

Novel chemolithotrophic and anoxygenic phototrophic genomes

2 extracted from ice-covered boreal lakes

Lucas Sinclair^{1,2}, Sari Peura^{3,4}, Pilar Hernández⁴, Martha Schattenhofer⁵, Alexander

4 Eiler^{6,7,*}.

1 Department of Genetics and Evolution, Molecular Systematics and Environmental

6 Genomics, Geneva University, Geneva, Switzerland

2 Sinclair.Bio - bioinformatics consultants, Geneva, Switzerland

8 3 Department of Forest Mycology and Plant Pathology, Science for Life Laboratory, Swedish
University of Agricultural Sciences, Uppsala, Sweden

10 4 Department of Ecology and Genetics, Limnology, Uppsala University, Uppsala, Sweden

5 Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University,

12 Uppsala, Sweden

6 Department of Chemistry and Molecular Biology, University of Gothenburg, Gothenburg,

14 Sweden

7 eDNA solutions AB, Mölndal, Sweden

16 * Corresponding author: A. Eiler alexander.eiler@icloud.com

* Address: Björkåsgatan 16, eDNA solutions AB, 43131 Mölndal, Sweden.

18 *Version: May 16, 2017*

Keywords: microbial diversity; 16S rRNA; metagenomics; freshwater; carbon cycle; green

20 house gases; anaerobe microbiology

Abstract

22 Although an important fraction of the world's lakes remains ice-covered during a large pro-
portion of the year, little is known about the microorganisms that govern the biogeochemical
24 processes occurring under-ice along the stratigraphic redox gradients. Reconstructed genomes
provide evidence for anoxygenic photosynthesis involving fixation of carbon using reduced sul-
26 phur and iron as an electron donor in the anoxic zone of the sampled lake systems. In addition to
anoxygenic photosynthesis, our molecular data reveals novel chemolithoautotrophic organisms
28 and supports the existence of methanotrophs in bottom anoxic waters. Reconstructed genomes
matched methanotrophs related to *Methylobacter tundripaludum*, phototrophic *Chloroflexi* and

30 *Chlorobia*, as well as lithoautotrophic genomes affiliated to the *Betaproteobacteria* class and
31 *Planctomycetes* phylum. Based on our in-depth characterization, complex metabolic interac-
32 tions emerge unique to each lake's redox tower and with sulfur, iron and carbon cycling tightly
intertwined through chemolithotrophy and anoxygenic photosynthesis.

34 **Introduction**

Lakes around the globe have a pronounced impact on the global carbon cycle [1] [2]. For ex-
35 ample, the estimated carbon losses through outgassing and burial from inland waters reach the
36 same magnitude as total global net ecosystem production [2]. Many of the processes related to
37 carbon cycling rest to a large extent on poorly understood microbial processes. Green-house gas
38 (GHG) emissions from seasonally ice-covered water systems at high latitude have been argued
39 to substantially contribute to global GHG production [3]. Moreover, these systems are thought
40 to be particularly sensitive to climate change as future reductions in the duration of ice-cover
41 are estimated to increase annual water body emissions by 20–54% before the end of the cen-
42 tury [4]. Still, scientific attention remains primarily focused on their ice-free state [5] while
43 lakes frozen temporarily are not put 'on hold' when ice-covered in the winter season. A recent
44 study [6] found that, in twelve small lakes in subarctic Sweden, the CO_2 emitted at ice-melt
45 accounted for 12 to 56% of the annual CO_2 emitted from these lakes. Understanding microbial
46 processes under ice is thus essential to construct accurate models and predict the lake systems'
47 role in global biogeochemical cycles such as those concerning the production of escaping GHGs
48 including CO_2 , CH_4 and N_2O .

49 Although the physical properties of ice-covered systems are well known and predictive mod-
50 els have been developed [7], how exactly this change of state affects the biogeochemistry and
51 microbial processes is poorly described and understood. Indeed, few studies have looked at pa-
52 rameters such as diversity, growth rate or metabolic capabilities of microbes under ice [5]. Work
53 so far suggests that under-ice conditions greatly influence the diversity of bacterial communities
54 via shifts in the availability and quality of organic matter and nutrients [8] [9].

55 Once ice-covered, hydrodynamic processes switch abruptly away from of the open-water
56 period as exchanges with the atmosphere are halted, including nutrient depositions and GHG
57 emissions. Light input is reduced, particularly when a layer of snow forms on top of the ice. In
58

addition to reduced mixing, the temperature gradient reverses in the opposite direction to that
60 of summer, with surface water switching from warmest to coolest layer, while the temperature
in the bottom layer is rather constant through the year [7]. Notably, the creation of redox-depth
62 gradients is expected to select for microbes with specific traits adapted towards the available
electron acceptors and donors [10].

64 Despite their global significance for GHG emissions, the microbial processes occurring in
the oxygen depleted or even anoxic ice-covered water column are as yet unknown and opaque.
66 To what extent nitrogen, sulfur and iron cycling or fermentation regulate the degradation of or-
ganic matter and how phototrophic and lithotrophic processes intertwine with organic matter
68 degradation by changing the availability of electron acceptors, are matters of ongoing debate.
Although numerous metabolic processes can possibly occur under ice, such as methanotrophy,
70 photoferrotrophy or the coupling of denitrification and iron oxidation, the molecular and geo-
chemical evidence has so far been slim, with few exceptions as in [9] [11].

72 The principal aims of this study were to retrieve the phylogenetic composition of the mi-
crobial communities proliferating under ice in high latitudinal lakes. Using a combination of
74 geochemical measurements and metagenomic techniques, we set out to disentangle the bio-
geochemical processes along depth profiles of five Swedish lakes, with certain resemblance
76 of the million boreal lakes in Fennoscandia, Siberia and the Canadian shield [12]. By recon-
structing the genomes of the abundant microbes we revealed the genomic basis of metabolic
78 traits previously undescribed in these poorly studied environments, such as photoferrotrophy
and methanotrophy. Finally, through linking the metabolic potential of the microbes to the cy-
80 cling of elements, we propose that sulfur, iron and carbon cycling are tightly linked through
anoxygenic photosynthesis and chemolithotrophy in these systems.

82 **Results**

Correspondence of bacterial diversity with lake characteristics

84 All five sampled lakes were thermally stratified with four of the systems displaying steep redox
gradients as indicated by oxygen, iron and sulfate profiles (figure 1, supplementary figure S2).
86 At the water-ice interface, methane bubbles were observed in two of the systems, LB and BT,

with CH_4 concentrations above 5 μ M. The profiles also indicated accumulation of methane in
88 lake RL with a reading of 150 nM CH_4 . Nevertheless, methane concentrations were highest
in the water column just above the sediment layer, reaching far beyond 50 μ M in two systems
90 (LB and KT), similar to the maximum concentrations previously reported for boreal lakes [13].
Carbon dioxide followed similar trends with concentrations increasing with depth and maximum
92 values ranging from 0.22 to 0.95 mM in the five systems.

Clear differences between the lakes concerning maximum concentrations and profiles of in-
94 organic nutrients and electron acceptors such as oxygen, sulfate and Fe(III) were evident. Lake
SB was the most oligotrophic system with very low total organic carbon (TOC) and nutrient
96 concentrations. There was no oxygen gradient, but the water column was oxygenated through-
out. Surface oxygen concentration close to saturation (13.8 mg/l at 2 °C) in RL and SB suggest
98 active oxygenic photosynthesis in these systems. Lakes BT and RL had oxygen concentration
below detection limit only in the deepest 1 m, while in LB and KT, which represent the most
100 productive systems as indicated by the bacterial cell numbers (supplementary figure S1), oxy-
gen was already depleted at 1 m depth from the surface. In the latter two, sulfate concentrations
102 decreased with depth, suggesting the presence of sulfate-based anaerobic respiration. Whereas
iron occurred in its reduced form (FeII), it occurred exclusively in its oxidized form (FeIII) in
104 KT with concentrations above 100 μ M even in oxygen-depleted layers.

These differences in chemical profiles among the studied systems were also reflected in the
106 unique bacterial communities present in individual lakes as assessed by 16S rRNA gene ampli-
con sequencing. As visualized in the ordination plot in figure 2A, depth profiles of individual
108 lakes grouped within lake with no convergence at any depths. This selection of microbial com-
munities by the specific conditions of each system is further emphasized by the observation that
110 environmental parameters significantly relate to community patterns. Indeed, pH, Fe(II), Fe(III)
and total organic nitrogen (TON) together explained 47% of the observed variability as inferred
112 by a redundancy analysis ($p < 0.001$). Further evidence is thus provided for the tight link be-
tween stratigraphic patterns in taxonomic composition and the depth-related variations in redox
114 potential.

Overall, 44 phyla of which fifteen candidate phyla were identified across all samples. Pro-
116 teobacteria, *Actinobacteria* and *Bacteroidetes* dominated in the five systems with differences

among the lakes and distinct stratigraphic patterns as seen in figure 2. The fifteen candidate phyla
118 recruited on average 12.8% of the reads ranging from 2.6 to 28.5% per sample and these per-
centages increased toward the lake bottoms, as previously observed [14]. The most prominent
120 vertical redox-gradient related changes were the increase in the relative abundance of *Chlorobia*
and *Chloroflexi* in KT, and a shift towards *Deltaproteobacteria*, Candidate division BSV13 and
122 *Lentisphaera* at intermediate depths in LB. Specific taxonomic groups identified were indicative
for aerobic (e.g. family *Comamonadaceae*) and anaerobic anoxygenic photosynthesis (e.g. phy-
124 lum *Chlorobia*, and genus *Oscillochloridaceae*), recalcitrant polymer degradation (e.g. genera
Paludibacter and *Chthoniobacter*) and methanotrophy (e.g. families *Methylococcales*, *Methy-*
126 *lophilaceae* and genus *Candidatus Methylacidiphilum*) as seen in figure 2 and in supplementary
figure S3. This detailed taxonomic analysis showed partial concordance with previously de-
128 scribed boreal lake communities (for comparison see [14] [15] [16] [8])

Microbial traits along three under ice redox gradients

130 We used the more detailed trait predictions offered by the shotgun-metagenomics technique
to infer the functional potential of a multitude of uncultured freshwater prokaryotes and gain
132 insight into their potential role in elemental cycles. Using the chemical information gathered
(in particular iron and sulfate concentrations) as well as the bacterial community depth profiles
134 built (in particular the proportions of candidate phyla and other poorly described deep branching
taxonomic groups), we chose three systems: LB, KT and BT. We obtained shotgun metagenome
136 libraries resulting in a total of almost 530 million reads ranging from 15.1 to 26.5 million reads
per sample. Assembly of individual depth profiles (i.e. individual lakes) resulted in almost
138 90'000 contigs with a minimum length of 1 kb, as seen in table 1. These contigs recruited 30.6%
of the total reads with a range from 22.3 to 44.7% in individual systems. Thus, on average,
140 70% of the lake's genomic content and the functions associated are not captured, introducing
uncertainties in our metabolic trait profiles. More details on the different stages of the sequence
142 processing are offered in online supplementary reports.

The final assemblies and mappings were used in combination with hidden Markov models
144 (HMM) similarity searches to provide more information on the functional repertoires along the
depth gradients. By screening the assembled metagenomes for genes encoding conserved pro-

146 tein family domains (PFAMs) of relevant enzymes (see supplementary table S2), we aimed to
determine the metabolic potential of the key microbial processes in these boreal lakes. Although
148 these annotations are derived from incomplete databases built on results from environments in
which the majority of taxa have not been well characterized [17], we were able to note that
150 the predicted functional diversity contained within the under-ice microbial communities was
congruent with the taxonomic diversity. We show a strong positive correlation between the
152 functional diversity, as assessed by PFAM annotations, and the taxonomic diversity derived by
amplicon sequencing of the 16S rRNA gene in each of the three systems' depth gradients. In-
154 deed, the correlation coefficients were 0.70 in LB, 0.78 in BT and 0.77 in KT with $p < 0.001$
as determined by a procrustes superimposition. Although there has been considerable debate
156 in the field of ecology as to how taxonomic diversity of communities relates to functional or
trait-level diversity [18], recent studies on microbial communities from soils [19] and aquatic
158 systems [20] have provided evidence for such a coupling, despite high functional redundancy
in microbial communities [21].

160 Reflecting the unique taxonomic profiles of the individual lakes, the three selected systems
were clearly separated when it came to their functional profiles as seen in supplementary figure
162 S4. Still, there were common features along the depth profiles as revealed by the progression
of genes encoding for methanotrophy, iron and sulfur oxidation-reduction reactions, phenolic
164 compound degradation, anoxygenic photosynthesis and methanogenesis as seen in figure 3. The
genes encoding sulfur and iron cycling were notably very abundant and genome equivalence in-
166 dicated potential for sulfur and iron oxidation or reduction in over 50% of the microorganisms in
all samples. The presence of sulfur oxidation genes (*sox* and *dsr*), together with steep gradients
168 in sulfate concentrations and identified taxa such as *Geobacter*, *Desulfobulbus* and *Desulfovib-*
rio, as well as *Chlorobium*, point to an important role of sulfur cycling in the water column of
170 these freshwater systems. Despite concentrations being much lower than those of marine sys-
tems, the importance of sulfur cycling has already been suggested for lake sediments [22]. Still,
172 LB and BT were shown to have low sulfate and high iron concentrations as do many lakes in
the boreal landscape.

174 Annotations also revealed genes implicated in the reduction of sulfate/sulfite (*dsr*) and Fe(III)
(*fer*), as well as denitrification genes (*nar*, *nir*, *nor* and *nos*) as part of chemolithotrophic redox

176 reactions or anaerobic respiration of organic matter. Screening for known genes containing
PFAM domains that catalyze the degradation of allochthonous organic matter revealed that the
178 potential to hydrolyze plant polymers (e.g. cellulose and hemicellulose) could be found through-
out the water column while organisms capable of degrading phenolic compounds (e.g. lignin)
180 were mainly found in the oxic portion of the water column.

Genes involved in several fermentative pathways were detected with a genome equivalence
182 value of approximately 1, indicating more than 1 copy per genome, in the deepest strata of
LB and KT, suggesting a high potential for the production of fermentation products in bottom
184 waters of boreal lakes. Likewise, genes encoding the terminal hydrogenase of H_2 -evolving fer-
mentations (hydA) increased with depth. The abundance of formyltetrahydrofolate synthetase
186 genes (fhs), which encodes the key enzyme of the Acetyl-CoA pathway of homoacetogene-
sis, was highest just below the ice and the oxycline as seen in figure 3 showing scaled abso-
188 lute genome equivalent values close to 0.5 [23]. Key genes indicative for methanogenesis in-
creased with depth together with the fraction of methanogenic *Archaea*. Three different orders
190 of methanogenic archaea were found, *Methanobacteriales*, *Methanomicrobiales* and *Methano-*
sarcinales. The first two are hydrogenotrophic, producing CH_4 from H_2 and CO_2 , whereas
192 *Methanosarcinales* is metabolically more versatile carrying out hydrogenotrophic, acetoclas-
tic and methylotrophic methanogenesis. The presence of these different types of methano-
194 gens was also verified by the homology to key enzymes described in these processes including
glutathione-independent formaldehyde dehydrogenase (FdhA), hydrogenase subunit A (EchA),
196 formylmethanofuran dehydrogenase subunit A (FmdA), formylmethanofuran-tetrahydromethanopterin
N-formyltransferase (FTR), methenyltetrahydromethanopterin cyclohydrolase (MCH), methylenete-
198 trahydromethanopterin dehydrogenase (MTD), coenzyme F420-dependent N5, N10-methenyltetrahydromethanopterin
reductase (MCH), tetrahydromethanopterin S-methyltransferase (MtrA), methyl-Co(III) methanol-
200 specific corrinoid protein coenzyme M methyltransferase (MtaA), methyl-coenzyme M reduc-
tase alpha subunit (McrA), acetate kinase (AckA), acetyl-CoA synthetase (ACCS), phosphate
202 acetyltransferase (PTA), heterodisulfide reductase subunit A (HdrA), acetyl-CoA decarboxy-
lase/synthase complex subunit beta (CdhC). While co-occurring in all three systems, alter-
204 nation from acetoclastic to hydrogenotrophic methanogens along the depth profiles differed
among the three systems. Differences in the distribution of acetotrophic, methylotrophic and

206 hydrogenotrophic methanogens pointed to a variability in the concentrations of fermentation
products. Such differences can be indicative of variable efficiency in the fermentation steps,
208 possibly owing to established interactions between fermentative syntrophic bacteria and their
methanogenic counterparts.

210 The genes indicative for methanotrophy were detected throughout the watercolumn of all
three lakes as seen in figure 3. Type I methanotrophs (RuMP based) of the families *Methy-*
212 *lococcaceae* and *Methylophilaceae* were abundant in all three lakes in bottom samples (maxi-
mum of 14.4 and 20.9%, respectively) as seen in figure 2. We did detect type II methanotrophs
214 (serine based), relatives of the *Verrucomicrobium Methylacidiphilium* [24] and *Proteobacteria*
Methylobacter tundripaludum [25]. Relatives to the anaerobic methanotroph *Candidatus Methy-*
216 *lomirabilis oxyfera* of the NC10 candidate phylum [26] were identified only in a few samples.

Coinciding with a depletion of oxygen, genome equivalence estimates indicate that in KT
218 and LB genes for fermentation were very abundant and they increased in the bottom layers
of the water columns, suggesting that organic substrates represented the main electron accep-
220 tors. Given the low contribution of tricarboxylic acid (TCA) cycle related genes and most
electron-transport chain complexes including terminal oxidases, we infer a strictly anaerobic
222 fermentation-based lifestyle in bottom layers of KT and LB. In all the samples, the key enzymes
for the glycolysis (Embden-Meyerhof-Parnas) pathway were present in addition to the potential
224 to convert pyruvate to acetyl-coenzyme A (acetyl-CoA). We identified that pyruvate-formate
lyase (PFAM:02901) and pyruvate ferredoxin oxidoreductase (PFAM:01855) had a larger pres-
226 ence when compared to pyruvate dehydrogenase.

Key findings from metagenome assembled genomes

228 To further validate the presence of metabolic potential such as methanotrophy and iron oxida-
tion, genomes were reconstructed for the members of the under ice communities – an approach
230 applied previously to numerous environments [27] [28] [29]. Contigs were binned into 462 bins
with 57 bins passing our quality criteria of more than 60% completeness and less than 10% con-
232 tamination as assessed by CheckM [30]. Each of these high-quality bins, from now on called
metagenome assembled genomes (MAGs), represented a minor portion of the community in
234 each sample as their average coverages ranged from 0.12 to 6.57% of the mapped reads. For

more details, see the online reports in supplementary material. The energy metabolism of the
236 lakes' microbes was a combination of using all different electron donors available in the wa-
ter column with the high quality bins representing a full range of microbial metabolisms from
238 chemolithoautotrophy to photoheterotrophy as seen in figure 4.

First, the 57 MAGs were taxonomically assigned to eleven phyla including Candidate phy-
240 lum *Saccharibacter* (TM7), as seen in figure 4. When present in the MAG, the large and small
subunit of the ribosomal RNA were assigned using the naive bayesian classifier [31] to obtain
242 a more detailed taxonomic association, allowing the link with our amplicon data. After this as-
signment, the 57 MAGs were sent through multiple annotation pipelines to obtain a metabolic
244 trait assignment, including their preferred electron acceptors and donors. Genomes of typi-
cal freshwater bacteria such as freshwater SAR11 (LD12) [32] and members of the acI clade
246 [33] were obtained from the various systems with a high genome completeness. Annotations
confirmed that, analogous to genomes from single cells and cultures [34] [35] [32], these con-
248 tained the genes necessary for carrying out the Embden–Meyerhof–Parnas (EMP) pathway,
the tricarboxylic acid (TCA) cycle, and a typical electron transport chain supplemented with
250 light-mediated ATP production (i.e. bacteriothodopsin). Additional genomes including mem-
bers of the phyla *Bacterioidetes*, *Acidobacteria* and *Plancomycetes*, contained a large variety
252 of carbohydrate-active enzyme families, as shown in figure 4, indicating their involvement in
allochthonous plant derived polymer degradation in the boreal lakes. For details, we refer the
254 reader to the “MicroScope” pipeline [36] [37] with automatic annotations presented in KEGG
maps and MetaCyc collections.

256 In addition to these heterotrophic degraders, we were able to obtain multiple MAGs re-
lated to previously described bacteria implicated in methanotrophy such as *Methylobacter tun-*
258 *dripaludum*. The multiple genomes obtained related to *M. tundripaludum* contained *pmoABC*
genes, and encoded for the complete reductive citrate cycle (Arnon-Buchanan cycle, rTCA) and
260 N_2 -fixation [25] [38]. Since several genomes contained nitrate reductase (*nar*), nitrite reductase
(*nir*) and nitric oxide reductase (*nor*) operons similar to genomes obtained from wetlands [25]
262 [38], it can be speculated that these organisms combine denitrification with anaerobic methane
oxidation, using a pathway similar to that of *Candidatus Methyloirabilis oxyfera* [26]. This
264 pathway is proposed to include a quinol-dependent nitric oxide reductase (*qnor*) and a nitric ox-

ide dismutase (*nod*) [39] with homologs identified in several of the *M. tundripaludum* MAGs.
266 Even though it was shown that under hypoxia such methanotrophs can also use nitrate as terminal electron acceptor, oxygen seems to be necessary to activate methane by *pmo*, since no
268 intrinsic oxygen production such as in *M. oxyfera* was observed [40]. Alternatively, the ability to grow heterotrophically was also implied by the presence of complete glycolysis and TCA
270 cycle in the genomes. Other methanotrophs representing distinct lineages within the bacterial phylum *Verrucomicrobia*, such as *Candidatus Methylacidiphilum*, were present in the amplicon
272 dataset but could not be retrieved in the genome reconstructions.

It is of note that a partial *Chlorobium* genome, with the closest relative identified to be
274 *Chlorobium ferroxidans*, reconstructed from LB possessed neither a complete set of genes required for sulfur and iron oxidation nor all encoded components required for a photosynthetic
276 electron transfer chain. In the iron-rich lake KT, the most representative bins were taxonomically assigned to the phylum *Chloroflexi*, most closely related to *Oscillochloris trichoides*. These
278 bins encoded the complete reductive pentose phosphate cycle (Calvin-Benson-Bassham cycle, CBB). The fact that *Chloroflexi* can fix carbon dioxide by employing the CBB rather than by the
280 3-hydroxypropionate cycle was also recently shown on isolated *O. trichoides* producing a type I ribulose-1,5-bisphosphate carboxylase–oxygenase (RubisCO) [41]. In addition to a complete
282 nitrogen fixing operon, several of the closely related Chloroflexi MAGs possessed a complete bacteriochlorophyll a and c synthesis pathways, allowing the prediction of phototrophic potential with an absorption maxima in the far red (750 and 860 nm) [42]. However, compared to previously described sulfite [41] and nitrite [43] oxidizing *Chloroflexi*, neither *sox*, *dsr* nor
286 *nir* genes were present in our genomes. Like *O. trichoides* and *Chloroflexus aurantiacus* two Chloroflexi MAGs, ‘kt_25’ and ‘kt_114’, contained homologs to genes and gene clusters for
288 the alternative complex III (ACIII), a cupredoxin (with redox potentials ranging from <200 mV (e.g aracyanin) to 700 mV in azurin [44]) and NADH:quinone oxidoreductase representing a
290 third type of photosynthetic electron transport complex [45]. In addition homologs to a Rieske protein representing a widely ranging electron reduction potential from -150 to 400 mV, could
292 be identified in ‘kt_25’ and ‘kt_41’. Considering that the redox potential of minerals such as FeS and FeCO₃ is about +200 mV at pH 7, this together with the essential pieces resembling
294 a ACIII reaction center with a membrane-spanning electron transfer chain terminating in Fe-S

centres of the the Rieske protein rather than dissociable quinones could hypothetically generate
296 reverse electron flow from Fe compounds (likely also Fe(II)) to NADH.

Under acidophilic conditions, few known microorganisms gain energy by the oxidation of
298 Fe(II) and oxygen to generate reverse electron flow from Fe(II) to NADH [46]. Homologs to
the redox cofactor pyrroloquinoline quinone, a protein with a transport function, iron oxida-
300 tion specific c-type cytochrome *foxE* with no significant similarity to other known proteins and
the cytochrome oxidase subunits I and II were identified in two representatives of the *Planco-*
302 *tomycetes* phylum. In addition, multiple genomes with strong signals of chemolithoautotrophy
could be derived from the three systems, including reduction of inorganic nitrogen and the oxida-
304 tion of inorganic sulfur and nitrogen as seen in figure 4. Other interesting findings were genomes
indicating high metabolic plasticity similar to *Rhodopseudomans palustris* as they seem to be
306 capable of using multiple redox reactions within their population. These were *Betaproteobacte-*
ria belonging to *Acetobacteraceae* and *Bradrhizobiaceae* with both chemolithoautotrophic and
308 chemoorganoheterotrophic capability using oxygen, nitrate, Fe(III), sulfate and even organic
compounds as electron acceptors. We also found an extremely coding-dense genome identified
310 belonging to the *Candidatus Saccharibacteria* (TM7) phylum. This phylum has been suggested
to contain epibionts or parasites of other bacteria due to their extreme auxotrophy [47].

312 **Discussion**

Gross metabolic properties of the boreal lakes have been studied extensively indicating their net-
314 heterotrophy, as inferred from partial pressure of carbon dioxide [48] [49] [50]. If we want to
predict and modulate ecosystem functions, knowledge of the quantities and types of organisms
316 as well as their functions that constitute a particular ecosystem is an essential first step. With
deep sequencing results presenting a detailed taxonomic inventory throughout the water column
318 of five ice-covered boreal lakes, we show that the bacterial diversity differed widely among the
lake systems. Taxonomic stratigraphy already shows some indications concerning the properties
320 of the microbial communities inhabiting the redox towers, such as that anoxygenic phototrophs
and methanotrophs are widely distributed in these ice-covered systems, similar to their ice-
322 free state [51]. Although sequences from well described branches of life, including certain
methanotrophic and sulfur reducing Proteobacteria were observed, most of identified taxa are

324 from less studied clades (e.g. representatives of phyla *Chloroflexi* and *Planctomycetes*). These
clades populate a large part of phylogenetic tree of life, yet are all in poorly documented areas
326 where we have very limited knowledge about the metabolic repertoire found in their genomes.

By exploring microbial genomes in a high throughput fashion we show that a wide range of
328 metabolic capabilities involving the use of multiple electron donors and acceptors appears to be
common in the microbial community in the studied lake systems. In spite of redox metabolic
330 plasticity occurring in microbial community of each system, we found that the majority of or-
ganisms probably lack the ability to perform multiple sequential redox transformations within
332 a pathway. Further, we revealed variations at the genomic level within and between systems
when it came to iron and sulfur cycling, which is further emphasized by specific taxonomic
334 groups proliferating in each lake system. Our genomic data exposes diverse metabolic traits,
such as photoferrotrophy, chemolithotrophy and methane oxidation, supporting alternative en-
336 ergy sources in these net-heterotrophic ecosystems, as seen in figure 5. From this first detailed
genetic characterization, a complex metabolic system emerges that can be hypothesized to de-
338 pend on the availability of light as determined by organic matter load, ice thickness, snow cover
and weather conditions, as well as the availability of nutrients (e.g. N, P, S, Fe) as determined
340 by the catchment, and, last but not least, by the physical properties (i.e. morphology and hy-
drology).

342 As we predict that organisms mediate individual reaction steps in redox pathways these
must be linked to form full or short-circuit cycles of elements. Short circuits are for example
344 elemental cycles where the most reduced and oxidized forms are not the most common reaction
products. Instead, interconversions of sulfide to elemental sulfur or denitrification from nitrate
346 to nitrite, resulting in products that can be oxidized back to sulfide and nitrate by phototrophs
or chemolithotrophs are very commonly encoded in the genomes. Such restricted metabolic
348 potential may give these highly specialized organisms an advantage under the redox conditions
at certain depth, even though organisms related to for example *Rhodopseudomans palustris* with
350 a wide range of metabolic capabilities do co-exist. Overall, the reconstructed depth profile of
the lakes allowed us to estimate the degree with which specialized metabolic niches are formed
352 and to quantify the proportion of genomes with the potential to obtain their energy from either
light, oxidation of organic molecules, or the oxidation of inorganic molecules.

354 Our findings emphasize that elemental cycles, such as those of sulfur and iron, have the
potential to play major roles in the systems and need to be taken into consideration when building
356 biogeochemical models. By contrast, nitrogen cycling seemed to be less widely distributed
among the microbes inhabiting frozen boreal lakes, as previously thought, while the importance
358 of methanotrophy-driven systems is corroborated [52]. Based on stable isotope experiments, a
pelagic food web largely supported by methane metabolism was already proposed previously
360 [52]. Photoferrotrophy has been shown to play a role in a handful of systems only very recently
[53] [54] [55]. These systems have been regarded as modern analogues of the Archaean ocean in
362 a time where oxygenic phototrophs had yet not evolved, as both photoferrotrophy and anaerobic
methane cycling are present. Our analyses based on genomic reconstructions in conjunction with
364 a very recent publication using isotope approaches [11] suggest that anoxic iron and methane
cycling may be widely distributed among the millions of lakes in the boreal zone.

366 To sum up, the insight gained from the reconstruction of the microbial genomes lead us to
formulate hypotheses concerning the biogeochemical cycling of elements in these ecosystems
368 of global significance. We provide an overview of the ongoing metabolic processes as shown
in figure 5, which resembles most redox stratified systems. Besides some general resemblance,
370 metabolic and taxonomic profiles emerged unique to each lake's redox tower and with sulfur,
iron and carbon cycling tightly intertwined through chemolithotrophy and anoxygenic photo-
372 synthesis.

Materials and Methods

374 Sample design and environmental properties

In total, thirty-nine water samples were obtained from five boreal lakes (seven or eight samples
376 per lake) all located in central Sweden. Our sampling took place in the area covered by N63°33'
to N63°59' and E12°27' to E14°46' during March 18th to 20th of 2014. Each lake was assigned
378 a two letter acronym for easy reference: RL, BT, KT (Lomtjärnen), LB (Liltjärnen) and SB
(Fröåtjärnen) as some of the lakes possess no authoritative names to our knowledge. These wa-
380 ter samples were obtained along a depth gradient by drilling small holes on the ice surface and
by using a rope-operated Limnos sampler. Samples were immediately processed: for CO_2 and

382 CH_4 , 25 ml of samples were directly taken from the sampler with 60 ml polypropylene syringes
equipped with three-way stopcocks and subsequently stored on ice. Upon return to the labora-
384 tory, after creating a 10 ml room-air headspace in each syringe, CH_4 was equilibrated between
the remaining water and the headspace by vigorous shaking for 3 minutes. The headspace was
386 then injected into N_2 filled 120 ml infusion vials with crimp aluminium seal secured 10-mm
butylrubber stoppers [56]. Within a week, gas concentrations were measured with a gas chro-
388 matograph (GC) (Agilent Technologies 7890A GC Systems) equipped with a flame ionization
detector by injecting 1 ml of gas from the infusion vials. CH_4 concentrations were calculated
390 according to Henry's law, correcting for temperature according to [57] and corrected for CH_4
in ambient air as measured on the GC.

392 Water temperature and oxygen concentrations were measured *in situ* using a YSI 55 com-
bined temperature and oxygen probe (Yellow Springs Instruments, Yellow Springs, Ohio, USA).
394 Total phosphorus (TP) and total nitrogen (TN) were measured using standard methods as previ-
ously described [58]. Total organic carbon (TOC) concentrations, also known as non-purgeable
396 organic carbon (NPOC), were obtained by analysis on a Shimadzu TOC-L with sample changer
ASI-L (Shimadzu Corporation, Japan).

398 Water for analyses of SO_4^{2-} analysis was first pre-filtered through rinsed 0.2 μ m membrane
filters (Pall Corporation) and then analyzed by ion chromatography on a Metrohm IC system
400 (883 Basic IC plus and 919 Autosampler Plus) fitted with a Metrosep A Supp 4/5 guard column
and a Metrosep A Supp 5 analytical column (150x4.0 mm). The concentrations of Fe(II) and
402 Fe(III) were measured with the ferrozine colorimetric method [59].

Samples for bacterial counts were fixed with 37% borax buffered formaldehyde (final con-
404 centration 2%) and stored at 4 °C prior to analyses. Cells were stained with the fluorescent nu-
cleic acid stain Syto13 (Molecular probes, Invitrogen, Carlsbad, CA, USA) according to the pro-
406 tocol of Del Giorgio and colleagues [60] and were then counted with a flow cytometer equipped
with a 488 nm blue solid state laser (Cyflow Space, Partec, Görlitz, Germany) using green flu-
408 orescence for triggered particle scoring. Cell counts were analyzed using Flowing Software
version 2.5 (Perttu Terho, Centre for Biotechnology, Turku, Finland).

410 DNA extraction and 16S rRNA gene amplicon sequencing

For DNA, between 0.5 and 1 liter of water was filtered through 0.2 µm Sterivex cartridges (Mil-
412 lipore), in duplicate, by using sterile syringes followed by on site liquid nitrogen freezing until
further analyses. After filters were removed from the cartridges, total DNA was extracted using a
414 Powersoil DNA Isolation Kit (MO BIO). Bacterial 16S rRNA gene amplicons were sequenced
on a MiSeq machine (Illumina) following procedures modified from Sinclair and colleagues
416 [61]. In short, each sample was first amplified in duplicate using primers targeting the variable
regions of the rRNA gene (V3/V4 region) and equipped with parts of the ThruPLEX Illumina se-
418 quencing adapter. After duplicates were pooled and purified using the Agencourt AMPure XP
system (Beckman Coulter) as recommended by the manufacturer, the pooled samples were used
420 as templates in a second PCR step using primers equipped with a 7bp index and the Illumina
sequencing adapters for multiplexing. After purifying the samples using the Agencourt AM-
422 Pure XP kit and quantifying by fluorescence with the PicoGreen assay (Quant-iT PicoGreen,
Invitrogen), samples were pooled in equimolar amounts. The pooled samples were sequenced
424 at the SciLifeLab SNP/SEQ sequencing facility (Uppsala University, Uppsala, Sweden) using
an Illumina MiSeq with a 2x300 bp chemistry. Finally, a 16S analysis and annotation pipeline,
426 previously described in [61], was used. The sequence processing comprises steps for pairing
reads, quality filtering, chimera checking, clustering with a 3% sequence dissimilarity thresh-
428 old, taxonomic assignment, diversity estimations and result visualization.

Shotgun-metagenomic sequencing

430 First, 10 ng of genomic DNA was sheared using a focused-ultrasonicator (Covaris E220). Next,
the sequencing libraries were prepared with the ThruPLEX FD Prep kit from Rubicon Genomics
432 according to the manufacturer's protocol (R40048-08, QAM-094-002). The library size selec-
tion was carried out with AMPure XP beads (Beckman Coulter) in a 1:1 ratio. The prepared
434 sample libraries were quantified by applying KAPA Biosystem's next-generation sequencing
library qPCR kit and run on a StepOnePlus (Life Technologies) real-time PCR instrument. The
436 quantified libraries were then prepared for sequencing on the Illumina HiSeq sequencing plat-
form utilizing a TruSeq paired-end cluster kit, v3, and Illumina's cBot instrument to generate a
438 clustered flowcell for sequencing. Two individual sequencing runs (an evaluation and re-run)

were performed with twenty-four samples. In both cases, sequencing of the flowcell was performed on the Illumina HiSeq2500 sequencer using Illumina TruSeq SBS sequencing kits, v3, with the exception of following a 2×100 bp indexed high-output run recipe for the evaluation and a 2×125 bp in the case of the re-run.

Shotgun-metagenome processing

Reads were filtered based on their PHRED quality scores using sickle (version 1.33) [62] and then assembled with Ray (version 2.3.1) [63]. Prior to generating a final assembly, we created a number of assemblies for optimization on the evaluation run. Assemblies with different k-mer sizes (31-81) were compared, based on different metrics such as N50. Assemblies of k-mer sizes of 51, 61, 71 and 81 were chosen to be applied on the re-run. Our tests further evidenced that when contigs were pooled and cut into 1'000 bp pieces and reassembled with Newbler (version 2.9) (454 Life Sciences, Roche Diagnostics) N50 was increased, therefore we used this technique to produce the final assembly on which all analyses were performed. Coverage was computed by running bowtie (version 2.2.5) [64] to map the reads back to the Newbler-produced assembly. Duplicates were removed by picard-tools (version 1.101). For computing coverage, bedtools (version 2.18.2) [65] was used. Once contigs and scaffolds shorter than 1 kb were discarded, concoct (version 0.3.0) [66] was run for binning. Bins were separated into low quality and high quality groups, based on results obtained from CheckM (version 0.9.7) [30]. A completeness value of over 60% and a contamination metric below 10% were the criteria chosen for classifying a bin as 'good' (high quality) and for it to obtain the denomination of "metagenome assembled genome" (MAG).

Functional trait profiles

Hmmsearch (version 3.1b2) [67] on the PFAM-A database at version 29.0 [68] was used to provide the annotation of the proteins predicted by Prodigal (version 2.6.2). Coverage information and abundance of proteins allowed us to estimate genome equivalence of individual PFAMs by normalizing each individual PFAM with the average quantity of 139 PFAMs predicted to occur in single copy in all genomes [69].

466 Genome analysis

Genome annotations were performed using “MicroScope” with automatic annotations assisted
468 by manual curation, as described in the integrated bioinformatics tools and the proposed an-
notation rules [36] [37]. In addition to the integrated annotation tools, which includes BlastP
470 homology searches against the full non-redundant protein sequence databank, UniProt [70] and
against the well-annotated 164 model organisms *Escherichia coli K-12* and *Bacillus subtilis* 168
472 [37], enzymatic classifications based on COG [71], InterPro [72], FIGFam [73] and PRIAM
[74] profiles, and prediction of protein localization using the TMHMM [75], SignalP [76] and
474 PSORTb [77] tools, we also used our local hmmsearch results. Synteny maps (i.e. conserva-
tion of local gene order) were used to validate the annotation of genes located within conserved
476 operons. Metabolic pathways were subsequently identified with the assistance of the integrated
MicroCyc database and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [78].
478 Genomes can be retrieved from “MicroScope” with their specific identifiers as given in the sup-
plementary material.

480 Statistical analyses

Statistical analyses were done using Python packages or the R language, including non-parametric
482 multidimensional scaling (NMDS) plots, permuted multivariate ANOVA (PERMANOVA), re-
dundancy analysis and the Procrustes superimpositions. The analyses were performed on rar-
484 eified OTU tables of amplicon data using the Bray-Curtis distance and PFAM tables standardized
to genome equivalents [69] using the Morisita-Horn distance. Particular attention was paid to
486 pathways linked to energy metabolism and carbon cycle when analyzing metabolic depth pro-
files. To plot the depth profiles of the traits, marker HMMs were selected as unique members
488 of specific traits and are listed in supplementary table S1.

Acknowledgments

490 We would like to thank the various institutions running the multiple computational resources that
were used in this project: (i) the Swedish National Infrastructure for Computing (SNIC) through
492 the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX), (ii) the

IT Center for Science in Finland (CSC) and (iii) the Cloud Infrastructure for Microbial Bioinformatics (CLIMB) funded by the UK's Medical Research Council (MRC).

We also would like to acknowledge the support given by the SciLifeLab SNP/SEQ facility hosted by Uppsala University in the molecular sequencing steps.

This research was made possible through a scholarship from the Olsson-Borgh foundation for limnological studies (Uppsala University foundation number 91173 to LS) as well as grants by the Swedish Research council (grant 2012-4592 to AE), the Swedish Foundation for strategic research (grant ICA10-0015 to AE) and the Academy of Finland (grant 265902 to SP).

This research was carried out at Uppsala University in the Department of Ecology and Genetics, Limnology and manipulations were done in the laboratory of the Evolution Biology Center at Norbyvägen 18E, Uppsala, Sweden.

We would like to thank Moritz Buck for his input on this manuscript and the bioinformatics methods used.

Author contributions

All authors participated in revising the manuscript.

LS : Designed the study. Participated in sampling. Processed the sequence data and did the bioinformatics. Carried out statistical analyses. Wrote this manuscript.

SP : Conducted and orchestrated the sampling campaign. Processed samples in the lab.

MB : Participated in sampling. Provided advice on bioinformatics procedures and methods.

PH : Participated in sampling. Processed samples in the lab.

MS : Participated in sampling.

AE : Designed the study. Participated in sampling. Processed the sequence data. Carried out statistical analyses. Wrote this manuscript.

The authors declare no competing financial interests or other conflicts of interest.

References

- 518 [1] W E Dean and E Gorham (1998) **Magnitude and significance of carbon burial in lakes,**
519 **reservoirs, and peatlands.** *Geology*, volume 26 (6), 535
- 520 [2] Lars J Tranvik, John A Downing, James B Cotner, Steven A Loiselle, Robert G Striegl,
521 Thomas J Ballatore et al. (2009) **Lakes and reservoirs as regulators of carbon cycling**
522 **and climate.** *Limnology and Oceanography*, volume 54 (6-2), 2298
- [3] Anthony K Aufdenkampe, Emilio Mayorga, Peter A Raymond, John M Melack, Scott C
523 Doney, Simone R Alin et al. (2011) **Riverine coupling of biogeochemical cycles between**
524 **land, oceans, and atmosphere.** *Frontiers in Ecology and the Environment*, volume 9 (1),
525 53
- [4] M Wik, R K Varner, K W Anthony, and S MacIntyre (2016) **Climate-sensitive north-**
526 **ern lakes and ponds are critical components of methane release.** *Nature Geoscience*,
527 volume 1 (9), 99
- 528 [5] Stefan Bertilsson, Amy Burgin, Cayelan C Carey, Samuel B Fey, Hans-Peter Grossart,
529 Lorena M Grubisic et al. (2013) **The under-ice microbiome of seasonally frozen lakes.**
530 *Limnology and Oceanography*, volume 58 (6), 1998
- [6] Jan Karlsson, Reiner Giesler, Jenny Persson, and Erik Lundin (2013) **High emission of**
531 **carbon dioxide and methane during ice thaw in high latitude lakes.** *Geophysical Re-*
532 *search Letters*, volume 40 (6), 1123
- 533 [7] K Salonen, M Leppäranta, M Viljanen, and R D Gulati (2009) **Perspectives in winter**
534 **limnology: closing the annual cycle of freezing lakes.** *Aquatic Ecology*, volume 43 (3),
535 609
- [8] Stefan Rösler, Martin Allgaier, and Hans-Peter Grossart (2012) **Long-Term Charac-**
536 **terization of Free-Living and Particle-Associated Bacterial Communities in Lake**
537 **Tiefwaren Reveals Distinct Seasonal Patterns.** *Microbial Ecology*, volume 64 (3), 571
- 538 [9] Monica Ricão Canelhas, Blaize A Denfeld, Gesa A Weyhenmeyer, David Bastviken, and

-
- 544 Stefan Bertilsson (2016) **Methane oxidation at the water-ice interface of an ice-covered lake**. *Limnology and Oceanography*, volume 61 (S1), S78
- 546 [10] R E Glatz, P W Lepp, B B Ward, and C A Francis (2006) **Planktonic microbial community composition across steep physical/chemical gradients in permanently ice-covered Lake Bonney, Antarctica**. *Geobiology*, volume 4 (1), 53
- 548 [11] Sherry Schiff, Jackson Tsuji, Lingling Wu, Jason Venkiteswaran, Lewis Molot, Richard Elgood et al. (2016) **Millions of Boreal Shield Lakes can be used to Probe the Evolution of Archaean Ocean Life**. *bioRxiv*, 054478
- 550
- [12] Pirkko Kortelainen, Hannu Pajunen, Miitta Rantakari, and Matti Saarnisto (2004) **A large carbon pool and small sink in boreal Holocene lake sediments**. *Global Change Biology*, volume 10 (10), 1648
- 552
- 554 [13] S Juutinen, M Rantakari, P Kortelainen, J T Huttunen, T Larmola, J Alm et al. (2009) **Methane dynamics in different boreal lake types**. *Biogeosciences*, volume 6 (2), 209
- 556 [14] Sari Peura, Alexander Eiler, Stefan Bertilsson, Hannu Nykänen, Marja Tirola, and Roger I Jones (2012) **Distinct and diverse anaerobic bacterial communities in boreal lakes dominated by candidate division OD1**. *The ISME Journal*, volume 6 (9), 1640
- 558
- [15] Jürg Brendan Logue, Silke Langenheder, Anders F Andersson, Stefan Bertilsson, Stina Drakare, Anders Lanzén et al. (2012) **Freshwater bacterioplankton richness in oligotrophic lakes depends on nutrient availability rather than on species area relationships**. *The ISME Journal*, volume 6 (6), 1127
- 560
- 562
- [16] Ashley Shade, Stuart E Jones, and Katherine D McMahon (2008) **The influence of habitat heterogeneity on freshwater bacterial community composition and dynamics**. *Environmental Microbiology*, volume 10 (4), 1057
- 564
- 566 [17] Jack A Gilbert, Ronald O'Dor, Nicholas King, and Timothy M Vogel (2011) **The importance of metagenomic surveys to microbial ecology: or why Darwin would have been a metagenomic scientist**. *Microbial Informatics and Experimentation*, volume 1 (5)
- 568

-
- [18] Helmut Hillebrand and Birte Matthiessen (2009) **Biodiversity in a complex world: consolidation and progress in functional biodiversity research.** *Ecology Letters*, volume 12 (12), 1405
- 570
- [19] N Fierer, J Ladau, J C Clemente, J W Leff, S M Owens, K S Pollard et al. (2013) **Reconstructing the Microbial Diversity and Function of Pre-Agricultural Tallgrass Prairie Soils in the United States.** *Science*, volume 342 (6158), 621
- 572
- 574
- [20] Alexander Eiler, Katarzyna Zaremba-Niedzwiedzka, Manuel Martinez-Garcia, Katherine D McMahon, Ramunas Stepanauskas, Siv G E Andersson et al. (2013) **Productivity and salinity structuring of the microplankton revealed by comparative freshwater metagenomics.** *Environmental Microbiology*, volume 16 (9), 2682
- 576
- 578
- [21] Steven D Allison and Jennifer B H Martiny (2008) **Colloquium paper: resistance, resilience, and redundancy in microbial communities.** *Proceedings of the National Academy of Sciences of the United States of America*, volume 105 (1), 11512
- 580
- [22] Marianne Holmer and Peter Storkholm (2001) **Sulphate reduction and sulphur cycling in lake sediments: a review.** *Freshwater Biology*, volume 46 (4), 431
- 582
- [23] Jeroen Raes, Jan O Korbel, Martin J Lercher, Christian von Mering, and Peer Bork (2007) **Prediction of effective genome size in metagenomic samples.** *Genome biology*, volume 8 (1), 1
- 584
- 586
- [24] Huub J M Op den Camp, Tajul Islam, Matthew B Stott, Harry R Harhangi, Alexander Hynes, Stefan Schouten et al. (2009) **Environmental, genomic and taxonomic perspectives on methanotrophic Verrucomicrobia.** *Environmental Microbiology Reports*, volume 1 (5), 293
- 588
- 590
- [25] I Wartiainen (2006) **Methylobacter tundripaludum sp. nov., a methane-oxidizing bacterium from Arctic wetland soil on the Svalbard islands, Norway (78 N).** *International Journal of Systematic and Evolutionary Microbiology*, volume 56 (1), 109
- 592
- [26] Katharina F Ettwig, Margaret K Butler, Denis Le Paslier, Eric Pelletier, Sophie Mangenot, Marcel M M Kuypers et al. (2010) **Nitrite-driven anaerobic methane oxidation by oxygenic bacteria.** *Nature Geoscience*, volume 464 (7288), 543
- 594
- 596

-
- [27] Gene W Tyson, Jarrod Chapman, Philip Hugenholtz, Eric E Allen, Rachna J Ram, Paul M
598 Richardson et al. (2004) **Community structure and metabolism through reconstruction
of microbial genomes from the environment.** *Nature Geoscience*, volume 428 (6978),
600 37
- [28] K C Wrighton, B C Thomas, I Sharon, C S Miller, C J Castelle, N C VerBerkmoes et al.
602 (2012) **Fermentation, Hydrogen, and Sulfur Metabolism in Multiple Uncultivated
Bacterial Phyla.** *Science*, volume 337 (6102), 1661
- 604 [29] Matthew L Bendall, Sarah LR Stevens, Leong-Keat Chan, Stephanie Malfatti, Patrick
Schwientek, Julien Tremblay et al. (2016) **Genome-wide selective sweeps and gene-
606 specific sweeps in natural bacterial populations.** *The ISME Journal*, volume 10 (7),
1589
- 608 [30] Donovan H Parks, Michael Imelfort, Connor T Skennerton, Philip Hugenholtz, and
Gene W Tyson (2015) **CheckM: assessing the quality of microbial genomes recovered
610 from isolates, single cells, and metagenomes..** *Genome Research*, volume 25 (7), 1043
- [31] Qiong Wang, George M Garrity, James M Tiedje, and James R Cole (2007) **Naive
612 Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial
taxonomy.** *Applied and Environmental Microbiology*, volume 73 (16), 5261
- 614 [32] Alexander Eiler, Rhiannon Mondav, Lucas Sinclair, Leyden Fernandez-Vidal, Douglas G
Scofield, Patrick Schwientek et al. (2016) **Tuning fresh: radiation through rewiring of
616 central metabolism in streamlined bacteria.** *The ISME Journal*, volume 10 (8), 1902
- [33] Sarahi L Garcia, Katherine D McMahon, Manuel Martinez-Garcia, Abhishek Srivastava,
618 Alexander Sczyrba, Ramunas Stepanauskas et al. (2012) **Metabolic potential of a single
cell belonging to one of the most abundant lineages in freshwater bacterioplankton.**
620 *The ISME Journal*, volume 7 (1), 137
- [34] Trevor W Ghylis, Sarahi L Garcia, Francisco Moya, Ben O Oyserman, Patrick Schwien-
622 tek, Katrina T Forest et al. (2014) **Comparative single-cell genomics reveals potential
ecological niches for the freshwater acI Actinobacteria lineage.** *The ISME Journal*,
624 volume 8 (12), 2503

- [35] Sarahi L Garcia, Moritz Buck, Katherine D McMahon, Hans-Peter Grossart, Alexander
626 Eiler, and Falk Warnecke (2015) **Auxotrophy and intrapopulation complementary in
the ‘interactome’ of a cultivated freshwater model community.** *Molecular Ecology*,
628 *volume 24* (17), 4449
- [36] D Vallenet, S Engelen, D Mornico, S Cruveiller, L Fleury, A Lajus et al. (2009) **Mi-
630 croScope: a platform for microbial genome annotation and comparative genomics.**
Database, *volume 2009*, 021
- 632 [37] D Vallenet, E Belda, A Calteau, S Cruveiller, S Engelen, A Lajus et al. (2012) **Micro-
Scope - an integrated microbial resource for the curation and comparative analysis
634 of genomic and metabolic data.** *Nucleic Acids Research*, *volume 41*, 636
- [38] M M Svenning, A G Hestnes, I Wartainen, L Y Stein, M G Klotz, M G Kalyuzhnaya et al.
636 (2011) **Genome Sequence of the Arctic Methanotroph *Methylobacter tundripaludum*
SV96.** *Journal of Bacteriology*, *volume 193* (22), 6418
- 638 [39] Katharina F Ettwig, Daan R Speth, Joachim Reimann, Ming L Wu, Mike S M Jetten, and
Jan T Keltjens (2012) **Bacterial oxygen production in the dark.** *Frontiers in microbiol-
640 ogy*, *volume 3*
- [40] K Dimitri Kits, Dustin J Campbell, Albert R Rosana, and Lisa Y Stein (2015) **Diverse
642 electron sources support denitrification under hypoxia in the obligate methanotroph
Methylobacterium album strain BG8.** *Frontiers in microbiology*, *volume 6* (74767), 68
- 644 [41] Donald A Bryant and Niels-Ulrik Frigaard (2006) **Prokaryotic photosynthesis and pho-
totrophy illuminated.** *Trends in Microbiology*, *volume 14* (11), 488
- 646 [42] Denis S Grouzdev, Ruslan N Ivanovsky, Boris B Kuznetsov, Natalia V Lebedeva, Elena N
Krasil’nikova, and Olga I Keppen (2015) **Reconstruction of bacteriochlorophyll biosyn-
648 thesis pathways in the filamentous anoxygenic phototrophic bacterium *Oscillochloris*
trichoides DG-6 and evolution of anoxygenic phototrophs of the order Chloroflexales.**
650 *Microbiology*, *volume 161* (1), 120
- [43] Dimitry Y Sorokin, Sebastian Lucker, Dana Vejmelkova, Nadezhda A Kostrikina, Robert

-
- 652 Kleerebezem, W Irene C Rijpstra et al. (2012) **Nitrification expanded: discovery, phys-**
iology and genomics of a nitrite-oxidizing bacterium from the phylum Chloroflexi.
654 *The ISME Journal*, volume 6 (12), 2245
- [44] Nicholas M Marshall, Dewain K Garner, Tiffany D Wilson, Yi-Gui Gao, Howard Robin-
656 son, Mark J Nilges et al. (2009) **Rationally tuning the reduction potential of a single**
cupredoxin beyond the natural range. *Nature Geoscience*, volume 462 (7269), 113
- 658 [45] Erica L W Majumder, Jeremy D King, and Robert E Blankenship (2013) **Alterna-**
tive Complex III from phototrophic bacteria and its electron acceptor auracyanin.
660 *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, volume 1827 (11-12), 1383
- [46] D B Johnson, K B Hallberg, and S Hedrich (2014) **Uncovering a Microbial Enigma: Iso-**
662 **lation and Characterization of the Streamer-Generating, Iron-Oxidizing, Acidophilic**
Bacterium "Ferrovum myxofaciens". *Applied and Environmental Microbiology*, vol-
664 *ume 80 (2)*, 672
- [47] Xuesong He, Jeffrey S McLean, Anna Edlund, Shibu Yooseph, Adam P Hall, Su-Yang
666 Liu et al. (2015) **Cultivation of a human-associated TM7 phylotype reveals a reduced**
genome and epibiotic parasitic lifestyle. *Proceedings of the National Academy of Sci-*
668 *ences of the United States of America*, volume 112 (1), 244
- [48] G W Kling, G W Kipphut, and M C Miller (1991) **Arctic lakes and streams as gas con-**
670 **duits to the atmosphere: Implications for tundra carbon budgets.** *Science*, volume
251, 298
- 672 [49] Gesa A Weyhenmeyer, Pirkko Kortelainen, Sebastian Sobek, Roger Müller, and Miitta
Rantakari (2012) **Carbon Dioxide in Boreal Surface Waters: A Comparison of Lakes**
674 **and Streams.** *Ecosystems*, volume 15 (8), 1295
- [50] Blaize A Denfeld, Pirkko Kortelainen, Miitta Rantakari, Sebastian Sobek, and Gesa A
676 Weyhenmeyer (2015) **Regional Variability and Drivers of Below Ice CO₂ in Boreal**
and Subarctic Lakes. *Ecosystems*, volume 19 (3), 461
- 678 [51] Sari Peura, Lucas Sinclair, Stefan Bertilsson, and Alexander Eiler (2015) **Metagenomic**

- 680 **insights into strategies of aerobic and anaerobic carbon and nitrogen transformation**
in boreal lakes. *Nature Publishing Group, volume 5* (12102), 1
- [52] David Bastviken, Jörgen Ejlertsson, Ingvar Sundh, and Lars Tranvik (2003) **Methane as**
682 **a source of carbon and energy for lake pelagic food webs.** *Ecology, volume 84* (4), 969
- [53] Sean A Crowe, CarriAyne Jones, Sergei Katsev, Cédric Magen, Andrew H O'Neill, Arne
684 Sturm et al. (2008) **Photoferrotrophs thrive in an Archean Ocean analogue.** *Proceed-*
ings of the National Academy of Sciences of the United States of America, volume 105 (41),
686 15938
- [54] Xavier A Walter, Antonio Picazo, Maria R Miracle, Eduardo Vicente, Antonio Camacho,
688 Michel Aragno et al. (2014) **Phototrophic Fe(II)-oxidation in the chemocline of a fer-**
ruginous meromictic lake. *Frontiers in microbiology, volume 5*, 713
- 690 [55] Marc Llirós, Tamara García Armisen, François Darchambeau, Cédric Morana, Xavier Tri-
adó Margarit, Özgül Inceoğlu et al. (2015) **Pelagic photoferrotrophy and iron cycling in**
692 **a modern ferruginous basin.** *Scientific Reports, volume 5* (13803), 1
- [56] David Bastviken, Jörgen Ejlertsson, and Lars Tranvik (2002) **Measurement of methane**
694 **oxidation in lakes: a comparison of methods.** *Environmental science & technology,*
volume 36 (15), 3354
- 696 [57] D R Lide and HPR Frederikse (1995) **CRC handbook of chemistry and physics: a**
ready-reference book of chemical and physical data
- 698 [58] A E Greenberg, M J Taras, and M C Rand (1976) *Standard methods for the examination*
of water and wastewater. 14th edition (Prepared and published jointly by American Pub-
700 lic Health Association, American Water Works Association, and Water Pollution Control
Federation.)
- 702 [59] E Viollier, P W Inglett, K Hunter, A N Roychoudhury, and P Van Cappellen (2000) **The**
ferrozine method revisited: Fe(II)/Fe(III) determination in natural waters. *Applied*
704 *Geochemistry, volume 15* (6), 785

-
- [60] Paul A del Giorgio, David F Bird, Yves T Prairie, and Dolors Planas (1996) **Flow cytometric determination of bacterial abundance in lake plankton with the green nucleic acid stain SYTO 13**. *Limnology and Oceanography*, volume 41 (4), 783
- 706
- [61] Lucas Sinclair, Omneya Ahmed Osman, Stefan Bertilsson, and Alexander Eiler (2015) **Microbial Community Composition and Diversity via 16S rRNA Gene Amplicons: Evaluating the Illumina Platform**. *PLoS ONE*, volume 10 (2), 0116955
- 708
- 710
- [62] NA Joshi and JN Fass (2011) *Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files*. Version 1.33
- 712
- [63] Sébastien Boisvert, Frédéric Raymond, Élénie Godzaridis, François Laviolette, and Jacques Corbeil (2012) **Ray Meta: scalable de novo metagenome assembly and profiling**. *Genome biology*, volume 13 (12), 122
- 714
- [64] Ben Langmead and Steven L Salzberg (2012) **Fast gapped-read alignment with Bowtie 2**. *Nature Methods*, volume 9 (4), 357
- 716
- [65] Aaron R Quinlan and Ira M Hall (2010) **BEDTools: a flexible suite of utilities for comparing genomic features**. *Bioinformatics*, volume 26 (6), 841
- 718
- [66] Johannes Alneberg, Brynjar Smári Bjarnason, Ino de Bruijn, Melanie Schirmer, Joshua Quick, Umer Z Ijaz et al. (2014) **Binning metagenomic contigs by coverage and composition**. *Nature Methods*, volume 11 (11), 1144
- 720
- 722
- [67] Sean R Eddy (1998) **Profile hidden Markov models**. *Bioinformatics*, volume 14 (9), 755
- [68] Robert D Finn, Alex Bateman, Jody Clements, Penelope Coggill, Ruth Y Eberhardt, Sean R Eddy et al. (2013) **Pfam: the protein families database**. *Nucleic Acids Research*, volume 42, 222
- 724
- 726
- [69] Christian Rinke, Patrick Schwientek, Alexander Sczyrba, Natalia N Ivanova, Iain J Anderson, Jan-Fang Cheng et al. (2013) **Insights into the phylogeny and coding potential of microbial dark matter**. *Nature Geoscience*, volume 499 (7459), 431
- 728
- [70] UniProt Consortium (2014) **Activities at the Universal Protein Resource (UniProt)**. *Nucleic Acids Research*, volume 42, D191
- 730

-
- 732 [71] Roman L Tatusov, Natalie D Fedorova, John D Jackson, Aviva R Jacobs, Boris Kiryutin,
Eugene V Koonin et al. (2003) **BMC Bioinformatics**. *BMC Bioinformatics*, volume 4 (1),
734 41
- [72] A Mitchell, H Y Chang, L Daugherty, M Fraser, S Hunter, R Lopez et al. (2015) **The**
736 **InterPro protein families database: the classification resource after 15 years**. *Nucleic*
Acids Research, volume 43, 213
- 738 [73] F Meyer, R Overbeek, and A Rodriguez (2009) **FIGfams: yet another set of protein**
families. *Nucleic Acids Research*, volume 37 (20), 6643
- 740 [74] C Claudel-Renard (2003) **Enzyme-specific profiles for genome annotation: PRIAM**.
Nucleic Acids Research, volume 31 (22), 6633
- 742 [75] E L Sonnhammer, G von Heijne, and A Krogh (1998) **A hidden Markov model for pre-**
dicting transmembrane helices in protein sequences. *International Conference on In-*
744 *telligent Systems for Molecular Biology*, volume 6, 175
- [76] Jannick Dyrlov Bendtsen, Henrik Nielsen, Gunnar von Heijne, and Søren Brunak (2004)
746 **Improved prediction of signal peptides: SignalP 3.0**. *Journal of molecular biology*,
volume 340 (4), 783
- 748 [77] J L Gardy, M R Laird, F Chen, S Rey, C J Walsh, M Ester et al. (2005) **PSORTb v.2.0:**
Expanded prediction of bacterial protein subcellular localization and insights gained
750 **from comparative proteome analysis**. *Bioinformatics*, volume 21 (5), 617
- [78] Minoru Kanehisa, Susumu Goto, Yoko Sato, Masayuki Kawashima, Miho Furumichi, and
752 Mao Tanabe (2013) **Data, information, knowledge and principle: back to metabolism**
in KEGG. *Nucleic Acids Research*, volume 42, D199
- 754 [79] Nicola Segata, Daniela Börnigen, Xochitl C Morgan, and Curtis Huttenhower (2013) **Phy-**
loPhlAn is a new method for improved phylogenetic and taxonomic placement of mi-
756 **crobes**. *Nature Communications*, volume 4, 2304
- [80] R Conrad (1999) **Contribution of hydrogen to methane production and control of hy-**

758 **drogen concentrations in methanogenic soils and sediments.** *FEMS Microbiology Ecology*, volume 28 (3), 193

760 **Figures**

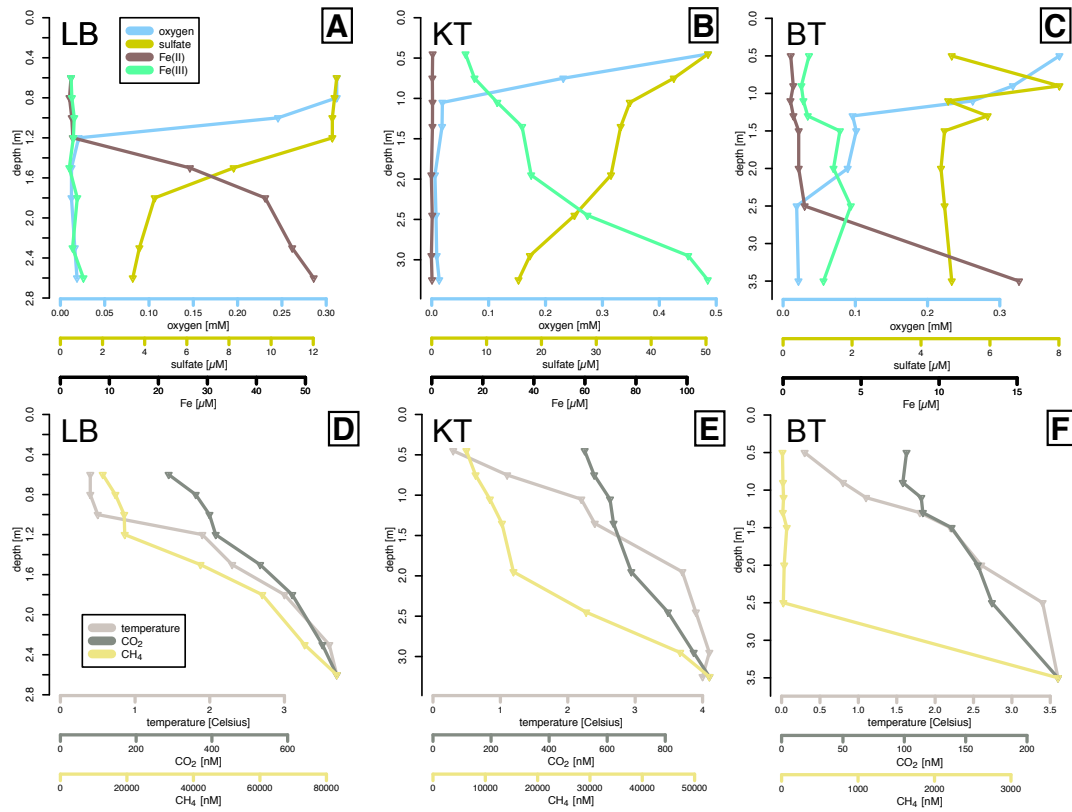


Figure 1. Summary of lake characteristics.

Profiles of water chemistry (A-C) and gas concentrations (D-F) in lakes LB (A,C), KT (B,D)

762 and BT (C,E). For RL and SB see supplementary figure S2.

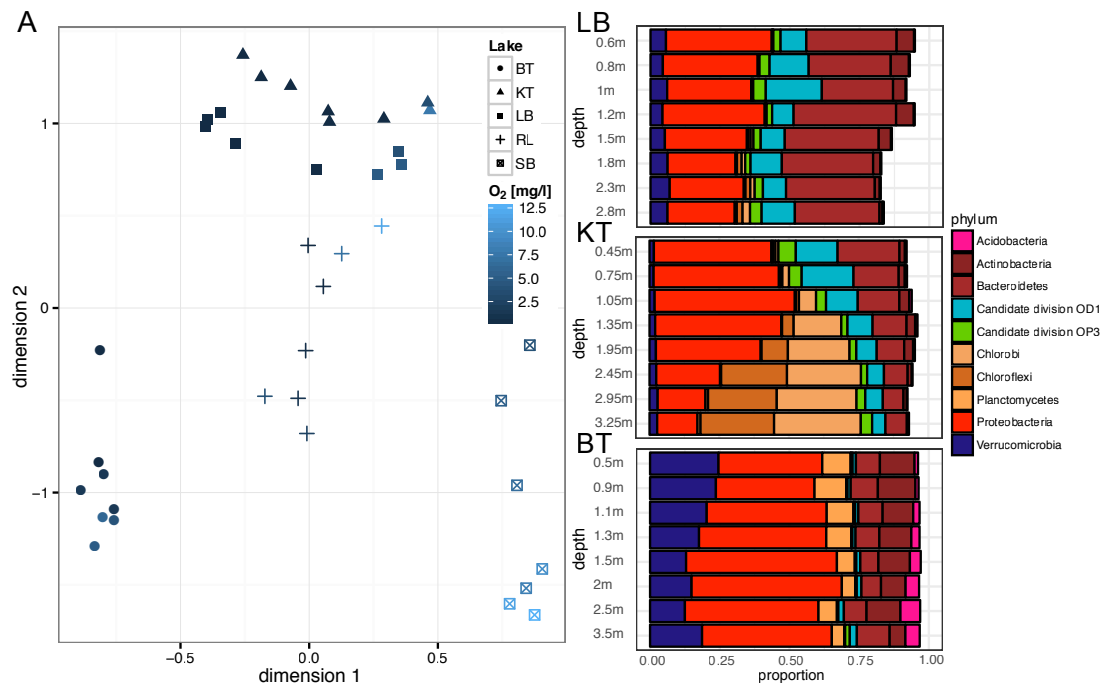


Figure 2. Bacterial community profiles by lake and by depth.

Non-parametric multidimensional scaling plot (A) and depth profiles of individual lakes LB,
764 KT and BT with the ten most abundant phyla. For RL and SB, see supplementary figure S3.

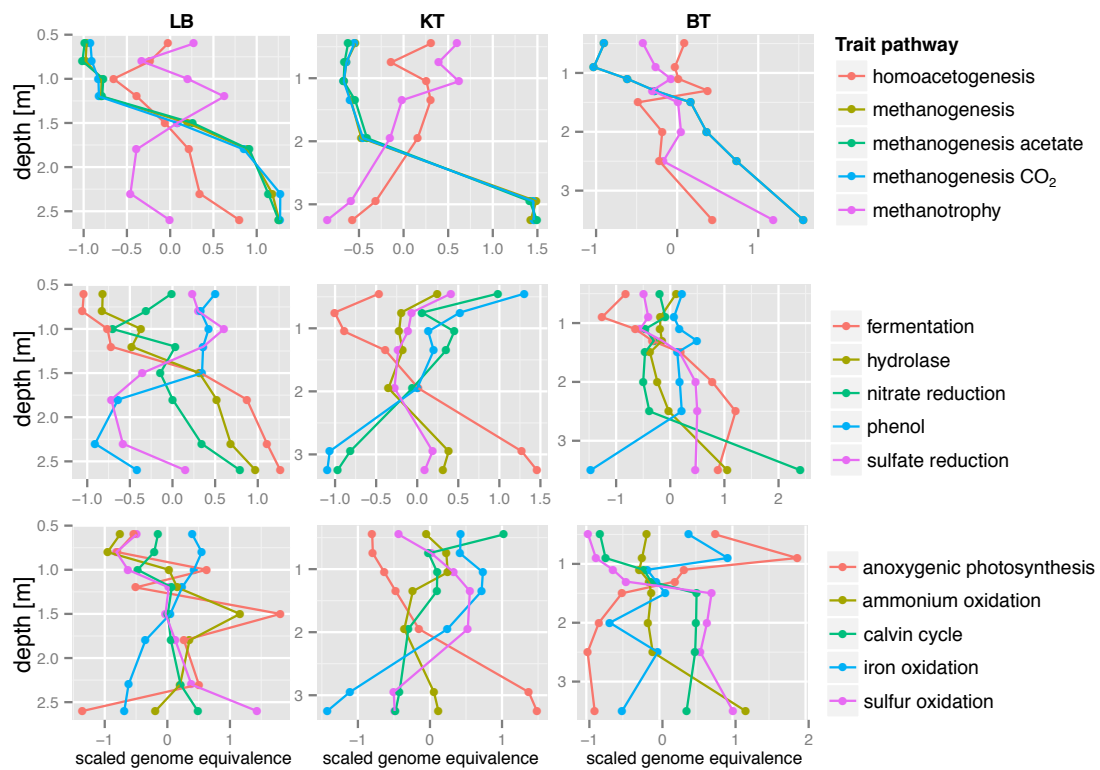


Figure 3. Genome encoded metabolic profiles by depth.

Depth profiles visualizing the abundances of PFAM markers related to carbon and energy
766 metabolism. The average of all scaled (z-score) HMM hits are indicative for particular metabolic
traits were plotted along with depth.

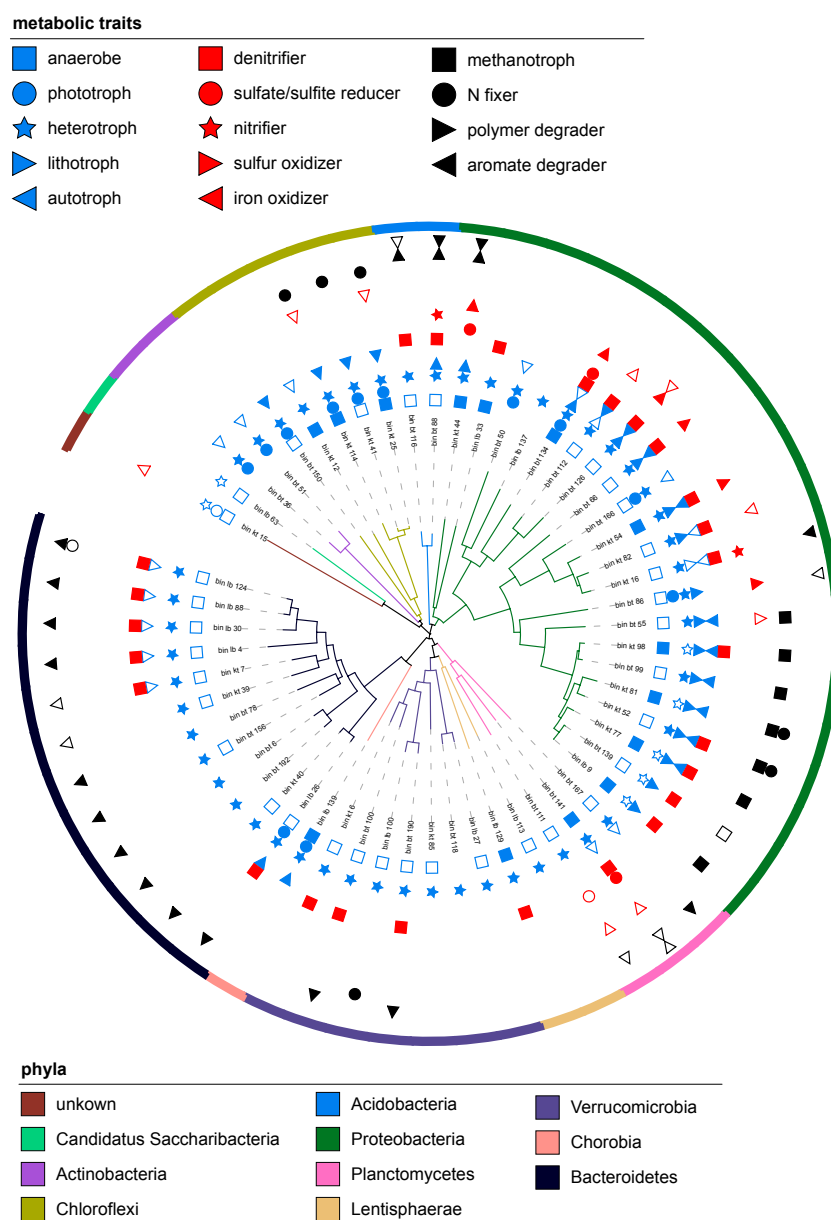


Figure 4. Traits, taxonomy and phylogeny of reconstructed genomes.

768 Phylogenomic tree of recovered genomes as computed by PhyloPhlAn [79]. The outer ring
 1770 delineates taxonomic affiliation. The inner circles show the metabolic traits as based on genome
 1770 annotations. Traits with high support are indicated by full symbols, while traits with low support
 have open symbols. In the case a symbol is missing, no indications were found.

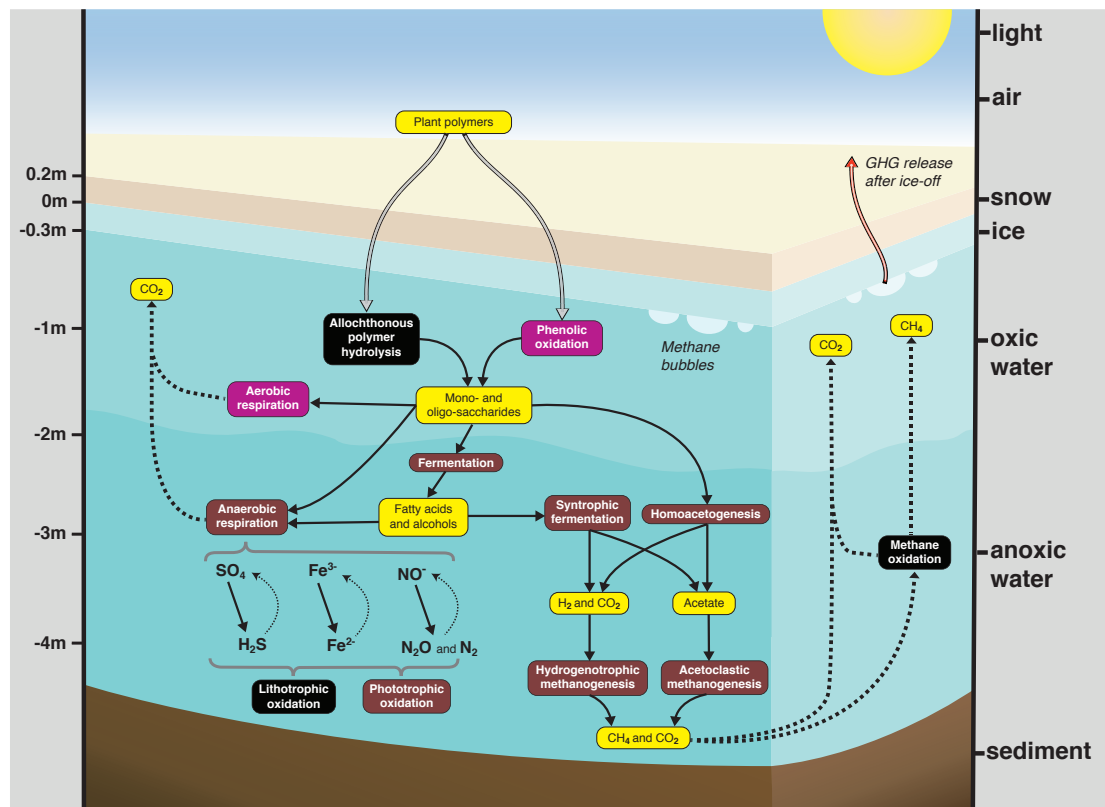


Figure 5. Main degradation pathways of allochthonous organic matter.

772 Schematic overview of the main degradation pathways of allochthonous organic matter oc-
curring under ice in boreal lakes. The pathways are divided into three categories: aerobic (pur-
774 ple), anaerobic (brown) and processes occurring under both conditions (black). Compounds
are represented in yellow. Metagenome assembled genomes (bins) encoding for the different
776 metabolic processes are presented in figure 4. The figure is adapted from figure 1 of Conrad
and colleagues [80].

778 **Tables**

	Latitude N	Longitude E	TOC [mg/l]	TOP [µg/l]	Amplicon reads	Shotgun reads	Contigs
BT	63.5821	12.2708	7.8 - 10.1	9.5 - 14.6	685'057	179'710'871	44'165
RL	63.5843	12.2743	7.8 - 7.0	5.4 - 14.1	811'181	n/a	n/a
LB	63.3382	12.5480	5.9 - 12.5	8.4 - 9.2	782'317	184'786'957	21'392
KT	63.3493	14.4587	15.8 - 26.6	12.7 - 111	842'872	165'342'647	23'668
SB	63.3961	13.1619	4.6 - 3.9	2.7 - 3.0	964'981	n/a	n/a

Table 1. Summary of sampling sites and sequencing results.

Supporting Information

Table S1. Selected PFAMs traits.

780 Supplementary table S1 is too long and not shown here. It is suited for online distribution
after publication and to be viewed within a spreadsheet application. It details the list of PFAMs
782 chosen to be unique to specific pathways.

Table S2. PFAMs found per bin.

Supplementary table S2 is too large to be shown here. It is best suited for online distribution
784 after publication and to be viewed with a spreadsheet application. It shows the results of the
hmmsearch against the PFAM database for lakes BT, KT and LB. Columns are bin identifier
786 numbers and rows are PFAMS. Each value indicates how many times each PFAM was found in
each bin.

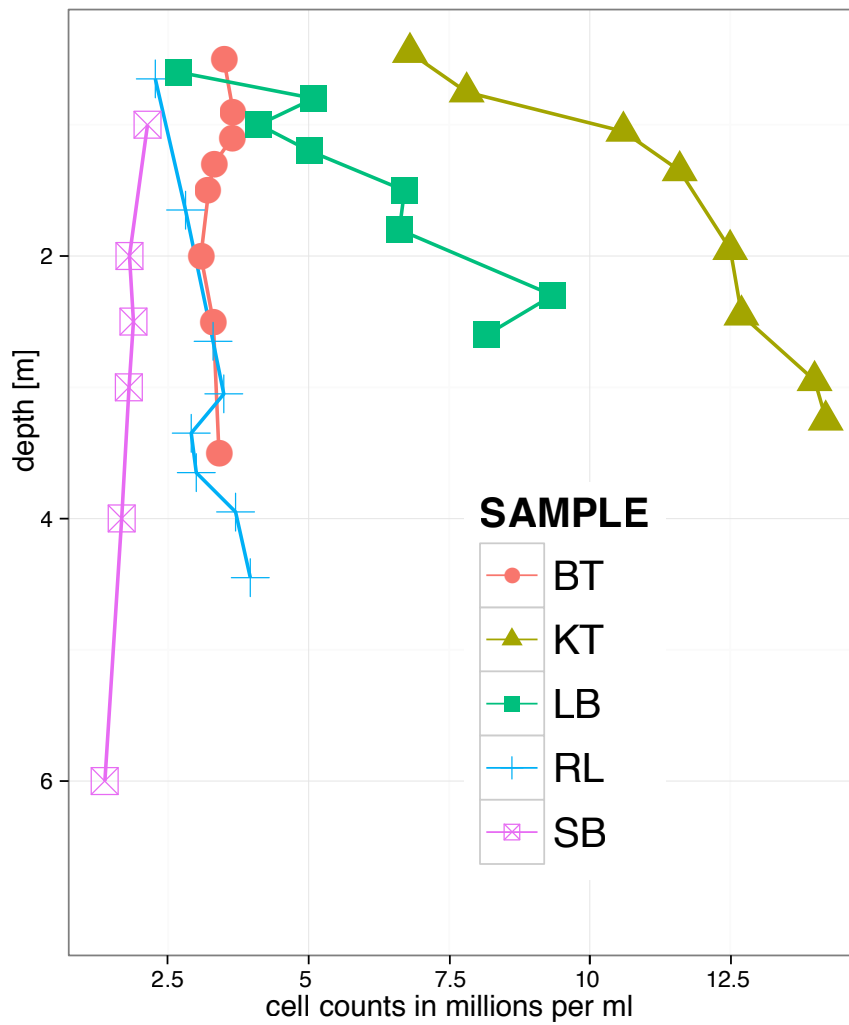


Figure S1. Cell counts per depth for all lakes.

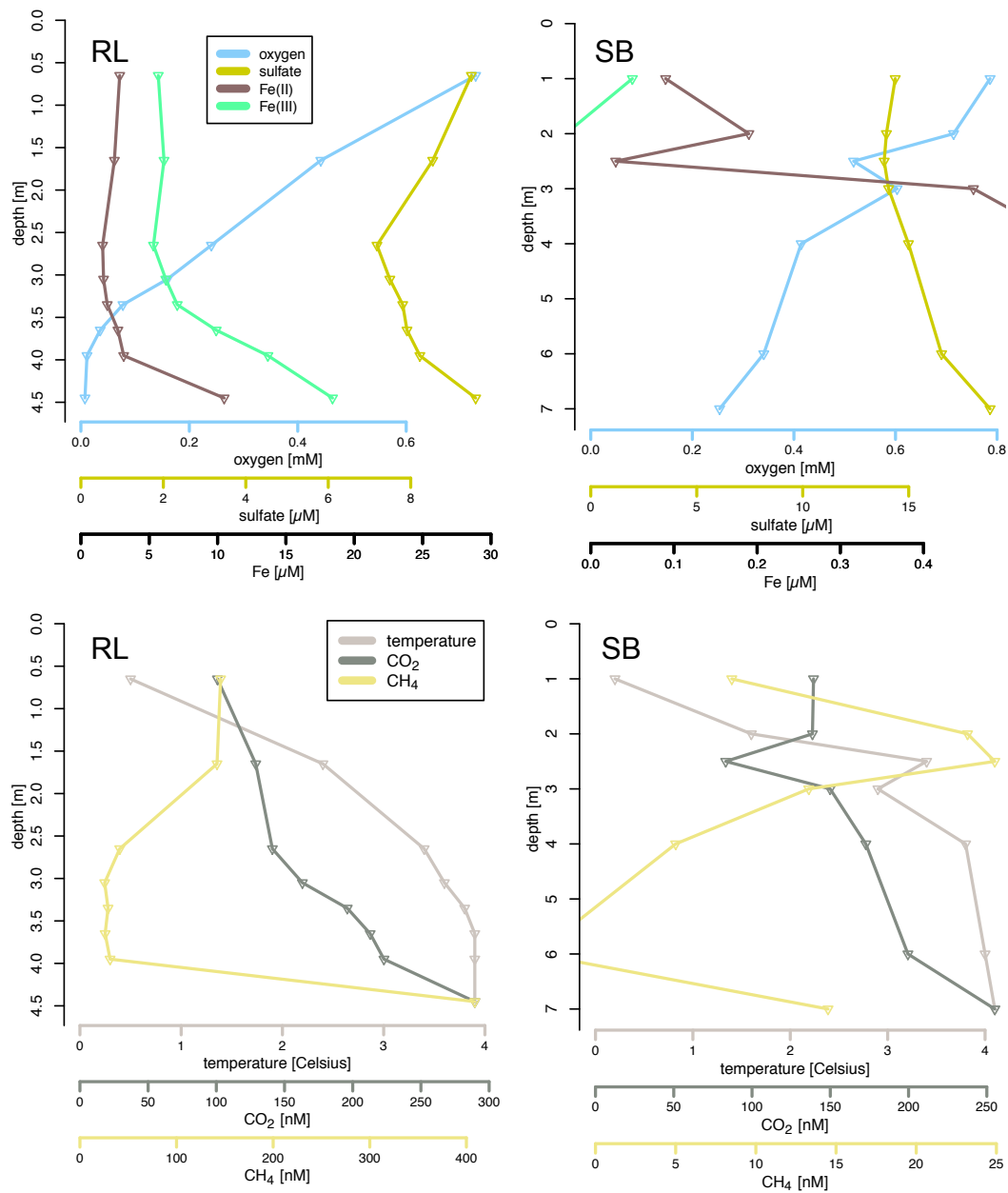


Figure S2. Chemistry and gas profiles in lakes RL and SB.

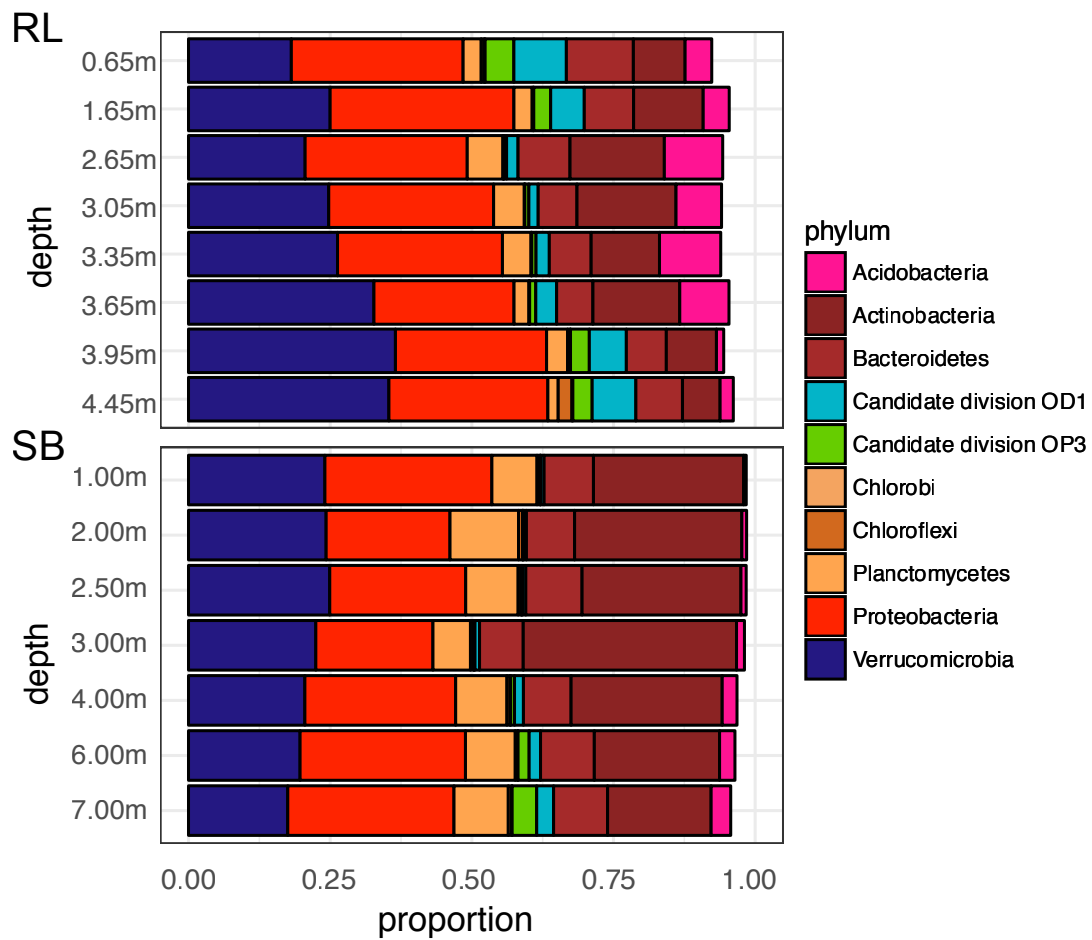


Figure S3. The distribution of the ten most abundant phyla in lakes RL and SB.

