacts, and phiX and quality trim both ends to quality score Q12. Reads with quality scores averaging less than 8 over the read before trimming or with a length of under 40 bp after trimming were discarded. The remaining reads were mapped to the human HG19 reference genome with the BBMap aligner tool, with all hits with over 93% identity being discarded. Trimmed, screened, paired-end Illumina reads were assembled using SOAPdenovo software (v1.05) (24) (default settings) and multiple k-mers (81, 85, 89, 93, 97, and 101 bases). The contigs resulting from each k-mer were dereplicated using in-house Perl scripts. Contigs smaller than 1,800 bp were assembled using Newbler software (v2.8; Life Technologies) to generate larger contigs (flags were -tr, -rip, -mi 98, and -ml 80). Newbler contigs and SOAPdenovo contigs larger than 1,800 bp were combined using the minimus (v2) genome assembler (25) (flags were -D MINID = 98 and -D OVERLAP = 80). Exact duplicate and fully contained sequences were removed using the dedupe tool (github.com/datamade/dedupe).

Full 16S rRNA gene reconstruction was performed with EMIRGE software (with default settings, except for a −l setting of 150, a −i setting of 200, a −s setting of 50, and a −j setting of 1.0), using as a reference the SILVA (release 119) database clustered at 97% using the USEARCH algorithm (26). The −j setting of 1.0 allows the highest possible phylogenetic resolution of the analysis, as only sequences that are 100% identical are combined during the iterative read recruitment process. Sequences matching 18S rRNA gene sequences were excluded.

To achieve improved population genome reconstructions, we used the GAM-NGS tool (27) to sequentially merge sample-specific assemblies from deep samples at the offshore station in spring, summer, and fall (see Table S2 in the supplemental material). All resulting contigs of >3,000 bp were clustered using ESOM tools, on the basis of the tetranucleotide frequency of the contigs (28). To help extract contigs of interest (i.e., those for *Chloroflexi*), contigs were phylogenetically classified by searching the

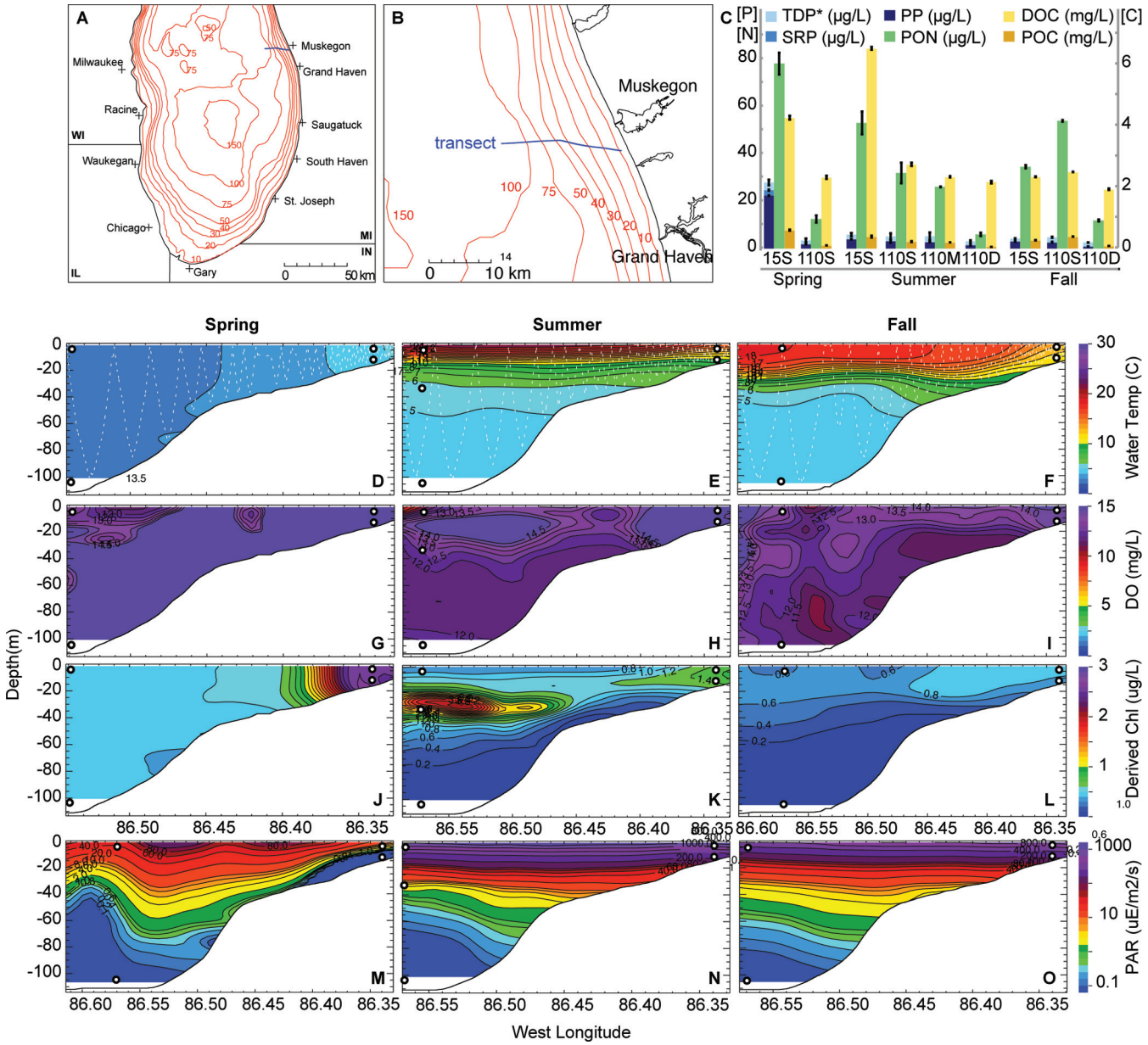

FIG 1 Muskegon transect location and physicochemical data from sampling sites. (A, B) Southern Lake Michigan with depth contours and transect locations. (C) Available geochemical data (N, P, C) from the sampling sites: S, 5 m below the surface; M, 35 m below the surface (the depth with the maximum chlorophyll concentration); D, 80 m below the surface. Error bars represent the standard errors between duplicate measurements. (D to O) Profiles determined by the plankton survey system tow (dotted lines indicate the travel path) at the time of sampling. Circles, microbiological field sampling sites. All data were collected during the nighttime transect, except for PAR (photosynthetically active radiation) data. The low surface PAR and the wavy shape of the PAR profiles across the spring transect result from overcast conditions plus rain showers on parts of the transect.

open reading frames identified by the Prodigal program (29) against those in the NCBI NR database using the RAPsearch2 program (30). We expanded the putative *Chloroflexi* ESOM bin with contigs of >1,000 bp that had a RAPsearch2 *Chloroflexi* classification and reduced it to the CL500-11-LM (where LM refers to Lake Michigan) sequence bin using the Maxbin method (31) and the mmgenome R library (32) (see Table S2 in the supplemental material). The latter relied on read recruitment with the Bowtie 2 program (default settings) (33) using 10 million paired-end reads from offshore spring surface and bottom water samples and summer bottom water samples. Analysis was performed with the CheckM tool to assess the completeness and purity of the final *Chloroflexi* CL500-11-LM sequence bin (34). We also performed a CheckM analysis on the genome of the closest sequenced isolate, *Anaerolinea thermophila* UNI-1 (35). The CL500-11-LM bin was submitted to the DOE Joint Genome Institute Integrated Microbial Genomes for annotation (GOLD analysis ID Ga0063436). To remove redundancy due to incomplete assembly merging, a self-BLASTP analysis of the CL500-11-LM proteins was performed to search for replicate contigs by identifying blocks of proteins (2 or more) that were duplicated at >99% sequence identity. Genes for redundant proteins were removed from the genome.

We used the STAMP program to identify genes for over- and under-represented protein functions in CL500-11-LM relative to their represen-

tation in a combined set of 6 freshwater heterotroph genomes (listed in Table S6 in the supplemental material), *A. thermophila*, and the deep-water summer offshore station metagenomic data set (36). Analyses were performed using the cluster of orthologous groups of proteins (COGs) and Pfam profiles obtained from the JGI IMG data web portal. All data sets were rarefied to the lowest number of profiles (typically, the number present in CL500-11-LM). Two-sample statistical analysis to identify genes for functions that were present in CL500-11-LM at relative abundances significantly different from their abundances in the reference genomes or metagenome was performed in STAMP using Fisher's exact test for Pfam profiles (due to low numbers of occurrences of many Pfam profiles) and the chi-square test for COGs.

Metatranscriptomic sequencing and analyses. The Joint Genome Institute generated metatranscriptomic data from RNA extracted from the offshore station samples (the 0.22- to 3-μm fraction only) collected in spring (at a depth of 5 m during the daytime [IMG GOLD accession number Ga0007745] and nighttime [IMG GOLD accession number Ga0007744] and at a depth of 108 m during the daytime [IMG GOLD accession number Ga0007746]), summer (at a depth of 108 m during the nighttime [IMG GOLD accession number Ga0007753] and at a depth of 35 m during the daytime [IMG GOLD accession number Ga0007754]), and fall (at a depth of 108 m during the nighttime [IMG GOLD accession number Ga0007762]). rRNA was removed from 10 ng of total RNA using a Ribo-Zero rRNA removal kit (*Bacteria*) (Epicentre). Stranded cDNA libraries were generated using an Illumina TruSeq stranded RNA LT kit. The rRNA-depleted RNA was fragmented and reverse transcribed using random hexamers and SuperScript II reverse transcriptase (Invitrogen) followed by second-strand synthesis. The fragmented cDNA was treated with end-pair, A-tailing, adapter ligation, and 15 cycles of PCR. Sequencing (two times for 150-bp sequences) was performed on the Illumina HiSeq 2000 sequencer using a TruSeq SBS sequencing kit (v3).

Raw sequences were (i) quality trimmed to Q10 and adapter trimmed using BBDuk (with options of ktrim = r k = 25 mink = 12 tpe = t tbo = t qtrim = r trimq = 10 maq = 10 maxns = 3 minlen = 50), (ii) filtered for process artifacts using BBDuk (with an option of k = 16), and (iii) mapped against a trimmed version of the SILVA database to remove remaining rRNA reads using BBMap (options were fast = t minid = 0.90 local = t), and (iv) human reads were removed using BBMap. Remaining reads were recruited to the curated CL500-11-LM coding genes using Bowtie (v2) (with options of -local -p 10 -D 5 -R 1 -N 0 -L 25 -i S,1,2.00) (33). For samples with a high abundance of CL500-11, 10 million paired-end read pairs were used (samples collected from deep water at night in summer and fall), while all reads were used for the remaining data sets (samples collected during the day in the summer for determination of the maximum chlorophyll concentration, samples collected from the surface during the day and at night in the spring, and samples collected from deep water during the day in the spring). Mapped reads were scaled to the data set with the lowest number of recruited reads as described previously (37), and differential expression was detected using the DEseq2 package, which identifies differentially expressed genes on the basis of empirical Bayes shrinkage of both gene count variance and the fold change estimate, two problematic issues of count-based data with a large dynamic range, such as transcriptomics data (38). To account for type I error, correction of the *P* values using the Benjamini-Hochberg false discovery rate correction available in the DeSeq2 package was performed. The overrepresentation of functions among highly expressed genes (the functions of proteins encoded by genes with the 5%, 10%, and 25% highest levels of expression relative to the functions of proteins encoded by genes present in the same proportion in the whole CL500-11-LM genome), as well as among genes differentially expressed between surface water and deep water samples, was determined by applying a chi-square test for COG categories and a Fisher exact test for Pfam categories (using the STAMP program). The *P* values were corrected using the Benjamini-Hochberg false discovery rate correction.

CARD-FISH. We used the CARD-FISH probes (CL500-11-specific horseradish peroxidase-labeled probe CLGNS-584 and helper probe CLGNS-567h to augment the signal strength) and the protocol developed by Okazaki et al. (14) on subsamples consisting of the 0.22- to 20-μm fraction from the water sampled for DNA and RNA extractions. Filters were examined by fluorescence microscopy by taking a photo and counting the number of DAPI (4′,6-diamidino-2-phenylindole)-stained and probe-tagged cells within the field; a minimum of 300 DAPI-stained cells (5 to 10 fields of view) was counted per sample. Cell size and the proportion of cell pairs were determined for ~50 cells per sample using the ZEN software measurement tools (Zeiss, Inc.). Because CL500-11 cells have a curved rod shape, the cell sizes were measured from the middle of the cell's width, following the cell shape length-wise. We identified the proportion of all cells that were part of dividing cells (cell pairs) and separately measured the length of each member of the cell pair. For all samples, cells were measured in either a DAPI- or green fluorescent protein-stained field, depending on the clarity of the cell boundaries.

RESULTS

Spatiotemporal distribution of CL500-11. All samples originated from the Muskegon transect on Lake Michigan, a long-term observatory operated by NOAA GLERL (Fig. 1). The water column was isothermal in spring and stratified in summer and fall (Fig. 1D to F), it remained oxygenated throughout the year (Fig. 1G to I), and at midday, PAR was high in surface waters and low in waters below depths of 60 to 80 m (Fig. 1M to O). N, P, and C levels were the highest in the near-shore water in spring and the lowest in the offshore deep water (Fig. 1C). The phytoplankton concentration was measured by determination of the chlorophyll *a* concentration and was low except near the shore in spring and offshore in summer, at the depth where chlorophyll concentrations were highest (35 m) (Fig. 1J to L).

rRNA sequencing data from the cDNA and DNA libraries were combined in one analysis, as DNA from summer offshore deep-water samples failed to be amplified for 16S rRNA iTag analyses. After rarefaction, 306 OTUs were classified as *Chloroflexi*, which contributed 1.8% of all rarefied sequencing reads across all samples (*Chloroflexi* was the 7th most abundant phylum; *Proteobacteria* was the most abundant phylum, accounting for ~40% of all rarefied reads). Only one of these OTUs reached a relative abundance of >3% in any of the individual sample data sets. This OTU was classified at the family level as *Anaerolineaceae* and contained ~75% of all *Chloroflexi* sequences identified in the entire data set. This population was primarily found in the free-living fraction of the deep offshore waters (Fig. 2) and contributed up to 20.7% of all sequences in a given sample (i.e., the 0.22- to 3-μm fraction from offshore, deep-water samples collected in the fall). To obtain more resolved taxonomic information for this OTU, we used the EMIRGE program to reconstruct small-subunit rRNA genes from the metagenomic data for the 0.22- to 3-μm fraction in samples collected in the summer at the offshore hypolimnion station. We reconstructed six unique *Chloroflexi* 16S rRNA genes (reconstructed from reference sequences of two members of the *Anaerolineaceae*, two members of the SL56 marine group of bacteria, one member of KD4-96, and one member of the genus *Roseiflexus*), and these 16S rRNA genes contained a total of 10.1% of all reads used to reconstruct bacterial sequences of >500 nucleotides. One sequence was generated using 7.3% of all recruited reads (see Table S1 in the supplemental material) and shared 99.6% identity to the sequence of clone CL500-11 from Crater Lake, OR, USA (GenBank accession number AF316759) and to a partial sequence

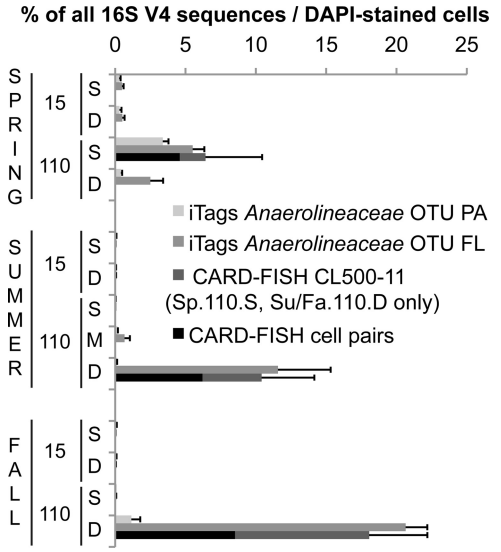

FIG 2 Spatiotemporal distribution of CL500-11-LM *Chloroflexi* along the Lake Michigan Muskegon transect. Relative abundance of the CL500-11-like population, assessed by 16S rRNA V4 region sequencing and CARD-FISH over time, along the Muskegon transect (at the near-shore [15] and offshore [110] stations) at different depths (surface water [S], deep water [D; i.e., the bottom], and the depth with the maximum chlorophyll concentration [M]) and in different filter fractions (particle-associated [PA]) and free-living [FL] fractions). For the CARD-FISH data, the fraction of all CL500-11 cells observed as part of a cell pair is indicated as well. Errors bars indicate 95% CIs. Sp., spring; Su, summer; Fa, fall; 110, 110-m-depth station; 15, 15-m-depth station.

retrieved from Lake Biwa, Japan (GenBank accession number AB686531).

While they were predominantly present in the 0.22- to 3-μm fraction, CL500-11-like 16S rRNA gene sequences were detected in the 3- to 20-μm fraction in fall, and CL500-11-like 16S rRNA gene sequences had nearly equal relative abundances in both fractions in the offshore surface waters in spring (Fig. 2). To obtain clues regarding this apparent shift in cell size, we performed CARD-FISH experiments with three offshore samples (the 0.22- to 20-μm fraction): surface water collected in the spring and bottom water collected in the summer and fall. The CARD-FISH data confirmed the patterns observed by 16S rRNA V4 region iTag sequencing (Fig. 2). The absolute abundance of CL500-11-like cells varied from $3.1 \times 10^4 \pm 1.5 \times 10^4$ cells/ml (95% confidence interval [CI]) in spring surface waters to $2.2 \times 10^4 \pm 0.8 \times 10^4$ and $4.6 \times 10^4 \pm 1.3 \times 10^4$ cells/ml in summer and fall deep-water samples, respectively. The average cell size was larger in spring surface waters (0.87 ± 0.06 μm [90% CI]) than in summer and fall surface waters (0.76 ± 0.04 μm) (see Fig. S1 in the supplemental material). In addition, the proportion of cells that appeared to be connected in cell pairs in spring was larger than that in summer and fall, though the difference was not significant (71.4% versus 50.1% of all cells; $P > 0.10$) (Fig. 2).

CL500-11-LM genome reconstruction and analysis. In an effort to maximize the completeness of the genome, we merged three Lake Michigan metagenomic data set assemblies and subsequently used tetranucleotide and differential coverage binning methods (see Fig. S2 to S4 in the supplemental material) to isolate contigs comprising the genomic data set for the CL500-11-LM population. Initially, we selected 503 contigs that had an average GC content of 60.9% and that carried 2,579 protein-coding genes (see Table S2 in the supplemental material). CheckM estimated the genome to be 84.5% complete with only 0.8% contamination not attributable to strain heterogeneity (see Table S3 in the supplemental material). After removal of redundancy resulting from the assembly merging process, we reduced the CL500-11-LM genome to 398 contigs carrying 2,153 protein-coding genes. CheckM indicated that this significantly reduced within-genome redundancy (see Table S3 in the supplemental material). The genome of the closest fully sequenced relative (*A. thermophila* UNI-1) was estimated to be 93.6% complete (CheckM). Corrected for the *A. thermophila* CheckM completeness estimate gap, we estimated that the CL500-11-LM bin was 90.2% complete.

The curated CL500-11-LM sequence bin contained genes for aerobic oxidative phosphorylation (see Table S4 in the supplemental material) and most genes required for glycolysis and the TCA cycle (see Table S4 in the supplemental material), including the gene for one of two key enzymes for the glyoxylate shunt (malate synthase; IMG locus tag Ga0063436_10633). Missing were genes for orthologs of the enzymes catalyzing the phosphorylation of glucose, though genes for multiple sugar kinases and the conversion of malate to oxaloacetate were present. We also identified multiple ABC transporters for sugars and peptides (see Tables S5 and S6 in the supplemental material). No evidence of a carbon fixation capability was found, but we did identify genes involved in C_1 compound oxidation: (i) type I and type II carbon monoxide dehydrogenases, on the basis of both motif and gene neighborhood analyses (see Table S7 in the supplemental material), and (ii) part of the pathway for the tetrahydrofolate-dependent oxidation of methanol, glycine, methylamines, and, potentially, betaine (four copies of trimethylamine:corrinoid methyltransferases [39]) (see Table S8 in the supplemental material). Missing genes were present in multiple copies in the metagenomic data set, though they were not part of the *Chloroflexi* bins. No genes encoding photosynthetic capabilities were found. However, a proteorhodopsin gene was encoded on a contig located on the margin of the *Chloroflexi* ESOM bin (see Fig. S3 in the supplemental material), and read mapping across multiple data sets supported its inclusion in the CL500-11-LM bin (see Fig. S4 in the supplemental material). The retinal biosynthesis pathway was only partially present, and genes encoding beta-carotene 15,15′-monooxygenase (*bcmo*), and lycopene cyclase (*crtY*) were missing from the CL500-11-LM bin on the basis of analysis of the gene annotation and a BLASTP search with the corresponding protein sequences from "*Candidatus* Pelagibacter ubique" HTCC1062 (E value, <1e−5).

Sixty-five percent of CL500-11-LM genes encoding proteins did not have an ortholog in the most closely related sequenced *Chloroflexi* isolate, *A. thermophila* UNI-1 (an ortholog was defined as the reciprocal best hit with a BLASTP score ratio of >0.3, which corresponds to 30% identity across 70% of the length of the protein [40]). The genome of CL500-11-LM encoded significantly fewer genes for regulatory processes and carbohydrate metabolism than the genome of *A. thermophila*, while genes for amino acid and coenzyme transport and metabolism and cell motility were overrepresented in CL500-11-LM relative to the representation of genes with the same functions in *A. thermophila* (Fig. 3; see also Table S9 in the supplemental material). The latter was the result of the unique presence in CL500-11-LM of genes for the flagellar apparatus and two methyl-accepting chemotaxis pro-

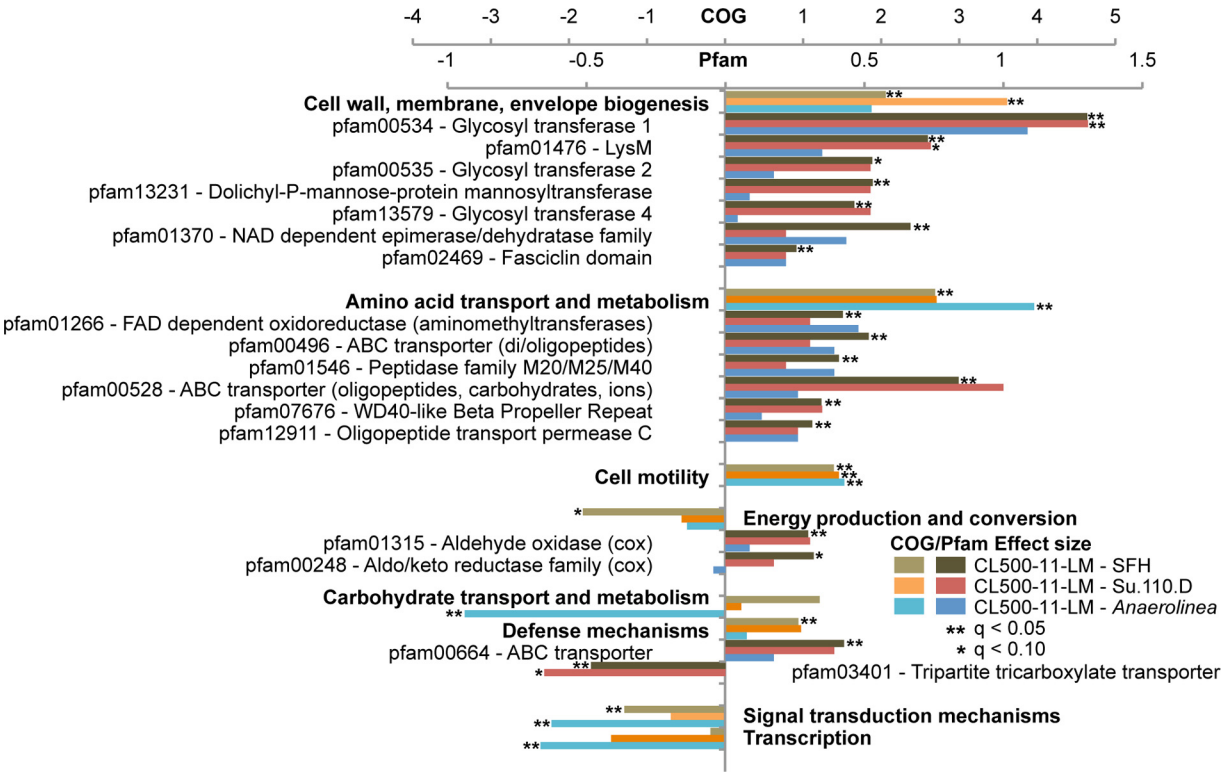

FIG 3 Differential representation of genes for different protein functions in the genome of CL500-11-LM relative to the genomes of other abundant freshwater heterotrophs (SFH) with streamlined genomes (see the list in Table S6 in the supplemental material), summer Lake Michigan offshore deep water metagenomic data, and the *A. thermophila* UNI-1 genome. The results represent a summary of STAMP analyses based on COG and Pfam profiles. The effect size is the difference in the relative abundances of protein functions in each genome or metagenomic data set. The *q* values are the modified *P* values after correction for multiple testing on the basis of the Benjamini-Hochberg false discovery rate. The complete data set is included in Table S9 in the supplemental material. Su, summer; 110 D, a sample from deep water at the 110-m-depth station.

teins. In line with limited regulatory mechanisms in CL500-11-LM, only five RNA polymerase sigma factors were identified (a sigma 28 factor, a sigma 54 factor, and 3 sigma 70 factors), whereas 12 sigma factors and 2 anti-sigma factors were identified in the genome of *A. thermophila* UNI-1.

Genes involved in cell motility, amino acid transport and metabolism, and cell wall and membrane biosynthesis were overrepresented in CL500-11-LM compared to their representation in the genomes of a set of six other abundant freshwater heterotrophic bacteria with relatively streamlined genomes (<2.5 Mbp; two acI-lineage genomes, two LD12-lineage genomes, and two *Polynucleobacter* [PnecC bacterium] genomes; see Tables S6 and S9 in the supplemental material) and the metagenome for deep water from the offshore station in summer (Fig. 3). A more detailed comparison using the Pfam profiles indicated a variety of peptide and carbohydrate transporters, peptidases, methyltransferases, and glycosyltransferases to be overrepresented in CL500-11-LM relative to their representation in its closest sequenced relative and the community that it dominates (Fig. 3; see also Table S9 in the supplemental material). The DOM substrate uptake of CL500-11-LM most closely resembled that of acI (see Table S6 in the supplemental material), though it was more diverse and marked by much higher levels of redundancy of di- and oligopeptide transporters (31 proteins for CL500-11-LM versus 5 for acI). The presence of multiple copies of CO dehydrogenase genes was another unique feature of CL500-11-LM.

CL500-11-LM *in situ* expression patterns. We recruited metatranscriptomic reads to the CL500-11-LM coding sequences for all data sets where iTag analysis indicated the presence of CL500-11-LM (see Tables S5 and S10 in the supplemental material). Average expression levels in samples from offshore stations (surface and deep waters in the spring, water from the depth with the maximum chlorophyll concentration and deep water in the summer, and deep water in the fall) indicated that among the top 5% most highly expressed genes, the following genes were overrepresented relative to their prevalence in the whole CL500-11-LM genome: genes for amino acid transport and metabolism (14.5% of all highly expressed genes relative to 8.3% of all CL500-11-LM genes; $q = 0.15$), particularly genes for di- and oligopeptide transporters (5.2% versus 0.5%; $q = 0.001$); genes for energy production (14.5% versus 4.5%; $q = 2.4 \times 10^{-5}$), which included genes for both the type I and the type II carbon monoxide dehydrogenases; and genes for protein turnover (10.0% versus 2.0%; $q = 2.2 \times 10^{-6}$) (see Table S11 in the supplemental material). Relatively few housekeeping genes were among the most highly expressed genes, though when the top 10% and top 25% most highly expressed genes were included, genes with housekeeping functions were more highly represented, as were genes involved in motility (see Table S11 in the supplemental material). Genes with transport functions for DOM were overrepresented in the CL500-11-LM transcriptome relative to the prevalence of genes with these

functions in the CL500-11-LM genome (11.4% of all mRNA reads; see Table S6 in the supplemental material).

The highest number of differentially expressed genes was identified when samples from deep water were compared to samples from surface water, regardless of the season (75 genes) (see Table S12 in the supplemental material). Only 67 genes were differentially expressed when samples from deep water in the spring were excluded; 60 of these were also identified to be differentially expressed when samples collected from deep water in the spring were included. When samples collected in spring were compared to samples from summer and fall, regardless of depth (samples collected from the surface [day and night] and deep water in the spring relative to samples collected from the depth with the maximum chlorophyll concentration and from deep water collected in the summer and fall), were compared, the result was only 7 differentially expressed genes. Finally, when samples collected from surface and deep water in the spring were compared to samples collected from deep water in the summer and fall, 26 genes were found to be differentially expressed, and 19 of these were shared by the samples collected from surface water in the spring and samples collected from deep water in the summer and fall. This indicated that the location in the water column was the factor that most influenced CL500-11-LM gene expression. Among measured environmental parameters, photosynthetically active radiation was the parameter that varied most between the surface and deep water (Fig. 1). More genes were more highly expressed in surface water than deep water, including genes for defense mechanisms, Fe-S assembly proteins (*suf* operon), Fe uptake, components of the electron chain, glycolysis and the tricarboxylic acid (TCA) cycle, proteorhodopsin, and proteins involved in carotenoid biosynthesis (see Table S12 in the supplemental material). Genes more highly expressed in deep water included those for multiple amino acid metabolic functions, including tryptophan biosynthesis, type III glutamine synthetase, part of a glycine cleavage complex, and part of the type II carbon monoxide dehydrogenase operon.

DISCUSSION

While high *Chloroflexi* levels are rarely reported in oxygenated freshwater lakes (8), the CL500-11 lineage of the *Chloroflexi* is emerging as a highly abundant taxon in the hypolimnion of deep stratified lakes around the world (14, 15). Our study adds Lake Michigan to this list of lakes, which means that CL500-11 has now been shown to reach relative abundances of up to 20% in the hypolimnia of all three upper Great Lakes (Lake Superior, Lake Huron, and Lake Michigan [17]). Analysis of its morphology and comparative genomic and metatranscriptomic analyses reveal adaptations typical of both oligotrophic and copiotrophic lifestyles, in line with their occurrence across a broad productivity gradient (14).

A recent summary of conditions under which CL500-11-lineage bacteria are found to be abundant indicates that they can thrive under both oligotrophic and mesotrophic conditions, as long as oxygen does not get depleted in the hypolimnion and hypolimnion temperatures remain below 10°C (see Table 1 in reference 14). The peak relative abundance of CL500-11 in Lake Michigan was similar to that in Lake Biwa, where the same CARD-FISH probe used in the present study was used, while it remained well below the 50% reported in Crater Lake (15, 16). Total bacterial and CL500-11 cell numbers were about 1 order of magnitude lower in Lake Michigan than in Lake Biwa, most likely reflecting the lower levels of DOC and nutrients in Lake Michigan, which is oligotrophic (Fig. 1); in contrast, Lake Biwa is mesotrophic (14). CL500-11 remained present at high numbers at the end of the spring mixed period in Lake Michigan, while it became undetected in Lake Biwa (14), potentially due to differences in the mixing regimes of these two lakes. All current observations of CL500-11 are consistent with its niche being oxygenated cold water of deep, though not necessarily large, lakes. The only findings that argue against cold adaptation were the identification of CL500-11-like sequences near hydrothermal vents in Lake Yellowstone. However, one study indicated multiple lines of evidence of dilution of the vent water samples with surrounding cold (<10°C) water, including the presence of multiple nonthermophilic microorganisms and the presence of significant levels of O_2 in the sample while the vent water was anoxic (41). The other study observed CL500-11-like sequences only in the coldest (~16°C) vent sample (42).

Our metabolic analysis as well as recent findings of amino sugar incorporation by a CL500-11-like population points to a heterotrophic lifestyle (43). In a previous study of Lake Biwa, it was shown that semilabile DOM produced by phytoplankton in the epilimnion is subsequently biologically remineralized in the hypolimnion (44). Considering its abundance, CL500-11 may play an important role in this process. Comparisons of the reconstructed CL500-11-LM genome to the genomes of other abundant freshwater heterotrophs with streamlined genomes identified traits that support this function and may help explain its success in the deep-lake hypolimnion habitat. While some of the genomes compared are even more streamlined than the genome of CL500-11-LM (e.g., the LD12 lineage [9] and the acI lineage [10, 11]), many similarities and some contrasts can be found. Similarly to other freshwater heterotrophs, CL500-11 dedicates ~3% of its genome to functions involved with the importation of diverse DOM substrates (45), but it stands out as having the most diverse substrate spectrum, particularly for di- and oligopeptides. In line with this observation, the overrepresentation of genes encoding metabolic functions to metabolize these peptides was observed. The low number of carboxylic acid transporters relative to the numbers in LD12 and *Polynucleobacter* could be a reflection of the preference of CL500-11 for water depths with lower PAR, where there is a lower availability of these compounds, which have been shown to be produced through the interaction between UV and DOC (46).

In addition to the overrepresentation of genes for transporters, the disproportionate number of genes encoding transporters among highly expressed genes confirmed the importance to CL500-11-LM of transporters, particularly those involved with peptide transport. The prevalence of genes encoding transporters among highly expressed genes contrasts to the findings of a recent metatranscriptomic survey of a mesotrophic reservoir that showed few genes encoding transporters among the 5% most highly expressed genes (47). However, our results are in line with community-level transcript measurements in coastal ocean systems, where up to 13% of all sequences assigned to COGs were involved in transport, and about half of those were involved in DOM transport (48). In the most oligotrophic areas of the open ocean, proteomics experiments have indicated that two-thirds of all protein originated from transport systems in SAR11 (49). Strikingly, the apparent importance of the CL500-11 lineage in organic nitrogen

transformation matches past findings that in the hypolimnion of Lake Biwa, during the mineralization of DOM, N-rich compounds are preferentially converted (50).

Carboxidovory and, likely, methylovory in CL500-11-LM allow additional energy generation through the oxidation of CO and methyl groups, respectively. Methylovory has been demonstrated for "*Candidatus* Pelagibacter ubique" (51), and we identified homologous genes in the SAR11 freshwater sister clade LD12. Carboxidovory in SAR11 has been suggested to be one mechanism by which it gains an edge in oligotrophic environments, such as the Sargasso Sea. Sargasso Sea DOC concentrations are 2 to 3 times lower than those in deep offshore Lake Michigan water, while Lake Michigan soluble reactive phosphorus levels are lower (52, 53). A lack of data precludes a comparison of inorganic nitrogen levels, though freshwater systems are typically more P limited than N limited (54). Evidence of methylovory (i.e., carbon monoxide dehydrogenase genes), on the other hand, was not identified in any of the other freshwater heterotrophs with streamlined genomes or SAR11. The gene for a carbon monoxide dehydrogenase is present in copiotrophic and often abundant freshwater *Limnohabitans* species (55) and has been observed to be highly expressed in a mesotrophic reservoir (47). The one-carbon oxidation pathways may be beneficial in both oligotrophic and more productive systems, where CL500-11 can be similarly abundant (14).

In addition to the large cell size, the overrepresentation of cell envelope biosynthesis genes and the energetically costly investment in motility are features that characterize a bacterial copiotrophic lifestyle (56). However, the relatively small genome size, the limited investment in transcriptional regulation (3.5% of CL500-11 genes are protein-coding genes) and signal transduction (1.5%), and the small proportion of secreted (3.2%) or membrane-bound (19.4%) proteins are consistent with adaptation to an oligotrophic lifestyle for CL500-11 (56). Genes for regulatory mechanisms are particularly reduced in number, even in comparison to the number of such genes in the genomes of other freshwater organisms with streamlined genomes. Relative to *A. thermophila* UNI-1, CL500-11-LM had 60% and 40% fewer genes involved in signal transduction and transcription, respectively, whereas the predicted decrease is 20% fewer genes in both categories on the basis of the difference in the sizes of the genomes alone (3,166 and 2,153 protein-coding genes in *A. thermophila* UNI-1 and CL500-11-LM, respectively) (57). Overall, the limited conservation in genome content and sequence composition between CL500-11-LM and the filamentous thermophile *A. thermophila* (35) are in line with the large differences in their morphology and ecology.

The cell envelope provides the interface that helps determine the outcome of interactions with viruses and predatory grazers, primarily nanozooplankton (58). Escape from grazers due to specific cell surface structures (S layer) has been shown for the acI lineage (59). Multiple envelope structures, including S layers, have been inferred from genomic data for *Chloroflexi* lineages (60). Although the composition of the cell envelope can contribute to multiple traits, the overrepresentation of cell envelope biosynthesis genes in CL500-11-LM relative to their representation in all other genomes may underpin the surface structures that allow it to escape from grazing predators, allowing it to maintain high relative abundance levels. Independently of the cell envelope composition, increased grazing resistance may be conferred by the large cell size and curved morphology of CL500-11-LM (61), as well as by the large proportion of cell pairs. It is notable that in filter-feeding experiments using water and invasive quagga mussel samples from Lake Michigan, the relative abundance of CL500-11-LM increased after 3.5 h of filter feeding (V. J. Denef, unpublished results).

The occurrence of CL500-11-LM in the surface waters in spring corresponded with increased cell size, which could explain its increased occurrence in the 3- to 20-μm fraction in spring. The larger cells and potentially increased abundance of cell pairs in the spring, which has previously been suggested to be indicative of cell division (14), could indicate more favorable conditions and a higher growth rate in spring (62). However, the metatranscriptomic data did not indicate increases in the numbers of genes for core functions, such as transcription, DNA replication, and cell division, but instead indicated exposure to stress in the surface waters. In particular, multiple genes indicated the presence of increased oxidative stress in surface water relative to that in deep water (the presence of the *suf* operon [63], the thioredoxin gene, *ahpC*, genes for several proteases involved in damaged protein turnover [64]), which is likely, considering the increased photosynthetically active radiation at the surface (65). Oxidative stress response mechanisms are commonly found in the genomes of other freshwater heterotrophs with streamlined genomes as well (10, 12), though different protein families appear to be involved, according to the genes found in CL500-11 and other freshwater organism genomes.

In light of the observed stress response, the higher level of expression of the gene for proteorhodopsin in the surface water, which is expected on the basis of the higher PAR, suggests that CL500-11 uses proteorhodopsin as a means of surviving under suboptimal conditions rather than as a means for increasing cell growth (66). However, the upregulation of genes for electron chain components, glycolysis, and TCA cycle enzymes is similar to the response seen during proteorhodopsin-induced growth in a marine flavobacterium (67). Considering its marginal classification as part of the CL500-11-LM bin, further verification of the presence of the gene for proteorhodopsin in the genome of CL500-11 and its ecological role will be required.

CL500-11-like populations predominate in the hypolimnia of at least three of the five Laurentian Great Lakes, which contain ~20% of the world's surface freshwater. In addition, they have been identified in multiple deep and often large lakes around the world and can likely be found in most deep lakes that maintain a cold and oxygenated hypolimnion. Their numerical abundance, large cell size, and low within-clade sequence variation (14) mean that populations with traits similar to those of CL500-11-LM likely contribute a significant proportion of the world's freshwater bacterial biomass. Our analyses suggest that they play an important role in the transformation of biologically derived organic matter, particularly nitrogen-rich DOM.

ACKNOWLEDGMENTS

We are grateful to the crew of the R/V *Laurentian*, NOAA Great Lakes Environmental Research Laboratory science staff, and Marian L. Schmidt and Ann McCarthy (University of Michigan) for sampling and nucleic acid extractions. We thank the anonymous reviewers for their help in improving our manuscript.

We declare that we do not have any competing financial interests in relation to this work.

V.J.D. was supported by the Community Sequencing Program (U.S.

Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, supported under contract no. DE-AC02-05CH11231) and the University of Michigan. E.C. was supported by an American Society for Microbiology undergraduate research fellowship and the Beckman Scholars Program.

V.J.D. designed the study; E.C., J.R.L., and H.A.V. performed experiments; V.J.D., R.S.M., H.A.V., and E.C. analyzed the data; and V.J.D. wrote the paper.

FUNDING INFORMATION

U.S. Department of Energy (DOE) provided funding to Vincent J. Denef under grant number DE-AC02-05CH11231.

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

REFERENCES

1. **Cole JJ, Prairie YT, Caraco NF, McDowell WH, Tranvik LJ, Striegl RG, Duarte CM, Kortelainen P, Downing JA, Middelburg JJ.** 2007. Plumbing the global carbon cycle: integrating inland waters into the terrestrial carbon budget. Ecosystems **10:**172–185. http://dx.doi.org/10.1007/s10021-006-9013-8.
2. **Tranvik LJ, Downing JA, Cotner JB, Loiselle SA, Striegl RG, Ballatore TJ, Dillon P, Finlay K, Fortino K, Knoll LB.** 2009. Lakes and reservoirs as regulators of carbon cycling and climate. Limnol Oceanogr **54:**2298–2314. http://dx.doi.org/10.4319/lo.2009.54.6_part_2.2298.
3. **Cole JJ, Findlay S, Pace ML.** 1988. Bacterial production in fresh and saltwater ecosystems: a cross-system overview. Mar Ecol Prog Ser **43:**1–10. http://dx.doi.org/10.3354/meps043001.
4. **Cotner JB, Biddanda BA.** 2002. Small players, large role: microbial influence on biogeochemical processes in pelagic aquatic ecosystems. Ecosystems **5:**105–121. http://dx.doi.org/10.1007/s10021-001-0059-3.
5. **McCallister SL, Del Giorgio PA.** 2008. Direct measurement of the d13C signature of carbon respired by bacteria in lakes: linkages to potential carbon sources, ecosystem baseline metabolism, and CO_2 fluxes. Limnol Oceanogr **53:**1204–1216. http://dx.doi.org/10.4319/lo.2008.53.4.1204.
6. **Cory RM, Ward CP, Crump BC, Kling GW.** 2014. Carbon cycle. Sunlight controls water column processing of carbon in arctic fresh waters. Science **345:**925–928. http://dx.doi.org/10.1126/science.1253119.
7. **Cole JJ, Carpenter SR, Kitchell JF, Pace ML.** 2002. Pathways of organic carbon utilization in small lakes: results from a whole-lake ^{13}C addition and coupled model. Limnol Oceanogr **47:**1664–1675. http://dx.doi.org/10.4319/lo.2002.47.6.1664.
8. **Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S.** 2011. A guide to the natural history of freshwater lake bacteria. Microbiol Mol Biol Rev **75:**14–49. http://dx.doi.org/10.1128/MMBR.00028-10.
9. **Zaremba-Niedzwiedzka K, Viklund J, Zhao W, Ast J, Sczyrba A, Woyke T, McMahon K, Bertilsson S, Stepanauskas R, Andersson SG.** 2013. Single-cell genomics reveal low recombination frequencies in freshwater bacteria of the SAR11 clade. Genome Biol **14:**R130. http://dx.doi.org/10.1186/gb-2013-14-11-r130.
10. **Garcia SL, McMahon KD, Martinez-Garcia M, Srivastava A, Sczyrba A, Stepanauskas R, Grossart HP, Woyke T, Warnecke F.** 2013. Metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton. ISME J **7:**137–147. http://dx.doi.org/10.1038/ismej.2012.86.
11. **Ghylin TW, Garcia SL, Moya F, Oyserman BO, Schwientek P, Forest KT, Mutschler J, Dwulit-Smith J, Chan LK, Martinez-Garcia M, Sczyrba A, Stepanauskas R, Grossart HP, Woyke T, Warnecke F, Malmstrom R, Bertilsson S, McMahon KD.** 2014. Comparative single-cell genomics reveals potential ecological niches for the freshwater acI Actinobacteria lineage. ISME J **8:**2503–2516. http://dx.doi.org/10.1038/ismej.2014.135.
12. **Hahn MW, Scheuerl T, Jezberová J, Koll U, Jezbera J, Šimek K, Vannini C, Petroni G, Wu QL.** 2012. The passive yet successful way of planktonic life: genomic and experimental analysis of the ecology of a free-living polynucleobacter population. PLoS One **7:**e32772. http://dx.doi.org/10.1371/journal.pone.0032772.
13. **Garcia SL, Buck M, McMahon KD, Grossart HP, Eiler A, Warnecke F.** 2015. Auxotrophy and intrapopulation complementary in the 'interactome' of a cultivated freshwater model community. Mol Ecol **24:**4449–4459. http://dx.doi.org/10.1111/mec.13319.
14. **Okazaki Y, Hodoki Y, Nakano S.** 2013. Seasonal dominance of CL500-11 bacterioplankton (phylum Chloroflexi) in the oxygenated hypolimnion of Lake Biwa, Japan. FEMS Microbiol Ecol **83:**82–92. http://dx.doi.org/10.1111/j.1574-6941.2012.01451.x.
15. **Urbach E, Vergin KL, Young L, Morse A, Larson GL, Giovannoni SJ.** 2001. Unusual bacterioplankton community structure in ultra-oligotrophic Crater Lake. Limnol Oceanogr **46:**557–572. http://dx.doi.org/10.4319/lo.2001.46.3.0557.
16. **Urbach E, Vergin KL, Larson GL, Giovannoni SJ.** 2007. Bacterioplankton communities of Crater Lake, OR: dynamic changes with euphotic zone food web structure and stable deep water populations. Hydrobiologia **574:**161–177. http://dx.doi.org/10.1007/s10750-006-0351-5.
17. **Rozmarynowycz MJ.** 2014. Spatio-temporal distribution of microbial communities in the Laurentian Great Lakes. Ph.D. thesis. Bowling Green State University, Bowling Green, MI.
18. **Vanderploeg HA, Ludsin SA, Ruberg SA, Höök TO, Pothoven SA, Brandt SB, Lang GA, Liebig JR, Cavaletto JF.** 2009. Hypoxia affects spatial distributions and overlap of pelagic fish, zooplankton, and phytoplankton in Lake Erie. J Exp Mar Biol Ecol **381:**S92–S107. http://dx.doi.org/10.1016/j.jembe.2009.07.027.
19. **Nalepa TF, Fahnenstiel GL, McCormick MJ, Johengen TH, Lang GA, Cavaletto JF, Goudy G.** 1996. Physical and chemical variables of Saginaw Bay, Lake Huron in 1991-93. NOAA technical memorandum ERL GLERL 91. Great Lakes Environmental Research Laboratory, Ann Arbor, MI.
20. **McCarthy A, Chiang E, Schmidt ML, Denef VJ.** 2015. RNA preservation agents and nucleic acid extraction method bias perceived bacterial community composition. PLoS One **10:**e0121659. http://dx.doi.org/10.1371/journal.pone.0121659.
21. **Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M.** 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J **6:**1621–1624. http://dx.doi.org/10.1038/ismej.2012.8.
22. **Schloss PD, Westcott SL.** 2011. Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. Appl Environ Microbiol **77:**3219–3226. http://dx.doi.org/10.1128/AEM.02810-10.
23. **Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO.** 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res **41:**D590–D596. http://dx.doi.org/10.1093/nar/gks1219.
24. **Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, Li S, Yang H, Wang J, Wang J.** 2010. De novo assembly of human genomes with massively parallel short read sequencing. Genome Res **20:**265–272. http://dx.doi.org/10.1101/gr.097261.109.
25. **Sommer DD, Delcher AL, Salzberg SL, Pop M.** 2007. Minimus: a fast, lightweight genome assembler. BMC Bioinformatics **8:**64. http://dx.doi.org/10.1186/1471-2105-8-64.
26. **Edgar RC,.** 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics **26:**2460–2461. http://dx.doi.org/10.1093/bioinformatics/btq461.
27. **Vicedomini R, Vezzi F, Scalabrin S, Arvestad L, Policriti A.** 2013. GAM-NGS: genomic assemblies merger for next generation sequencing. BMC Bioinformatics **14**(Suppl 7):S6. http://dx.doi.org/10.1186/1471-2105-14-S7-S6.
28. **Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP, Banfield JF.** 2009. Community-wide analysis of microbial genome sequence signatures. Genome Biol **10:**R85. http://dx.doi.org/10.1186/gb-2009-10-8-r85.
29. **Hyatt D, LoCascio PF, Hauser LJ, Uberbacher EC.** 2012. Gene and translation initiation site prediction in metagenomic sequences. Bioinformatics **28:**2223–2230. http://dx.doi.org/10.1093/bioinformatics/bts429.
30. **Zhao Y, Tang H, Ye Y.** 2012. RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. Bioinformatics **28:**125–126. http://dx.doi.org/10.1093/bioinformatics/btr595.
31. **Wu YW, Tang YH, Tringe SG, Simmons BA, Singer SW.** 2014. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome **2:**26. http://dx.doi.org/10.1186/2049-2618-2-26.
32. **Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH.** 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nat Biotechnol **31:**533–538. http://dx.doi.org/10.1038/nbt.2579.

33. **Langmead B, Salzberg SL.** 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods **9:**357–359. http://dx.doi.org/10.1038/nmeth.1923.
34. **Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW.** 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res **25:**1043–1055. http://dx.doi.org/10.1101/gr.186072.114.
35. **Sekiguchi Y, Yamada T, Hanada S, Ohashi A, Harada H, Kamagata Y.** 2003. Anaerolinea thermophila gen. nov., sp. nov. and Caldilinea aerophila gen. nov., sp. nov., novel filamentous thermophiles that represent a previously uncultured lineage of the domain Bacteria at the subphylum level. Int J Syst Evol Microbiol **53:**1843–1851. http://dx.doi.org/10.1099/ijs.0.02699-0.
36. **Parks DH, Tyson GW, Hugenholtz P, Beiko RG.** 2014. STAMP: statistical analysis of taxonomic and functional profiles. Bioinformatics **30:**3123–3124. http://dx.doi.org/10.1093/bioinformatics/btu494.
37. **McMurdie PJ, Holmes S.** 2014. Waste not, want not: why rarefying microbiome data is inadmissible. PLoS Comput Biol **10:**e1003531. http://dx.doi.org/10.1371/journal.pcbi.1003531.
38. **Love MI, Huber W, Anders S.** 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol **15:**550. http://dx.doi.org/10.1186/s13059-014-0550-8.
39. **Ticak T, Kountz DJ, Girosky KE, Krzycki JA, Ferguson DJ.** 2014. A nonpyrrolysine member of the widely distributed trimethylamine methyltransferase family is a glycine betaine methyltransferase. Proc Natl Acad Sci U S A **111:**E4668–E4676. http://dx.doi.org/10.1073/pnas.1409642111.
40. **Rasko DA, Myers GS, Ravel J.** 2005. Visualization of comparative genomic analyses by BLAST score ratio. BMC Bioinformatics **6:**2. http://dx.doi.org/10.1186/1471-2105-6-2.
41. **Clingenpeel S, Macur RE, Kan J, Inskeep WP, Lovalvo D, Varley J, Mathur E, Nealson K, Gorby Y, Jiang H, LaFracois T, McDermott TR.** 2011. Yellowstone Lake: high-energy geochemistry and rich bacterial diversity. Environ Microbiol **13:**2172–2185. http://dx.doi.org/10.1111/j.1462-2920.2011.02466.x.
42. **Yang T, Lyons S, Aguilar C, Cuhel R, Teske A.** 2011. Microbial communities and chemosynthesis in Yellowstone Lake sublacustrine hydrothermal vent waters. Front Microbiol **2:**130. http://dx.doi.org/10.3389/fmicb.2011.00130.
43. **Tada Y, Grossart HP.** 2014. Community shifts of actively growing lake bacteria after N-acetyl-glucosamine addition: improving the BrdU-FACS method. ISME J **8:**441–454. http://dx.doi.org/10.1038/ismej.2013.148.
44. **Maki K, Kim C, Yoshimizu C, Tayasu I, Miyajima T, Nagata T.** 2010. Autochthonous origin of semi-labile dissolved organic carbon in a large monomictic lake (Lake Biwa): carbon stable isotopic evidence. Limnology **11:**143–153. http://dx.doi.org/10.1007/s10201-009-0299-z.
45. **Salcher MM, Posch T, Pernthaler J.** 2013. In situ substrate preferences of abundant bacterioplankton populations in a prealpine freshwater lake. ISME J **7:**896–907. http://dx.doi.org/10.1038/ismej.2012.162.
46. **Bertilsson S, Tranvik LJ.** 2000. Photochemical transformation of dissolved organic matter in lakes. Limnol Oceanogr **45:**753–762. http://dx.doi.org/10.4319/lo.2000.45.4.0753.
47. **Tsementzi D, Poretsky R, Rodriguez-R LM, Luo C, Konstantinidis KT.** 2014. Evaluation of metatranscriptomic protocols and application to the study of freshwater microbial communities. Environ Microbiol Rep **6:**640–655. http://dx.doi.org/10.1111/1758-2229.12180.
48. **Poretsky RS, Sun S, Mou X, Moran MA.** 2010. Transporter genes expressed by coastal bacterioplankton in response to dissolved organic carbon. Environ Microbiol **12:**616–627. http://dx.doi.org/10.1111/j.1462-2920.2009.02102.x.
49. **Sowell SM, Wilhelm LJ, Norbeck AD, Lipton MS, Nicora CD, Barofsky DF, Carlson CA, Smith RD, Giovanonni SJ.** 2009. Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. ISME J **3:**93–105. http://dx.doi.org/10.1038/ismej.2008.83.
50. **Kim C, Nishimura Y, Nagata T.** 2006. Role of dissolved organic matter in hypolimnetic mineralization of carbon and nitrogen in a large, monomictic lake. Limnol Oceanogr **51:**70–78. http://dx.doi.org/10.4319/lo.2006.51.1.0070.
51. **Sun J, Steindler L, Thrash JC, Halsey KH, Smith DP, Carter AE, Landry ZC, Giovannoni SJ.** 2011. One carbon metabolism in SAR11 pelagic marine bacteria. PLoS One **6:**e23973. http://dx.doi.org/10.1371/journal.pone.0023973.
52. **Carlson CA, Ducklow HW, Michaels AF.** 1994. Annual flux of dissolved organic carbon from the euphotic zone in the northwestern Sargasso Sea. Nature **371:**405–408. http://dx.doi.org/10.1038/371405a0.
53. **Wu J, Sunda W, Boyle EA, Karl DM.** 2000. Phosphate depletion in the western North Atlantic Ocean. Science **289:**759–762. http://dx.doi.org/10.1126/science.289.5480.759.
54. **Hecky RE, Kilham P.** 1988. Nutrient limitation of phytoplankton in freshwater and marine environments: a review of recent evidence on the effects of enrichment. Limnol Oceanogr **33:**796–822. http://dx.doi.org/10.4319/lo.1988.33.4_part_2.0796.
55. **Zeng Y, Kasalický V, Šimek K, Koblížeka M.** 2012. Genome sequences of two freshwater betaproteobacterial isolates, Limnohabitans species strains Rim28 and Rim47, indicate their capabilities as both photoautotrophs and ammonia oxidizers. J Bacteriol **194:**6302–6303. http://dx.doi.org/10.1128/JB.01481-12.
56. **Lauro FM, McDougald D, Thomas T, Williams TJ, Egan S, Rice S, DeMaere MZ, Ting L, Ertan H, Johnson J.** 2009. The genomic basis of trophic strategy in marine bacteria. Proc Natl Acad Sci U S A **106:**15527–15533. http://dx.doi.org/10.1073/pnas.0903507106.
57. **Konstantinidis KT, Tiedje JM.** 2004. Trends between gene content and genome size in prokaryotic species with larger genomes. Proc Natl Acad Sci U S A **101:**3160–3165. http://dx.doi.org/10.1073/pnas.0308653100.
58. **Jürgens K, Matz C.** 2002. Predation as a shaping force for the phenotypic and genotypic composition of planktonic bacteria. Antonie Van Leeuwenhoek **81:**413–434. http://dx.doi.org/10.1023/A:1020505204959.
59. **Tarao M, Jezbera J, Hahn MW.** 2009. Involvement of cell surface structures in size-independent grazing resistance of freshwater Actinobacteria. Appl Environ Microbiol **75:**4720–4726. http://dx.doi.org/10.1128/AEM.00251-09.
60. **Hug LA, Castelle CJ, Wrighton KC, Thomas BC, Sharon I, Frischkorn KR, Williams KH, Tringe SG, Banfield JF.** 2013. Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. Microbiome **1:**22. http://dx.doi.org/10.1186/2049-2618-1-22.
61. **Grujčić V, Kasalický V, Šimek K.** 2015. Prey-specific growth responses of freshwater flagellate communities induced by morphologically distinct bacteria from the genus Limnohabitans. Appl Environ Microbiol **81:**4993–5002. http://dx.doi.org/10.1128/AEM.00396-15.
62. **Schaechter M, Maaløe O, Kjeldgaard NO.** 1958. Dependency on medium and temperature of cell size and chemical composition during balanced growth of Salmonella typhimurium. J Gen Microbiol **19:**592–606. http://dx.doi.org/10.1099/00221287-19-3-592.
63. **Nachin L, Loiseau L, Expert D, Barras F.** 2003. SufC: an unorthodox cytoplasmic ABC/ATPase required for [Fe-S] biogenesis under oxidative stress. EMBO J **22:**427–437. http://dx.doi.org/10.1093/emboj/cdg061.
64. **Cabiscol E, Tamarit J, Ros J.** 2000. Oxidative stress in bacteria and protein damage by reactive oxygen species. Int Microbiol **3:**3–8.
65. **Blough NV, Zepp RG.** 1995. Reactive oxygen species in natural waters. *In* Foote CS, Valentine JS (ed), Active oxygen in chemistry. Chapman and Hall, New York, NY.
66. **DeLong EF, Béjà O.** 2010. The light-driven proton pump proteorhodopsin enhances bacterial survival during tough times. PLoS Biol **8:**e1000359. http://dx.doi.org/10.1371/journal.pbio.1000359.
67. **Kimura H, Young CR, Martinez A, DeLong EF.** 2011. Light-induced transcriptional responses associated with proteorhodopsin-enhanced growth in a marine flavobacterium. ISME J **5:**1641–1651. http://dx.doi.org/10.1038/ismej.2011.36.