AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

ENVIRONMENTAL MICROBIOLOGY

Temperature and Nutrient Levels Correspond with Lineage-Specific Microdiversification in the Ubiquitous and Abundant Freshwater Genus *Limnohabitans*

Ruben Props,[a,b] Vincent J. Denef[b]

[a]Center for Microbial Ecology and Technology, Ghent University, Ghent, Belgium
[b]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan, USA

ABSTRACT Most freshwater bacterial communities are characterized by a few dominant taxa that are often ubiquitous across freshwater biomes worldwide. Our understanding of the genomic diversity within these taxonomic groups is limited to a subset of taxa. Here, we investigated the genomic diversity that enables *Limnohabitans*, a freshwater genus key in funneling carbon from primary producers to higher trophic levels, to achieve abundance and ubiquity. We reconstructed eight putative *Limnohabitans* metagenome-assembled genomes (MAGs) from stations located along broad environmental gradients existing in Lake Michigan, part of Earth's largest surface freshwater system. *De novo* strain inference analysis resolved a total of 23 strains from these MAGs, which strongly partitioned into two habitat-specific clusters with cooccurring strains from different lineages. The largest number of strains belonged to the abundant LimB lineage, for which robust *in situ* strain delineation had not previously been achieved. Our data show that temperature and nutrient levels may be important environmental parameters associated with microdiversification within the *Limnohabitans* genus. In addition, strains predominant in low- and high-phosphorus conditions had larger genomic divergence than strains abundant under different temperatures. Comparative genomics and gene expression analysis yielded evidence for the ability of LimB populations to exhibit cellular motility and chemotaxis, a phenotype not yet associated with available *Limnohabitans* isolates. Our findings broaden historical marker gene-based surveys of *Limnohabitans* microdiversification and provide *in situ* evidence of genome diversity and its functional implications across freshwater gradients.

IMPORTANCE *Limnohabitans* is an important bacterial taxonomic group for cycling carbon in freshwater ecosystems worldwide. Here, we examined the genomic diversity of different *Limnohabitans* lineages. We focused on the LimB lineage of this genus, which is globally distributed and often abundant, and its abundance has shown to be largely invariant to environmental change. Our data show that the LimB lineage is actually comprised of multiple cooccurring populations for which the composition and genomic characteristics are associated with variations in temperature and nutrient levels. The gene expression profiles of this lineage suggest the importance of chemotaxis and motility, traits that had not yet been associated with the *Limnohabitans* genus, in adapting to environmental conditions.

KEYWORDS metagenomics, strain resolved, metatranscriptomics, Great Lakes, *Limnohabitans*, environmental adaptation, chemotaxis, freshwater, microbial ecology, microdiversity, strain delineation

In natural and managed environments, bacterial taxa partition across habitats at both coarse (1, 2) and fine (e.g., >97 to 99% 16S rRNA gene similarity, or >~96.5% genome-wide nucleotide identity) (3, 4) taxonomic scales. This genetic diversification,

Citation Props R, Denef VJ. 2020. Temperature and nutrient levels correspond with lineage-specific microdiversification in the ubiquitous and abundant freshwater genus *Limnohabitans*. Appl Environ Microbiol 86:e00140-20. https://doi.org/10.1128/AEM.00140-20.

Editor Hideaki Nojiri, University of Tokyo

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

Received 21 January 2020
Accepted 10 March 2020
Accepted manuscript posted online 13 March 2020
Published 5 May 2020

also labeled microdiversity, emerges from both mutational and gene gain/loss events (4–8) and represents an important trait-modifying process by which a bacterial taxon can achieve ubiquity and thus maximize its niche coverage in an ecosystem (9). Usually, this microdiversity is masked by the consensus similarity thresholds used to define operational taxonomic units (OTUs) in marker gene surveys or hidden within consensus metagenome-assembled genomes (MAGs). In recent years several computational methods have been described to robustly infer microdiversity from 16S rRNA amplicon gene surveys (10, 11), as well as from metagenomic data (12–14). This has led to numerous observations that small or even no differences in 16S rRNA gene identity can lead to substantial alterations in, for example, optimal growth temperature (6), carbon substrate utilization (15), pH tolerance (16), and light preference (17). These trait differences are often reflected in the habitat partitioning across environmental gradients within the contiguous environment from which these taxa were sampled. Understanding how microdiversification enables bacterial taxa to adapt to changing environmental conditions can therefore facilitate our understanding and possibly mitigation of the short- and long-term impacts of global change (18, 19).

Freshwater lakes are known hot spots of global biogeochemical cycles and act as important environmental "sentinels" for local and global environmental change (20). This is the case for Lake Michigan, which is part of the Laurentian Great Lakes system that contains an estimated 21% of the world's surface freshwater. This ecosystem has been rapidly changing as a consequence of various anthropogenic stressors (21). The invasion of dreissenid mussels is causing drastic changes in the food web structure by decimating offshore primary production and indirectly causing nearshore harmful algal blooms (22–26). In addition to natural temperature and light gradients across seasons and depths, invasive mussel disturbances have led to steep estuary-to-offshore nutrient gradients across relatively small spatial scales.

Previous surveys of Lake Michigan have shown that the *Limnohabitans* genus (family *Burkholderiaceae*, order *Burkholderiales*, class *Gammaproteobacteria*; GTDB taxonomy) (27), and in particular the so-called "Lhab-A1 tribe" (28), is abundant and that its abundance is largely unaffected by environmental gradients of temperature, trophic state, and light availability (29). *Limnohabitans* is a metabolically versatile, fast-growing, morphologically diverse bacterioplankton genus that has been observed in nearly every lake system worldwide in high abundance (~12% [30]) and plays an important role in funneling carbon from primary producers to higher trophic levels (28, 31–35). The extensive collection of cultures classified as this genus has revealed a wide metabolic versatility, ranging from photoheterotrophy (36) to putative ammonium and sulfur oxidation (37). In light of their biogeochemical importance, it is important to understand why *Limnohabitans* lineages are both abundant and ubiquitous across diverse environments. This knowledge can in turn be leveraged to better predict how their abundance and their functional contributions will respond to ongoing environmental changes.

Here, we studied microdiversification among core genes, differences in accessory genes, and *in situ* differential gene expression of populations belonging to the *Limnohabitans* genus across the sampling stations of a well-studied Lake Michigan transect. For more than 2 decades, this transect has been used to characterize the spatiotemporal changes in the pelagic food web and relate these changes to anthropogenic pressures and impacts of global change (38). It consists of three stations that represent the meso- to eutrophic Muskegon Lake freshwater estuary, the oligotrophic to mesotrophic nearshore (M15), and the ultraoligotrophic offshore waters (M110). We hypothesized that in response to environmental changes microdiversity would be an important adaptation to enable *Limnohabitans* to maintain its prevalence in the freshwater food web. To address this hypothesis, we used gene- and genome-centric metagenomics and metatranscriptomics to infer the phylogeny, abundance, genome properties, gene expression, and inferred phenotypic traits (i.e., growth rate) of putative *Limnohabitans* populations. We then performed variant detection and *de novo* strain

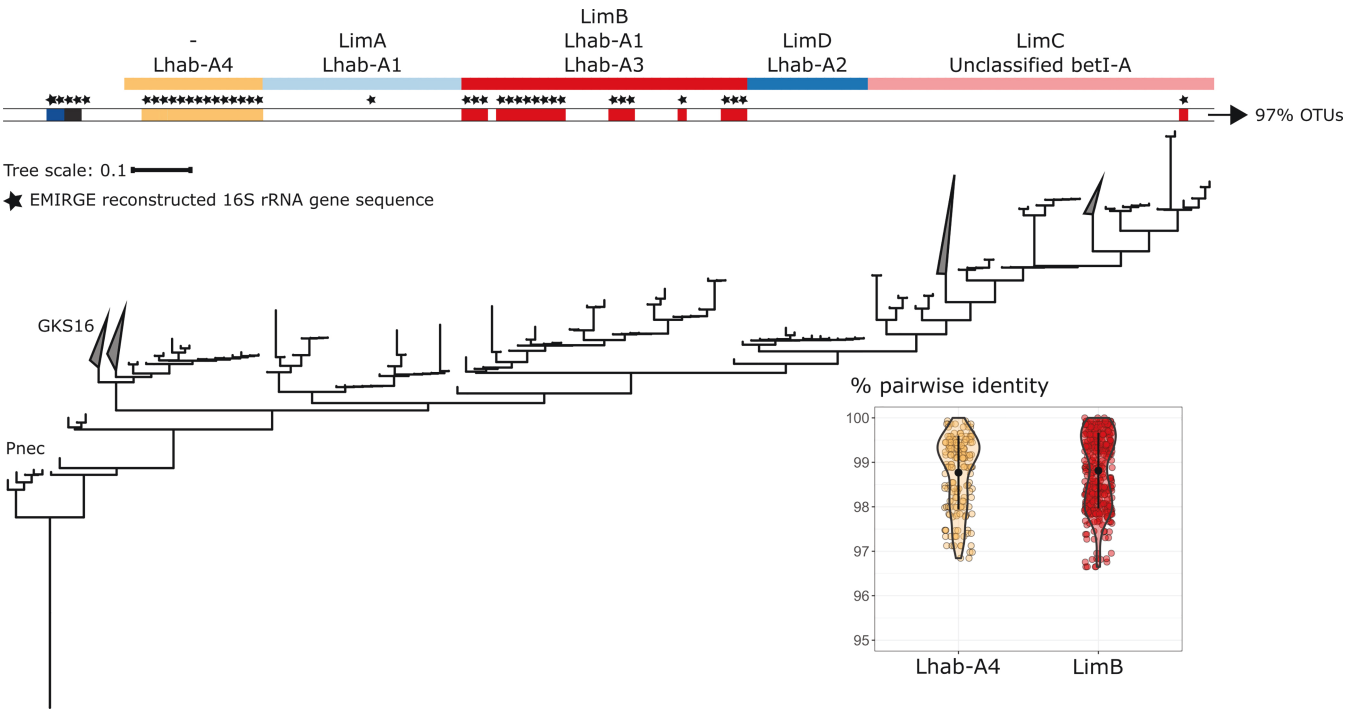

FIG 1 Phylogenetic tree of EMIRGE reconstructed 16S rRNA gene sequences and reference betaproteobacterium 16S rRNA gene sequences based on previous findings (39). Color-coded bars at the top indicate *Limnohabitans* lineages. Sequences within these lineages were also classified according to the Newton et al. (28) taxonomic framework into "Lhab" tribes. The second row of color bars corresponds to the clustering of the EMIRGE reconstructed sequences at 97% average similarity. Pairwise percent identity distributions are provided for the LimB and Lhab-A4 metagenome reconstructed 16S rRNA gene sequences in the inset figure. The tree was rooted using sequences of *Enterobacter cancerogenus* LMG 2693 (Z96078.1) and *Escherichia vulneris* (AF530476.1) as outgroups.

inference on the reconstructed MAGs to evaluate the level of microdiversity within them.

RESULTS AND DISCUSSION

We used publicly available metagenomic and metatranscriptomic data of the 0.22- to 3-μm plankton fraction that were previously collected at three stations (i.e., labeled MLB, M15, and M110) spanning a productivity gradient across Lake Michigan and one of its freshwater estuaries, Muskegon Lake (29). This sampling campaign consisted of 24 samples taken at seven spatially separated sampling locations spread across the three stations and their station-dependent sampling depths. The samples were taken at three time points, divided over three seasons (i.e., spring, summer, and fall of 2013), at different depths (5 m up to 108 m), and by time of day (day/night). At the M110 station, additional seasonal sampling was conducted at the deep chlorophyll maximum (DCM). The environmental data for these three stations spanned large environmental gradients: temperature, 3 to 28°C; depth, 1 to 110 m; photosynthetically active radiation (PAR), 0 to 1,042 W m^2; total phosphorus, 3 to 30 μg liter^{-1}; and chlorophyll *a*, 0.2 to 13 μg liter^{-1}. A more detailed representation of the study site is described in Denef et al. (24). We provide a summary of its environmental characteristics in Fig. S1 in the supplemental material.

Marker gene diversity. Full-length putative *Limnohabitans* 16S rRNA gene sequences were reconstructed from metagenomic data and incorporated in a phylogenetic tree together with publicly available isolate and environmental 16S rRNA gene sequences of the *Limnohabitans* genus (39). We reconstructed 45 unique sequences putatively classified as *Limnohabitans*, 35 of which were phylogenetically placed within the *Limnohabitans* genus (Fig. 1). The majority of sequences were assigned to the LimB lineage ($n = 18$) (39) and the Lhab-A4 tribe ($n = 14$) (28). This was in agreement with an earlier survey that found abundant LimB-related clone sequences in Lake Michigan (40) but is in contrast to the majority of isolate and clone reference sequences from other

TABLE 1 Assembly statistics and genome properties of the putative *Limnohabitans* MAGs described in this study[a]

Putative *Limnohabitans* MAG	Total length (Mbp)	No. of contigs	N_{50} (bp)	Completeness estimate (%)[b]	Mean coverage ± SD	No. of genes	G+C content (%)	Putative lineage	No. of inferred strains	IMG taxon ID
MAG1.FA-MLB-DN	2.39	406	7,798	80.7 (1.8)	8.7 ± 9.4	2,665	60.0	LimDEA	2	2757320395
MAG2.FA-MLB-SN	3.18	207	26,825	97.2 (2.9)	6.2 ± 3.5	3,237	58.0	−[c]	3	2757320402
MAG3.FA-MLB-SN	2.59	313	11,543	90.3 (1.4)	4.2 ± 4.3	2,722	61.7	LimDEA	2	2757320401
MAG4.FA-M110-DN	2.35	188	18,374	91.2 (0.6)	14.5 ± 16.6	2,524	55.1	LimDEA	3	2757320403
MAG5.SP-M110-DD	1.34	290	5,067	66.9 (2.3)	6.9 ± 5.4	1,584	61.4	−	3	2757320396
MAG6.SP-M15-SD	1.99	453	4,874	61.4 (0.2)	11.9 ± 14.6	2,362	59.3	LimDEA	2	2757320404
MAG7.SU-MLB-SD	1.47	390	4,009	59.4 (1.2)	9.4 ± 7.7	1,767	58.3	LimC	2	2757320397
MAG8.SU-M110-DCMD	2.04	425	5,726	78.7 (0.9)	163.5 ± 90.7	2,340	51.8	LimB	6	2757320398
MAG9.SU-M15-SN	2.58	550	5,293	79.0 (1.3)	11.6 ± 9.2	2,930	59.9	LimC	3	2757320399
MAG10.SU-M15-SN	2.76	438	7,688	90.1 (2.3)	12.2 ± 8.1	3,075	59.1	LimDEA	3	2757320400

[a]MAGs were labeled according to the environmental conditions of the sample from which they were reconstructed; putative lineages were based on placement in the phylogenomic tree (see Fig. 2). MAG labels (MAGx.XX-YYY-AB) are comprised of a MAG identifier (MAGx) and environmental information on the sample from which it was reconstructed (XX, season; YYY, site, A, depth; B, day/night). DCM, deep chlorophyll maximum. N_{50}, minimum contig length needed to cover 50% of the MAG.
[b]The percent estimated contamination is indicated in parentheses.
[c]−, uncertain taxonomy within the family *Comamonadaceae* (NCBI) or *Burkholderiaceae* (GTDB).

systems that belong primarily to the LimA, LimD, and LimC lineages (39). This suggests that the LimB lineage may be difficult to cultivate using standard dilution-to-extinction cultivation methods. All reconstructed *Limnohabitans* sequences could be clustered into three OTUs (clustered at >97% similarity), which corresponds to the observations in a previous 16S rRNA gene amplicon survey (29). We found strong indications of microdiversity in the LimB and Lhab-A4 reconstructed sequences since within their respective clades the reconstructed sequences were on average 98.8% similar, which is higher than the currently adopted species threshold of 98.5% (41). Such a large number of highly similar marker gene sequences has been observed before in several other aquatic genera (e.g., *Vibrio* [42]), of which not all the observed genotypes have been shown to clearly correspond to distinct environmental adaptations. In addition, *Limnohabitans* isolates are known to host multiple 16S rRNA gene copies (rrnDB database, accessed on 7 March 2020) (43). Therefore, care must be taken to equate the number of reconstructed 16S rRNA gene sequences with the number of ecologically cohesive populations in the environment. Interestingly, the sole sequence phylogenetically placed in the LimC lineage was 98.0 ± 0.6% similar to those in the LimB lineage and was clustered together with the LimB sequences into a single OTU. These close similarities between lineages in the 16S rRNA gene have led other research groups to evaluate additional marker genes for delineating *Limnohabitans* taxonomic groups (33). Using this multimarker approach, the majority of reported microdiversity has primarily been found in the LimC and LimA lineages, although the LimB lineage has been found to be less diverse but consistently the most abundant and ubiquitous (39, 44, 45). Jezberová et al. argue that the *Limnohabitans* lineage-specific probes may not be sensitive enough to delineate the microdiversity in the LimB lineage (45), suggesting that a genome-resolved approach may be necessary.

Genomic diversity. (i) Lake Michigan and Muskegon Lake contain diverse *Limnohabitans* taxa. We performed a genome-resolved study of the genetic diversity within the *Limnohabitans* taxa of Lake Michigan and its eutrophic estuary Muskegon Lake. Using a genome-centric metagenomic approach, we were able to reconstruct 10 metagenome-assembled genomes (MAGs) of medium-to-high estimated completeness (59 to 97%) and low estimated contamination (<3%). Eight of these were confidently classified into the *Limnohabitans* genus using the MiGA taxonomic annotation pipeline (Table 1) (46). Several attempts to increase the quality of the medium-quality MAGs by modifying the assembly input (i.e., coverage-based normalization, randomized downsampling), assembly set-up (i.e., assembler software Megahit [-meta preset settings]/IDBA-UD and kmer range), or binning parameterization (i.e., minimum contig length and binning parameters) were not successful. The MAG coverage varied over more than an order of magnitude, with MAG8.SU-M110-DCMD, henceforth referred to as MAG8,

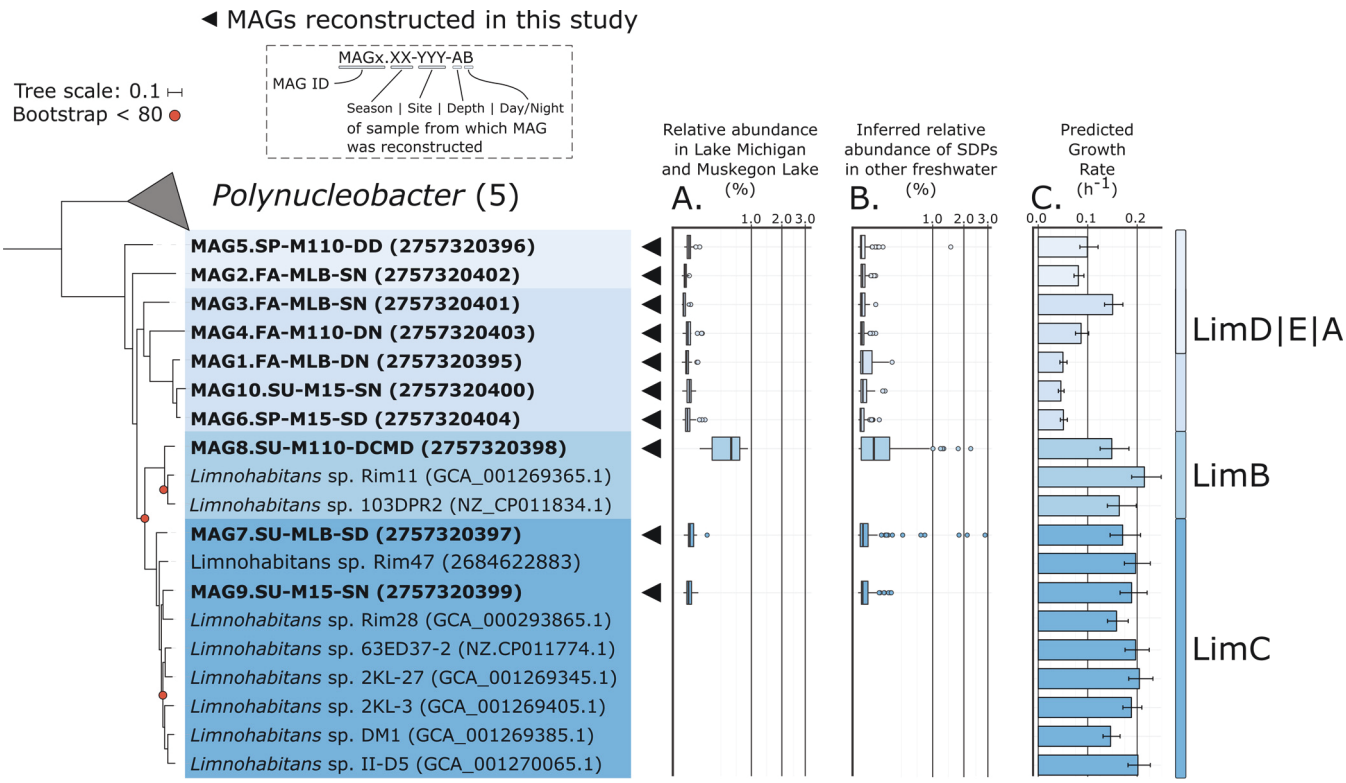

FIG 2 Phylogenomic tree of putative *Limnohabitans* sp. MAGs and available reference *Limnohabitans* genomes. Only bootstrap values of <80 are shown. (A) Boxplots show the normalized relative abundance of each MAG across the 24 samples (square root scale). (B) Inferred (normalized) relative abundances of closely related populations to the reconstructed MAGs in 117 freshwater metagenomic data sets publicly available from NCBI (94.5% identity cutoff). (C) For each genome or MAG, the growth rate was predicted from growth-imprinted genomic traits. The tree was rooted in *Chitinophaga niabensis* (IMG TaxID 2636416022). MAG labels (MAGx.XX-YYY-AB) are comprised of a MAG identifier (MAGx) and environmental information on the sample from which it was reconstructed (XX, season; YYY, site; A, depth; B, day/night). DCM, deep chlorophyll maximum.

being the overall best-covered MAG. Most MAGs had a mean coverage of <10×, highlighting the need for deep sequencing to capture the rare *Limnohabitans* taxa.

A phylogenomic tree based on 37 single-copy conserved core genes was made with the putative *Limnohabitans* MAGs and publicly available *Limnohabitans* isolate genomes (Fig. 2). Two MAGs were placed into the LimC lineage and one in the LimB lineage (≤80% average nucleotide identity [ANI] with reference genomes). The other seven were undefined as there were no reference genomes available for the LimDEA lineages, and two (i.e., MAG5 and MAG2) were part of a separate non-*Limnohabitans* taxonomic group that was not included in the tree. The LimB-placed MAG8 was the most abundant taxon, while all other MAGs were comparable in mean abundance across the sampling stations (Fig. 2A). LimB-associated marker gene sequences have been detected in many river and lake systems and have been typically classified as "generalist" taxa since their (relative) abundance was reported to be invariant to environmental gradients (45).

To assess how prevalent and abundant populations closely related to our reconstructed MAGs in other freshwater systems are, we competitively recruited reads from 117 publicly available metagenomic data sets of American and European river, lake, and reservoir systems to our set of MAGs (Fig. 2B). In concordance with previous findings from marker gene surveys and our Lake Michigan metagenomic survey, populations belonging to the LimB lineage (i.e., closely related to MAG8) are in most freshwater systems the most abundant (45). However, in some systems, populations closely related to other MAGs were also abundant (e.g., MAG7.SU-MLB-SD and MAG1.FA-MLB-DN in Dexter reservoir [Oregon] and Lake Ontario [Ontario, Canada]; see Fig. S2 in the supplemental material for a detailed description of the inferred *Limnohabitans* popu-

lation abundances across freshwater environments). Despite its abundance and ubiquity, the underlying functional and genomic adaptations responsible for the success of the LimB lineage remain unclear.

On average, more than 50% of each putative *Limnohabitans* population was replicating at the time of sampling, as assessed by the index of replication (iRep $>$ 1.5), an inferred measure for the replication state of the genome (47) (Fig. S3). This suggests that both the abundant and rare populations were actively growing and thus metabolically active across the entire estuary-to-pelagic zone gradient. The maximum growth rates of the *Limnohabitans* taxa, as predicted from growth-imprinted genome features, also suggest the conservation of specific growth rates within currently defined *Limnohabitans* lineages (Fig. 2C) (48). Both LimC and LimB lineages had high predicted maximum growth rates, hypothesized to be a necessary trait to survive protozoan grazing and to quickly respond to carbon and nutrient pulses (e.g., as a result of phytoplankton blooms) (39, 49). To assess whether carbon source preference differentiated the MAGs, we evaluated the diversity in known dissolved organic carbon (DOC) transporters in each MAG (50). While we found a large diversity in DOC transporters in all MAGs, only minor differences existed between the genomes (Fig. S4A).

(ii) Accessory genomes of *Limnohabitans* MAGs are enriched in environmental sensing genes. We performed a pangenome analysis using a set of *Limnohabitans* isolate genomes and the Lake Michigan/Muskegon Lake MAGs ($n = 19$) to pinpoint the functional differences between the abundant LimB MAG8, the other MAGs, and available *Limnohabitans* reference genomes. We observed that most MAGs, even those with high completeness, were missing fragments of the core genome present in the available *Limnohabitans* isolate genomes (Fig. S5). We hypothesize that this may be due to a combination of (i) the (known) strain heterogeneity in *Limnohabitans* MAGs, (ii) the rather short insert size of the Nextera library prep (~200 bp), (iii) the high diversity of the microbial community from which the MAGs were reconstructed, and (iv) eco-evolutionary drivers (e.g., streamlining) (51, 52). All of these factors have been shown to directly affect the assembly quality and estimated completeness of MAGs. This necessitates a rather conservative interpretation of the pangenome analysis, including what are core and accessory genes, as we cannot fully account for unassembled MAG gene content.

We tested for gene set enrichments in the inferred accessory genome of each MAG and found that functional genes involved in environmental sensing (e.g., [ABC-type] transport systems) and secretion systems (e.g., type II and VI secretion systems) were the categories almost exclusively enriched in five out of ten MAGs and thus may be a differentiator between each MAG's functional repertoire (Fig. S6 and Table S1). The LimC accessory genomes had no significant enrichment in functional gene categories relative to their genome-wide functional categories. The major difference in the inferred accessory gene content of MAG8 (288 genes) relative to other cooccurring but less abundant MAGs was its specific enrichment in genes for bacterial motility (*pilEF* genes for twitching motility) and chemotaxis (*mcp* and *cheABRW* genes), which were not detected in all other MAGs and reference genomes except for MAG7 (LimC) and *Limnohabitans* sp. strain 63ED37-2 (LimC). The *cheABRW* genes were all located on a single 8.6-kb contig (Ga0224460_1280), while flagellar assembly genes (*flg*, *flh*, and *fli* gene sets) were present in the *Limnohabitans* core genome, enabling active cellular motility. Currently, all available *Limnohabitans* cultures have been phenotyped as nonmotile and nonchemotactic. We found no evidence of chemotaxis genes in other LimC isolate genomes but did find flagellar assembly pathways in the overall core *Limnohabitans* genome, showing that most *Limnohabitans* populations may exhibit motility under certain environmental conditions. Our results thus suggest that the LimB lineage, and possibly also members of the LimC lineage, may consist of motile and chemotactic populations that have yet to be cultured. The *cheABRW* genes in MAG8 were identified as genes putatively acquired through lateral gene transfer according to the Integrated Microbial Genomes annotation workflow, and independent blastx searches against the NCBI nonredundant database identified the closest homology of

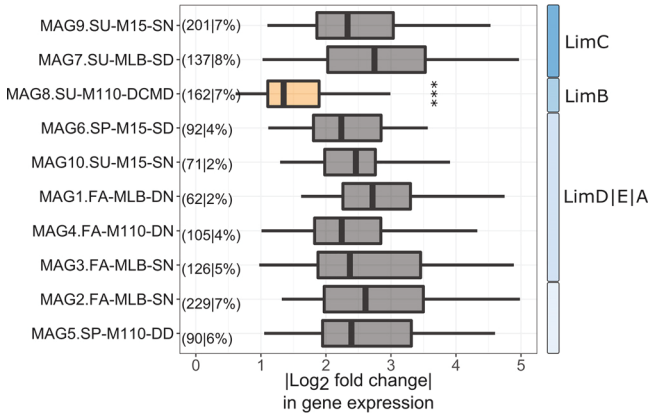

FIG 3 Absolute values of $\log_2$-fold changes of genes differentially expressed (adjusted *P* value of <0.01) between the spring and fall seasons for each MAG (controlled for sampling station). The number and percentage of differentially expressed genes relative to the total DESeq selected genes of each MAG are indicated between parentheses. ***, pairwise Wilcoxon rank sum test (adjusted *P* value of ≤ 0.001) of MAG8 compared to all other MAGs. MAGs were ordered according to their position in the phylogenomic tree. MAG labels (MAGx.XX-YYY-AB) are comprised of a MAG identifier (MAGx) and environmental information on the sample from which it was reconstructed (XX, season; YYY, site; A, depth; B, day/night). DCM, deep chlorophyll maximum.

these genes to those of a freshwater *Chloroflexi* genome (ZSMay80m-chloro-G1, *Anaerolinea* cluster CL500-11) reconstructed from freshwater lakes (Table S2) (53). The genes had lower homology to chemotaxis genes that were detected in *Limnohabitans* sp. 63ED37-2 and MAG7.SU-MLB-SD, highlighting a different origin of these genes in the LimB MAG. We verified the correct binning of this contig to MAG8 and found no evidence of it being a contaminant based on GC content and kmer signatures (Fig. S7). Similarly to *Limnohabitans*, *Chloroflexi* populations have been shown to be abundant in both the hypo- and epilimnion (up to 26% of total bacterial cells). Their abundance and large cell volume suggest that, similarly to *Limnohabitans*, they may be important contributors to freshwater carbon cycling. Our results suggest that future research should investigate the importance of *Chloroflexi-Limnohabitans* interactions to freshwater carbon flow.

MAG8's accessory genome was also significantly enriched, compared to its whole genome, in high-affinity ABC-type transporters for a variety of (in)organic components, such as iron(III), di- and oligopeptides, polyamines, branched-chain amino acids, phosphate, phosphonate, and sulfate. This suggests that the ability to quickly detect and scavenge resource pulses may confer a competitive advantage to this LimB population. A motile lifestyle has been postulated to play an important role in efficient scavenging of substrates with microscale patchiness, which can be found during events of high primary productivity and *Limnohabitans* abundance (54).

(iii) Gene expression by *Limnohabitans* MAG8s across environmental conditions. To gain further insights into what sets the populations represented by MAG8 apart from other, less-abundant ones, we tested whether the *Limnohabitans* populations represented by the reconstructed MAGs were regulating their overall gene expression differently across the sampled environments. Overall, most MAGs showed both season- and site-specific trends in gene expression profiles (Fig. S8), which corresponded to changes in the expression of (branched-chain) amino acid and carbohydrate transporters (Fig. S4B and C).

We assessed changes in gene expression at the surface water across the spring and fall seasons between which the largest environmental changes occurred (while controlling for sampling location [$n = 15$; Data Set S1]). We found that compared to all other cooccurring MAGs, MAG8 exhibited a significantly smaller change in gene expression across 7% of its gene content ($n = 162$, Benjamini-Hochberg adjusted *P* value of ≤ 0.001 [pairwise Wilcoxon rank sum test] [Fig. 3]). Similarly, the genes differ-

entially expressed between the eutrophic (Muskegon Lake) and ultraoligotrophic (M110) stations, when controlled for seasonal effects, also showed a smaller change in gene expression for MAG8 than the other MAGs ($n = 159$, Benjamini-Hochberg adjusted P value of <0.001 [pairwise Wilcoxon rank sum test]). Interestingly, we found that the previously identified chemotaxis genes (*mcp* and *cheABR* genes) were significantly upregulated ($\log_2$-fold change between 2.0 and 3.4) in the low-nutrient environment of the M110 and M15 sampling sites, suggesting that more active motile behavior is facilitating its ability to thrive under the oligotrophic conditions encountered in Lake Michigan. Phosphorus-dependent chemotaxis has been reported to be an important adaptation to oligotrophic conditions (55, 56) and, overall, chemotaxis has recently been found to be an important trait to enable range expansions of microbial populations (57).

In addition, we detected 271 genes in MAG8 (adjusted P value of <0.01) that were differentially expressed between the surface and bottom of Lake Michigan station M110 (summer and fall samples, $n = 6$; Data Set S1). The majority of differentially expressed genes of MAG8 were upregulated ($n = 185$) at the bottom of Lake Michigan. The aerobic carbon monoxide (CO) dehydrogenase medium subunit gene (CoxM) was upregulated at the bottom of Lake Michigan, suggesting that energy generation from aerobic CO oxidation (i.e., methylovory) may be an important trait in this environment. The source of CO in the deep regions of Lake Michigan is unknown, but it could be a result of the incomplete decomposition of humic acids and phenolic compounds in the sediment (58). Our findings regarding C_1 oxidation expression are in line with the expression patterns of a *Chloroflexi* population along the water column at the offshore site in Lake Michigan (59). The expression of several high-affinity branched-chain amino acid genes was also upregulated at the bottom of Lake Michigan during thermal stratification. Although this may indicate that a shift in amino acid preference to branched-chain amino acids could be important, laboratory experiments, although limited in taxonomic resolution, found that amino acid specificity is rather rare in freshwater taxa (60). In combination with the lower (in)organic transporter gene expression for phosphonate, iron(III), and sulfate, these results suggest a lower investment in (in)organic nutrient uptake by the LimB lineage at the bottom of Lake Michigan.

Genomic microdiversity. (i) Sequence discrete populations of *Limnohabitans*. MAGs represent consensus population genomes and can encompass additional strain-level variation (10, 13, 61). In addition to testing the level of functional differentiation between the MAGs, we screened the MAGs for the potential presence of microdiversity based on their metagenomic recruitment profiles (Fig. 4; detailed recruitment profiles are provided in Fig. S9 and S10 in the supplemental material). The presence of distinct local maxima at $<\sim$99% identity indicated the presence of sequence variants in the metagenomic data for which we were unable to reconstruct the genome and are commonly referred to as sequence discrete populations (SDPs) (61, 62). Several MAGs appeared to have multiple SDPs but the strongest support was found in MAG8, which was congruent with the large number of reconstructed 16S rRNA genes for the LimB lineage. In almost every sample we found SDPs of MAG8 confined to the 90 to 97% nucleotide identity range (Fig. S9). The Muskegon Lake site, which was the most diverging environment in terms of nutrient levels, had primarily read recruitment at $<$99% identity, suggesting that in this environment LimB populations more evolutionarily distant from the one represented by MAG8 were present. In addition, this microdiversity may explain the difficulty in fully assembling this high-coverage MAG, as many assembly tools struggle in resolving the genomes from related strains from metagenomic data (63). In our case, normalizing to a lower and more even sample coverage or downsampling to a fixed number of reads did not improve the assembly of MAG8, again suggesting that microdiversity and not excessive coverage is the cause of the poor assembly. The presence of SDPs at various sites and during different seasons indicated the presence of microdiversity in this MAG. However, care must be taken with the interpretation of these density profiles, since the level of smoothing dictated by the

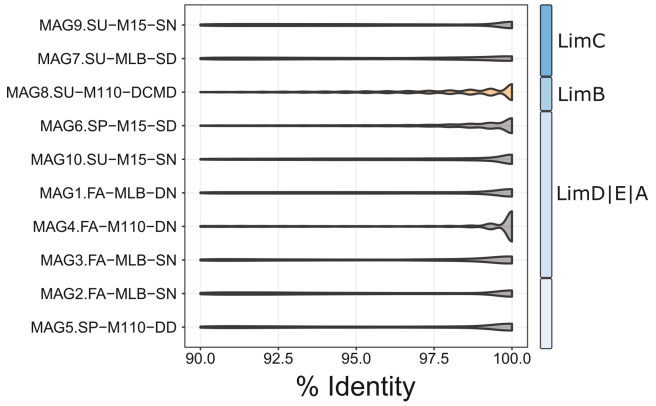

FIG 4 Competitive metagenomic read recruitment to each MAG visualized using violin plots with the bw.nrd0 rule-of-thumb bandwidth. For each MAG, the identity profiles of all sample reads were pooled into a single violin plot. The area of each violin plot was fixed in order to allow visual comparison between MAGs. MAGs were ordered according to their position in the phylogenomic tree. MAG labels (MAGx.XX-YYY-AB) are comprised of a MAG identifier (MAGx) and environmental information of the sample from which it was reconstructed (XX, season; YYY, site; A, depth; B, day/night). DCM, deep chlorophyll maximum.

bandwidth parameter can either overemphasize spurious local maxima (undersmoothing) or instead confound closely related populations (oversmoothing) (64). Therefore, we used and recommend other researchers to use only these profiles as qualitative indicators for the presence of microdiversity and/or closely related populations.

(ii) Microdiversification in *Limnohabitans* MAGs across environmental conditions. In order to determine whether these SDPs were actually indicative of the presence of different strains, we used the DESMAN workflow to infer the number of strains and their abundance based on variant base frequencies and coverage patterns of 36 single-copy core genes (12). A total of 23 putative strains could be resolved from the eight initial *Limnohabitans* MAGs (Table 1), with most MAGs representing at least two strains and MAG8 representing a total of six individual strains. Within each MAG, the strains exhibited both positive and negative correlations with environmental parameters previously associated with *Limnohabitans* marker-gene-based subtype abundances (Fig. 5A) (45). Several strain frequencies were strongly correlated with the total phosphorus concentration, but the majority showed strong correlations with water temperature and thus indirectly with seasonality and sampling site as well. Light availability, as measured by photosynthetically active radiation (PAR), did not strongly correlate with strain frequencies. Two MAGs of the LimDEA lineages (MAG4 and MAG10), two MAGs of the LimC lineage (MAG7 and MAG9), and the LimB MAG (MAG8) had strains that were correlated either positively or negatively with the water temperature. Only strains within the LimB and LimC lineages showed strong correlations with the total phosphorus concentration in the water. In the case of the LimA lineage, preferences toward higher water temperature and specific carbon sources (i.e., allochthonous DOM) have previously been shown (33), which is congruent with our observations on the LimDEA inferred strain abundance. Using network analysis, we found that the delineated strains clustered into two distinct modules of approximately equal numbers of cooccurring strains (Fig. 5B). One network module contained primarily strains with a seasonal frequency increase in Muskegon Lake, while the other contained strains with a seasonal frequency increase in Lake Michigan. The differences in environmental conditions associated with these sampling sites (e.g., trophic state and temperature) appear to have given rise to the existence of two subcommunities of *Limnohabitans* populations with correlating seasonal abundance profiles. Certain MAGs only had strains in a single network module, such as Muskegon Lake for MAG7 and Lake Michigan for MAG3 and MAG10, showing that environmental specificity may already be dictated at the MAG level. The other MAGs had a more balanced distribution over the

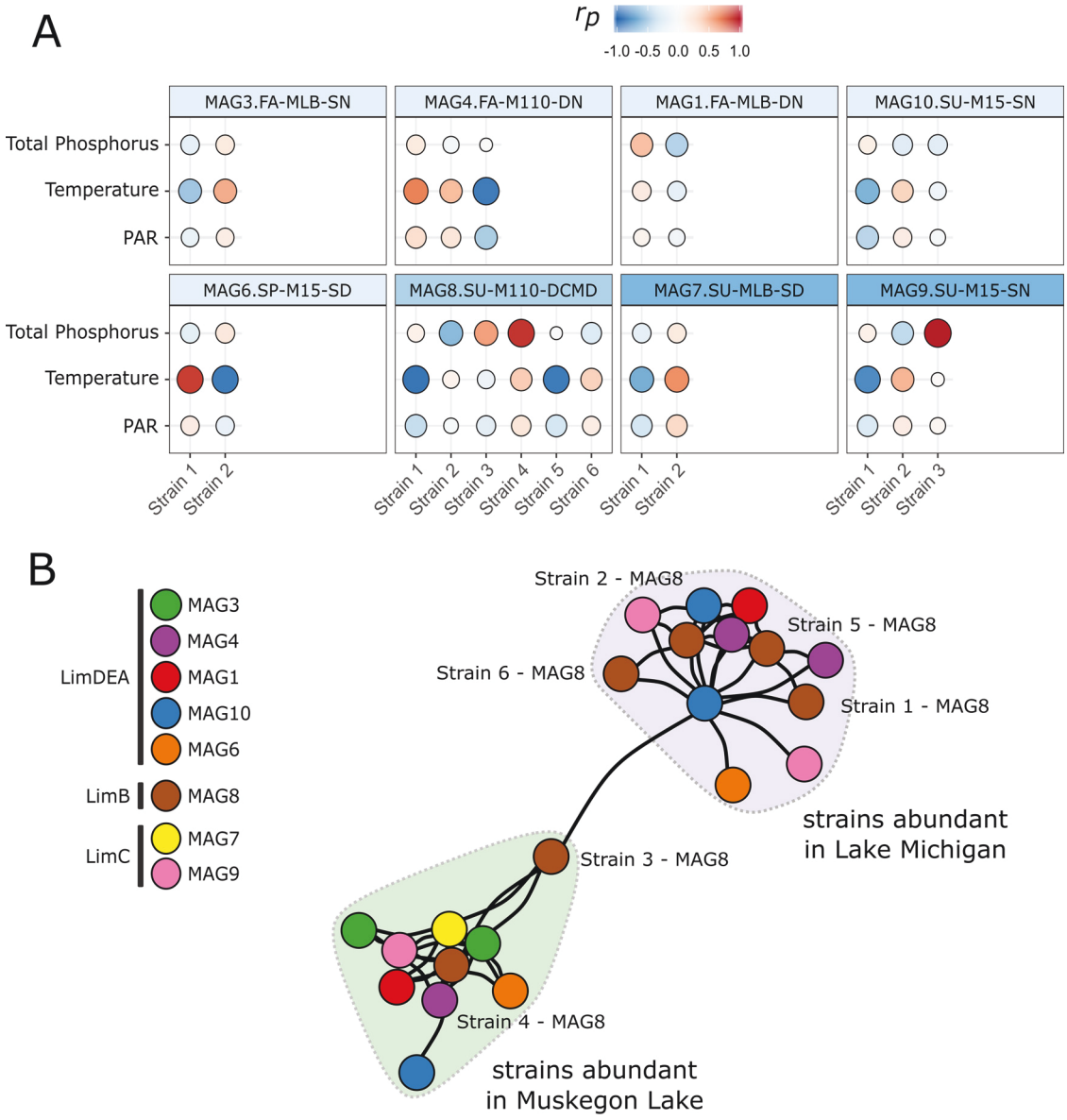

FIG 5 (A) Pearson's correlation (r_p) between *Limnohabitans* MAG frequencies and primary environmental parameters differentiating the environmental gradients of the studied transect. The size of labels is proportional to the correlation strength. PAR, photosynthetically active radiation. (B) Cooccurrence network of all *Limnohabitans* strains using a Fruchterman-Reingold layout. Colored regions highlight identified network modules ($n = 24$).

network modules, and MAG8 had the majority of its strains in the Lake Michigan-associated network module.

Focusing on the strain diversity of MAG8, we found that it was strongly associated with the spatiotemporal components of the data (Fig. 6). The six inferred LimB strains displayed strong environmental specificity, with two strains dominating under the eutrophic conditions found in the Muskegon Lake watershed and four other strains dominating across the more oligotrophic Lake Michigan transect. Based on the SDP analysis, the strains abundant in Muskegon Lake appeared to be the most evolutionarily distant from the reconstructed consensus MAG (i.e., majority of read recruitment at ≤95% identity; Fig. S9). This highlights that the largest evolutionary distance existed between the sets of strains adapted to the different trophic conditions found in Lake Michigan and Muskegon Lake. In addition, all strains had clearly distinct seasonal abundance profiles that were either positively or negatively correlated with the season.

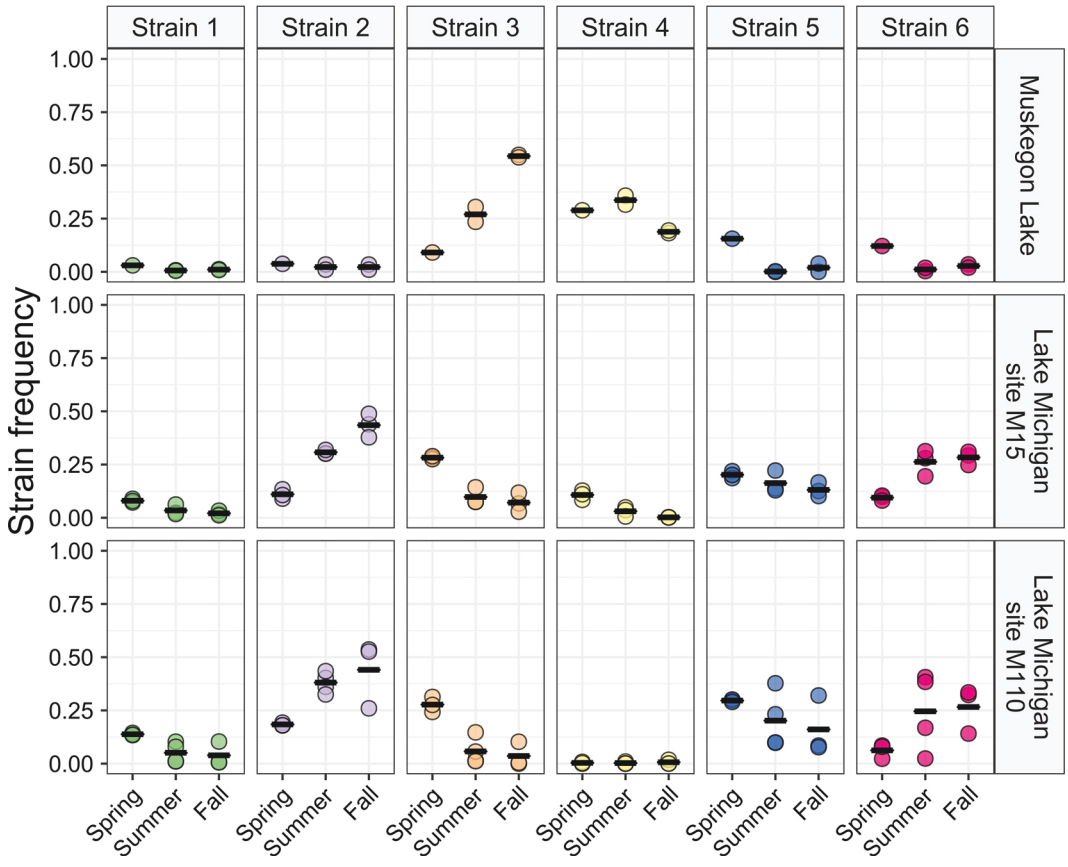

FIG 6 Spatiotemporal strain frequency dynamics of MAG8.SU-M110-DCMD as inferred from DESMAN. Only strains for which inference was robust are shown, and mean frequencies per site and season are indicated by horizontal bars.

For example, five strains were inferred to be present in the spring at both the surface and the deepest region of the oligotrophic M110 site, where temperature and overall environmental conditions remain stable. During the summer and fall seasons, five of the six strains were still present at deepest region of the oligotrophic M110 site, while only three remained at the surface (Fig. S11).

By means of permutational multivariate analysis of variance (PERMANOVA), 16.1% of the variation in strain composition could be explained by the season (spring, summer, or fall), with an additional 25.4% of seasonal variation being conditional on the sampling site (Fig. S12). A total of 12.6% of the variation was explained solely by the sampling site. The strain alpha diversity, as estimated by the inverse Simpson index, was among the highest at the offshore oligotrophic M110 sampling station and, in contrast to other sampling locations, was largely invariant to seasonality (see Fig. S13). Although this is in concordance with the relatively stable environmental conditions at the M110 site, such a broad abundance distribution across depth and nutrient gradients has not yet been reported for other abundant freshwater taxa (e.g., *Actinobacteria* [65] and *Polynucleobacter* [5, 66]). Data from extensive marker gene surveys have shown that other *Limnohabitans* taxa, such as those in the LimA lineage, have a preference for surface waters but can be found at greater depths as well (33, 44). The alpha diversity of the strains was highest in spring and lowest in the fall (approximately 50% reduction). The maximum diversity during spring is in line with previous studies which have reported increased levels of microdiversification during peaks of primary production (45) and higher community diversity levels that have been attributed to the increased resource heterogeneity that may occur early in the season (29). The strain diversity of LimB, as well as that of the other MAGs, appeared to follow the overall taxonomic diversity patterns across the seasons.

It is important to note that the MAGs used in the DESMAN analysis were of medium-to-high completion but also quite fragmented (i.e., large number of contigs), which could impact strain inference accuracy. However, DESMAN is not directly affected by fragmentation, since for the strain inference it considers variant positions on single-copy core genes, which can be located on both short and long contigs. It is well known that strain heterogeneity negatively affects assembly quality (67–69), thereby making it by default more difficult to achieve high-quality MAGs of populations with large strain heterogeneity. The observed level of fragmentation in our MAGs was also inflated by the inclusion of contigs down to 2,000 nucleotides. As indicated by a recent study from our group, it is important to take into account these smaller contigs for correct functional inference (70). We concluded that the judicious binning of the contigs is the most critical step in performing this *de novo* strain inference workflow, which is why we performed and strongly recommend manual curation of MAGs in Anvi'o (71, 72).

Conclusions. We investigated the genomic and functional diversity of the ubiquitous and prevalent *Limnohabitans* taxa across a transect of Lake Michigan, part of the largest freshwater ecosystem in the world. Our findings show that thermal adaptation may be a more important factor in the overall microdiversification within the *Limnohabitans* genus but that specifically for the LimB and LimC lineages, nutrient levels are associated with significantly larger genomic divergence. These findings are corroborated by previous work on *Vibrio* and other *Limnohabitans* taxa showing that closely related strains can have remarkably different environmental adaptations (6, 39, 45). However, due to the strong correlation of environmental conditions and sampling sites described in this study, additional supporting studies with more extensive environmental data will be necessary to avoid missing other potential drivers of microdiversity. Lineage-level expression analysis indicated that, compared to other *Limnohabitans* lineages, LimB displayed smaller shifts in gene expression across the sampled gradients, possibly due to the genomic heterogeneity in strains optimized to specific conditions, and that a shift to methylovory and increased chemotaxis were possible adaptations for handling depth and/or nutrient gradients. The importance of the *Limnohabitans* genus for freshwater food webs has been proposed to be equivalent to that of the SAR11 taxon for marine food webs (45). While marker gene surveys have taught us a great deal about this important group of bacterioplankton, recent genome-centric research on *Limnohabitans* has just started to uncover the extensive metabolic portfolio of this genus (36). Future *in situ* metagenomics studies, targeted cultivation efforts, and synthetic ecology experiments on this genus are needed to enable a detailed understanding of the functional, phenotypic, and ecological implications that are associated with its microdiversification.

MATERIALS AND METHODS

Metagenomic and metatranscriptomic data. Sampling and DNA/RNA extraction from these samples have been described elsewhere (59). Libraries were prepared using the Nextera (summer and fall samples) or TruSeq (spring samples and all metatranscriptomic samples) preparation kits (Illumina, Inc.) and sequenced on a 2-bp × 150-bp paired-end HiSeq 2000 sequencing system. *Sickle* (v1.33.6; https://github.com/najoshi/sickle) was used for removing erroneous and low-quality reads from the data. *Scythe* (v0.993; https://github.com/vsbuffalo/scythe) was used for removing adapter contaminant sequences. Denoised reads were evaluated using FastQC.

16S rRNA gene reconstruction. We reconstructed full-length 16S rRNA gene sequences from quality-trimmed reads using EMIRGE (v0.60.3) (73). EMIRGE was run using the nonredundant 97%-clustered Silva database (v123 [74]) as reference. EMIRGE was run for 65 iterations with the quality-trimmed reads and with an insert size of 200 and a standard deviation of 50, since these were known parameters. The EMIRGE reconstructed sequences were dereplicated, and sequences that fell outside the 1,000- to 1,700-bp range or contained ambiguous bases were discarded. We classified the sequences according to the *Limnohabitans*-specific framework of Šimek et al. (34) by evaluating their phylogenetic placement relative to the reference sequences, as well as by the freshwater classification framework of Newton et al. (28) through the TaxAss pipeline (28, 39, 75). Classification of the sequences using the TaxAss pipeline combines both the Silva v123 database and a manually curated freshwater taxonomy database (FWDB) (75). FWDB classification was favored over the Silva classification if the length-corrected identity to a FWDB sequence was >97% (75). Sequences were clustered into OTUs at an average similarity of 97% using the average nearest-neighbor method in mothur (v1.37) (76).

Genome reconstruction. The reads were dereplicated using a custom Perl script and interleaved into a single sequence file for subsequent assembly (77). Interleaved sequences were digitally normalized using *bbnorm* to a target 60× coverage, and reads with a coverage less than 5× were discarded in order to improve the subsequent assembly step (78, 79). The interleaved reads of all 24 samples were assembled on a per-sample basis into contigs using IDBA-UD (v1.1.3) with kmer lengths varying from 41 to 101 in steps of 10 (80). Contigs were taxonomically classified at the order level by a diamond search against the nonredundant NCBI protein database using DESMAN scripts (12) (*classify_contigNR.pl* with MIN_FRACTION = 0.1), after which the classification output files were formatted into an annotation file compatible with the Vizbin binning tool (81). Initial metagenome-assembled genomes (MAGs) of all classified *Betaproteobacteria* (following NCBI taxonomy) were retrieved through a manual binning strategy in Vizbin (v0.9, default settings, minimum contig length = 2,000 bp) (81). Quality-trimmed raw reads were mapped to each individual assembly using *bwa-mem* (v0.7.8) on default settings (82). SAMtools (v1.3.1) was used to convert, sort, and index the SAM files (83). From this initial set of 92 population genomes, we extracted the unique representative genomes by taking into account the average nucleotide identity (ANI; *pyani* v0.2.7; ANIb method) between MAGs, as well as the completeness statistics inferred from a CheckM analysis (v1.07) (84). If the ANI was >99% between genomes and the contamination <10%, the MAG with the highest completeness was chosen as representative. This approach reduced the number of MAGs from 92 betaproteobacterium MAGs to 10 putative *Limnohabitans* MAGs. Next, all sample reads were competitively mapped to the putative *Limnohabitans* MAGs. The Anvi'o platform (v2.3.0) was then used to manually refine the unique MAGs identified through Vizbin by evaluating differential coverage patterns across the samples (71). Completeness and contamination estimates of the final MAGs were estimated by CheckM. Mean relative MAG abundances, normalized for observed genome size, were calculated from the mean coverages exported from Anvi'o as follows:

$$\text{Mean relative abundance} = \frac{\text{mean coverage}}{\text{read length (bp)} \times \text{no. of sample reads} \times \text{bin size (bp)}}$$

The presence of closely related populations (sequence discrete populations [SDPs]) in publicly available data sets ($n = 117$) used by Neuenschwander et al. (65) was assessed by blast searching one million reads of each data set to the 10 putative *Limnohabitans* MAGs and assessing the read alignment across the contigs of the MAGs. The relative abundance of SDPs was inferred by normalizing the number of reads that aligned with >94.5% identity, with the genome sizes of the corresponding MAGs (in Mbp), and the total number of reads mapped (i.e., one million).

The refined MAGs were submitted for gene calling and annotation to the Joint Genome Institute's Integrated Microbial Genomes isolate annotation pipeline (85). For each MAG and selected publicly available freshwater genomes, the minimal generation time (MGT) and optimal growth temperature were predicted based on "growth-imprinted" genome features by means of the Growthpred (v1.07) software (48). The predicted specific growth rates (μ_{spec}) were then calculated by using the following formula:

$$\mu_{spec} = \frac{\ln(2)}{\text{MGT}}$$

In addition, we calculated the index of replication (iRep), a measure of the population-averaged replication status at the time of sampling, for each MAG (47). Sequence discrete populations were examined by competitively mapping one million reads per sample (blastn v2.25 [86]) and evaluating the nucleotide identity profiles of the best hits for each MAG.

Phylogenetic and phylogenomic tree construction. A 16S rRNA gene phylogenetic tree was constructed using a set of publicly available betI-clade, *Limnohabitans* genus, and other related lineage sequences. Sequences were aligned using the SINA aligner (87). The tree was constructed with Fasttree (v2.1.9) using the GTR+CAT evolutionary model and run in sensitive mode (-spr 4 -mlacc 2 –slownni options) (88). The phylogenomic tree was constructed based on the codon alignment of a set of 37 conserved marker genes through Phylosift (89). The tree was generated from the codon alignment by means of RAxML (v8.2.8) with the GTRGAMMA model and 1,000 bootstraps (90). Both trees were visualized and annotated in iTOL (91), exported, and further annotated in Inkscape (v0.91).

Strain inference. We used the standard workflow of the DESMAN software for (i) inferring strain frequencies and (ii) assigning accessory genomes to robust strains in the MAG of interest (12). Briefly, gene calling was performed with *prodigal* (v2.6.3) (92), and single-copy core genes (SCCGs) were detected by means of a reversed-position-specific blast (RPS-blast) search against a SCCG *E. coli* reference database provided with DESMAN. Variant positions were then found using the base count frequencies of these SCCGs with the *Variant_filter.py* script (*-p* for 1d optimization of minor variant frequency detection). Samples were required to have a coverage >1× (*-m*). SCCGs were filtered based on their median coverage (*-c*) and a maximum divergence of 2.5 on the median coverage was allowed for each SCCG. The initial coverage cutoff was set at 25.0 (*-f* flag) and SCCGs were filtered if <80% of the samples passed the coverage filter (*-sf* flag). The identified variant positions were then used as input for the DESMAN haplotype inference tool. We ran 10 replicate runs of DESMAN for up to 12 haplotypes. The model performance for each run is available in Fig. S14 in the supplemental material. The final robust number of strains was determined by using the *resolvenhap.py* script; the average single-nucleotide variation error was <6% (see Table S3). The scripts used to perform this analysis are available at https://github.com/rprops/MetaG_analysis_workflow/wiki/21.-DESMAN. For inference on the abundance of each inferred strain, the Gamma output files were used. Beta-diversity analysis was conducted by principal coordinate analysis (PCoA), and PERMANOVA was used to associate environmental variables

with shifts in strain composition. Strain alpha diversity was estimated by the Hill diversity of order 2 (inverse Simpson index).

Network analysis. Cooccurrence networks on presumed *Limnohabitans* strains were inferred using SparCC as implemented in the SpiecEasi package (v1.0.6) (93, 94). The input data consisted of the strain frequency data multiplied by the normalized abundance of each MAG. A total of 100 iterations were run on both the outer and inner loops, and a correlation threshold of 0.2 on the Pearson's correlation values was put on the inner loop. The graphs were visualized using the *igraph* package (v1.2.4.1) for all correlations with an absolute value of $>$0.3, and empty nodes were discarded. Network modules were identified via a spin-glass model and simulated annealing as implemented in *igraph* under default settings.

Expression analysis. Denoised metatranscriptomic reads were competitively mapped to the set of unique *Betaproteobacteria* MAGs. The count table was extracted from the mapping files using *pileup.sh*, available from bbtools (https://jgi.doe.gov/data-and-tools/bbtools/). Transcript counts were normalized using the TMM with singleton pairing method (TMMwsp) prior to principal component analysis (95). Testing for differential expression of genes in the *Limnohabitans* MAGs was performed with DESeq2 using the raw count data (v1.18.1) on default settings (96). Genes were considered differentially expressed if their Benjamini-Hochberg-corrected *P* value was $<$0.01.

Pangenome analysis. We applied the default pangenome analysis workflow available in Anvi'o (v2.3.0) to all *Limnohabitans* genomes (71). Gene calling was performed with *prodigal* (v2.6.2), amino acid sequences were aligned with *muscle* (v3.8.31) (97), amino acid similarities were calculated with blastp (v2.2.29), and sequences were clustered with *mcl* (v14-137) (98). Accessory genomes were manually binned in Anvi'o (Fig. S5). Since there were no up-to-date IMG annotation projects available for the other *Limnohabitans* genomes, the gene calls from Anvi'o were exported and annotated with KEGG orthology identifiers using the BlastKOALA web server (TaxID 665874; species_prokaryotes database) to facilitate direct comparison between genomes (99). Enrichment of functional categories (e.g., KEGG subsystems) was conducted with hypergeometric tests available in the *clusterProfiler* package (v3.6.0) (100). *P* values were adjusted for multiple testing with the Benjamini-Hochberg correction.

Data availability. Supplementary information is also available at https://doi.org/10.6084/m9.figshare.10265165. Raw sequence reads are available from the JGI portal under the project IDs specified in Table S4 posted at https://doi.org/10.6084/m9.figshare.10265165. MAGs and their genome annotations are available from IMG under the taxon IDs specified in Table 1. The data analysis workflow is publicly available at https://github.com/rprops/AEM_Props2020. Pangenome analysis input and output files are available from https://doi.org/10.6084/m9.figshare.7547159. The Anvi'o profile database (v2.3.0) used to refine the MAGs is available from https://doi.org/10.6084/m9.figshare.7547852.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 3.2 MB.

SUPPLEMENTAL FILE 2, CSV file, 5.3 MB.

ACKNOWLEDGMENTS

We thank the crew on the R/V *Laurentian* and Ann McCarthy and Marian Schmidt for sampling of Lake Michigan and Muskegon Lake, and we thank Ann McCarthy for the DNA and RNA extractions.

Part of this work was supported by funding to V.J.D. by the Community Sequencing Program (U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, supported under contract DE-AC02-05CH11231), the National Science Foundation (award 1737680), and the University of Michigan. R.P. was supported by Ghent University (BOFDOC2015000601) and a Sofina Gustave-Boël grant from the Belgian American Educational Foundation.

The authors declare that there are no conflicts of interest.

REFERENCES

1. Schmidt ML, White JD, Denef VJ. 2016. Phylogenetic conservation of freshwater lake habitat preference varies between abundant bacterioplankton phyla. Environ Microbiol 18:1212–1226. https://doi.org/10.1111/1462-2920.13143.
2. Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine MC, Kane M, Krumins JA, Kuske CR, Morin PJ, Naeem S, Øvreås L, Reysenbach A-L, Smith VH, Staley JT. 2006. Microbial biogeography: putting microorganisms on the map. Nat Rev Microbiol 4:102–112. https://doi.org/10.1038/nrmicro1341.
3. Schmidt VT, Reveillaud J, Zettler E, Mincer TJ, Murphy L, Amaral-Zettler LA. 2014. Oligotyping reveals community level habitat selection within the genus *Vibrio*. Front Microbiol 5:563. https://doi.org/10.3389/fmicb.2014.00563.
4. Hunt DE, David LA, Gevers D, Preheim SP, Alm EJ, Polz MF. 2008. Resource partitioning and sympatric differentiation among closely related bacterioplankton. Science 320:1081–1085. https://doi.org/10.1126/science.1157890.
5. Sangwan N, Zarraonaindia I, Hampton-Marcell JT, Ssegane H, Eshoo TW, Rijal G, Negri MC, Gilbert JA. 2016. Differential functional constraints cause strain-level endemism in polynucleobacter populations. mSystems 1:e00003-16. https://doi.org/10.1128/mSystems.00003-16.
6. Yung CM, Vereen MK, Herbert A, Davis KM, Yang J, Kantorowska A, Ward CS, Wernegreen JJ, Johnson ZI, Hunt DE. 2015. Thermally adaptive tradeoffs in closely related marine bacterial strains. Environ Microbiol 17:2421–2429. https://doi.org/10.1111/1462-2920.12714.

7. Fuhrman JA, Campbell L. 1998. Microbial microdiversity. Nature 393:410–411. https://doi.org/10.1038/30839.
8. Larkin AA, Martiny AC. 2017. Microdiversity shapes the traits, niche space, and biogeography of microbial taxa. Environ Microbiol Rep 9:55–70. https://doi.org/10.1111/1758-2229.12523.
9. Jezbera J, Jezberová J, Brandt U, Hahn MW. 2011. Ubiquity of *Polynucleobacter necessarius* subspecies asymbioticus results from ecological diversification. Environ Microbiol 13:922–931. https://doi.org/10.1111/j.1462-2920.2010.02396.x.
10. Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, Morrison HG, Sogin ML. 2013. Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. Methods Ecol Evol 4:1111–1119. https://doi.org/10.1111/2041-210X.12114.
11. 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.
12. Quince C, Delmont TO, Raguideau S, Alneberg J, Darling AE, Collins G, Eren AM. 2017. DESMAN: a new tool for de novo extraction of strains from metagenomes. Genome Biol 18:181. https://doi.org/10.1186/s13059-017-1309-9.
13. Luo C, Knight R, Siljander H, Knip M, Xavier RJ, Gevers D. 2015. ConStrains identifies microbial strains in metagenomic datasets. Nat Biotechnol 33:1045–1052. https://doi.org/10.1038/nbt.3319.
14. Nayfach S, Rodriguez-Mueller B, Garud N, Pollard KS. 2016. An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. Genome Res 26:1612–1625. https://doi.org/10.1101/gr.201863.115.
15. Jones SE, Newton RJ, McMahon KD. 2009. Evidence for structuring of bacterial community composition by organic carbon source in temperate lakes. Environ Microbiol 11:2463–2472. https://doi.org/10.1111/j.1462-2920.2009.01977.x.
16. Newton RJ, Jones SE, Helmus MR, McMahon KD. 2007. Phylogenetic ecology of the freshwater Actinobacteria acI lineage. Appl Environ Microbiol 73:7169–7176. https://doi.org/10.1128/AEM.00794-07.
17. Moore LR, Rocap G, Chisholm SW. 1998. Physiology and molecular phylogeny of coexisting Prochlorococcus ecotypes. Nature 393:464–467. https://doi.org/10.1038/30965.
18. Shade A. 2017. Diversity is the question, not the answer. ISME J 11:1–6. https://doi.org/10.1038/ismej.2016.118.
19. Shade A, Peter H, Allison SD, Baho DL, Berga M, Burgmann H, Huber DH, Langenheder S, Lennon JT, Martiny JBH, Matulich KL, Schmidt TM, Handelsman J. 2012. Fundamentals of microbial community resistance and resilience. Front Microbiol 3:417. https://doi.org/10.3389/fmicb.2012.00417.
20. Williamson CE, Dodds W, Kratz TK, Palmer MA. 2008. Lakes and streams as sentinels of environmental change in terrestrial and atmospheric processes. Front Ecol Environ 6:247–254. https://doi.org/10.1890/070140.
21. Allan JD, McIntyre PB, Smith SDP, Halpern BS, Boyer GL, Buchsbaum A, Burton GA, Campbell LM, Chadderton WL, Ciborowski JJH, Doran PJ, Eder T, Infante DM, Johnson LB, Joseph CA, Marino AL, Prusevich A, Read JG, Rose JB, Rutherford ES, Sowa SP, Steinman AD. 2013. Joint analysis of stressors and ecosystem services to enhance restoration effectiveness. Proc Natl Acad Sci U S A 110:372–377. https://doi.org/10.1073/pnas.1213841110.
22. Turschak BA, Bunnell D, Czesny S, Hook TO, Janssen J, Warner D, Bootsma HA. 2014. Nearshore energy subsidies support Lake Michigan fishes and invertebrates following major changes in food web structure. Ecology 95:1243–1252. https://doi.org/10.1890/13-0329.1.
23. Vanderploeg HA, Bunnell DB, Carrick HJ, Höök TO. 2015. Complex interactions in Lake Michigan's rapidly changing ecosystem. J Great Lakes Res 41:1–6. https://doi.org/10.1016/j.jglr.2015.11.001.
24. Denef VJ, Carrick HJ, Cavaletto J, Chiang E, Johengen TH, Vanderploeg HA. 2017. Lake bacterial assemblage composition is sensitive to biological disturbance caused by an invasive filter feeder. mSphere 2:e00189-17. https://doi.org/10.1128/mSphere.00189-17.
25. Props R, Schmidt ML, Heyse J, Vanderploeg HA, Boon N, Denef VJ. 2018. Flow cytometric monitoring of bacterioplankton phenotypic diversity predicts high population-specific feeding rates by invasive dreissenid mussels. Environ Microbiol 20:521–534. https://doi.org/10.1111/1462-2920.13953.
26. Vanderploeg HA, Liebig JR, Carmichael WW, Agy MA, Johengen TH, Fahnenstiel GL, Nalepa TF. 2001. Zebra mussel (*Dreissena polymorpha*) selective filtration promoted toxic *Microcystis* blooms in Saginaw Bay (Lake Huron) and Lake Erie. Can J Fish Aquat Sci 58:1208–1221. https://doi.org/10.1139/f01-066.
27. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol 36:996–1004. https://doi.org/10.1038/nbt.4229.
28. 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. https://doi.org/10.1128/MMBR.00028-10.
29. Fujimoto M, Cavaletto J, Liebig JR, McCarthy A, Vanderploeg HA, Denef VJ. 2016. Spatiotemporal distribution of bacterioplankton functional groups along a freshwater estuary to pelagic gradient in Lake Michigan. J Great Lakes Res 42:1036–1048. https://doi.org/10.1016/j.jglr.2016.07.029.
30. Šimek K, Kasalický V, Jezbera J, Jezberová J, Hejzlar J, Hahn MW. 2010. Broad habitat range of the phylogenetically narrow R-BT065 cluster, representing a core group of the betaproteobacterial genus *Limnohabitans*. Appl Environ Microbiol 76:631–639. https://doi.org/10.1128/AEM.02203-09.
31. Salcher MM. 2014. Same same but different: ecological niche partitioning of planktonic freshwater prokaryotes. J Limnol 73(s1):74–87. https://doi.org/10.4081/jlimnol.2014.813.
32. Simek K, Kasalický V, Zapomělová E, Hornák K. 2011. Alga-derived substrates select for distinct betaproteobacterial lineages and contribute to niche separation in *Limnohabitans* strains. Appl Environ Microbiol 77:7307–7315. https://doi.org/10.1128/AEM.05107-11.
33. Shabarova T, Kasalický V, Šimek K, Nedoma J, Znachor P, Posch T, Pernthaler J, Salcher MM. 2017. Distribution and ecological preferences of the freshwater lineage LimA (genus *Limnohabitans*) revealed by a new double hybridization approach. Environ Microbiol 19:1296–1309. https://doi.org/10.1111/1462-2920.13663.
34. Šimek K, Kasalický V, Jezbera J, Horňák K, Nedoma J, Hahn MW, Bass D, Jost S, Boenigk J. 2013. Differential freshwater flagellate community response to bacterial food quality with a focus on *Limnohabitans* bacteria. ISME J 7:1519–1530. https://doi.org/10.1038/ismej.2013.57.
35. Horňák K, Kasalický V, Šimek K, Grossart H-P. 2017. Strain-specific consumption and transformation of alga-derived dissolved organic matter by members of the *Limnohabitans*-C and *Polynucleobacter*-B clusters of *Betaproteobacteria*. Environ Microbiol 19:4519–4535. https://doi.org/10.1111/1462-2920.13900.
36. Kasalicky V, Zeng Y, Piwosz K, Simek K, Kratochvilova H, Koblizek M. 2018. Aerobic anoxygenic photosynthesis is commonly present within the genus *Limnohabitans*. Appl Environ Microbiol 84:e02116-17. https://doi.org/10.1128/AEM.02116-17.
37. Zeng Y, Kasalický V, Šimek K, Koblížek 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. https://doi.org/10.1128/JB.01481-12.
38. Bundy MH, Vanderploeg HA, Lavrentyev PJ, Kovalcik PA. 2005. The importance of microzooplankton versus phytoplankton to copepod populations during late winter and early spring in Lake Michigan. Can J Fish Aquat Sci 62:2371–2385. https://doi.org/10.1139/f05-111.
39. Kasalicky V, Jezbera J, Hahn MW, Simek K. 2013. The diversity of the *Limnohabitans* genus, an important group of freshwater bacterioplankton, by characterization of 35 isolated strains. PLoS One 8:0058209. https://doi.org/10.1371/journal.pone.0058209.
40. Mueller-Spitz SR, Goetz GW, McLellan SL. 2009. Temporal and spatial variability in nearshore bacterioplankton communities of Lake Michigan. FEMS Microbiol Ecol 67:511–522. https://doi.org/10.1111/j.1574-6941.2008.00639.x.
41. Rodriguez-R LM, Castro JC, Kyrpides NC, Cole JR, Tiedje JM, Konstantinidis KT. 2018. How much do rRNA gene surveys underestimate extant bacterial diversity? Appl Environ Microbiol 84:e00014-18. https://doi.org/10.1128/AEM.00014-18.
42. Thompson JR, Pacocha S, Pharino C, Klepac-Ceraj V, Hunt DE, Benoit J, Sarma-Rupavtarm R, Distel DL, Polz MF. 2005. Genotypic diversity within a natural coastal bacterioplankton population. Science 307:1311–1313. https://doi.org/10.1126/science.1106028.
43. Stoddard SF, Smith BJ, Hein R, Roller BR, Schmidt TM. 2015. rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. Nucleic Acids Res 43(Database Issue):D593–D598. https://doi.org/10.1093/nar/gku1201.
44. Jezbera J, Jezberová J, Kasalický V, Šimek K, Hahn MW. 2013. Patterns

of *Limnohabitans* microdiversity across a large set of freshwater habitats as revealed by reverse line blot hybridization. PLoS One 8:e58527. https://doi.org/10.1371/journal.pone.0058527.
45. Jezberova J, Jezbera J, Znachor P, Nedoma J, Kasalicky V, Simek K. 2017. The *Limnohabitans* genus harbors generalistic and opportunistic subtypes: evidence from spatiotemporal succession in a canyon-shaped reservoir. Appl Environ Microbiol 83:e01530-17. https://doi.org/10.1128/AEM.01530-17.
46. Rodriguez RL, Gunturu S, Harvey WT, Rossello-Mora R, Tiedje JM, Cole JR, Konstantinidis KT. 2018. The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of *Archaea* and *Bacteria* at the whole genome level. Nucleic Acids Res 46:W282–W288. https://doi.org/10.1093/nar/gky467.
47. Brown CT, Olm MR, Thomas BC, Banfield JF. 2016. Measurement of bacterial replication rates in microbial communities. Nat Biotechnol 34:1256–1263. https://doi.org/10.1038/nbt.3704.
48. Vieira-Silva S, Rocha E. 2010. The systemic imprint of growth and its uses in ecological (meta)genomics. PLoS Genet 6:e1000808. https://doi.org/10.1371/journal.pgen.1000808.
49. Simek K, Kasalický V, Hornák K, Hahn MW, Weinbauer MG. 2010. Assessing niche separation among coexisting *Limnohabitans* strains through interactions with a competitor, viruses, and a bacterivore. Appl Environ Microbiol 76:1406–1416. https://doi.org/10.1128/AEM.02517-09.
50. 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. https://doi.org/10.1111/j.1462-2920.2009.02102.x.
51. Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, Bibbs L, Eads J, Richardson TH, Noordewier M, Rappé MS, Short JM, Carrington JC, Mathur EJ. 2005. Genome streamlining in a cosmopolitan oceanic bacterium. Science 309:1242–1245. https://doi.org/10.1126/science.1114057.
52. Konstantinidis KT, Tiedje JM. 2005. Genomic insights that advance the species definition for prokaryotes. Proc Natl Acad Sci U S A 102:2567–2572. https://doi.org/10.1073/pnas.0409727102.
53. Mehrshad M, Salcher MM, Okazaki Y, Nakano S-i, Šimek K, Andrei A-S, Ghai R. 2018. Hidden in plain sight: highly abundant and diverse planktonic freshwater *Chloroflexi*. Microbiome 6:176. https://doi.org/10.1186/s40168-018-0563-8.
54. Pernthaler J. 2017. Competition and niche separation of pelagic bacteria in freshwater habitats. Environ Microbiol 19:2133–2150. https://doi.org/10.1111/1462-2920.13742.
55. Props R, Monsieurs P, Vandamme P, Leys N, Denef VJ, Boon N. 2019. Gene expansion and positive selection as bacterial adaptations to oligotrophic conditions. mSphere 4:e00011-19. https://doi.org/10.1128/mSphereDirect.00011-19.
56. Hutz A, Schubert K, Overmann J. 2011. *Thalassospira* sp. isolated from the oligotrophic eastern Mediterranean Sea exhibits chemotaxis toward inorganic phosphate during starvation. Appl Environ Microbiol 77:4412–4421. https://doi.org/10.1128/AEM.00490-11.
57. Cremer J, Honda T, Tang Y, Wong-Ng J, Vergassola M, Hwa T. 2019. Chemotaxis as a navigation strategy to boost range expansion. Nature 575:658–663. https://doi.org/10.1038/s41586-019-1733-y.
58. Hoshino T, Inagaki F. 2017. Distribution of anaerobic carbon monoxide dehydrogenase genes in deep subseafloor sediments. Lett Appl Microbiol 64:355–363. https://doi.org/10.1111/lam.12727.
59. Denef VJ, Mueller RS, Chiang E, Liebig JR, Vanderploeg HA. 2015. *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. https://doi.org/10.1128/AEM.03014-15.
60. Ricão Canelhas M, Eiler A, Bertilsson S. 2016. Are freshwater bacterioplankton indifferent to variable types of amino acid substrates? FEMS Microbiol Ecol 92:fiw005. https://doi.org/10.1093/femsec/fiw005.
61. Garcia SL, Stevens SLR, Crary B, Martinez-Garcia M, Stepanauskas R, Woyke T, Tringe SG, Andersson SGE, Bertilsson S, Malmstrom RR, McMahon KD. 2018. Contrasting patterns of genome-level diversity across distinct co-occurring bacterial populations. ISME J 12:742–755. https://doi.org/10.1038/s41396-017-0001-0.
62. Caro-Quintero A, Konstantinidis KT. 2012. Bacterial species may exist, metagenomics reveal. Environ Microbiol 14:347–355. https://doi.org/10.1111/j.1462-2920.2011.02668.x.
63. Gregor I, Schönhuth A, McHardy AC. 2016. Snowball: strain aware gene assembly of metagenomes. Bioinformatics 32:i649–i657. https://doi.org/10.1093/bioinformatics/btw426.
64. Heidenreich N-B, Schindler A, Sperlich S. 2013. Bandwidth selection for kernel density estimation: a review of fully automatic selectors. AStA Adv Stat Anal 97:403–433. https://doi.org/10.1007/s10182-013-0216-y.
65. Neuenschwander SM, Ghai R, Pernthaler J, Salcher MM. 2018. Microdiversification in genome-streamlined ubiquitous freshwater *Actinobacteria*. ISME J 12:185–198. https://doi.org/10.1038/ismej.2017.156.
66. Hoetzinger M, Schmidt J, Jezberová J, Koll U, Hahn MW. 2017. Microdiversification of a pelagic polynucleobacter species is mainly driven by acquisition of genomic islands from a partially interspecific gene pool. Appl Environ Microbiol 83:e02266-16. https://doi.org/10.1128/AEM.02266-16.
67. Nelson WC, Maezato Y, Wu Y-W, Romine MF, Lindemann SR. 2016. Identification and resolution of microdiversity through metagenomic sequencing of parallel consortia. Appl Environ Microbiol 82:255–267. https://doi.org/10.1128/AEM.02274-15.
68. Martinez-Hernandez F, Fornas O, Lluesma Gomez M, Bolduc B, de la Cruz Peña MJ, Martínez JM, Anton J, Gasol JM, Rosselli R, Rodriguez-Valera F, Sullivan MB, Acinas SG, Martinez-Garcia M. 2017. Single-virus genomics reveals hidden cosmopolitan and abundant viruses. Nat Commun 8:15892. https://doi.org/10.1038/ncomms15892.
69. Allen EE, Banfield JF. 2005. Community genomics in microbial ecology and evolution. Nat Rev Microbiol 3:489–498. https://doi.org/10.1038/nrmicro1157.
70. Jackrel SL, White JD, Evans JT, Buffin K, Hayden K, Sarnelle O, Denef VJ. 2019. Genome evolution and host-microbiome shifts correspond with intraspecific niche divergence within harmful algal bloom-forming *Microcystis aeruginosa*. Mol Ecol 28:3994–4011. https://doi.org/10.1111/mec.15198.
71. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ 3:e1319. https://doi.org/10.7717/peerj.1319.
72. Shaiber A, Eren AM. 2019. Composite metagenome-assembled genomes reduce the quality of public genome repositories. mBio 10:e00725-19. https://doi.org/10.1128/mBio.00725-19.
73. Miller CS, Baker BJ, Thomas BC, Singer SW, Banfield JF. 2011. EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. Genome Biol 12:R44. https://doi.org/10.1186/gb-2011-12-5-r44.
74. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 35:7188–7196. https://doi.org/10.1093/nar/gkm864.
75. Rohwer RR, Hamilton JJ, Newton RJ, McMahon KD. 2017. TaxAss: leveraging custom databases achieves fine-scale taxonomic resolution. bioRxiv https://www.biorxiv.org/content/10.1101/214288v2.
76. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75 https://doi.org/10.1128/AEM.01541-09.
77. Li M, Baker BJ, Anantharaman K, Jain S, Breier JA, Dick GJ. 2015. Genomic and transcriptomic evidence for scavenging of diverse organic compounds by widespread deep-sea archaea. Nat Commun 6:8933. https://doi.org/10.1038/ncomms9933.
78. Titus Brown C, Howe A, Zhang Q, Pyrkosz AB, Brom TH. 2012. A reference-free algorithm for computational normalization of shotgun sequencing data. arXiv:1203.4802.
79. Hug LA, Thomas BC, Sharon I, Brown CT, Sharma R, Hettich RL, Wilkins MJ, Williams KH, Singh A, Banfield JF. 2016. Critical biogeochemical functions in the subsurface are associated with bacteria from new phyla and little studied lineages. Environ Microbiol 18:159–173. https://doi.org/10.1111/1462-2920.12930.
80. Peng Y, Leung HC, Yiu SM, Chin FY. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28:1420–1428. https://doi.org/10.1093/bioinformatics/bts174.
81. Laczny CC, Sternal T, Plugaru V, Gawron P, Atashpendar A, Margossian HH, Coronado S, der Maaten L, Vlassis N, Wilmes P. 2015. VizBin: an application for reference-independent visualization and human-augmented binning of metagenomic data. Microbiome 3:1. https://doi.org/10.1186/s40168-014-0066-1.
82. Li H, Durbin R. 2010. Fast and accurate long-read alignment with

Burrows-Wheeler transform. Bioinformatics 26:589–595. https://doi.org/10.1093/bioinformatics/btp698.
83. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
84. 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. https://doi.org/10.1101/gr.186072.114.
85. Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Tennessen K, Palaniappan K, Szeto E, Pillay M, Chen IMA, Pati A, Nielsen T, Markowitz VM, Kyrpides NC. 2016. The standard operating procedure of the DOE-JGI Metagenome Annotation Pipeline (MAP v.4). Stand Genomic Sci 11:17. https://doi.org/10.1186/s40793-016-0138-x.
86. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. https://doi.org/10.1093/nar/25.17.3389.
87. Pruesse E, Peplies J, Glöckner FO. 2012. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics 28:1823–1829. https://doi.org/10.1093/bioinformatics/bts252.
88. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2: approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490.
89. Darling AE, Jospin G, Lowe E, Matsen F, Bik HM, Eisen JA. 2014. PhyloSift: phylogenetic analysis of genomes and metagenomes. PeerJ 2:e243. https://doi.org/10.7717/peerj.243.
90. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
91. Letunic I, Bork P. 2011. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res 39:W475–W478. https://doi.org/10.1093/nar/gkr201.
92. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119.
93. Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. 2015. Sparse and compositionally robust inference of microbial ecological networks. PLoS Comput Biol 11:e1004226. https://doi.org/10.1371/journal.pcbi.1004226.
94. Friedman J, Alm EJ. 2012. Inferring correlation networks from genomic survey data. PLoS Comput Biol 8:e1002687. https://doi.org/10.1371/journal.pcbi.1002687.
95. Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 11:R25. https://doi.org/10.1186/gb-2010-11-3-r25.
96. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8.
97. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340.
98. van Dongen S, Abreu-Goodger C. 2012. Using MCL to extract clusters from networks. Methods Mol Biol 804:281–295. https://doi.org/10.1007/978-1-61779-361-5_15.
99. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 428:726–731. https://doi.org/10.1016/j.jmb.2015.11.006.
100. Yu G, Wang L-G, Han Y, He Q-Y. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics 16:284–287. https://doi.org/10.1089/omi.2011.0118.