bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Novel chemolithotrophic and anoxygenic phototrophic genomes extracted from ice-covered boreal lakes

Lucas Sinclair[1,2], Sari Peura[3,4], Pilar Hernández[4], Martha Schattenhofer[5], Alexander Eiler[6,7,*].

1 Department of Genetics and Evolution, Molecular Systematics and Environmental Genomics, Geneva University, Geneva, Switzerland

2 Sinclair.Bio - bioinformatics consultants, Geneva, Switzerland

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

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

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

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

7 eDNA solutions AB, Mölndal, Sweden

∗ Corresponding author: A. Eiler `alexander.eiler@icloud.com`

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

Version: May 16, 2017

Keywords: microbial diversity; 16S rRNA; metagenomics; freshwater; carbon cycle; green house gases; anaerobe microbiology

Abstract

Although an important fraction of the world's lakes remains ice-covered during a large proportion of the year, little is known about the microorganisms that govern the biogeochemical processes occurring under-ice along the stratigraphic redox gradients. Reconstructed genomes provide evidence for anoxygenic photosynthesis involving fixation of carbon using reduced sulphur 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 and supports the existence of methanotrophs in bottom anoxic waters. Reconstructed genomes matched methanotrophs related to *Methylobacter* tundripaludum, phototrophic *Chloroflexi* and

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

Introduction

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

addition to reduced mixing, the temperature gradient reverses in the opposite direction to that 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 gradients is expected to select for microbes with specific traits adapted towards the available electron acceptors and donors [10].

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. To what extent nitrogen, sulfur and iron cycling or fermentation regulate the degradation of organic matter and how phototrophic and lithotrophic processes intertwine with organic matter 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, photoferrotrophy or the coupling of denitrification and iron oxidation, the molecular and geochemical evidence has so far been slim, with few exceptions as in [9] [11].

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

Results

Correspondence of bacterial diversity with lake characteristics

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). At the water-ice interface, methane bubbles were observed in two of the systems, LB and BT,

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

with CH_4 concentrations above 5 µM. The profiles also indicated accumulation of methane in 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 µM in two systems (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 values ranging from 0.22 to 0.95 mM in the five systems.

Clear differences between the lakes concerning maximum concentrations and profiles of inorganic 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 concentrations. There was no oxygen gradient, but the water column was oxygenated throughout. Surface oxygen concentration close to saturation (13.8 mg/l at 2 °C) in RL and SB suggest 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 productive systems as indicated by the bacterial cell numbers (supplementary figure S1), oxygen was already depleted at 1 m depth from the surface. In the latter two, sulfate concentrations 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 KT with concentrations above 100 µM even in oxygen-depleted layers.

These differences in chemical profiles among the studied systems were also reflected in the unique bacterial communities present in individual lakes as assessed by 16S rRNA gene amplicon sequencing. As visualized in the ordination plot in figure 2A, depth profiles of individual lakes grouped within lake with no convergence at any depths. This selection of microbial communities by the specific conditions of each system is further emphasized by the observation that 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 by a redundancy analysis ($p < 0.001$). Further evidence is thus provided for the tight link between stratigraphic patterns in taxonomic composition and the depth-related variations in redox potential.

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

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

among the lakes and distinct stratigraphic patterns as seen in figure 2. The fifteen candidate phyla recruited on average 12.8% of the reads ranging from 2.6 to 28.5% per sample and these percentages increased toward the lake bottoms, as previously observed [14]. The most prominent 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 *Lentisphaera* at intermediate depths in LB. Specific taxonomic groups identified were indicative for aerobic (e.g. family *Comamonadaceae*) and anaerobic anoxygenic photosynthesis (e.g. phylum *Chlorobia*, and genus *Oscillochloridaceae*), recalcitrant polymer degradation (e.g. genera *Paludibacter* and *Chthoniobacter*) and methanotrophy (e.g. families *Methylococcales*, *Methylophilaceae* and genus *Candidatus Methylacidiphilum*) as seen in figure 2 and in supplementary figure S3. This detailed taxonomic analysis showed partial concordance with previously described boreal lake communities (for comparison see [14] [15] [16] [8])

Microbial traits along three under ice redox gradients

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 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 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 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 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, 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 processing are offered in online supplementary reports.

The final assemblies and mappings were used in combination with hidden Markov models (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-

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 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 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 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. Indeed, 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 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 systems [20] have provided evidence for such a coupling, despite high functional redundancy in microbial communities [21].

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 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 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 indicated 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 in sulfate concentrations and identified taxa such as *Geobacter*, *Desulfobulbus* and *Desulfovibrio*, as well as *Chlorobium*, point to an important role of sulfur cycling in the water column of these freshwater systems. Despite concentrations being much lower than those of marine systems, the importance of sulfur cycling has already been suggested for lake sediments [22]. Still, LB and BT were shown to have low sulfate and high iron concentrations as do many lakes in the boreal landscape.

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

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 potential to hydrolyze plant polymers (e.g. cellulose and hemicellulose) could be found throughout the water column while organisms capable of degrading phenolic compounds (e.g. lignin) were mainly found in the oxic portion of the water column.

Genes involved in several fermentative pathways were detected with a genome equivalence 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 waters of boreal lakes. Likewise, genes encoding the terminal hydrogenase of H_2-evolving fermentations (hydA) increased with depth. The abundance of formyltetrahydrofolate synthetase genes (fhs), which encodes the key enzyme of the Acetyl-CoA pathway of homoacetogenesis, was highest just below the ice and the oxycline as seen in figure 3 showing scaled absolute genome equivalent values close to 0.5 [23]. Key genes indicative for methanogenesis increased with depth together with the fraction of methanogenic *Archaea*. Three different orders of methanogenic archaea were found, *Methanobacteriales*, *Methanomicrobiales* and *Methanosarcinales*. The first two are hydrogenotrophic, producing CH_4 from H_2 and CO_2, whereas *Methanosarcinales* is metabolically more versatile carrying out hydrogenotrophic, acetoclastic and methylotrophic methanogenesis. The presence of these different types of methanogens was also verified by the homology to key enzymes described in these processes including glutathione-independent formaldehyde dehydrogenase (FdhA), hydrogenase subunit A (EchA), formylmethanofuran dehydrogenase subunit A (FmdA), formylmethanofuran-tetrahydromethanopterin N-formyltransferase (FTR), methenyltetrahydromethanopterin cyclohydrolase (MCH), methylenetetrahydromethanopterin dehydrogenase (MTD), coenzyme F420-dependent N5, N10-methenyltetrahydromethanopter reductase (MCH), tetrahydromethanopterin S-methyltransferase (MtrA), methyl-Co(III) methanol-specific corrinoid protein coenzyme M methyltransferase (MtaA), methyl-coenzyme M reductase alpha subunit (McrA), acetate kinase (AckA), acetyl-CoA synthetase (ACCS), phosphate acetyltransferase (PTA), heterodisulfide reductase subunit A (HdrA), acetyl-CoA decarbonylase/synthase complex subunit beta (CdhC). While co-occurring in all three systems, alternation from acetoclastic to hydrogenotrophic methanogens along the depth profiles differed among the three systems. Differences in the distribution of acetotrophic, methylotrophic and

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

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 *Methylococcaceae* and *Methylophilaceae* were abundant in all three lakes in bottom samples (maximum of 14.4 and 20.9%, respectively) as seen in figure 2. We did detect type II methanotrophs (serine based), relatives of the *Verrucomicrobium Methylacidiphilium* [24] and *Proteobacteria Methylobacter tundripaludum* [25]. Relatives to the anaerobic methanotroph *Candidatus Methylomirabilis* 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 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 acceptors. 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 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 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 presence when compared to pyruvate dehydrogenase.

Key findings from metagenome assembled genomes

To further validate the presence of metabolic potential such as methanotrophy and iron oxidation, genomes were reconstructed for the members of the under ice communities – an approach 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% contamination 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 each sample as their average coverages ranged from 0.12 to 6.57% of the mapped reads. For

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

First, the 57 MAGs were taxonomically assigned to eleven phyla including Candidate phylum *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 a more detailed taxonomic association, allowing the link with our amplicon data. After this assignment, the 57 MAGs were sent through multiple annotation pipelines to obtain a metabolic trait assignment, including their preferred electron acceptors and donors. Genomes of typical freshwater bacteria such as freshwater SAR11 (LD12) [32] and members of the acI clade [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 contained 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 light-mediated ATP production (i.e. bacteriothodopsin). Additional genomes including members of the phyla *Bacteriodetes*, *Acidobacteria* and *Plancomycetes*, contained a large variety 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 reader to the "MicroScope" pipeline [36] [37] with automatic annotations presented in KEGG maps and MetaCyc collections.

In addition to these heterotrophic degraders, we were able to obtain multiple MAGs related to previously described bacteria implicated in methanotrophy such as *Methylobacter tundripaludum*. The multiple genomes obtained related to *M. tundripaludum* contained *pmoABC* genes, and encoded for the complete reductive citrate cycle (Arnon-Buchanan cycle, rTCA) and 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] [38], it can be speculated that these organisms combine denitrification with anaerobic methane oxidation, using a pathway similar to that of *Candidatus* Methylomirabilis oxyfera [26]. This pathway is proposed to include a quinol-dependent nitric oxide reductase (q*nor*) and a nitric ox-

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

ide dismutase (*nod*) [39] with homologs identified in several of the *M. tundripaludum* MAGs. 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 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 cycle in the genomes. Other methanotrophs representing distinct lineages within the bacterial phylum *Verrucomicrobia*, such as *Candidatus Methylacidiphilum*, were present in the amplicon 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 *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 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 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 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 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 *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 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 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 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 a ACIII reaction center with a membrane-spanning electron transfer chain terminating in Fe-S

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

Under acidophilic conditions, few known microorganisms gain energy by the oxidation of 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 oxidation 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 *Planctomycetes* phylum. In addition, multiple genomes with strong signals of chemolithoautotrophy could be derived from the three systems, including reduction of inorganic nitrogen and the oxidation 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 capable of using multiple redox reactions within their population. These were *Betaproteobacteria* belonging to *Acetobacteraceae* and *Bradrhizobiaceae* with both chemolithoautotrophic and chemoorganoheterotrophic capability using oxygen, nitrate, Fe(III), sulfate and even organic compounds as electron acceptors. We also found an extremely coding-dense genome identified 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].

Discussion

Gross metabolic properties of the boreal lakes have been studied extensively indicating their net-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 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 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 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-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

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 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 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 plasticity occurring in microbial community of each system, we found that the majority of organisms probably lack the ability to perform multiple sequential redox transformations within 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 groups proliferating in each lake system. Our genomic data exposes diverse metabolic traits, such as photoferrotrophy, chemolithotrophy and methane oxidation, supporting alternative energy 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 depend 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 by the catchment, and, last but not least, by the physical properties (i.e. morphology and hydrology).

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 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 to nitrite, resulting in products that can be oxidized back to sulfide and nitrate by phototrophs or chemilithotrophs are very commonly encoded in the genomes. Such restricted metabolic 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 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 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.

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 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 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 [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 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 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.

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 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, 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 photosynthesis.

Materials and Methods

Sample design and environmental properties

In total, thirty-nine water samples were obtained from five boreal lakes (seven or eight samples 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 a two letter acronym for easy reference: RL, BT, KT (Lomtjärnen), LB (Liltjärnen) and SB (Fröåtjärnen) as some of the lakes posses no authoritative names to our knowledge. These water 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

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 laboratory, 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 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 chromatograph (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 according to Henry's law, correcting for temperature according to [57] and corrected for CH_4 in ambient air as measured on the GC.

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

Water for analyses of SO_4^{2-} analysis was first pre-filtered through rinsed 0.2 µm membrane filters (Pall Corporation) and then analyzed by ion chromatography on a Metrohm IC system (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 Fe(III) were measured with the ferrozine colorimetric method [59].

Samples for bacterial counts were fixed with 37% borax buffered formaldehyde (final concentration 2%) and stored at 4 °C prior to analyses. Cells were stained with the fluorescent nucleic acid stain Syto13 (Molecular probes, Invitrogen, Carlsbad, CA, USA) according to the protocol 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 fluorescence for triggered particle scoring. Cell counts were analyzed using Flowing Software version 2.5 (Perttu Terho, Centre for Biotechnology, Turku, Finland).

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 (Millipore), 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 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 [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 sequencing 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 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 AMPure 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 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, previously described in [61], was used. The sequence processing comprises steps for pairing reads, quality filtering, chimera checking, clustering with a 3% sequence dissimilarity threshold, taxonomic assignment, diversity estimations and result visualization.

Shotgun-metagenomic sequencing

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 according to the manufacture's protocol (R40048-08, QAM-094-002). The library size selection was carried out with AMPure XP beads (Beckman Coulter) in a 1:1 ratio. The prepared 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 quantified libraries were then prepared for sequencing on the Illumina HiSeq sequencing platform utilizing a TruSeq paired-end cluster kit, v3, and Illumina's cBot instrument to generate a clustered flowcell for sequencing. Two individual sequencing runs (an evaluation and re-run)

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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].

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Genome analysis

Genome annotations were performed using "MicroScope" with automatic annotations assisted by manual curation, as described in the integrated bioinformatics tools and the proposed annotation rules [36] [37]. In addition to the integrated annotation tools, which includes BlastP 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 [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 PSORTb [77] tools, we also used our local hmmsearch results. Synteny maps (i.e. conservation of local gene order) were used to validate the annotation of genes located within conserved 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]. Genomes can be retrieved from "MicroScope" with their specific identifiers as given in the supplementary material.

Statistical analyses

Statistical analyses were done using Python packages or the R language, including non-parametric multidimensional scaling (NMDS) plots, permuted multivariate ANOVA (PERMANOVA), redundancy analysis and the Procrustes superimpositions. The analyses were performed on rarefied 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 pathways linked to energy metabolism and carbon cycle when analyzing metabolic depth profiles. To plot the depth profiles of the traits, marker HMMs were selected as unique members of specific traits and are listed in supplementary table S1.

Acknowledgments

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 the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX), (ii) the

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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.

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

References

[1] W E Dean and E Gorham (1998) **Magnitude and significance of carbon burial in lakes, reservoirs, and peatlands**. *Geology*, *volume 26* (6), 535

[2] Lars J Tranvik, John A Downing, James B Cotner, Steven A Loiselle, Robert G Striegl, Thomas J Ballatore et al. (2009) **Lakes and reservoirs as regulators of carbon cycling and climate**. *Limnology and Oceanography*, *volume 54* (6-2), 2298

[3] Anthony K Aufdenkampe, Emilio Mayorga, Peter A Raymond, John M Melack, Scott C Doney, Simone R Alin et al. (2011) **Riverine coupling of biogeochemical cycles between land, oceans, and atmosphere**. *Frontiers in Ecology and the Environment*, *volume 9* (1), 53

[4] M Wik, R K Varner, K W Anthony, and S MacIntyre (2016) **Climate-sensitive northern lakes and ponds are critical components of methane release**. *Nature Geoscience*, *volume 1* (9), 99

[5] Stefan Bertilsson, Amy Burgin, Cayelan C Carey, Samuel B Fey, Hans-Peter Grossart, Lorena M Grubisic et al. (2013) **The under-ice microbiome of seasonally frozen lakes**. *Limnology and Oceanography*, *volume 58* (6), 1998

[6] Jan Karlsson, Reiner Giesler, Jenny Persson, and Erik Lundin (2013) **High emission of carbon dioxide and methane during ice thaw in high latitude lakes**. *Geophysical Research Letters*, *volume 40* (6), 1123

[7] K Salonen, M Leppäranta, M Viljanen, and R D Gulati (2009) **Perspectives in winter limnology: closing the annual cycle of freezing lakes**. *Aquatic Ecology*, *volume 43* (3), 609

[8] Stefan Rösel, Martin Allgaier, and Hans-Peter Grossart (2012) **Long-Term Characterization of Free-Living and Particle-Associated Bacterial Communities in Lake Tiefwaren Reveals Distinct Seasonal Patterns**. *Microbial Ecology*, *volume 64* (3), 571

[9] Monica Ricão Canelhas, Blaize A Denfeld, Gesa A Weyhenmeyer, David Bastviken, and

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Stefan Bertilsson (2016) **Methane oxidation at the water-ice interface of an ice-covered lake**. *Limnology and Oceanography*, *volume 61* (S1), S78

[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

[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

[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

[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

[14] Sari Peura, Alexander Eiler, Stefan Bertilsson, Hannu Nykänen, Marja Tiirola, 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

[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

[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

[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)

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[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

[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

[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

[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

[22] Marianne Holmer and Peter Storkholm (2001) **Sulphate reduction and sulphur cycling in lake sediments: a review**. *Freshwater Biology*, *volume 46* (4), 431

[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

[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

[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

[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

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[27] Gene W Tyson, Jarrod Chapman, Philip Hugenholtz, Eric E Allen, Rachna J Ram, Paul M Richardson et al. (2004) **Community structure and metabolism through reconstruction of microbial genomes from the environment**. *Nature Geoscience*, *volume 428* (6978), 37

[28] K C Wrighton, B C Thomas, I Sharon, C S Miller, C J Castelle, N C VerBerkmoes et al. (2012) **Fermentation, Hydrogen, and Sulfur Metabolism in Multiple Uncultivated Bacterial Phyla**. *Science*, *volume 337* (6102), 1661

[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-specific sweeps in natural bacterial populations**. *The ISME Journal*, *volume 10* (7), 1589

[30] Donovan H Parks, Michael Imelfort, Connor T Skennerton, Philip Hugenholtz, and Gene W Tyson (2015) **CheckM: assessing the quality of microbial genomes recovered 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 Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy**. *Applied and Environmental Microbiology*, *volume 73* (16), 5261

[32] Alexander Eiler, Rhiannon Mondav, Lucas Sinclair, Leyden Fernandez-Vidal, Douglas G Scofield, Patrick Schwientek et al. (2016) **Tuning fresh: radiation through rewiring of central metabolism in streamlined bacteria**. *The ISME Journal*, *volume 10* (8), 1902

[33] Sarahi L Garcia, Katherine D McMahon, Manuel Martinez-Garcia, Abhishek Srivastava, Alexander Sczyrba, Ramunas Stepanauskas et al. (2012) **Metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton**. *The ISME Journal*, *volume 7* (1), 137

[34] Trevor W Ghylin, Sarahi L Garcia, Francisco Moya, Ben O Oyserman, Patrick Schwientek, Katrina T Forest et al. (2014) **Comparative single-cell genomics reveals potential ecological niches for the freshwater acI Actinobacteria lineage**. *The ISME Journal*, *volume 8* (12), 2503

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[35] Sarahi L Garcia, Moritz Buck, Katherine D McMahon, Hans-Peter Grossart, Alexander Eiler, and Falk Warnecke (2015) **Auxotrophy and intrapopulation complementary in the 'interactome' of a cultivated freshwater model community**. *Molecular Ecology*, *volume 24* (17), 4449

[36] D Vallenet, S Engelen, D Mornico, S Cruveiller, L Fleury, A Lajus et al. (2009) **MicroScope: a platform for microbial genome annotation and comparative genomics**. *Database*, *volume 2009*, 021

[37] D Vallenet, E Belda, A Calteau, S Cruveiller, S Engelen, A Lajus et al. (2012) **MicroScope - an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data**. *Nucleic Acids Research*, *volume 41*, 636

[38] M M Svenning, A G Hestnes, I Wartiainen, L Y Stein, M G Klotz, M G Kalyuzhnaya et al. (2011) **Genome Sequence of the Arctic Methanotroph Methylobacter tundripaludum SV96**. *Journal of Bacteriology*, *volume 193* (22), 6418

[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 microbiology*, *volume 3*

[40] K Dimitri Kits, Dustin J Campbell, Albert R Rosana, and Lisa Y Stein (2015) **Diverse electron sources support denitrification under hypoxia in the obligate methanotroph Methylomicrobium album strain BG8**. *Frontiers in microbiology*, *volume 6* (74767), 68

[41] Donald A Bryant and Niels-Ulrik Frigaard (2006) **Prokaryotic photosynthesis and phototrophy illuminated**. *Trends in Microbiology*, *volume 14* (11), 488

[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 biosynthesis pathways in the filamentous anoxygenic phototrophic bacterium Oscillochloris trichoides DG-6 and evolution of anoxygenic phototrophs of the order Chloroflexales**. *Microbiology*, *volume 161* (1), 120

[43] Dimitry Y Sorokin, Sebastian Lücker, Dana Vejmelkova, Nadezhda A Kostrikina, Robbert

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

652 Kleerebezem, W Irene C Rijpstra et al. (2012) **Nitrification expanded: discovery, physiology 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) **Alternative 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 atmosphrere: 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**

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 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 Sturm et al. (2008) **Photoferrotrophs thrive in an Archean Ocean analogue**. *Proceedings of the National Academy of Sciences of the United States of America*, *volume 105* (41), 15938

[54] Xavier A Walter, Antonio Picazo, Maria R Miracle, Eduardo Vicente, Antonio Camacho, Michel Aragno et al. (2014) **Phototrophic Fe(II)-oxidation in the chemocline of a ferruginous meromictic lake**. *Frontiers in microbiology*, *volume 5*, 713

[55] Marc Llirós, Tamara García Armisen, François Darchambeau, Cédric Morana, Xavier Triadó Margarit, Özgül Inceoğlu et al. (2015) **Pelagic photoferrotrophy and iron cycling in a modern ferruginous basin**. *Scientific Reports*, *volume 5* (13803), 1

[56] David Bastviken, Jörgen Ejlertsson, and Lars Tranvik (2002) **Measurement of methane oxidation in lakes: a comparison of methods.**. *Environmental science & technology*, *volume 36* (15), 3354

[57] D R Lide and HPR Frederikse (1995) **CRC handbook of chemistry and physics: a ready-reference book of chemical and physical data**

[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 Public Health Association, American Water Works Association, and Water Pollution Control Federation.)

[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 Geochemistry*, *volume 15* (6), 785

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[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

[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

[62] NA Joshi and JN Fass (2011) *Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files*. Version 1.33

[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

[64] Ben Langmead and Steven L Salzberg (2012) **Fast gapped-read alignment with Bowtie 2**. *Nature Methods*, *volume 9* (4), 357

[65] Aaron R Quinlan and Ira M Hall (2010) **BEDTools: a flexible suite of utilities for comparing genomic features**. *Bioinformatics*, *volume 26* (6), 841

[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

[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

[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

[70] UniProt Consortium (2014) **Activities at the Universal Protein Resource (UniProt)**. *Nucleic Acids Research*, *volume 42*, D191

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[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), 41

[72] A Mitchell, H Y Chang, L Daugherty, M Fraser, S Hunter, R Lopez et al. (2015) **The InterPro protein families database: the classification resource after 15 years**. *Nucleic Acids Research*, *volume 43*, 213

[73] F Meyer, R Overbeek, and A Rodriguez (2009) **FIGfams: yet another set of protein families**. *Nucleic Acids Research*, *volume 37* (20), 6643

[74] C Claudel-Renard (2003) **Enzyme-specific profiles for genome annotation: PRIAM**. *Nucleic Acids Research*, *volume 31* (22), 6633

[75] E L Sonnhammer, G von Heijne, and A Krogh (1998) **A hidden Markov model for predicting transmembrane helices in protein sequences**. *International Conference on Intelligent Systems for Molecular Biology*, *volume 6*, 175

[76] Jannick Dyrløv Bendtsen, Henrik Nielsen, Gunnar von Heijne, and Søren Brunak (2004) **Improved prediction of signal peptides: SignalP 3.0**. *Journal of molecular biology*, *volume 340* (4), 783

[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 from comparative proteome analysis**. *Bioinformatics*, *volume 21* (5), 617

[78] Minoru Kanehisa, Susumu Goto, Yoko Sato, Masayuki Kawashima, Miho Furumichi, and Mao Tanabe (2013) **Data, information, knowledge and principle: back to metabolism in KEGG**. *Nucleic Acids Research*, *volume 42*, D199

[79] Nicola Segata, Daniela Börnigen, Xochitl C Morgan, and Curtis Huttenhower (2013) **PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes**. *Nature Communications*, *volume 4*, 2304

[80] R Conrad (1999) **Contribution of hydrogen to methane production and control of hy-**

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figures

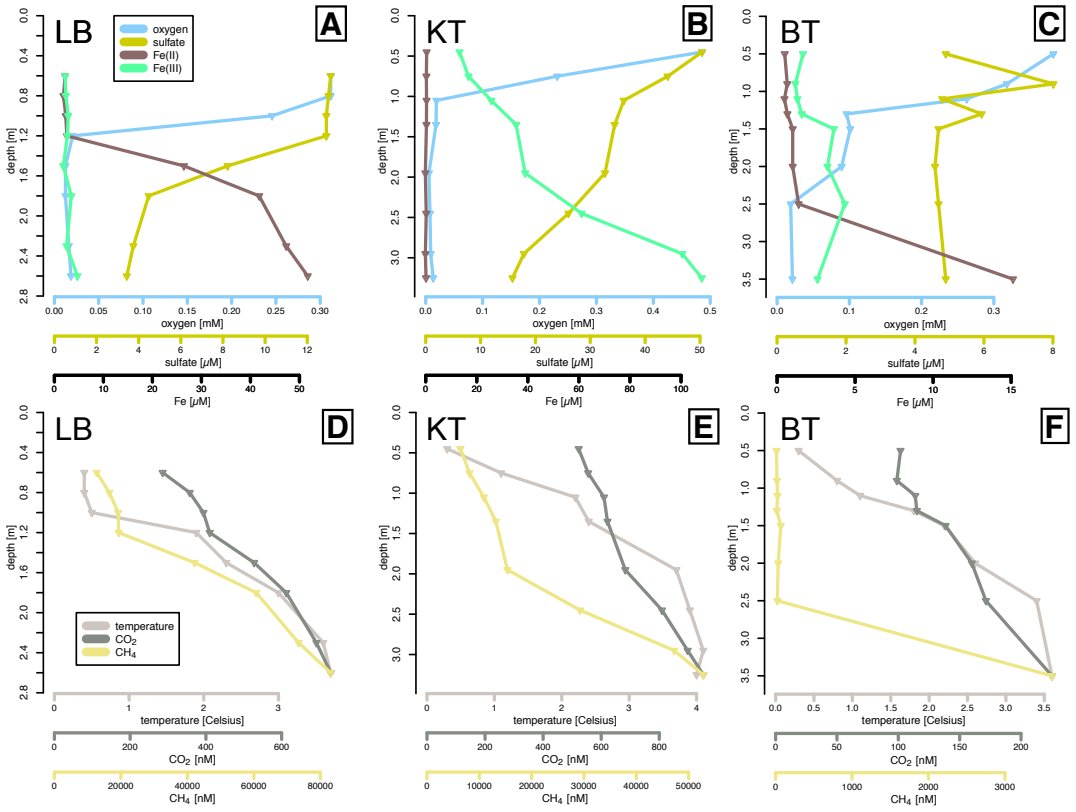

Figure 1. Summary of lake characteristics.

Profiles of water chemistry (A-C) and gas concentrations (D-F) in lakes LB (A,C), KT (B,D) and BT (C,E). For RL and SB see supplementary figure S2.

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

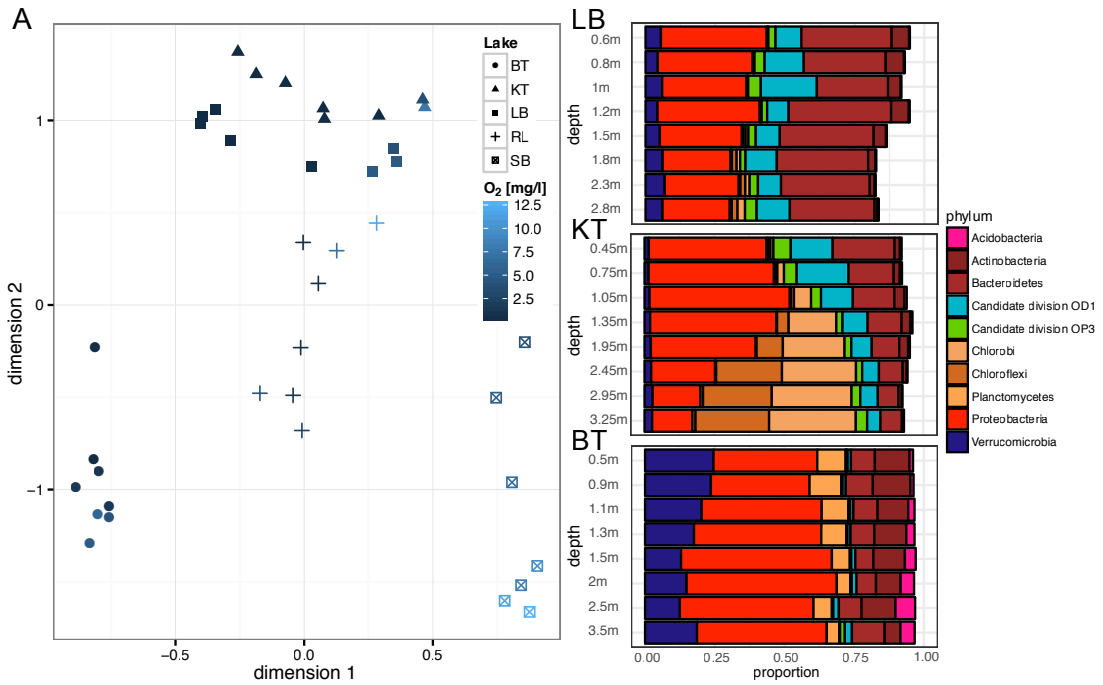

Figure 2. Bacterial community profiles by lake and by depth.

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

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

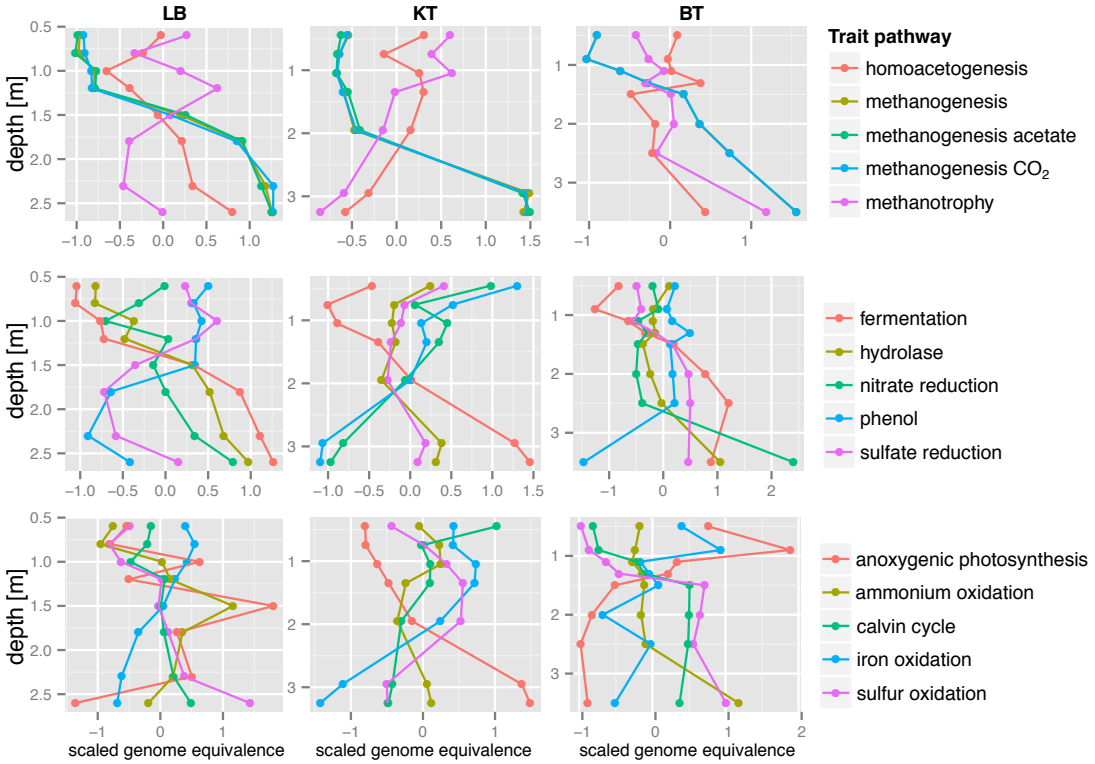

Figure 3. Genome encoded metabolic profiles by depth.

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

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

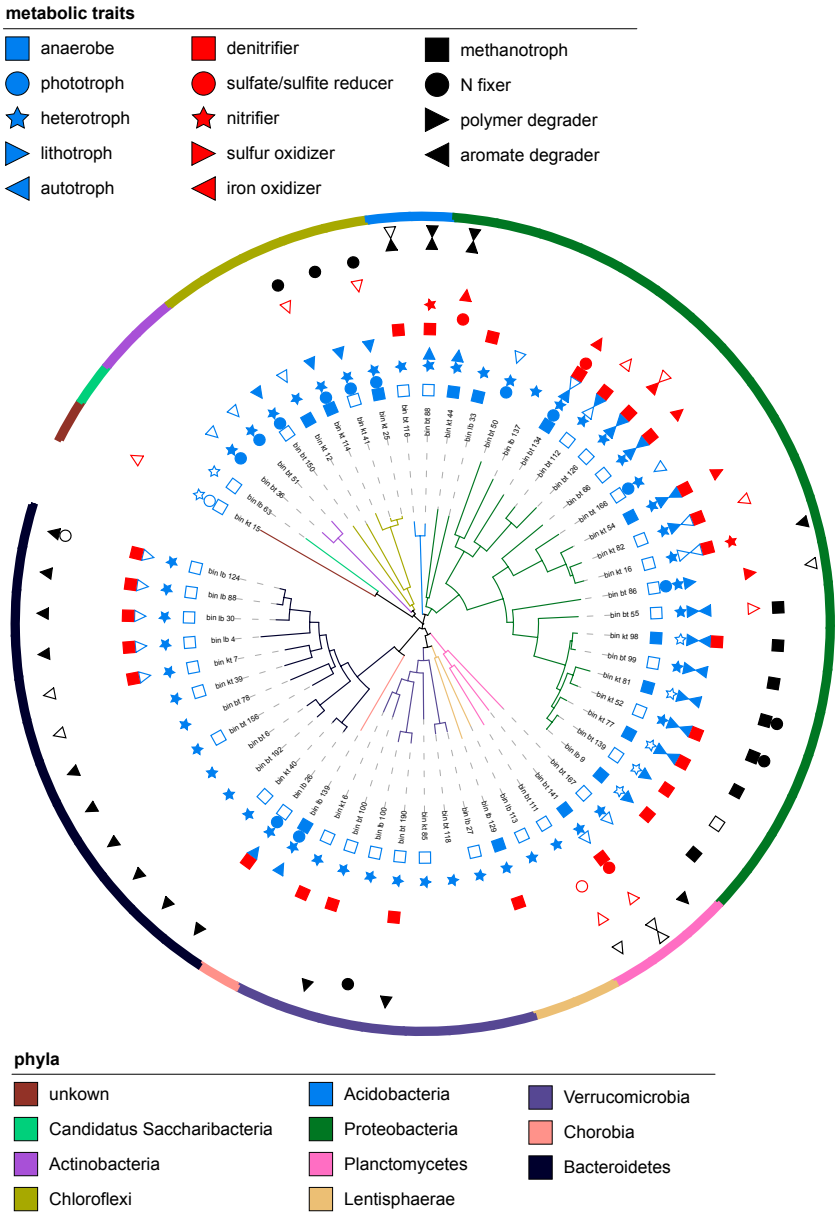

Figure 4. Traits, taxonomy and phylogeny of reconstructed genomes.

768 Phylogenomic tree of recovered genomes as computed by PhyloPhlAn [79]. The outer ring delineates taxonomic affiliation. The inner circles show the metabolic traits as based on genome 770 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.

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

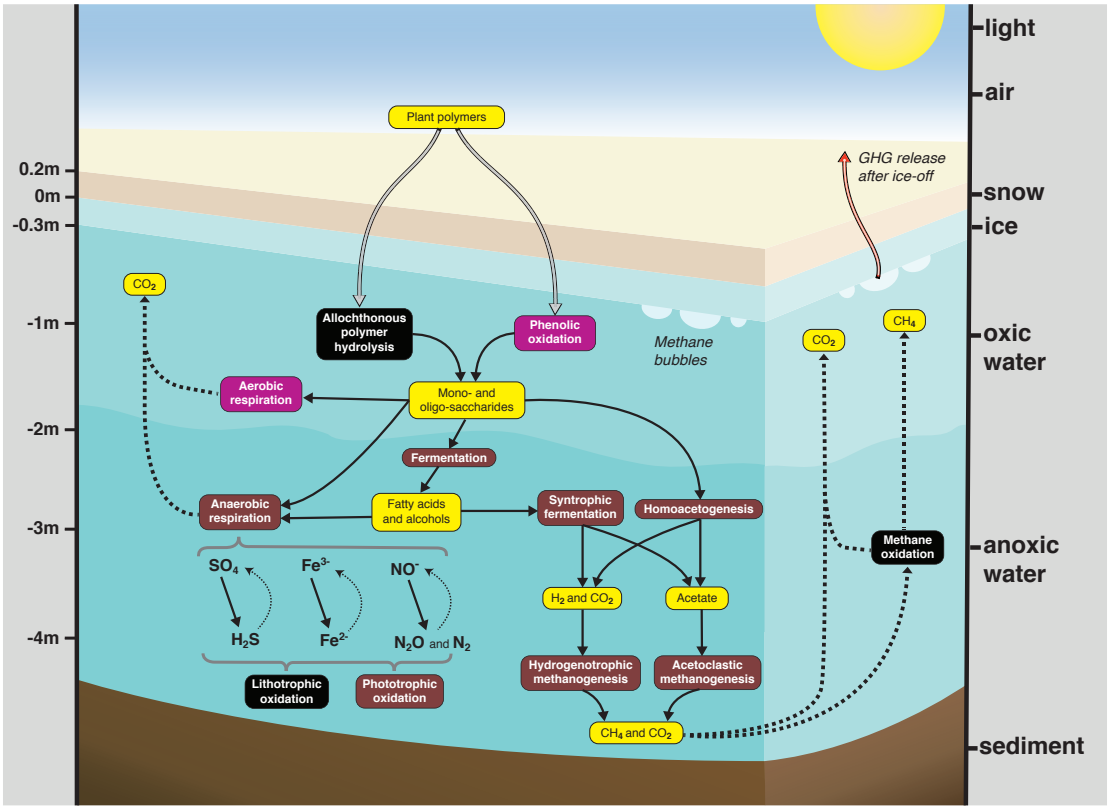

Figure 5. Main degradation pathways of allochthonous organic matter.

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].

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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.

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Supporting Information

Table S1. Selected PFAMs traits.

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 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 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 numbers and rows are PFAMS. Each value indicates how many times each PFAM was found in each bin.

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

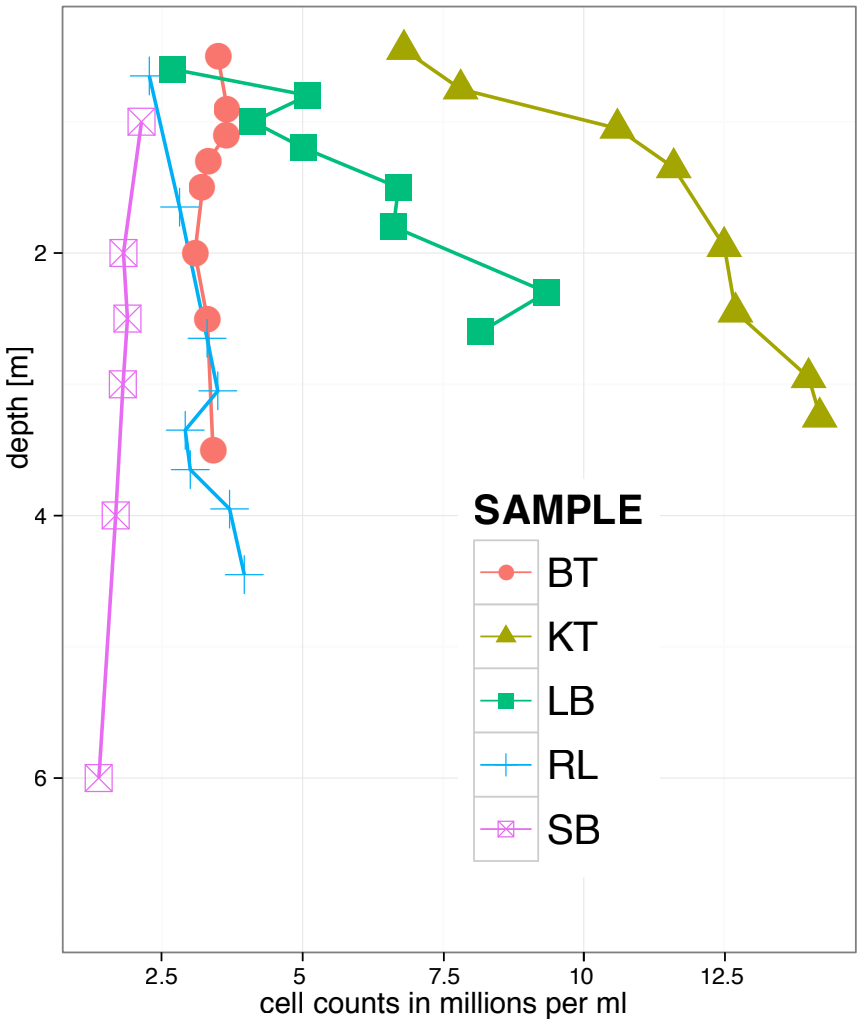

Figure S1. Cell counts per depth for all lakes.

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

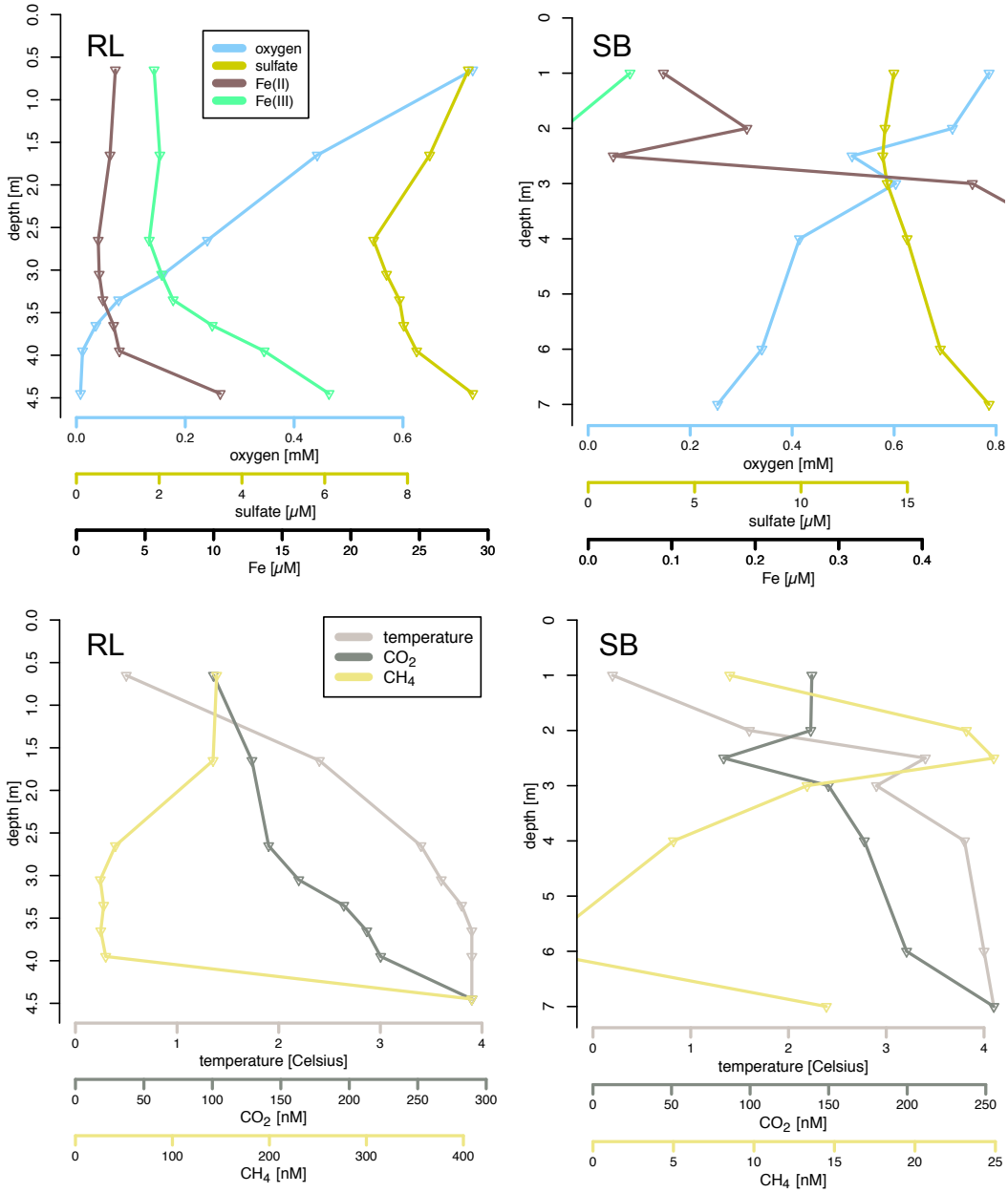

Figure S2. Chemistry and gas profiles in lakes RL and SB.

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

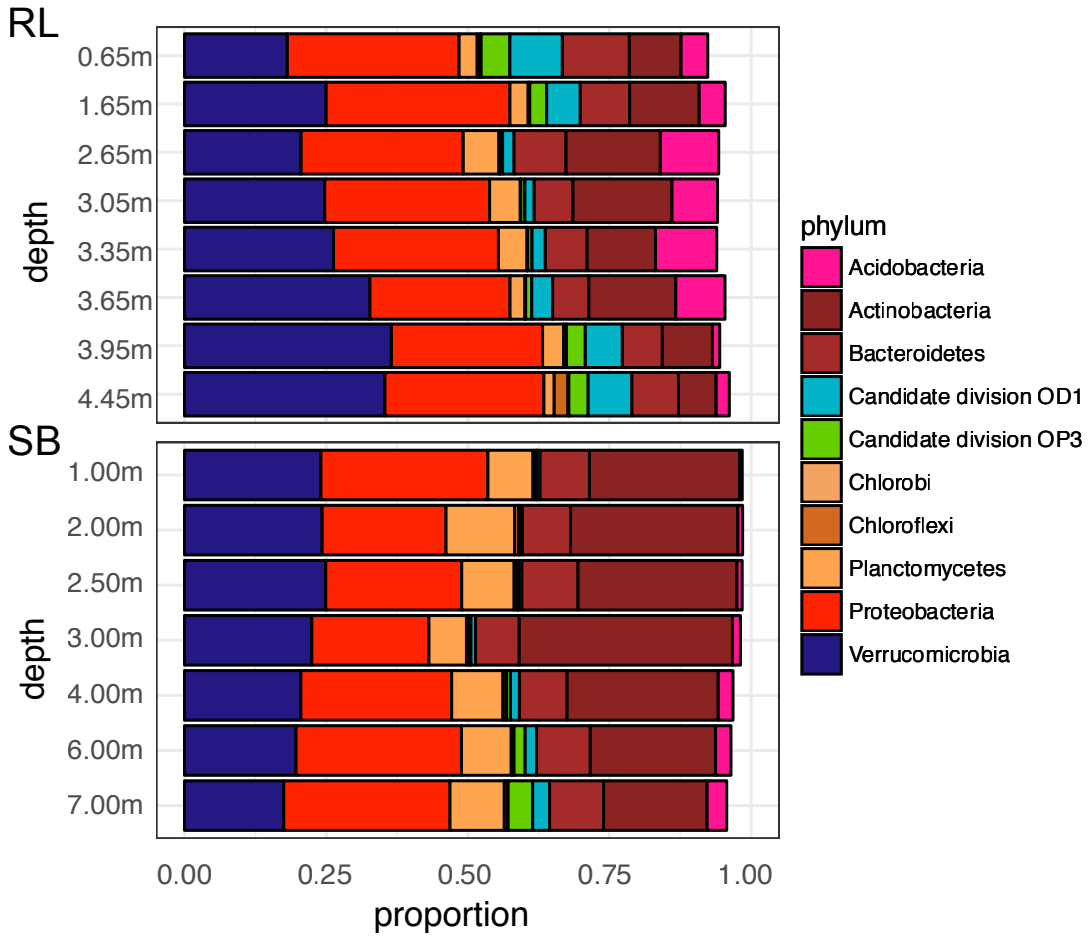

Figure S3. The distribution of the ten most abundant phyla in lakes RL and SB.

bioRxiv preprint doi: https://doi.org/10.1101/139212; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

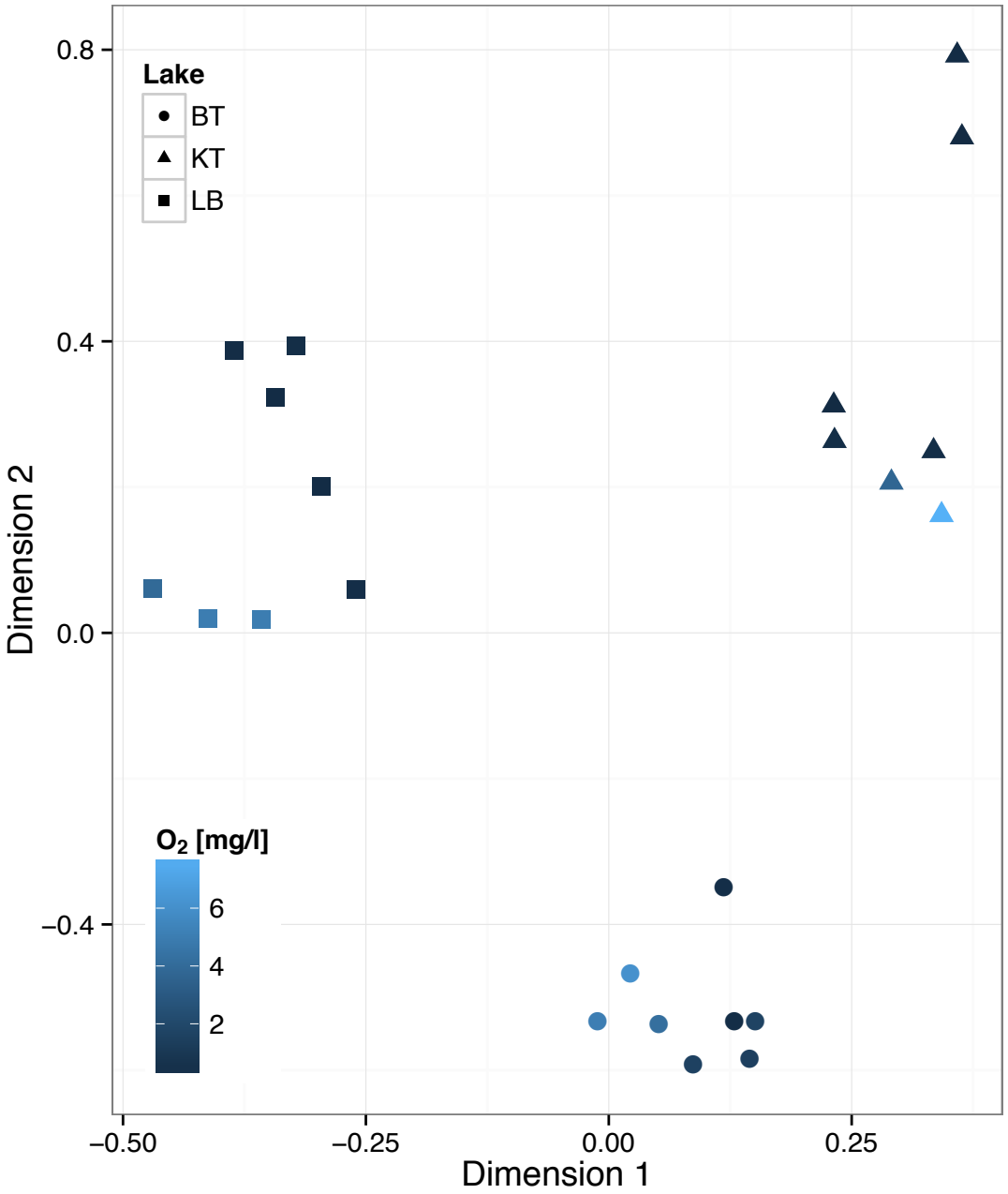

Figure S4. NMDS based on PFAMs found in three lakes.