Title: Archaea in boreal Swedish lakes are diverse, dominated by Woesearchaeota and follow deterministic community assembly

Heli Juottonen[1,2], Laurent Fontaine[3], Christian Wurzbacher[4,5], Stina Drakare[6], Sari Peura[1,7], Alexander Eiler*[1,3,8]

1 Limnology, Department of Ecology and Genetics, Uppsala University, Norbyvägen 18D, 75234 Uppsala, Sweden

2 Department of Biological and Environmental Science, University of Jyväskylä, P.O. Box 35, 40014 University of Jyväskylä, Finland

3 Section for Aquatic Biology and Toxicology, Centre for Biogeochemistry in the Anthropocene, Department of Biosciences, University of Oslo, Blindernv. 31, 0371 Oslo, Norway

4 Department of Biological and Environmental Sciences, University of Gothenburg, Box 463, 405 30 Göteborg, Sweden

5 Chair of Urban Water Systems Engineering, Technical University of Munich, Am Coulombwall 3, 85748 Garching, Germany

6 Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, SLU, Box 7050, 750 07 Uppsala, Sweden

7 Department of Forest Mycology and Plant Pathology, Science for Life Laboratory, Swedish University of Agricultural Sciences, Almas allé 5, 75007 Uppsala, Sweden

8 eDNA solutions AB, Björkåsgatan 16, 43131 Mölndal, Sweden

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1462-2920.15058

Corresponding author: Alexander Eiler, Section for Aquatic Biology and Ecotoxicology, Center for BioGeoChemistry in the Anthropocene, Department of Biosciences, University of Oslo, Blindernv. 31, 0371 Oslo, Norway

Email: alexander.eiler@ibv.uio.no; Telephone: +47-22854527

Originality-Significance Statement

Archaea perform many key functions in freshwater environments from the oxidation of ammonium to the production of methane. Here, we show that archaea, including members of the DPANN superphyla, represent a diverse part of the epilimnic planktonic community in lakes. Archaeal diversity was shown to be structured by deterministic processes specifically environmental variables related to nutrient status. We further expand knowledge about the ecological range of key archaea and link taxa and individual sequence variants to hypotheses about processes governing biogeochemical cycles in lakes.

Summary

Despite their key role in biogeochemical processes, particularly the methane cycle, archaea are widely underrepresented in molecular surveys because of their lower abundance compared to bacteria and eukaryotes. Here, we use parallel high-resolution small subunit rRNA gene sequencing to explore archaeal diversity in 109 Swedish lakes and correlate archaeal community assembly mechanisms to large-scale latitudinal, climatic (nemoral to arctic), and nutrient (oligotrophic to eutrophic) gradients. Sequencing with universal primers showed the contribution of archaea was on average 0.8% but increased up to 1.5% of the three domains in forest lakes. Archaea-specific sequencing revealed that freshwater archaeal diversity could be partly explained by lake variables associated with nutrient status. Combined with deterministic co-occurrence patterns this finding suggests that ecological drift is

overridden by environmental sorting, as well as other deterministic processes such as biogeographic and evolutionary history, leading to lake-specific archaeal biodiversity. Acetoclastic, hydrogenotrophic and methylotrophic methanogens as well as ammonia-oxidizing archaea were frequently detected across the lakes. Archaea-specific sequencing also revealed representatives of Woesearchaeota and other phyla of the DPANN superphylum. This study adds to our understanding of the ecological range of key archaea in freshwaters and links these taxa to hypotheses about processes governing biogeochemical cycles in lakes.

Introduction

Lakes around the globe receive, transport and transform sizable amounts of carbon (Battin et al., 2009; Tranvik et al., 2009). Yet the magnitude of their role in global carbon cycling remains uncertain and may rest to a large extent on poorly understood microbes that drive ecosystem-scale processes. Moreover, these freshwater systems are thought to be particularly sensitive to climate warming (Schneider and Hook 2010) enhancing microbial productivity (Prowse and Stephenson 1986, Rouse et al. 1997) and ultimately biogeochemical processes (Wik et al. 2016, Thornton et al. 2015) and water quality (Weyhenmeyer et al. 2016, Roulet and Moore 2006).

To construct accurate models for water quality and predict the role of lakes in global biogeochemical cycles, such as the production of greenhouse gases (GHGs) including carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O), it is essential to understand the microbes underpinning these processes. Most studies specific to freshwater archaea have focused on sediments and wetlands, because these habitats are important CH_4 sources (Bastviken et al., 2004; Bridgham et al., 2013). Pelagic and oxygenated freshwaters are, however, also a source of CH_4 (Bogart et al., 2014; Grossart et al., 2011; Angle et al., 2017) and show considerable archaeal diversity (Auguet et al., 2011; Hugoni et al., 2013).

In contrast to extensive studies of bacterial (e.g. Newton et al., 2011; Eiler et al., 2012; Savio et al., 2015; Ruiz-Gonzalez et al., 2017) and to a lesser extent eukaryotic

microbial diversity (Debroas et al., 2017) in freshwater systems, archaeal diversity is largely under sampled in freshwaters. Because archaea form only a minor fraction of the microbial community, their diversity is poorly represented in studies using PCR primers that amplify both bacterial and archaeal 16S rRNA genes (Caporaso et al., 2012; Wang et al., 2012; Klindworth et al., 2013; Thompson et al., 2017). Despite their low abundance, archaea are expected to be a critical component of the microbial community in freshwaters, due to their role in CH_4 and nitrogen cycling and the unknown metabolic potential of the recently described archaeal lineages.

The best-known archaea in the pelagic zones of freshwaters are CH_4 producing archaea (methanogens) and ammonia-oxidizing archaea (AOA). Activity of methanogens is the main source of biogenic CH_4, and nitrification by AOA contributes to N_2O production. Methanogens often found in lakes include Methanobacteriales, Methanosarcinales, and Methanomicrobiales (Earl et al, 2003; Borrel et al, 2011) that are particularly abundant in the anoxic bottom waters of dystrophic lakes (Peura et al. 2015). AOA comprise the phylum Thaumarchaeota and groups such as Nitrosoarchaeum-like (group I.1a) and Nitrosotalea-like (SAGMGC-1) archaea. AOA have been shown to inhabit in particular oligotrophic surface freshwaters (Pouliot et al. 2009; Hu et al. 2010; Auguet and Casamayor 2013; Hugoni et al. 2013; Berdjeb et al. 2013; Mukherjee et al. 2016). Much less is known about the distribution and drivers of planktonic freshwater archaea other than methanogens or AOA. In a survey of high-altitude lakes, the main archaeal groups were

Woesearchaeota and Parvarchaeota, recently described lineages with poorly defined functions (Ortiz-Alvarez and Casamayor 2016). In a comparison of two alpine lakes differing in trophic status, environmental factors explained considerably less spatio-temporal variation of the archaeal community than in a parallel study on bacteria (Berdjeb et al. 2013). In summary, there is a need for more extensive phylogenetic sampling and characterization of the habitat preferences of archaea in freshwaters, as well as to develop hypotheses about the assembly of the archaeal community in freshwaters.

The structure of the archaeal community can be assumed to depend on the balance between stochastic and deterministic processes. Stochastic processes (i.e. ecological drift) will result in random combinations of taxa, whereas, if deterministic processes dominate, predictable patterns of taxa distributions and abundances will emerge. Deterministic processes have been shown to override random processes in macroorganisms (Gotelli and McCabe, 2002), and this has also been demonstrated for bacteria and eukaryotes in a number of habitats (Horner-Devine et al. 2007; Otiferu et al. 2010; Caruso et al. 2011; Eiler et al. 2011; Vanwonterghem et al. 2014). Important deterministic processes that determine community composition are thought to be environmental filtering and species interactions (Diamond, 1975; Gotelli and McCabe, 2002), as well as biogeographic and evolutionary history (Vuilleumier and Simberloff, 1980; Cracraft, 1988). However, to our knowledge it has not been tested

if the dominance of deterministic processes also applies to freshwater archaeal communities.

Here, we use high-throughput amplicon sequencing of i) a universal region of the small subunit (SSU) rRNA gene (V6-V8) covering all three domains, and ii) the V3-V5 region of the archaeal 16S rRNA gene (Gantner et al. 2011) to obtain a detailed measure of archaeal diversity in surface water of 109 boreal lakes. We compare phylogenetic diversity of lake archaea to environmental variables such as catchment land use/cover, hydrological, and geochemical properties to determine what variables best correlate with the distribution of archaeal taxa, and the roles random (i.e. ecological drift) and deterministic (i.e. environmental sorting and dispersal) factors in archaeal community assembly.

Results and Discussion

Characteristics of surveyed lakes

We explored and quantified variation in the diversity of epilimnic freshwater archaea at latitudes ranging from 55.4° to 68.3° (Figure S1) representing globally the latitudes with the highest concentration, area, and perimeter of inland water bodies (Verpoorter et al., 2014). The sampled lakes span from nemoral to arctic (subalpine) vegetation zones and represent summer conditions as they were all sampled during August 2014. Metadata associated with each of the 119 sampled lakes (from 109 sufficient archaeal sequences were retrieved), or at least a substantial subset thereof, included latitude and longitude, temperature, chlorophyll concentration, nutrients, and catchment characteristics (for summary statistics see Table S1).

Besides varying in latitude, catchment characteristics and temperature (range 10.3 - 24.7 ºC), the lakes varied in nutrient content. Total organic carbon (TOC) in the lakes ranged from 0.6 to 31 mg l^{-1} (median 13 mg l^{-1}), total phosphorus (TP) from 2 to 136 µg l^{-1} (median 10 µg l^{-1}), total nitrogen (TN) from 60 to 1280 µg l^{-1} (median 360 µg l^{-1}), and pH from 4.8 to 9.0 (median 6.7). TOC, TP and TN correlated positively with each other and with turbidity and negatively with latitude (Figure 1). Lake size varied from 0.03 to 14 km^2 with a median of 0.44 km^2 (Table S1).

Archaeal contribution and diversity in freshwater lakes

Based on universal sequence reads of archaeal, bacterial and eukaryotic SSU rRNA genes, archaeal SSU rRNA genes comprised from 0.03% up to 1.5% of the overall diversity (Figure S2). This range with the average of 0.8% of archaeal reads corresponds to the lower ranges of previous studies of archaeal relative abundances in lakes (Pernthaler et al., 1998; Glöckner et al., 1999; Keough et al., 2003; Auguet and Casamayor, 2008; Ortiz-Alvarez et al., 2016). Partial least squares modelling revealed that the relative abundance of archaeal in relation to bacterial and eukaryotic SSU rRNA genes increased with catchment land cover classified as forests and wetlands (Figure 1). The highest relative abundances were found in dystrophic lakes. Many dystrophic lakes are characterized by anoxic bottom waters (hypolimnion) where electron acceptors for respiration are highly depleted. Accordingly, the genomes of hypolimnion microbes show potential for fermentation and methanogenesis (Peura et al. 2015; Peura et al. 2018). A high prevalence of archaea also in the surface waters (epilimnion) of these net-heterotrophic, greenhouse-gas-emitting systems can be speculated to be the result of anoxic microenvironments (Gossart et al., 2011) and their transitory occurrence in oxygenated waters.

In silico analysis of the selected PCR primers against the SILVA 16S rRNA database predicted 86.8% coverage of archaea for the archaea-specific primers and 71.5% coverage of archaea for the universal primers. The archaeal primers had only minor cross-domain amplification predicted which was confirmed in the sequencing: on average, 95.2% (range 72.2 - 100% per sample) of the archaeal sequencing reads

mapped to sequences of the target domain. While our approach targeted most of the archaeal diversity deposited to databases at the time of analysis, recently discovered archaeal groups emerging through random shotgun metagenomic sequencing were potentially missed by our amplicon sequencing approach. Only 4.9% of Asgardaeota in SILVA matched the archaea-specific primers, and we determined an incomplete matching to Altiarchaeota (40.7%), Diapherotrites (52.5%) and Nanoarchaeota (58.5%). The selection against the unknown diversity is a well-known limitation of primer-based sequencing approaches (Karst et al., 2018), and thus uncertainty in the taxonomic coverage needs to be accounted for in data interpretation.

The archaea-specific amplification detected in total 119,483 archaeal amplicon sequence variants (ASVs) with an average of 1309 ASVs per sample (range 107 - 4024). Rarefaction analysis suggested that a large part of the diversity in individual lakes was recovered (Figure 3A). Most of the archaeal ASVs occurred in single lakes (90.6%) with another 7.1% occurring in two or three lakes. The remaining 2.0% of the ASVs (n=2488) were found in more than three lakes. In these lakes these ASVs were the dominant ASVs with a combined relative abundance ranging from 22.7 to 82.4% per lake and an average of 50.1% across the lakes (Figure S3A). There were 346 ASVs that occurred in more than 10 lakes, and six ASVs (including a Woesearchaeia, a *Methanobacterium*, two *Methanosaeta* and two *Methanoregula* ASVs) that occurred in more than 50 lakes. The most prevalent ASV (a Woesearchaeia) occurred in 76 lakes (prevalence distribution of the ASVs is shown in

Figure S3B). The high number of ASVs that were restricted to a single lake together with the low prevalence of most ASVs (Figure S3 C) indicates high among-lake richness. High richness was confirmed using species accumulation curves (SAC) (Figure 3B). Unsaturation strongly suggests that the ASV pool is widely undersampled along the broad environmental gradients sampled, despite sequencing effort with almost 120,000 unique ASVs detected. However, although these results are based on denoised data, which aims to remove artificial diversity introduced by PCR amplification and sequencing, outputs likely include incorrect sequences which can inflate richness and distort community composition.

Community-level biogeography corresponds with lake characteristics

We examined trends in archaeal diversity and richness in relation to lake physicochemical variables, latitude and catchment land cover. Archaeal diversity measures (phylogenetic PD, inverse Simpson, Shannon, Fisher indices, ACE, Chao1, observed ASVs) were highly intercorrelated ($r > 0.7$, $P > 0.001$) and positively correlated with TOC, TP, TN, chlorophyll a, conductivity (ion concentrations) and turbidity (Figure 1). All of these environmental variables are indicative of productivity, suggesting that high productivity enhances archaeal richness. Increased richness with increased productivity, as predicted by ecological theory (Cardinale et al., 2009; Duffy et al., 2017), was also corroborated by positive correlations with the combined bacterial, eukaryotic and archaeal diversity in the universal primer data set.

Archaeal beta diversity across the lakes showed a similar pattern of community composition when assessed either by Bray-Curtis distance or by UNIFRAC distance (Procrustes superimposition; R = 0.261, p = 0.001). Among-lake community dissimilarity based on either of the two distance measures was poorly explained by the spatial distance between the lakes (Partial Mantel test controlling for environmental variables, $R_{unifrac}$ = 0.03, $p_{unifrac}$ = 0.19, R_{bray} = 0.08, p_{bray} = 0.011).This result does not support a distance-decay relationship of the archaeal species distribution, similar to what has been described for bacteria (Bell, 2010). Controlling for the effects of between-lake geographic distance revealed that archaeal community structure was weakly correlated with the measured environmental variables (Partial Mantel test, $R_{unifrac}$ = 0.18, $p_{unifrac}$ = 0.04, R_{bray} = 0.17, p_{bray} = 0.001). Correlation showed that archaeal phylogenetic composition was associated with lake productivity (nutrients, TN and TP and chlorophyll concentrations), as well as with pH, conductivity and other lake physicochemical variables (Table 1).

As observed in other large-scale spatial studies on bacteria and protists (e.g. Lima-Mendez et al., 2015; Thompson et al., 2017), archaeal beta diversity was significantly related to many potential explanatory variables, indicating that environmental sorting contributes to structuring archaeal communities over large regional scales. However, ecological drift, a stochastic process, should be the driving force of archaeal community assembly when the basic entities of diversity are assessed at the ASV level. Populations of ASVs are expected to be small and inhabit geographically

isolated habitats (i.e. lake systems), a pattern that is supported by the low abundance and narrow distribution of most archaeal ASVs in our study lakes. Regarding the low abundant ASVs, demographic stochasticity is expected to play a strong role. In addition, drift processes may dictate the likelihood of population detection when dormant life stages, such as a seed bank, dominate the community (Lennon and Jones, 2011). Consequences of ecological drift are that ASV abundances fluctuate randomly, increasing the differences among otherwise equivalent communities (Gilbert and Levine, 2017). Such fluctuations potentially explain the high number of ASVs with low prevalence, often occurring only in single lakes.

To assess the importance of random vs. deterministic factors shaping the archaeal communities, we performed Monte Carlo simulations on the co-occurrence patterns of the ASVs in the study lakes. We observed a positive Standard Effective Size (5.34) significantly different from random ($p < 0.001$) indicative of low co-occurrence and deterministic processes being important in community assembly. Although random ecological drift is suggested to be ever present, important deterministic processes that determine archaeal community composition can be environmental filtering and species interactions (Diamond, 1975; Gotelli and McCabe, 2002), as well as biogeographic and evolutionary history (Vuilleumier and Simberloff, 1980; Cracraft, 1988). According to the size-dispersal hypothesis, small organisms such as archaea are more likely affected by species sorting than dispersal limitation, because organisms with a size on the µm-scale can widely disperse (Cottenie, 2005; Beisner et

al., 2006; Shurin et al., 2009). Thus, the distribution is mainly a reflection of the environmental properties (Farjalla et al., 2012). Lack of distance decay in the archaeal community composition along our geographic gradient, as suggested by the partial Mantel tests, further emphasizes that environmental sorting plays a more prominent role than deterministic dispersal processes such as mass effects or biogeographic history in community assembly of archaea. Our finding corroborates results from study on bacteria (Lindström and Langenheder 2011).

Taxonomical distribution of archaea across freshwater lakes

The most abundant archaeal classes across the successfully sequenced 109 lakes were Woesearchaeia, Methanomicrobia and Nitrosophaeria (Fig. 4A). As there is a current revolution in archaeal taxonomy, we also linked SILVA taxonomy (v. 132) with other current taxonomic hierarchies (Castelle et al., 2015). The Woesearchaeia (previously DHVEG-6 and Parvarchaea, also termed Woesearchaeota phylum), which are recently discovered members of the DPANN superphylum (Castelle et al., 2015), were the most dominant class in our dataset in total number of reads and the number of unique ASVs (n > 58 000). While most lakes were dominated by Woesearchaeia, in a fifth of the lakes the dominant archaeal class was Methanomicrobia or Nitrosphaera (Figure 4A). Redundancy analysis indicated that nutrient status and the aromatic character of the dissolved organic matter are important explanatory variables underpinning the taxonomic shift at the class level (Figure 4B). However, if this shift reflects the metabolic predictions from the available Woesearchaeota genomes such

as potentially fermentative metabolism (Castelle et al., 2015; Lazar et al., 2017) or symbiotic lifestyle (Castelle et al., 2015) is still unknown.

Methane-cycling archaea and their distribution

Lakes are sources of greenhouse gases such as CH_4, CO_2 and N_2O, with archaea playing a particularly important role in CH_4 cycling. Consistent with other freshwater systems such as wetlands (Borrel et al., 2011; Bridgham et al., 2013; Narrowe et al., 2017), the second most dominant archaeal class was methanogenic Methanomicrobia. We identified multiple methanogen groups including hydrogenotrophic methanogens (*Methanoregula, Methanobacterium*), acetoclastic methanogens (*Methanosaeta*) and as minor groups methylotrophic methanogens that use methylated compounds such as methylamines and methanol to produce CH_4 (Methanomassiliicoccales, *Candidatus* Methanofastidiosa; Nobu et al., 2016). The most abundant methanogen genera in our dataset were hydrogenotrophic *Methanoregula* and acetoclastic *Methanosaeta* (Figure 5A). As in previous studies of freshwater lakes and wetlands, these two genera were frequently identified together. While their relative dominance has been suggested to depend on factors such as pH, season, and carbon availability (Kotsyurbenko et al., 2007; Sun et al., 2012; He et al., 2016), our study points to the importance of nutrient status and catchment land cover as important determinants. For example, *Methanoregula* showed the highest relative abundances in lakes with a forested catchment, while in eutrophic (phosphorus and nitrogen rich) lakes *Methanobacterium* and *Methanolinea* had their highest relative abundances (Figure

5B). In two lakes, the dominant methanogen genus was *Ca.* Methanoperedens. Members of this taxon have been shown to conduct anaerobic oxidation of CH_4 using nitrate (Raghoebarsing et al., 2006; Haroon et al., 2013; Arshad et al., 2015) or Fe(III) and Mn(IV) (Ettwig et al., 2016) as electron acceptors. Related sequences have also been linked to anaerobic oxidation of CH_4 coupled to sulfate reduction in freshwaters (Schubert et al., 2011; Timmers et al., 2015). As mentioned earlier, it can be speculated that the detected methanogens may either represent transient members originating from anoxic environments or inhabit anoxic microenvironments (Grossart et al., 2011) in the oxygenic part of the water column. Either way their presence and taxonomic composition suggests an active role in CH_4 cycling of freshwater lakes.

Ammonia-oxidizing archaea

Thaumarchaeota classes detected in the lakes included Group 1.1c, SCGC_AB-179-E04 and most prominently Nitrosphaeria including *Ca.* Nitrosotalea and *Ca.* Nitrosoarchaeum which are described as AOA. In freshwater ecosystems, especially those with high allochthonous inputs such as dystrophic lakes, increased nitrogen supply could promote the occurrence of ammonium oxidizers. Ammonia-oxidizing archaea, however, are favored at low levels of nitrate, low light and low pH (French et al., 2012; Hatzenpichler, 2012). Accordingly, Nitrosphaeria showed the highest relative numbers in lakes with low pH and TN (Figure 5B). Recent genome analysis of a *Ca*. Nitrosotalea devanaterra revealed genes encoding both a predicted high-affinity substrate acquisition system and potential pH homeostasis mechanisms which

were expressed during acidophilic growth (Lehtovirta-Morley et al., 2016). While *Ca.* Nitrosotalea were abundant in dystrophic lakes, *Ca.* Nitrosoarchaeum, previously found in low-salinity habitats worldwide (Blainey et al., 2011), were the abundant members of Nitrosophaeria in most other lake types.

Novel taxa with potentially versatile metabolic roles

In addition to canonical methanogens, we detected Bathyarchaeota and Verstraetearchaeota (*Ca.* Methanomethylicus; Vanwonterghem et al., 2016) which are hypothesized to represent methylotrophic methanogens. The class Bathyarchaeia, contributed highly to the archaeal community in lakes with agricultural land use in their catchments, as indicated by redundancy analysis (Figure 4B). The detected Bathyarchaeia were highly diverse, represented by more than 7761 ASVs that were collectively abundant across the lakes. Bathyarchaeia are among the most abundant organisms reported in marine and freshwater sediments globally (Biddle et al., 2006; Borrel et al., 2012; Lloyd et al., 2013; Fillol et al., 2015; Lazar et al., 2015, Wurzbacher et al., 2017). Verstraetearchaeia, occurring as a minor archaeal class in our dataset, appear to be capable of fermentation utilizing sugars as a carbon source and generating acetyl-CoA via the Embden–Meyerhof–Parnas (EMP) pathway and pyruvate-ferredoxin oxidoreductase (Vanwonterghem et al., 2016). Combined with the taxonomic results discussed above, our findings support the diverse functional roles of archaea in freshwater systems, beyond CH_4 and ammonium metabolism.

CONCLUSION

Monte Carlo simulations on archaeal ASV patterns suggested that freshwater archaeal communities were shaped by deterministic processes such as environmental sorting as well as biogeographic and evolutionary history, overriding stochastic processes such as ecological drift. Large among-lake variability was reflected in the cumulative rarefaction curves of archaeal ASVs that did not saturate over the lake gradient, although many rarefaction curves of archaeal ASVs in individual lakes did reach an asymptote. Furthermore, our observations of a coupling between productivity indicators and archaeal richness provide evidence for the generality of the productivity–diversity relationship across all organismal kingdoms.

We revealed a marked phylogenetic diversity of archaea in freshwater lakes, not restricted to functionally characterized groups such as canonical methanogens or AOA. As such our study expands the knowledge of archaeal diversity inhabiting freshwater environments and shaping carbon and nitrogen cycling and CH_4 emissions of lakes. Future studies should aim to estimate the mass and energy fluxes through the archaeal compartment in these aquatic environments of regional (water quality) and global (GHG emissions) significance.

Experimental procedures

Sampling

Samples of freshwater microbes were collected from 95 Swedish lakes in August 2014 as part of a national lake monitoring programe (Figure S1). In addition, 24 additional lakes were sampled in two separate campaigns during the same time period (Figure S1). Water samples for analysis of water chemistry were collected from the depth of 0.5 m with a 0.5 m long tube sampler (Ruttner-type). Water samples for analysis of microbes were sampled from the depth of the whole epilimnion (usually down to between 2 and 8 m) with a 2-m long tube sampler. The epilimnion samples were pooled in a large bucket and a subsample of 100 and 500 ml was filtered with a peristaltic pump on 142 mm Millipore (Billerica, MA, USA) polycarbonate filters with 0.2-μm pore size until the filters clogged. Filters were immediately stored at − 80 °C until further processing.

Environmental data

Geographical information, such as lake area and catchment area, was derived from the database Svenskt Vattenarkiv (Swedish Water Archive, SMHI 2012). Catchment land use/cover for each lake (e.g. % forest, % agriculture, % urban) were downloaded from the Swedish Land Cover Data database (Naturvårdsverket 2014), part of the CORINE Land Cover data (European Environment Agency 2014).

Water physicochemical data from the Swedish national lake monitoring program at the time of sampling is publicly available from the national data host (https://www.slu.se/en/departments/aquatic-sciences-assessment/data-host/). The

physicochemical variables were measured according to international (ISO) or European (EN) standards.

Large surveys, like in our case, result in datasets containing missing values for various reasons, often encoded as NaNs, blanks or any other placeholders. One way to handle this problem is to get rid of the observations that have missing data. However, such an approach will risk losing overall statistical power and data points with valuable information, as well as introducing systematic biases. In addition, different subsets of the metadata for different analysis results in different values which can lead to biased comparisons among analysis results and most importantly does not allow the application of multivariate statistical analyses. Thus, we used a strategy where we factored in the missing values (n = 425; 16% of total metadata entries) and learned the best imputation values for the missing data (see below for further description). Prior to analyses two samples were entirely removed as they contained highly incomplete metadata. In addition, 10 variables were removed because they were redundant or contained high numbers of NAs.

DNA extraction, amplification of SSU rRNA genes and Illumina MiSeq sequencing

Filters were cut into small pieces and DNA was extracted following MoBio Power Soil kit protocol. For both PCR assays (i.e. archaeal and universal) used in this study, a two-step PCR protocol was applied to minimize PCR biases, such as chimera and

heteroduplex formation (Thompson et al., 2002) and to add barcodes, Illumina handles and adapters to the amplicons of individual samples. In the archaeal protocol, we added 1 µl of DNA extract to duplicate PCR tubes containing dNTPs (0.2 mM), the archaeal primers 340f (5′- ACACTCTTTCCCTACACGACGCTCTTCCGATCT-NNNN- CCCTAYGGGGYGCASCAG -3′) and 1000r (5′-AGACGTGTGCTCTTCCGATCT-GGCCATGCACYWCYTCTC -3′) (modified from Gantner et al. (2011) with added Illumina adapters) at 0.25 µM, as well as Q5 Polymerase (0.4 unit), Q5 enhancer (4 µl) and PCR buffer (1 ×) in a final volume of 20 µl. Amplicon size was 650-750 bp. The first PCR consisted of an initial denaturation step of 98 °C for 30 s and then 30 cycles of 10 s at 98°C, 30 s at 63°C, 30 s at 72°C and a final elongation of 2 min at 72 °C.

Using universal primers 926f (5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-AC-AAACTYRAAGRAATWGRCGG-3′) and 1392r (5′-AGACGTGTGCTCTTCCGATCT-CA-GACGGGCGGTGWGTRC-3′) at 0.4 µM, we added 1 µl of DNA extract to duplicate PCR tubes containing dNTPs (0.25 mM), as well as Herculase II Fusion Polymerase (Agilent Technologies, 0.6 µl) and PCR buffer (1 ×) in a final volume of 40 µl. Amplicon size was around 500 bp. The first PCR consisted of an initial denaturation step of 95 °C for 3 min and then 30 cycles of 30 s at 95 °C, 45 s at 50°C, 90 s at 70 °C and a final elongation of 5 min at 72 °C.

In both cases, duplicate amplicons were pooled and purified using Agencourt AMPure XP purification system. Then, we did a second PCR step (using Q5

polymerase) of 15 cycles (archaea) or 12 cycles (universal) (initial denaturation 30 s at 98°C; 10 s at 98°C, 30 s at 66°C, 30 s at 72°C; with a final step of 2 min at 72 °C) using 1 µl of the previous PCR product as template. This second PCR added barcodes and complete Thruplex adapters for Illumina sequencing (forward primer 5'-AATGATACGGCGACCACCGAGATCTACAC-[i5 index]-ACACTCTTTCCCTACACGACG -3' and reverse primer 5'-CAAGCAGAAGACGGCATACGAGAT-[i7 index]-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3').

After once again purifying the samples using the Agencourt AMPure XP kit and quantification by a PicoGreen assay (Quant-iT PicoGreen, Invitrogen), 16S rRNA gene and universal SSU rRNA gene samples were pooled separately in equimolar amounts. Samples were sequenced on a MiSeq using v3 chemistry and software version 2.6.1.1 (Science for Life Laboratory, Uppsala). Final sequencing results were obtained from 103 samples (94 samples from the national monitoring program and 9 from the separate sampling campaign) using the universal primer pair and for 109 samples (88 from the national program and 21 from the separate campaigns) using the archaea-specific primer pair.

Amplicon sequences produced in this study are publicly available at the NCBI-SRA under accession numbers SRR10321796-10321906 (archaea) and SRR10965090 - SRR10965180 (universal).

Amplicon data processing

After raw sequence data had been trimmed of primers with CUTADAPT (Martin 2011) and sequences without matching primers removed, they were analyzed with R package dada2 (Callahan et al., 2016) for de-replicating, denoising and sequence-pair assembly. After manual inspection of quality score plots, the forward and reverse reads of the universal amplicons were trimmed at 280 and 260 bp length, respectively and the archaeal forward and reverse reads at 230 and 180 bp length, respectively. Additional quality filtering removed any sequences with unassigned base pairs and reads with a single phred score below 10. The universal SSU rRNA gene amplicons were assembled by merging the read pairs. The archaeal 16S rRNA gene amplicon reads, for which the read pairs did not overlap, were concatenated. After reads were dereplicated, forward and reverse error models were created in dada2 with a subset of the sequences (10^7 reads). Chimeras were removed using 'removeBiomeraDenova' in the dada2 package, which resulted in the final taxon table.

Phylogenetic and taxonomic analysis

MAFFT (Katoh and Standley, 2013) version 7.305 was used for sequence alignment. FastTree 2 (Price et al., 2010) was used for generating a phylogenetic tree of the aligned sequences. The function assignTaxonomy of the DADA2 package was used to assign taxonomy using version 132 of the Silva database (Quast et al., 2013). Chloroplast and mitochondrial reads were removed from the universal data set and non-archaeal reads from the archaeal data set. Taxonomic coverage of the archaeal

primers was tested *in silico* with Silva TestPrime (Klindworth et al., 2013; https://www.arb-silva.de/search/testprime/) allowing 1 mismatch and no mismatches in the 3' end.

Statistical analysis

All statistical analyses and plotting were performed in R version 3.5.2 (R Core Team, 2016) using multiple R packages with R code deposited to https://gitlab.com/eiler_lab/scandinavian-archaea-diversity. Missing values in the metadata were approximated using multiple imputation with Fully Conditional Specification (FCS) implemented by the MICE algorithm as described in Van Buuren and Groothuis-Oudshoorn (2011). This approach uses variables as predictors that (i) appear in the complete data model, (ii) are related to the none-response, and (iii) explain a considerable amount of the variance while (iv) removing variables that have too many missing values within the subgroup of incomplete cases. If variables autocorrelated, a subset of the variables was retained for downstream statistical analysis. Partial least squares regression models were used to infer environmental variables that could predict the contribution of archaeal rRNA gene reads to the total read abundance in the dataset obtained with universal rRNA gene primers. Prior to alpha and beta diversity analyses, ASV data was rarefied to the smallest sample size (4649 reads) using the function rrarefy in the vegan package (Oksanen et al. 2019). Alpha diversity measures were obtained using EstimateR on rarefied community data (i.e. ACE, Chao1, Shannon, Fisher and Simpson index) and picante (phylogenetic

diversity; PD). Beta diversity matrices were calculated using UniFrac and Bray-Curtis distances on the rarefied and Hellinger transformed community data. We used distance-based redundancy analysis (dbRDA) with the function capscale in vegan to identify environmental variables co-varying with archaeal beta diversity.

Model significance of dbRDA was tested by permutational analysis (function anova.cca) for the overall model and for significance of individual factors (marginal effects). Redundancy analysis on the relative proportions of archaeal classes and methanogen genera was carried out with the function rda to identify environmental variables co-varying with the classes and genera. The importance of random vs. deterministic factors was assessed with EcoSim (Gotelli et al., 2015). EcoSim performs Monte Carlo randomizations to create "pseudo-communities" (Pianka 1986), then statistically compares the patterns in these randomized communities with those in the real data matrix.

Contributions

The research was conceptualized by HJ and AE. Sample collection was coordinated by SD while molecular and data analyses as well as writing was coordinated by AE. Molecular analyses were performed by HJ and CW. The main analyses and visualization of the data was performed by HJ and AE, with additional analyses performed by LF and CW. The first version of the manuscript was drafted by AE with help from HJ. All authors provided comments and were involved in writing the final

version of the manuscript. Financial support for the project was acquired by SD, SP and AE.

The authors confirm no conflict of interest.

Acknowledgements

We thank the Trend Lake national monitoring program in Sweden for taking extra samples for us during the regular sampling and Pilar López Hernández for performing DNA extraction. We also want to thank Richard K Johnson for critically reviewing the language of the manuscript. This research was funded by the Carl Tryggers Foundation (grant CTS:13-113 to AE), Academy of Finland (grant 265902 to SP), internal funds by the University of Oslo to AE, the Swedish Research Council VR (grant 2012-4592 to AE), and internal funds by the Swedish University of Agricultural Sciences to SD.

The authors have no conflict of interest.

References

Angle, J.C., Morin, T.H., Solden, L.M., Narrowe, A.B., Smith, G.J., Borton, M.A., et al. (2017) Methanogenesis in oxygenated soils is a substantial fraction of wetland methane emissions. Nat Comm 8:1567.

Arshad, A., Speth, D.R., de Graaf, R.M., Op den Camp, H.J.M., Jetten, M.S.M., and Welte, C.U. (2015) A Metagenomics-based metabolic model of nitrate-dependent anaerobic oxidation of methane by Methanoperedens-like archaea. Front Microbiol 6: 1423.

Auguet, J. and Casamayor, E. O. (2008) A hotspot for cold crenarchaeota in the neuston of high mountain lakes. Environ Microbiol 10: 1080-1086.

Auguet, J.C., Nomokonova, N., Camarero, L., Casamayor, E.O. (2011) Seasonal changes of freshwater ammonia-oxidizing archaeal assemblages and nitrogen species in oligotrophic alpine lakes. Appl Environ Microbiol 77: 1937-1945.

Auguet, J.C., and Casamayor, E.O. (2013) Partitioning of Thaumarchaeota populations along environmental gradients in high mountain lakes. FEMS Microbiol Ecol 84: 154–164.

Bååth, E., Kritzberg, E. (2015) pH tolerance in freshwater bacterioplankton: trait variation of the community as measured by leucine incorporation. Appl Environ Microbiol 81: 7411-7419

Bahram, M., Hillebrand, F., Forslund, S.K., Anderson, J.L., Soudzilovskaia, N.A., Nodehom, P.M., et al. (2018) Structure and function of the global topsoil microbiome. Nature 560: 233-237.

Bastviken, D., Cole, J., Pace, M., and Tranvik, L. (2004) Methane emissions from lakes: Dependence of lake characteristics, two regional assessments, and a global estimate, Global Biogeochem Cycles 18, GB4009

Bastviken, D. Tranvik, L.J., Downing, J.A., Crill, P.M., and Enrich-Prast, A. (2011) Freshwater methane emissions offset the continental carbon sink. Science 331: 6013.

Battin, T.J., Luyssaert, S., Kaplan, L.A., Aufdenkampe, A.K., Richter, A., and Tranvik, LJ. (2009) The boundless carbon cycle. Nat Geosci 22: 598–600.

Beisner, B. E., Peres-Neto, P. R., Lindström, E. S., Barnett, A., and Longhi, M. L. (2006) The role of environmental and spatial processes in structuring lake communities from bacteria to fish. Ecology 87: 2985-2991.

Bell, T. (2010) Experimental tests of the bacterial distance–decay relationship. ISME J 4: 1357-1365.

Berdjeb, L., Pollet, T., Chardon, C., and Jacquet S. (2013) Spatio-temporal changes in the structure of archaeal communities in two deep freshwater lakes. FEMS Microbiol Ecol 86: 215–230.

Biddle, J.F., Lipp, J.S., Lever, M. A., Lloyd, K.G., Sørensen, K.B., Anderson, R., et al. (2006) Heterotrophic Archaea dominate sedimentary subsurface ecosystems off Peru. Proc Natl Acad Sci USA 103: 3846–3851.

Bižić, M, Klintzsch, T., Ionescu, D., Hindiyeh, M., Günthel, M., Muro-Pastor, A.M., et al. (2020) Aquatic and terrestrial cyanobacteria produce methane. Science Advances 6: eaax5343.

Blainey, P.C., Mosier, A.C., Pntanina, A., Francis, C.A., and Quake, S.R. (2011) Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analysis. PLoS ONE 6: e16626.

Bogard, M.J., del Giorgio, P.A., Boutet, L., Chaves, M.C.G., Prairie, Y.T., Merante, A., Derry, A.M. (2014) Oxic water column methanogenesis as a major component of aquatic CH_4 fluxes. Nat Comm 5:5350.

Borrel, G., Jézéequel, D., Biderre-Petit, C., Morel-Desrosiers, N., Morel, J.P., Peyret, P., et al. (2011) Production and consumption of methane in freshwater lake ecosystems. Res Microbiol 162: 832–847.

Bridgham, S.D., Cadillo-Quiroz, H., Keller, J.K., and Zhuang, Q. (2013) Methane emissions from wetlands: biogeochemical, microbial, and modeling perspectives from local to global scales. Glob Change Biol 19: 1325–1346.

Cai, Y., Zheng, Y., Bodelier, P.L.E., Conrad, R., Jia, Z., Conrad, R., et al. (2016) Conventional methanotrophs are responsible for atmospheric methane oxidation in paddy soils. Nat Commun 7: 11728.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johanson, A.J.A, and Holmes, S.P. (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nat Meth 13: 581-583.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6: 1621-1624.

Cardinale, B.J., Hillebrand, H., Harpole, W.S., Gross, K., and Ptacnik, R. (2009) Separating the influence of resource 'availability' from resource 'imbalance' on productivity–diversity relationships. Ecol Lett 12: 475-487.

Caruso, T., Chan, Y., Lacap, D.C., Lau, M.C.Y., McKay, C.P., and Pointing, S.B. (2011) Stochastic and deterministic processes interact in the assembly of desert microbial communities on a global scale. ISME J 5: 1406–1413.

Castelle, C.J., Wrighton, K.C., Thomas, B.C., Hug, L.A., Brown, C.T., Wilkins, M.J., et al. (2015) Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. Curr Biol 25: 690–701.

Cottenie, K. (2005) Integrating environmental and spatial processes in ecological community dynamics. Ecol Lett 8: 1175-1182.

Cracraft, J. (1988) Deep-history biogeography: retrieving the historical pattern of evolving continental biotas. Syst Zool 37: 221– 236.

Debroas, D., Domaizon, I., Humbert, J.F., Jardillier, L., Lepère, C., Oudart, A., and Taïb, N. (2017) Overview of freshwater microbial eukaryotes diversity: a first analysis of publicly available metabarcoding data. FEMS Microb Ecol 93: fix023

de Vargas, C, Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., et al. (2015) Eukaryotic plankton diversity in the sunlit ocean. Science 348: 1261605 doi:10.1126/science.1261605

Diamond, J.M. 1975. Assembly of species communities. In Ecology and evolution of communities. M.L. Cody, and J.M. Diamond (eds). Cambridge, Massachusetts, USA: Harvard University Press, pp. 342–444.

Duffy, J.E., Godwin, C.M., and Cardinale, B.J. (2017) Biodiversity effects in the wild are common and as strong as key drivers of productivity. Nature 549: 261–264.

Earl, J., Hall, G., Pickup, R. W., Ritchie, D. A., and Edwards, C. (2003) Analysis of methanogen diversity in a hypereutrophic lake using PCR-RFLP analysis of mcr sequences. Microb Ecol 46: 270-8.

Eiler, A., Zaremba-Niedzwiedzka, K., Andersson, S.G.E., Martinez-Garcia, M., McMahon, K.D., Stepanauskas, R., and Bertilsson, S. (2014) Productivity and salinity structuring of the microplankton revealed by comparative freshwater metagenomics. Environ Microbiol 18: 2682–2698.

Eiler, A., Hayakawa, D.H., and Rappé, M. S. (2011). Non-random assembly of bacterioplankton communities in the subtropical North Pacific Ocean. Front Microbiol 2: 140.

Ettwig, K.F., Zhu, B., Speth, D., Keltjens, J.T., Jetten, M.S.M., and Kartal, B. (2016) Archaea catalyze iron-dependent anaerobic oxidation of methane. Proc Natl Acad Sci 113: 12792–12796.

European Environment Agency. (2014) Corine land cover 2006 raster data. European Environment Agency, Copenhagen, Denmark. Available from: http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster-3

Evans, P.N., Parks, D.H., Chadwick, G.L., Robbins, S.J., Orphan, V.J., Golding, S.D, and Tyson, G.W. (2015) Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. Science 350: 434-438.

Farjalla, V. F., Srivastava, D. S., Marino, N. A., Azevedo, F. D., Dib, V., Lopes, P. M., Rosado, A.S, Bozelli, R.L., and Esteves, F. A. (2012) Ecological determinism increases with organism size. Ecology, 93: 1752-1759.

Fillol, M., Auguet, J.C., Casamayor, E.O., and Borrego, C.M. (2015) Insights in the ecology and evolutionary history of the Miscellaneous Crenarchaeotic Group lineage. ISME J 10: 665–677.

French, E., Kozlowski, J.A., Mukherjee, M., Bullerjahn, G., Bollmann, A. (2012) Ecophysiological characterization of ammonia-oxidizing archaea and bacteria from freshwater. Appl Environ Microbiol 78: 5773-5780.

Fuhrman, J.A., Steele, J.A., Hewson, I, Schwalbach M.S., Brown, M.V., Green, J.L., and Brown, J.H. (2008) A latitudinal diversity gradient in planktonic marine bacteria. Proc Natl Acad Sci USA 105: 7774-7778.

Gantner, S., Andersson, A.F., Alonso-Sáez, L., and Bertilsson, S. (2011) Novel primers for 16S rRNA-based archaeal community analyses in environmental samples. J Microbiol Methods 84: 12-18.

Gilbert, B., and Levine, J.M. (2017) Ecological drift and the distribution of species diversity. Proc. R. Soc. B 284: 20170507.

Glöckner, F. O., Fuchs, B. M., and Amann, R. (1999) Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence in situ hybridization. Appl Environ Microbiol 65: 3721-3726.

Gotelli, N.J., Hart, E.M., and Ellison, A.M. (2015) EcoSimR: Null model analysis for ecological data. R package version 0.1.0. http://github.com/gotellilab/EcoSimR doi:10.5281/zenodo.16522

Gotelli, N.J., and McGill, B.J. (2006) Null versus neutral Mmodels: what's the difference? Ecography 29: 793-800.

Gotelli, N.J., and McCabe, D. J. (2002) Species co-occurrence: a meta- analysis of J.M. Diamond's assembly rules model. Ecology 83: 2091-2096.

Grossart, H.P., Frindte, K., Dziallas, D., Eckert, W., and Tang, K.W. (2011) Microbial methane production in oxygenated water column of an oligotrophic lake. Proc Natl Acad Sci USA 108: 19657-19661.

Haroon, M.F., Hu, S., Shi, Y., Imelfort, M., Keller, J., Hugenholtz, P., Yuan, Z., and Tyson, G.W. (2013) Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. Nature 500: 567-570.

Hatzenpichler, R. (2012) Diversity, physiology, and niche differentiation of ammonia-oxidizing archaea. Appl Environ Microbiol 78: 7501-7510.

He, Y., Li, M., Perumal, V., Feng, X., Fang, J., Xie, J., et al. (2016) Genomic and enzymatic evidence for acetogenesis among multiple lineages of the archaeal phylum Bathyarchaeota widespread in marine sediments. Nat Microbiol 1: 16035.

Horner-Devine, M.C., Silver, J.M., Leibold, M.A., Bohannan, B.J.M., Colwell, R.K., Fuhrman, J.A., et al. (2007) A comparison of taxon co-occurrence patterns for macro-and microorganisms. Ecology 88: 1345 -1353.

Hu, A., Yao, T., Jiao, N., Liu, Y., Yang, Z. and Liu, X. (2010) Community structures of ammonia-oxidising archaea and bacteria in high- altitude lakes on the Tibetan Plateau. Freshw Biol 55: 2375-2390.

Hugoni, M., Etien, S., Bourges, A., Lepère, C., Domaizon, I., Mallet, C., et al. (2013) Dynamics of ammonia-oxidizing Archaea and Bacteria in contrasted freshwater ecosystems. Res Microbiol 164: 360-370.

Karst, S.M., Dueholm, M.S., McIlroy, S.J., Kirkegaard, R.H., Nielsen, P.H., and Albertsen, M. (2018) Retrieval of a million high-quality, full-length microbial 16S and 18S rRNA gene sequences without primer bias. Nat Biotechnol 36: 190-195.

Katoh, K., and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30: 772-780.

Keough, B.P., Schmidt, T.M., and Hicks, R.E. (2003) Archaeal nucleic acids in picoplankton from great lakes on three continents. Microbial Ecol 46: 238-248.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., and Glöckner, F.O. (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res 41: e1.

Kotsyurbenko, O.R., Friedrich, M.W., Simankova, M.V., Nozhevnikova, A.N., Golyshin, P.N., Timmis, K.N., and Conrad, R. (2007) shift from acetoclastic to H2-dependent methanogenesis in a west Siberian peat bog at low pH values and isolation of an acidophilic Methanobacterium strain. Appl Environ Microbiol 73: 2344–2348.

Lazar, C.S., Baker, B.J., Seitz, K.W., and Teske, A.P. (2017) Genomic reconstruction of multiple lineages of uncultured benthic archaea suggests distinct biogeochemical roles and ecological niches. ISME J 11: 1118-1129.

Lazar, C.S., Biddle, J.F., Meador, T.B., Blair, N., Hinrichs, K.U., and Teske, A.P. (2015) Environmental controls on intragroup diversity of the uncultured benthic archaea of the miscellaneous Crenarchaeotal group lineage naturally enriched in

anoxic sediments of the White Oak River estuary (North Carolina, USA). Environ Microbiol 17: 2228–2238.

Lehtovirta-Morley, L.E., Sayavedra-Soto, L.A., Gallois, N., Schouten, S., Stein, L.Y., Prosser, J.I., and Nicol, G.W. (2016) Identifying potential mechanisms enabling acidophily in the ammonia-oxidizing archaeon "*Candidatus* Nitrosotalea devanaterra." Appl Environ Microbiol 82: 2608 –2619.

Lennon, J.T., and Jones, S.E. (2011) Microbial seed banks: the ecological and evolutionary implications of dormancy. Nat Rev Microbiol 9: 119-130.

Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., et al. (2015) Determinants of community structure in the global plankton interactome. Science 348: 1262073

Lindström, E.S. and Langenheder, S. (2011) Local and regional factors influencing bacterial community assembly. Env Microbiol Rep 4: 1-9.

Linz, A.M., He, S., Stevens, S.L.R., Anantharaman, K., Rohwer, R.R., Malmstrom, R.R., et al. (2018) Connections between freshwater carbon and nutrient cycles revealed through reconstructed population genomes. bioRxiv doi: https://doi.org/10.1101/365627

Llirós, M., Gich, F., Plasencia, A., Auguet, J.C.C., Darchambeau, F., Casamayor, E.O., et al. (2010) Vertical distribution of ammonia-oxidizing crenarchaeota and methanogens in the epipelagic waters of lake Kivu (Rwanda-Democratic Republic of the Congo). Appl Environ Microbiol 76: 6853–6863.

Lloyd, K.G., Schreiber, L., Petersen, D.G., Kjeldsen, K.U., Lever, M.A., Steen, A.D., et al. (2013) Predominant archaea in marine sediments degrade detrital proteins. Nature 496: 215–218.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17:200 DOI: https://doi.org/10.14806/ej.17.1.200

Mukherjee, M., Ray, A., Post, A.F., McKay, R.M., and Bullerjahn, G.S. (2016) Identification, enumeration and diversity of nitrifying planktonic archaea and bacteria in trophic end members of the Laurentian Great Lakes. J Great Lakes Res 42: 39-49.

Narrowe, A.B., Angle, J.C., Daly, R.A., Stefanik, K.C., Wrighton, K.C., and Miller C.S. (2017) High-resolution sequencing reveals unexplored archaeal diversity in freshwater wetland soils. Environ Microbiol 19: 2191-2209.

Newton, R.J., Jones, S.E., Eiler, A., McMahon, K.D., and Bertilsson, S. (2011) A guide to the natural history of freshwater lake bacteria. Microbiol Mol Biol Rev 75: 14-49.

Nobu, M.K., Narihiro, T., Kuroda, K., Mei, R., and Liu, W.T. (2016) Chasing the elusive Euryarchaeota class WSA2: genomes reveal a uniquely fastidious methyl-reducing methanogen. ISME J 10: 2478–2487.

Ofiteru, I.D., Lunn, M., Curtis, T.P., Wells, G.F., Criddle, C.S., Francis, C.A., and Sloan, W.T. (2010) Combined niche and neutral effects in a microbial wastewater treatment community. Proc Natl Acad Sci USA 107: 15345–15350.

Ortiz-Alvarez, R., and Casamayor, E.O. (2016) High occurrence of Pacearchaeota and Woesearchaeota (Archaea superphylum DPANN) in the surface waters of oligotrophic high-altitude lakes. Environ Microbiol Rep 8: 210-217.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2019) vegan: Community Ecology Package. R package version 2.5-6. https://CRAN.R-project.org/package=vegan

Pernthaler, J., Glöckner, F. O., Unterholzner, S., Alfreider, A., Psenner, R., and Amann, R. (1998) Seasonal community and population dynamics of pelagic bacteria and archaea in a high mountain lake. Appl Environ Microbiol 64: 4299-4306.

Pianka, E.R. (1986) Ecology and Natural History of Desert Lizards. Analyses of the Ecological Niche and Community Structure. Princeton Legacy Library, USA.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE 5: e9490.

Pouliot, J. , Galand, P. E., Lovejoy, C. and Vincent, W. F. (2009) Vertical structure of archaeal communities and the distribution of ammonia monooxygenase A gene variants in two meromictic High Arctic lakes. Environ Microbiol 11: 687-699.

Prowse, T.D., and Stephenson, R.L. (1986) The relationship between winter lake cover, radiation receipts and the oxygen deficit in temperate lakes. Atmos Ocean 24: 386–403.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucl Acid Res 41: D590-D596.

Raghoebarsing, A.A., Pol, A., van de Pas-Schoonen, K.T., Smolders, A.J.P., Ettwig, K.F., Rijpstra, W.I.C., et al. (2006) A microbial consortium couples anaerobic methane oxidation to denitrification. Nature 440: 918–921.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Reed, D.C., Algar, C.K., Huber, J.A. and Dick G.J. (2014) Gene-centric approach to integrating environmental genomics and biogeochemical models. Proc Natl Acad Sci USA 111: 1879-1884.

Roulet, N., and Moore, T.R. (2006) Browning the waters. Nature 444: 283-284.

Rouse, W.R., Douglas, M.V., Hecky, R.E., Heshey, A.E., Kling, G.W., Lesack, L., et al. (1997) Effects of climate change on the freshwaters of Arctic and subarctic North America. Hydrol Process 11: 873–902.

Ruiz-González, C., Niño-Garcia, J.P., Kembel, S.W., and del Giorgio P.A. (2017) Identifying the core seed bank of a complex boreal bacterial metacommunity. ISME J 11: 2012-2021.

Saarenheimo, J., Rissanen, A.J., Arvola, L., Nykänen, H., Lehmann, M.F., Tiirola, M. (2015) Genetic and environmental controls on nitrous oxide accumulation in lakes. PLoS ONE 10(3): e0121201.

Savio, D., Sinclair, L., Iljaz, U.Z., Blaschke, A.P., Reischer, G.H., Blöschl, G., et al. (2015) Bacterial diversity along a 2600 km river continuum. Environ Microbiol 12: 4994-5007.

Schneider, P., and Hook, S.J. (2010) Space observations of inland water bodies show rapid surface warming since 1985. Geophys Res Lett 37: L22405.

Schubert, C.J., Vazquez, F., Lösekann-Behrens, T., Knittel, K., Tonolla, M., and Boetius, A. (2011) Evidence for anaerobic oxidation of methane in sediments of a freshwater system (Lago di Cadagno). FEMS Microbiol Ecol 76: 26–38.

Seitz, K.W., Dombrowski, N., Eme, L., Spang, A., Lombard, J., Sieber, J.R., et al. (2019) Asgard archaea capable of anaerobic hydrocarbon cycling. Nat Comm 10: 1822.

Segarra, K.E.A., Schubotz, F., Samarkin, V., Yoshinaga, M.Y., Hinrichs, K.U., and Joye, S.B. (2015) High rates of anaerobic methane oxidation in freshwater wetlands reduce potential atmospheric methane emissions. Nat Commun 6: 7477.

Shurin, J. B., Cottenie, K., and Hillebrand, H. (2009) Spatial autocorrelation and dispersal limitation in freshwater organisms. Oecologia 159: 151-159.

SMHI (Swedish Meteorological and Hydrological Institute). (2012) Svenskt Vattenarkiv 2012_2. Swedish Meteorological and Hydrological Institute, Norrköping, Sweden. (Available from: http://www.smhi.se/klimatdata/hydrologi/sjoar-och-vattendrag/ladda-ner-data-fran-svenskt-vattenarkiv-1.20127

Spang, A., Hatzenpichler, R., Brochier-Armanet, C., Rattei, T., Tischler, P., Spieck, E., et al. (2010) Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum Thaumarchaeota.Trends Microbiol 18: 331-340.

Spang, A., Stairs, C.W., Dombrowski, N., Eme, L., Lombard, J., Caceres, E.F., et al. (2019) Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. Nat Microbiol 4: 1138–1148 doi: 10.1038/s41564-019-0406-9.

Sun, C.L., Brauer, S.L., Cadillo-Quiroz, H., Zinder, S.H., and Yavitt, J.B. (2012) Seasonal changes in methanogenesis and methanogenic community in three peatlands, New York state. Front Microbiol 3: 81.

Naturvårdsvärket (Swedish EPA) (2014) Svenska Marktäckedata. Available from: http://gpt.vicmetria.nu/data/land/SMD_produktbeskrivning_20140627.pdf

Tan, Z., and Zhuang, Q. (2015) Arctic lakes are continuous methane sources to the atmosphere under warming conditions. Environ Res Lett 10: 1–9.

Thompson, J.R., Marcelino, L.A., and Polz, M.F. (2002) Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. Nucl Acid Res 30: 2083-2088.

Thompson, L.R., Sanders, J.G., McDonald, D., Amir, A., Ladau, J., Locey, K.J., et al. (2017) A communal catalogue reveals Earth's multiscale microbial diversity. Nature 551: 457-463.

Thornton, B.F., Wik, M., and Crill, P.M. (2015) Climate-forced changes in available energy and methane bubbling from subarctic lakes. Geophys Res Lett 42: 1936–1942.

Timmers, P.H., Suarez-Zuluaga, D.A., van Rossem, M., Diender, M., Stams, A.J., and Plugge, C.M. (2015) Anaerobic oxidation of methane associated with sulfate reduction in a natural freshwater gas source. ISME J 10: 1400–1412.

Tranvik, L., Downing, J., Cotner, J., Loiselle, S., et al. (2009) Lakes and reservoirs as regulators of carbon cycling and climate. Limnol Oceanogr 54: 2298–2314.

Van Buren, S., and Groothuis-Oudshoorn, K. (2011) mice: Multivariate Imputation by Chained Equations in R. J Stat Software doi 10.18637/jss.v045.i03

Vanwonterghem, I., Evans, P.N., Parks, D.H., Jensen, P.D., Woodcroft, B.J., Hugenholtz, P., and Tyson, G.W. (2016) Methylotrophic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. Nat Microbiol 1: 16170.

Vanwonterghem, I., Jensen, P., Dennis, P., Hugenholtz, P., Rabaey, K., Tyson, G.W. (2014) Deterministic processes guide long-term synchronised population dynamics in replicate anaerobic digesters. ISME J 8: 2015–2028.

Verpoorter, C., Kutser, T., Seekell, D.A., and Tranvik, L.J. (2014) A global inventory of lakes based on high-resolution satellite imagery. Geophys Res Lett 41: 6396 - 6402.

Vuilleumier, F., and Simberloff, D. (1980) Ecology versus history as determinants of patchy and insular distribution in high andean birds. Evol Biol 12: 235– 379.

Wang, Y., Sheng, H.F., He, Y., Wu, J.Y., Jiang, Y.X., Tam, N.F.Y., and Zhou, H.W. (2012) Comparison of the levels of bacterial diversity in freshwater, intertidal wetland, and marine sediments by using millions of illumina tags. Appl Environ Microbiol 78: 8264–8271.

Weyhenmeyer, G.A., Müller, R.A., Norman, M., and Tranvik, L.J. (2016) Sensitivity of freshwaters to browning in response to future climate change. Climate Change 134: 225-239.

Wik, M., Varner, R.K., Anthony, K.W., MacIntyre, S., and Bastviken, D. (2016) Climate-sensitive northern lakes and ponds are critical components of methane release. Nat Geosci 9: 99-105.

Wurzbacher, C., Fuchs, A., Attermeyer, K., Frindte, K., Grossart, H.P., Hupfer, M., Casper, P., and Monaghan, M.T. (2017) Shifts among Eukaryota, Bacteria, and Archaea define the vertical organization of a lake sediment. Microbiome 5: 41.

Figures and Tables

Figure 1. Correlation matrix between alpha diversity measures of archaea and environmental variables of the study lakes. The size of the symbol is inversely related to the p value (correlations with p values > 0.05 are not shown), while color coding indicates the correlation coefficient (r>0 in blue, r<0 in red). Abbreviations: nitrite and nitrate nitrogen (NO_2+NO_3), silicate (Si), ammonium nitrogen (NH_4), total phosphorous (TP), total organic carbon (TOC), total nitrogen (TN), phosphate phosphorus (PO_4-P); specific ultraviolet absorbance (SUVA); diversity measures: Simpson diversity index (Simpson, InvSimpson), Shannon index (Shannon), phylogenetic diversity (PD), species richness (SR), Chao 1 richness estimator (Chao1), -abundance based coverage estimator (ACE); oxygen concentration (O_2); catchment characteristics: percentage of Water, Agriculture, and Wetland.

Figure 2. Plot representing the results from a partial least squares regression model that predicts archaeal contribution (in orange) to the microbial community from environmental data (in blue). Environmental variable arrows with the smallest angle (i.e. closest) to the archaeal arrow have the highest correlation with archaeal contribution. The model is based on data from 103 Swedish lakes.

Figure 3. Sample (lake) specific (A) and region wide (B, Sweden) accumulation curves of archaeal sequence variants (ASVs).

Figure 4. Taxonomic composition of the archaeal classes (following SILVA 132 taxonomy) (A) and their number of reads (relative abundance in %) in relation to environmental parameters as inferred by redundancy analysis (B) with the 15 most abundant classes represented. In lake system names, 'slu' refers to the lakes of the national sampling campaign and 'jam' and 'asa' to separate sampling campaigns.

Figure 5. Taxonomic composition of the archaeal methanogenic genera (following SILVA 132 taxonomy) (A) and their number of reads (relative abundance in %) in relation to environmental variables as inferred by redundancy analysis (B) with the 15 most abundant genera represented. In lake system names, 'slu' refers to the lakes of the national sampling campaign and 'jam' and 'asa' to separate sampling campaigns.

Table 1. Environmental variables co-varying with archaeal beta diversity in the study lakes based on Bray-Curtis and Unifrac distance measures as provided by distance-based redundancy analysis (dbRDA). (NS = not significant)

Figure S1. Map of the sampled Swedish lakes.

Figure S2. Proportion of domain-specific rRNA gene reads resulting from amplicon sequencing with universal primers.

Figure S3. The contribution of archaeal amplicon sequence variants (ASVs) unique to individual study lakes to the overall community of each lake as well as the contribution of archaeal ASVs occurring in two to three lakes and more to each lake community (A). The prevalence of archaeal ASVs cumulation curve (B) and (C).

Table S1. Summary statistics of metadata including 119 sampled lakes.

Table 1. Environmental variables co-varying with archaeal beta diversity in freshwater lakes based on Bray-Curtis and Unifrac distance measures as provided by dbRDA. (NS = not significant)

Variable	Bray-Curtis pseudo-F	Bray-Curtis P	Unifrac pseudo-F	Unifrac P
CaKMgNa	2.01	0.005 **	2.31	0.005 **
Alkalinity	1.98	0.005 **	2.29	0.005 **
Conductivity	1.96	0.005 **	2.27	0.005 **
pH	1.68	0.005 **	1.87	0.005 **
Total nitrogen	1.66	0.005 **	1.88	0.005 **
Turbidity	1.62	0.005 **	1.96	0.005 **
Total phosphorus	1.58	0.005 **	1.69	0.005 **
Total organic carbon	1.45	0.005 **	1.53	0.010 **
Aluminium	1.42	0.005 **	1.28	0.010 **
Chlorophyll	1.42	0.005 **	1.53	0.005 **
Secchi depth	1.40	0.005 **	1.59	0.005 **
Lake area	1.28	0.010 **	1.50	0.010 **
Catchment-to-Lake-ratio	1.21	0.015 *	-	NS
Latitude	-	NS	1.24	0.045 *