

ORIGINAL ARTICLE

Spatial structure and persistence of methanogen populations in humic bog lakes

Kim Milferstedt¹, Nicholas D Youngblut and Rachel J Whitaker

Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Patterns of diversity within methanogenic archaea in humic bog lakes are quantified over time and space to determine the roles that spatial isolation and seasonal mixing play in structuring microbial populations. The protein encoding gene *mcrA* is used as a molecular marker for the detection of fine-scale differences between methanogens in four dimictic bog lakes in which the water column is mixed twice a year and one meromictic lake that is permanently stratified. Although similar sequences are observed in each bog lake, each lake has its own characteristic set of persisting sequence types, indicating that methanogen populations are delimited either by low migration between the anaerobic hypolimnia or by lake-specific selection. The meromictic lake is differentiated from all other lakes and contains sequences with a higher degree of microdiversity than the dimictic lakes. By relating the structure of diversity to the depth of each bog lake, we propose the hypothesis that the deeper parts of the water column favor microdiversification of methanogens, whereas the periodically disturbed water column of shallower dimictic lakes promote genetically more diverse methanogen communities.

The ISME Journal (2010) 4, 764–776; doi:10.1038/ismej.2010.7; published online 25 February 2010

Subject Category: microbial population and community ecology

Keywords: methanogens; *mcrA*; disturbance; humic bog lakes; population structure; biogeography

Introduction

Fine-scale variations of conserved 16S rRNA sequences are commonly observed in microbial populations (Rocap *et al.*, 2003; Klepac-Ceraj *et al.*, 2004; Hunt *et al.*, 2008). However, the ecological relevance and factors promoting diversification are not well understood. Within these closely related groups of microorganisms, an even greater level of diversity has been observed when analyzing variation among sequences of protein-encoding genes (for example, Whitaker *et al.*, 2003; Hunt *et al.*, 2008). This level of sequence variation is considered the population-scale diversity (within a species) and is called microdiversity. The amount of microdiversity within a population is driven by population genetic forces such as mutation, recombination, migration, selection and genetic drift. Investigating the spatial and temporal distribution of microdiversity within populations is necessary to elucidate its ecological importance, mechanisms of microbial speciation,

and diversification patterns (Achtman and Wagner, 2008; Fraser *et al.*, 2009).

Small freshwater lakes are an ideal system to investigate microbial population dynamics because the spatial boundaries are easily identified and seasonal ecosystem dynamics have been well characterized. At the community level, bacteria inhabiting freshwater lakes are highly dynamic (Eiler and Bertilsson, 2007; Shade *et al.*, 2007; Nelson, 2009) and diverse (Zwart *et al.*, 2002; Shade *et al.*, 2007). Differences in community composition are driven in time by seasonal fluctuations in extrinsic and intrinsic variables (Eiler and Bertilsson, 2004; Shade *et al.*, 2007), and in space by regional factors and landscape position (Yannarell and Triplett, 2005). Recently it has been shown that the bacterial community composition in the two thermal layers of the water column in a stratified lake (epilimnion and hypolimnion) experiences different levels of variation and dynamics during one summer season (Shade *et al.*, 2008). The epilimnion is the upper, warmer water layer during summer stratification. The availability of light in this layer initiated primary production and interactions between phytoplankton and the bacterial community (Kent *et al.*, 2004). In contrast to the epilimnion, the bacterial community in the cooler and deeper water layer, the hypolimnion, was largely unaffected by extrinsic drivers (Shade *et al.*, 2008). The apparent stability of the hypolimnion community during the summer

Correspondence: RJ Whitaker, Department of Microbiology, University of Illinois at Urbana-Champaign, 601 S. Goodwin Avenue, Urbana, IL 61801, USA.

E-mail: rwhitaker@life.illinois.edu

¹Current address: INRA Laboratoire de Biotechnologie de l'Environnement, Avenue des Etangs, F-11100 Narbonne, France. E-mail: milferst@supagro.inra.fr

Received 25 September 2009; revised 17 December 2009; accepted 11 January 2010; published online 25 February 2010

months may result from the absence of a strong driver for successional change.

Over the course of the seasons, turnover of the water column can induce variation in the microbial communities of freshwater lakes. The water column in dimictic lakes mixes in spring and fall as a consequence of differential warming and cooling of the epilimnion versus the hypolimnion. Mixing events disrupt established chemical gradients in the water column and combine two otherwise isolated layers of water. Following the mixing event, the water column re-stratifies during summer and winter. Jones *et al.* (2008) showed repeated re-sets in the temporal development of the bacterial community in the epilimnion of a Taiwanese freshwater lake after a series of typhoons mixed and destratified the lake. Shade *et al.* (2008) argue that disturbance by mixing may also have an effect on hypolimnion communities. In contrast to dimictic lakes, meromictic lakes are permanently stratified and do not experience seasonal mixing. The variation in the frequency of mixing events between meromictic and dimictic lakes allows us to test the effects of the disturbance introduced by mixing on microbial populations.

In this study, we examine the spatial and temporal dynamics of closely related methanogen populations in a series of humic bog lakes in northern Wisconsin, USA. Methanogens are strictly anaerobic members of Euryarchaeota. We hypothesize that these strict anaerobes would be sensitive to mixing events that introduce oxygen traces into the anaerobic hypolimnion of bog lakes. In methanogens, the functional gene *mcrA* encodes the α -subunit of methyl-coenzyme M reductase. This methanogen-specific reductase is a key enzyme in methanogenesis (Ermler *et al.*, 1997) and useful for the non-culture-based assessments of diversity. Using the *mcrA* gene, the distributions of methanogen species have been surveyed in anaerobic digesters (Leclerc *et al.*, 2004), peat bogs (Juottonen *et al.*, 2006; Bräuer *et al.*, 2006b), soils (Lueders *et al.*, 2001; Luton *et al.*, 2002), sediments from freshwater lakes (Banning *et al.*, 2005) and marine environments (Dhillon *et al.*, 2005). Methanogens have also been identified in the water column of freshwater lakes (Earl *et al.*, 2003; Lehours *et al.*, 2005, 2007)

where sequences of the orders Methanosarcinales and Methanomicrobiales dominated clone libraries. Galand *et al.* (2002) characterized a Methanomicrobiales-associated cluster of sequences that appears to be globally significant in nutrient-poor acidic environments like peat bogs and the humic bogs investigated here. Bräuer *et al.* (2006a) have sequenced the genome of the first isolate from this cluster, *candidatus* Methanoregula boonei 6A in which the *mcrA* gene is found in single copy.

Here we test (1) how methanogen populations differ between five humic bog lakes, (2) whether methanogen populations change across seasons of 1 year including mixing events and (3) how the population structure compares between the lakes with different mixing regimes.

Materials and methods

Description of the sampling sites

The methanogen populations in five dystrophic humic bog lakes in Vilas County in northern Wisconsin, USA, were investigated. Two sets of bog lakes separated by 30 km were studied; 'Mary Lake' (MA) and 'Rose Lake' (RL) from the Adelaide-Yolanda group of lakes (Juday and Birge, 1941) and 'South Sparkling Bog' (SSB), 'North Sparkling Bog' (NSB) and 'Trout Bog' (TB) in the Trout Lake area of the North Temperate Lakes Long Term Ecological Research Station. The five lakes represent two mixing regimes: RL, NSB, SSB and TB are dimictic (mixing twice a year); MA is meromictic (permanently stratified and never mixed). Lake characteristics are specified in Table 1. All lakes contain darkly stained, acidic water with an approximate average Secchi depth of 1 m. The geology of the area is dominated by quaternary sediments deposited during the Wisconsin glaciation (Clayton *et al.*, 2002). The immediate shore vegetation is predominantly *Sphagnum* spp. and *Vaccinium* spp. or adjacent coniferous forests.

Sample collection

Samples from different subsets of lakes were collected at four time points between October 2007

Table 1 Lake characteristics

Name of lake	Latitude	Longitude	Surface area [ha]	Max. depth [m]	Distance [km]	Mixing regime	Average pH ^a
Mary lake (MA)	46°15'2	89°54'1	1.2	21.5	31.9	m	5.6
Rose lake (RL)	46°15'8	89°54'3	1.43	13	32.1	d	6.5
North sparkling bog (NSB)	46°00'0	89°42'0	0.47	4.5	0	d	4.6
South sparkling bog (SSB)	46°00'0	89°42'0	0.44	8	0.2	d	4.3
Trout bog (TB)	46°02'31	89°41'20	1.01	7.9	4.7	d	4.4

Distance is the distance between the lakes and North Sparkling Bog as an arbitrary reference point. Dimictic lakes are indicated by the letter 'd' in the column 'Mixing regime'. 'm' denotes meromictic.

^apH measured from 1 May 2008 to 19 Aug 2008 except RL, which was measured only on 1 July 2008.

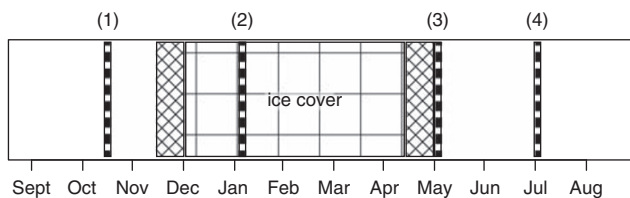


Figure 1 Timeline indicating sampling times (1–4), ice cover and approximate times for lake mixing (hatched) in 2007–2008. Lake mixing only applies to dimictic lakes. NSB mixed earlier than the other dimictic lakes in fall 2007 before sampling date (1).

and July 2008 (Figure 1). Integrated water samples were collected at the deepest point of the lakes from the epi- and hypolimnion using connected 2 m PVC pipe segments with stop valves at 1 m intervals. To determine the location of the thermocline separating the hypolimnion and epilimnion, temperature and dissolved oxygen measurements were taken at 0.5–1 m intervals throughout the water column using a YSI 550 dissolved oxygen meter (Yellow Springs, OH, USA) (Supplementary Figures S1–S5, Supplementary materials). Samples were collected in ethanol-washed Nalgene containers and immediately transported to the lab for further processing.

Particulate matter, including microbial cells in the water samples, was collected from 150–200 ml sample volume by vacuum filtration in ethanol-washed Nalgene filter holder assemblies onto 47 mm 0.2 µm Supor-200 nylon membrane filters (Pall Life Sciences, Ann Arbor, MI, USA). The filters were temporarily stored at -20°C until transported on ice to storage at -80°C .

DNA extraction

Two extractions were performed per membrane filter ($2 \times$ one half of a filter) and kept separated as replicates. Each nylon membrane was aseptically cut into pieces of an approximate size of less than 10 mm^2 per piece and subjected to 10 min bead beating on a flat bed vortex followed by DNA extraction using the column-purification MO BIO UltraClean Fecal DNA kit (Mo Bio Laboratories Inc., Carlsbad, CA, USA) according to the manufacturer's instructions.

PCR, cloning and sequencing

The primer pair 1AF and 1100 AR (Hales *et al.*, 1996) with good coverage of methanogenic archaea (Cadillo-Quiroz *et al.*, 2006) was used for 16S rRNA gene amplification in a PCR containing final concentrations of 4 mM MgCl_2 , 0.3 mM dNTPs (each), 300 nM primers and 1.25 U of GoTaq flexi polymerase (Promega Corporation, Madison, WI, USA). Initial denaturation was done for 3 min at 94°C ; melting for 1 min at 94°C , annealing at 54°C for

1.5 min, extension at 72°C for 1.5 min and a final extension at 72°C for 10 min after 29 cycles.

To amplify the *mcrA* gene, primers *mcrF* and *mcrR* (Luton *et al.*, 2002) were used in a PCR containing final concentrations of 1.5 mM MgCl_2 , 0.2 mM dNTPs, 200 nM primers and 0.6 U of Phusion High Fidelity DNA polymerase F-530 (Finnzymes, distributed by New England Biolabs Inc., Ipswich, MA, USA). Initial denaturation was done for 3 min at 98°C ; melting for 30 s at 98°C , annealing at 59°C for 40 s, extension at 72°C for 1 min and a final extension at 72°C for 10 min after 30 cycles.

Phusion PCR products were cloned into Invitrogen pCR II-Blunt TOPO vector using the Zero Blunt TOPO PCR Cloning Kit (Invitrogen, Carlsbad, CA, USA) at the maximum recommended incubation time of 30 min. Taq PCR products were cloned into an Invitrogen TOPO TA Cloning Kit (Invitrogen) according to the manufacturer's instruction. All clones were directly submitted for high-throughput sequencing to the WM Keck Center for Comparative and Functional Genomics at UIUC.

qPCR for specific *mcrA* sequence types

Three primer pairs designated 'MA high F' (5'-3' TA TACCAGCTACGGTGTGGACTAC), 'MA high R' (5'-3' AGTCGGGTACTCCTCGTACTGCT); 'SSB high F' (5'-3' TCGACGACTACACCTACTACGG), 'SSB high R' (5'-3' CATAGAGCGTGACTTTCAGTTGC); 'MA.SSB low2.0 F' (5'-3' AGCAGCATCAGGTATCTCCTGT), 'MA.SSB low2.0 R' (5'-3' ATACAGCCTTCGTCGGG TCT) were designed to target highly abundant genotypes in the May clone libraries of MA and SSB.

To test the specificity of the qPCR primer pairs, PCR were performed using as DNA templates a selection of clones from our clone libraries with a range of mismatches to the target sequences. PCR mixes with a total volume of $30\ \mu\text{l}$ contained final concentrations of $1 \times$ GoTaq PCR buffer, 2 mM MgCl_2 , 0.2 mM dNTPs, 0.2 mM of both forward and reverse primers, and 0.03 U GoTaq Flexi DNA polymerase (Promega). Cycling conditions consisted of 3 min of incubation at 94°C followed by 30 cycles of alternating temperatures of 94°C for 30 seconds, 63°C for 30 s, and 72°C for 1 min followed by a final elongation at 72°C for 10 min. Amplification was only detected for clones with less than five mismatches near the 5' end, and less than one mismatch near the 3' end of either primer. All sequence types in the clone libraries for the May samples of MA and SSB that fell within the above mismatch criteria were considered to be potentially amplified by the primer pairs when estimating sequence abundance based on clone libraries as in Table 3.

qPCRs were carried out using IQ SYBR Green Supermix (Bio-Rad, Hercules, CA, USA) in a Mastercycler ep Realplex (Eppendorf, Westbury, NY, USA). Clones from each representative sequence were used to generate DNA standards for standard curves using PureLink Quick Plasmid

Miniprep Kit (Invitrogen) and quantified using a Nanodrop ND-1000 (Thermo Scientific, Wilmington, DE, USA). Ten-fold serial dilutions from 10^9 copies to 10^3 copies of uncut plasmid were used for standard curve generation. qPCR was run with the conditions described above for 40 cycles. The subsequent melting curve generation consisted of 94 °C for 30 s followed by a 20 min stepwise increase of temperature from 63 °C to 94 °C. Each sample was run in triplicate wells. All samples were run on the same 96-well plate.

Sequence analysis

Sequences were automatically trimmed of vectors, manually checked for sequencing errors using Sequencher 4.7 (Gene Codes Corporation, Ann Arbor, MI, USA), and manually aligned. We determined the combined error rate for Phusion PCR and sequencing in our hands by re-processing a known clone. After PCR amplification of a single clone with the Phusion enzyme, the product was cloned once more and 80 clones were sequenced. We observed no PCR and sequencing error when re-sequencing 30 kbp. For the total number of 507 kbp sequenced in this study, assuming a conservatively estimated error rate of 1/30 kbp, approximately 17 erroneous bases resulting from PCR and sequencing error can be expected. We observed 25 sequences containing one unique single nucleotide polymorphisms. These individual bases were changed to an 'N' in order to avoid counting possible PCR and sequence error as diversity. Sequences containing multiple unique bases were left unchanged because the likelihood of multiple PCR and sequence errors in the same sequence is low. In addition, Bellerophon (Huber, 2004) was used to check for chimeric sequences that could form during PCR. No potential chimeras with 100% identity to either parental sequence were identified in our dataset indicating an absence of detectable PCR based chimeras. The R software environment (R development core team, 2009) with the package APE (Paradis *et al.*, 2004) was used to calculate a distance matrix for all unique sequences based on the proportion of nucleotide sites that differ (p-distance). Gaps and ambiguous positions (N) were deleted in each pairwise comparison.

The R function hclust was used for hierarchical clustering according to the furthest neighbor clustering criteria applied to the distance matrix. We used Arlequin 3.1 (Excoffier *et al.*, 2005) for performing F_{ST} to test for differentiation among samples (Wright, 1951; Hudson *et al.*, 1992). DOTUR (Schloss and Handelsman, 2005) was used for calculating rarefaction curves for each lake after calculating p-distance matrices in R.

Construction of phylogeny

ARB (Ludwig *et al.*, 2004) was used for aligning 16S rRNA gene sequences. A maximum likelihood phylogeny was constructed from 74 unique 16S

rRNA gene sequences and 9 reference sequences using Garli version 0.96 with the default settings (Zwickl, 2008). The analysis was repeated six times and similar trees were obtained. Bootstrap support was based on more than 1500 replicates.

Distribution of diversity within populations

Subsampling was done to adjust for the difference in available clone sequences per bog lake. Sequences for all time points were pooled for each lake and then replicate subsamples of 100 sequences were drawn without replacement. All pairwise genetic distances of the subsamples with 100 sequences were calculated and binned at intervals of 0.001. Gaps and ambiguous positions (N) were excluded from the overall length in all pairwise comparisons. The frequency of each binned pairwise distance value was calculated as a fraction of the total number of binned pairwise distances. The subsampling was repeated 1000 times, and the mean frequencies and standard deviations are recorded. The distribution of genetic distance within and between clades (A and B in Figure 6) were determined as the maximum and minimum nucleotide identity from all pairwise comparisons within each lake using MEGA 4 (Kumar *et al.*, 2008).

Nucleotide sequences accession numbers

All 263 unique *mcrA* gene sequence types are available at GenBank under Accession numbers GU084829–GU085091. Abundance data for the sequence types and the sequences in fasta format are available in separate files in the Supplementary Materials.

Results

16S rDNA and *mcrA* sequences of methanogens from humic bogs

Ninety 16S rDNA clones of an average length of 1041 bp were sequenced from the October hypolimnion samples from MA, TB and SSB. Out of 90 sequences, 74 were unique sequence types. The phylogenetic relationships between the unique types and select reference sequences are shown in Figure 2. Based on this phylogeny, *candidatus* Methanoregula boonei 6A8 (Bräuer *et al.*, 2006a) is the most closely related cultured representative to the majority of the sequences retrieved from the bog lakes with $96.3\% \pm 0.6$ (mean \pm standard deviation) nucleotide identity between it and the clones within its nearest bootstrap-supported clade.

To resolve population structure at a finer scale, we sequenced 1222 clones of the protein-coding methanogen gene *mcrA* that we obtained from 15 water samples (Table 2). The average length of these sequences was 415 nucleotides. Each hypolimnion

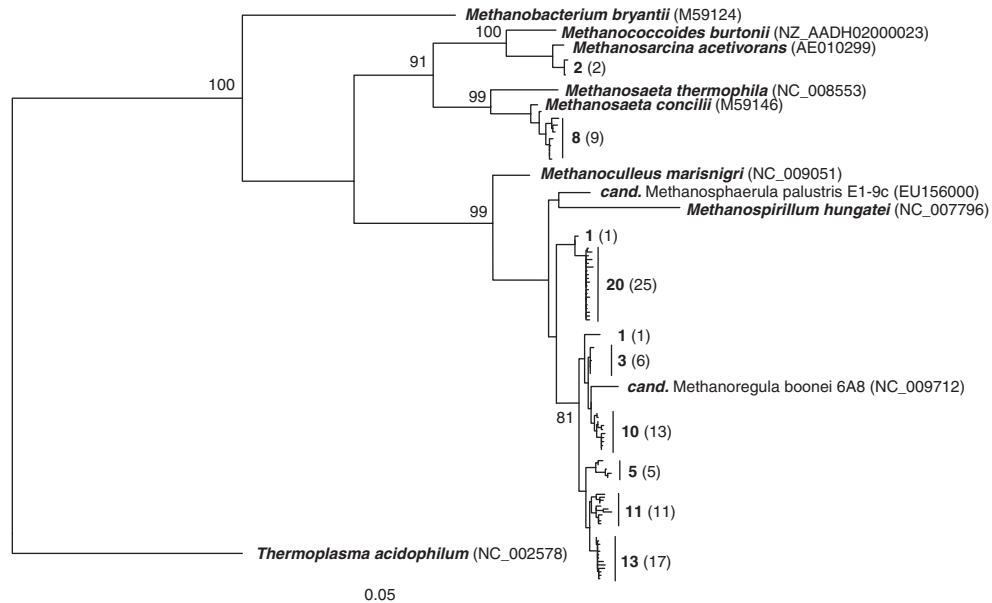


Figure 2 Likelihood tree for 16S rDNA reference sequences (accession numbers in parentheses) and environmental sequences from October samples of MA, TB, and SSB. For the environmental sequences, the number of unique sequence types per clade is given in bold followed in parentheses by the total abundance of sequences in that clade. Bootstrap support (%) is added to major nodes when greater than 70%. Scale bar represents 0.05 changes per nucleotide position.

Table 2 Characteristics of *mcrA* clone libraries.

Sampling time	Lake			No. of clones	No. of unique sequence types	Percentage of dominant sequence type ^a	Avg. p-dist within sample	No. of persisting genotypes ^b	Percentage of clones from persisting genotypes ^c
Oct 2007	MAH	Meromictic	Adelaide area	202	72	23.3	0.09	6	49
Jan 2008	MAE			79	19	20.3 ^d	0.09		
Jan 2008	MAH			39	21	30.8	0.10		
May 2008	MAH			69	29	29.0	0.07		
July 2008	MAH			86	46	22.1	0.13		
July 2008	RLH			81	35	12.5 ^d	0.15		
Oct 2007	NSBH	Dimictic	Trout area	112	49	17.0	0.22	3	33
Jan 2008	NSBH			46	21	15.2	0.20		
July 2008	NSBH			20	7	30.0	0.17		
Oct 2007	SSBH			87	14	59.8	0.09	5	76
May 2008	SSBH			116	17	55.2	0.11		
July 2008	SSBH			51	12	33.3 ^d	0.10		
Oct 2007	TBH			85	22	20.0	0.14	8	52
Jan 2008	TBH	59	17	27.1 ^d	0.25				
May 2008	TBH	90	33	16.7 ^d	0.20				

^aMeromictic' and 'dimictic' describe the mixing regime of the lakes. 'Adelaide area' and 'Trout area' give the relative geographic locations between the two groups of lakes. The letter 'H' behind the lake acronym indicates hypolimnetic samples; one MA sample is taken from the anoxic epilimnion ('E').

^bRelative abundance [%] of the dominant sequence type per sample.

^cListed once for each lake sample.

^dCalculated for each lake when all time points are combined.

^eSamples contain a second sequence type within 5% abundance difference from the dominant one.

sample contained partial *mcrA* sequences indicating the presence of methanogen DNA in the anaerobic water column. Amplification of epilimnion DNA yielded no PCR product with the exception of one MA sample in January when the epilimnion was anaerobic.

Figure 3a shows the sampling curve calculated for each set of lake samples with no binning of sequences. Sampling was incomplete when considering each unique sequence type. However, when sequences were binned at the 97% nucleotide identity level, curves began to become saturated,

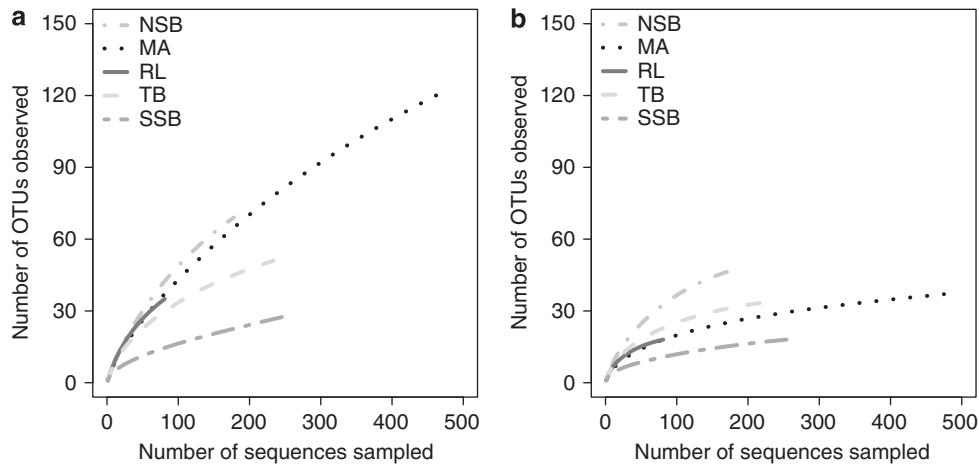


Figure 3 Rarefaction curves for pooled samples from each lake, binned at levels of nucleotide identities of 100% (a) 97% (b).

indicating that methanogen diversity was well sampled at this scale (Figure 3b).

Despite the larger information content of the more variable *mcrA* gene, we could not resolve a stable phylogeny within the *candidatus* *M. boonei* clade containing the majority of our sequences. Although many clusters of sequences could be resolved, their relationships to each other could not be determined. Because relationships among clades could not be resolved, we clustered sequences using hierarchical clustering by nucleotide identity. In Figure 4 we show the hierarchical clustering of those 66 of the 263 unique sequence types that were observed more than twice in our clone libraries to qualitatively identify patterns in space and time for the most dominant sequence types. This set of 66 unique sequence types represents 79% of the total 1222 clone sequences sampled. As we observed in the phylogenetic analysis using the 16S rRNA gene, the majority of the *mcrA* clone sequences (clusters 1, 2 and 3 in Figure 4) were most closely related to the *candidatus* *Methanoregula boonei* 6A8 *mcrA* sequence with an average nucleotide identity of $83.6\% \pm 1.3$ (mean \pm standard deviation).

Spatial and temporal patterns

As shown in Figure 4, three primary clusters of sequence types were observed that are differently distributed among the five lakes. The majority of sequences from the bog lakes of the Trout area were found within clusters 1 and 2, whereas the majority of MA sequences populate cluster 3. Sequences in cluster 3 were most similar to the *mcrA* gene sequences of the Methanomicrobiales-associated *Methanoregula*-cluster in Bräuer *et al.* (2006a). Because only one sample was analyzed for RL, the majority of unique sequences from this bog (16 of 35) are not shown in Figure 4. These sequence types are accounted for in the inset table in Figure 4 and can be found mostly in cluster 2. Only three sequences were shared by one or more of the Trout

area bog lakes and both of the Adelaide–Yolanda lakes.

Each bog lake contained a unique set of 3–8 persisting sequence types that were present at all sampling times (Table 2, Figure 4). The persisting types contained the most frequently sampled sequence types in each lake (Table 2). Only NSB and SSB shared persisting sequence types with another lake (TB) but not with each other. The six persisting sequence types in MA accounted for 49% of all MA clone sequences and were more closely related to each other (95.9% nucleotide identity) than the persisting sequences within the other lakes (78.8% TB, 86.2% NSB, 87.4% SSB). The most abundant persisting type in SSB dominated each clone library by contributing more than 50% of all clone sequences from this lake (Table 2). The persistence of sequence types through time also indicates a spatially homogenous distribution of methanogen sequences within a single lake.

qPCR was used to independently verify that the abundance patterns in our clone libraries did not result from biases in PCR amplification or cloning. We selected two clone libraries (May samples from MA and SSB) in which the highly abundant sequence type in MA was absent in SSB and vice versa. Two primer pairs ('MA_{high} F/R' and 'SSB_{high} F/R') target the highly abundant genotypes in MA and SSB, but in addition also amplify a range of closely related sequence types from the respective libraries that occur in low abundance according to the clone libraries. A third primer pair was designed to target sequences that were at low abundance in both MA and SSB (MA.SSB_{low}). qPCR results confirmed that highly abundant sequences in May samples of MA and SSB were not abundant in the other respective clone library. For both bog lakes, the highly abundant sequence types were present at approximately 5-fold higher concentrations than low abundant sequence types in the same sample (Table 3). This relationship was similar to the relative abundances inferred from the clone libraries

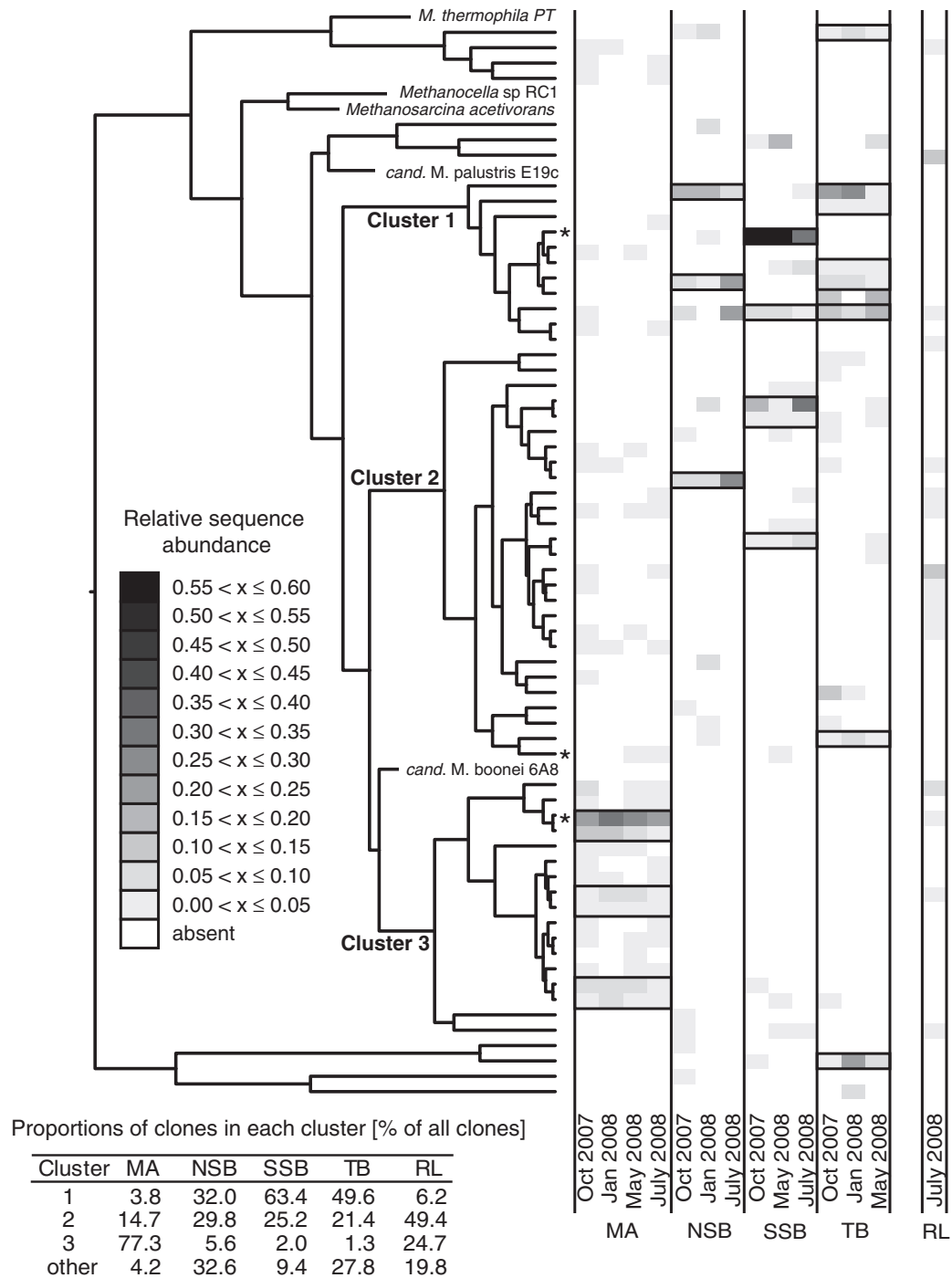


Figure 4 Distribution of *mcrA* sequence types and abundance among clone libraries. The dendrogram to the left hierarchically clusters the sequence types by their genetic distance. The heatmap to the right indicates the relative abundance of each unique sequence type according to the scale. Persisting genotypes are defined as present at all time points in specific lakes and appear boxed in the heatmap. Only clones with an absolute abundance of more than 2 clones are displayed in this figure. The inset table at the bottom of the dendrogram shows the relative frequency of clones belonging to each cluster from each lake in the complete data set. *Indicate target sequences selected for qPCR the assay. *Methanosaeta thermophila* PT (Mthe_0569), *Methanocella* sp RC1 (YP_686530), *Methanosarcina acetivorans* (AAM07885), *can.* *M. palustris* E19c (EU296536), and *can.* *M. boonei* 6A8 (Mboo_0582).

(Table 3). Results from the qPCR assay therefore support the relative abundance pattern we observe from the clone libraries.

Using F_{ST} we tested for differentiation between lakes and among samples taken at different times

from the same lake. The matrix of multiple pairwise F_{ST} values is displayed as a heatmap in Figure 5. Theoretically, F_{ST} values range between 0 (no differentiation) and 1 (complete differentiation). To test the importance of clone abundance for driving

differentiation, we calculated F_{ST} values based on the presence–absence of unique sequence types (Figure 5 top), in addition to the whole data sets including clone abundance information (Figure 5 bottom). All F_{ST} values >0.04 were significant ($P < 0.05$). The numeric data for Figure 5 is available in Supplementary Table S1 in the Supplementary materials. Differentiation was also measured using UniFrac (Lozupone *et al.*, 2006). This analysis agreed with our results from the F_{ST} despite a poorly supported phylogeny (data not shown).

F_{ST} quantitatively confirmed the qualitative patterns observed in Figure 4. MA was differentiated from all dimictic bog lakes including RL with an average F_{ST} of 0.32 ± 0.10 when accounting for the abundance distribution of sequence types. Very low F_{ST} values were observed among the Trout area bog lakes (0.07 ± 0.06) with differentiation resulting primarily from differences in abundance of indivi-

dual clones rather than the types of sequence present in the samples (average F_{ST} 0.01 ± 0.06 calculated only with presence/absence of sequences). F_{ST} values between MA and RL and between RL and the bog lakes from the Trout area were similar and lower than between MA and the Trout area bog lakes indicating that dimictic RL is intermediate in its level of differentiation. This is because RL is dominated by sequences from cluster 2 (Figure 4), which shares members with the Trout area bog lakes while in MA sequences from cluster 3 predominate.

Comparing temporal samples from each bog lake, F_{ST} values ranged from 0.0 to 0.07 when accounting for abundance information, indicating no differentiation among samples taken at different times. The exceptions were three comparisons within SSB and TB, where F_{ST} values were relatively high (0.11, 0.12 and 0.17, respectively, marked with asterisks in Figure 5 bottom). Considering only composition (presence–absence information, Figure 5 top), the respective comparisons showed average F_{ST} values between 0.00 and 0.01 and thus no signal for differentiation. Therefore, changes over time within SSB and TB related to differences in relative sequence abundance but not in observed sequence types.

Table 3 Ratios based on qPCR and clone libraries between high and low abundance sequence types

Comparison	Ratio based on qPCR	Ratio based on number of clones ^a
MA _{high} /MA _{low}	5.6	5.8
SSB _{high} /SSB _{low}	4.8	6.6
MA _{low} /SSB _{low}	1.5	0.9

^aCorrected for the actual amplification range of qPCR primers determined as described in methods

Distribution of diversity within lake populations

In Figure 6, we show the frequency distribution of nucleotide identity for all pairwise comparisons of

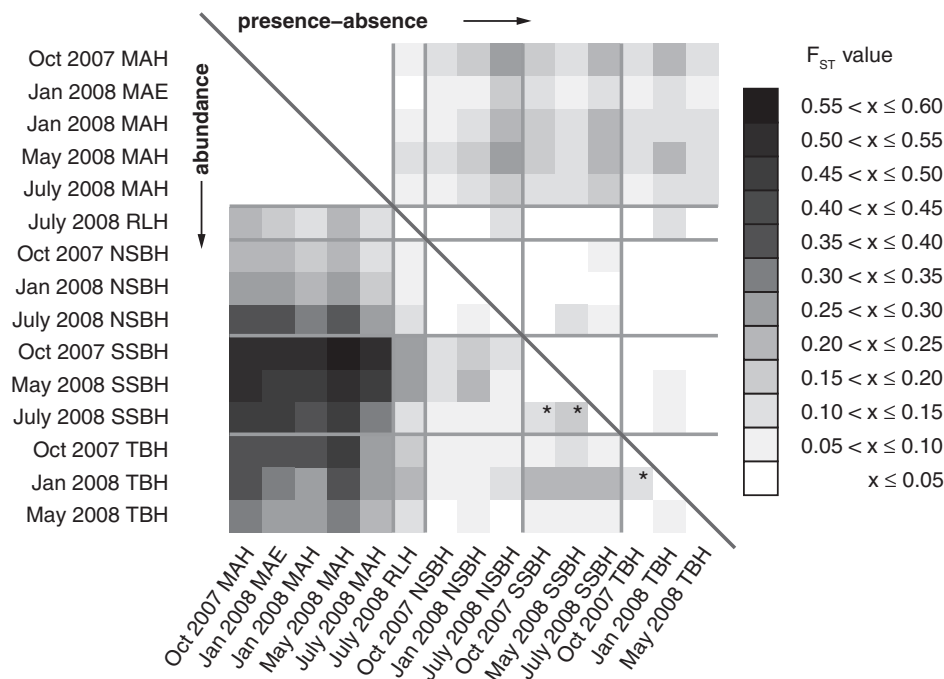


Figure 5 F_{ST} for all multiple pairwise comparisons of 15 samples. The lower triangle of the matrix represents F_{ST} values for the analysis of the clone abundance pattern. The upper triangle gives F_{ST} data when only presence–absence of unique sequence types was considered. The numeric matrix of the F_{ST} values can be found in Supplementary Table S1 in the Supplementary materials. *Three comparisons within SSB and TB, where F_{ST} values were relatively high.

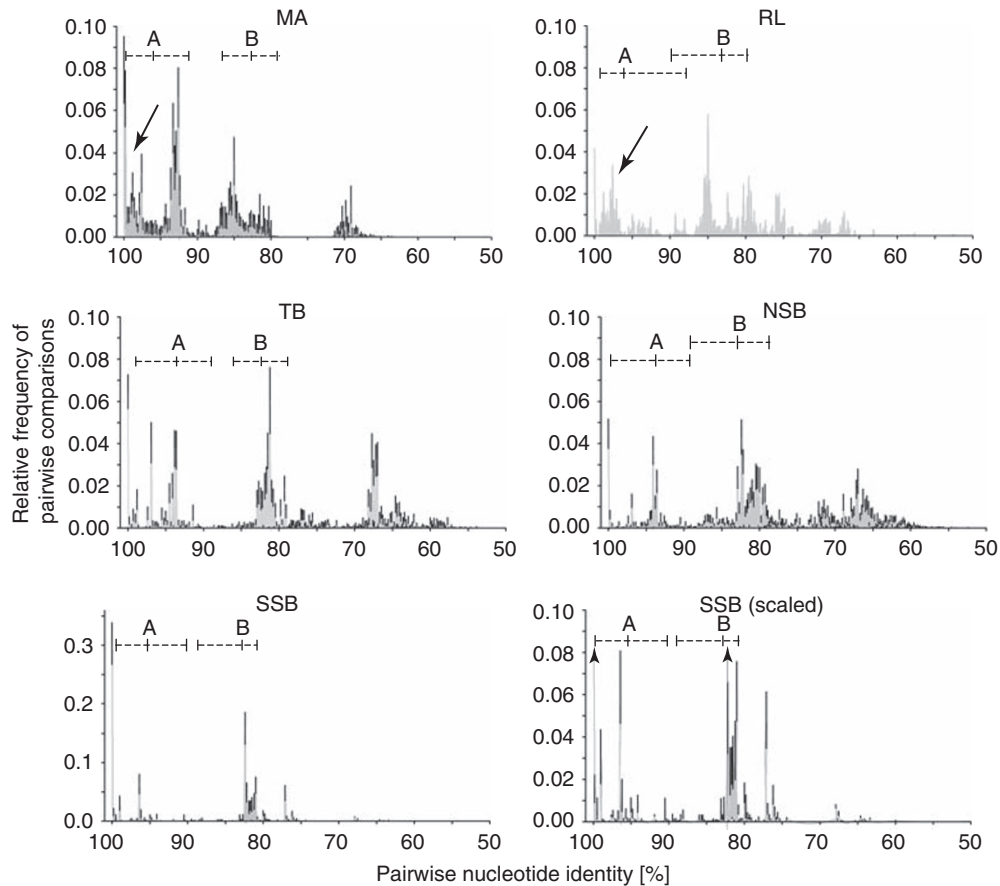


Figure 6 Distribution of pairwise nucleotide identity for the five lakes and a scaled plot for SSB. The frequency of each binned pairwise nucleotide identity value is shown as a fraction of the total number of pairwise distance values. Dark bars indicate the standard deviations of each frequency value of a random subsample of 100 clones over 1000 repetitions. The graph for RL does not contain bars for standard deviations due to a low number of clones sampled from only one time point. The y axis of SSB (scaled) is scaled to a maximum value of 0.1 to better show the smaller peaks. Arrows on two of the peaks in SSB (scaled) indicate values that extend beyond the scaling of the graph. The brackets indicate the range of pairwise identities seen when comparing sequences within each of the three clusters in Figure 4 ('A') and between the three clusters ('B'). The arrows in MA and RL point out the higher level of microdiversity seen in these lakes.

sequence types within each bog lake. Mean and standard deviation were estimated using random subsampling of 100 sequences to account for differences in the number of clone sequences per sample. Sequences from different time points were pooled to represent the diversity of the lake through time because of the temporal consistency observed in Figures 4 and 5. From this frequency distribution we observed differences in the structure of diversity among the different lakes. For MA (Figure 6), an arrow in the figure identifies a region of high microdiversity resulting from highly similar but not identical sequences with nucleotide identities of greater than 96%. Several nucleotide changes within this group were non-synonymous resulting in an average amino acid identity of 97%. For MA, 22% of all pairwise comparisons were greater than 96% nucleotide identity as opposed to 8%, 3% and 7% for TB, NSB and SSB, respectively. In the Trout area bog lakes (TB, NSB and SSB; Figures 6), we observed a higher frequency of pairwise genetic distances with lower percent identity resulting from

comparisons between clusters 1, 2 and 3 (Figure 4) and with sequences outside of the three clusters (<83% nucleotide identity, B bracket in Figure 6). The dimictic RL from the Adelaide area showed a composite structure with high levels of microdiversity (arrow in Figure 6 18% of comparisons with greater than 96% nucleotide identity) as well as more divergent types. The distribution for SSB was unique with a high frequency of a single dominant clone sequence. The observed microdiversity in MA and RL samples were not due to PCR and sequencing errors because single nucleotide polymorphisms were removed from the analysis, and because PCR and sequencing errors should be consistent among samples from all lakes. The microdiversity observed in Figure 6 did not result from the higher number of clones sequenced for MA because we compensated for sample size by displaying the mean frequencies of pairwise distances in subsamples of 100 clones. The difference in microdiversity between the methanogen populations in the five bog lakes was also prominent in the change between the

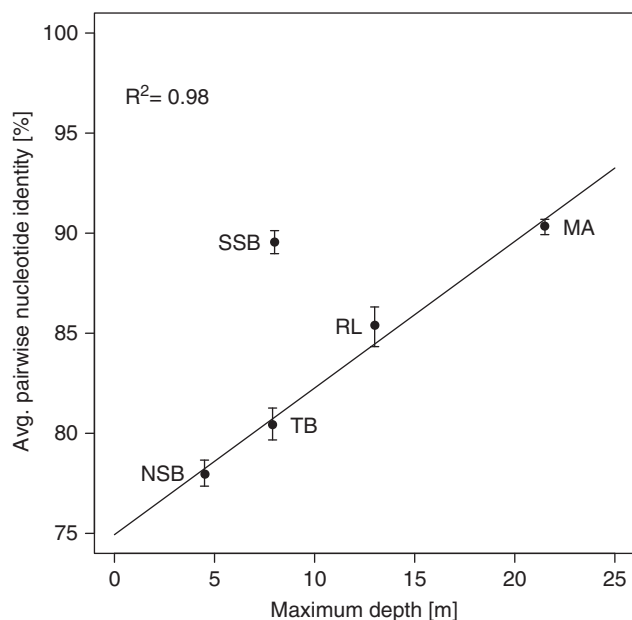


Figure 7 Average pairwise nucleotide identity within each lake for methanogen *mcrA* sequences versus the maximum depth. SSB is an outlier because of the high abundance of a single clone sequence. Error bars are standard errors for the estimates of genetic distance. The R^2 of 0.98 does not include the point for SSB.

rarefaction curves for MA and RL when closely related sequences are binned (Figures 3a and b). Even though the observed number of unique genotypes in NSB, MA and RL was similar (Table 2), the structure of the methanogen populations for these three lakes was different (Figure 6).

The distribution of diversity observed in Figure 6 is reflected in the average nucleotide identity shown in Table 2 and plotted as a function of lake depth in Figure 7. The deepest lakes RL and MA had the highest average nucleotide identity while the shallower dimictic lakes had higher levels of divergence. As shown in Figure 7 there was a strong positive relationship between depth and diversity in these lakes ($R^2 = 0.98$), where the single-clone-dominated SSB is considered an outlier and disregarded for the linear fit.

Discussion

Spatial structure among bog lakes

Methanogen populations in each of five humic bog lakes are unique, differentiated by persisting genotypes in each population. The observed differentiation may result from selection due to differing environmental conditions in the bog lakes, low migration rates or a combination of these two forces. Despite attempts to sample bog lakes with similar environmental conditions, the environmental characteristics we could measure differentiate bog lakes in the Adelaide area from the Trout area. The Adelaide area bog lakes MA and RL have a higher

maximum depth and a higher average pH (Table 1). In addition, the quaternary deposits surrounding the Trout and Adelaide area bog lakes are slightly different (Clayton *et al.*, 2002). These and other potentially important environmental differences, for example in quality and quantity of dissolved organic matter, may drive the differentiation among Adelaide and the Trout Bog area lakes. Consequently, as in many other studies in microbial ecology, it is difficult to quantify the relevant physical and chemical parameters that result in differentiation on the microbial scale. Despite the geographic similarities between RL and MA, the dimictic RL has an intermediate level of differentiation by F_{ST} between MA and the Trout area bog lakes caused by the dominance of sequences from cluster 2. In addition, RL contains both microdiverse sequences and a relatively higher number of divergent sequences than MA. This hybrid pattern of diversity in RL suggests that mixing regime, in conjunction with other physical or biogeochemical characteristics (geographic distance, pH and depth), influences the methanogen population that is present.

Even in the absence of selective drivers, differentiation of methanogens between bog lakes may occur simply by genetic drift with low migration. Given the toxicity of oxygen to methanogens, migration rates between the anaerobic hypolimnia of any two bog lakes are likely lower than for epilimnetic aerobic microorganisms as discussed by Jones and McMahon (2009). The difference between lakes could be interpreted as an indicator of random colonization from a larger and diverse metapopulation (Curtis and Sloan, 2004). However, temporal persistence of different genotypes in each of the five bog lakes suggests that colonization is not an ongoing process but instead, low migration rates decouple local bog populations from one another.

Isolation among nearby natural populations from similar environments is found in soil environments as well. Grundmann and Debouzie (2000) showed spatial separation of ammonia and nitrite oxidizing communities on the millimeter scale in soil samples. Chantratita *et al.* (2008) found in soil samples taken only meters apart from each other uniquely composed populations of *Burkholderia pseudomallei* differing in composition as well as abundance distribution. Our data in accordance with these studies show that physical separation and low migration may be important drivers for diversification at a local spatial scale.

Temporal persistence of methanogen sequence types in dynamic environments

Persistence, defined as the continuous occurrence of a sequence type despite environmental changes, has also been observed at the community level in dynamic ecosystems. For example, persistent methanogen 16S rDNA sequences were observed in dynamic soil and engineered systems (Lueders and

Friedrich, 2000; Zumstein *et al.*, 2000; Leclerc *et al.*, 2001; Collins *et al.*, 2003; Pender *et al.*, 2004). In contrast, several studies have shown population-level changes in response to seasonal variation in conditions such as temperature (Wu and Hahn, 2006; Hunt *et al.*, 2008), viral lysis (Middelboe *et al.*, 2001), and resource availability (Rocap *et al.*, 2003). Our results show stability at the population scale in bog lakes as we find several persisting *mcrA* sequence types continuously present at all sampling times (Figure 4). Our current sampling interval of approximately 3 months would not have been sufficient to accurately quantify an immediate short-term response of the methanogen populations to a mixing disturbance if it occurs. Small changes in rare sequence types cannot be detected even at a higher temporal resolution with limited sequencing capacity. However, the similar methanogen populations at a seasonal sampling interval show that if disturbances caused by lake mixing occurred, populations are resilient, quickly reverting to the original abundance distributions.

Although it does not affect the validity of the spatial and temporal patterns described here, the activity of methanogen populations in the water column should be considered. With exceptions as described for the meromictic Knaack Lake in central Wisconsin, USA (Winfrey and Zeikus, 1979), methane measurements based on concentration profiles through the water column generally suggest little or no methanogenic activity outside the sediment. Therefore, the presence of *mcrA* sequences in the hypolimnion waters may be explained by passive transport into the water column for example through ebullition of biogas formed in the sediment body or resuspension of sediment particles. In this case, hypolimnetic methanogen populations may then mirror methanogen populations in the sediments of the bog lakes. The persistence of sequence types in our data would then point towards a temporally stable population structure of methanogens in the sediment. It should be noted that PCRs with methanogen-specific primers yielded no product with aerobic epilimnion samples, which would be expected if the presence of methanogens in the water column was indeed only because of transport phenomena.

If we have sampled an active methanogen population in the water column, the relative stability over time suggests that methanogens in dynamic environments of bog lakes have developed mechanisms to survive the short-term effects of disturbances, such as oxygen spikes in the hypolimnion. Physiological adaptations to low oxygen concentrations have been suggested for various methanogens. For example, the genomes of *Methanocella sp* RC-1 (Erkel *et al.*, 2006), *Methanosarcina barkeri* (Maeder *et al.*, 2006), *candidatus* Methanosphaerula palustris E1-9c (Cadillo-Quiroz *et al.*, 2008) and *candidatus* Methanoregula boonei 6A8 (Bräuer *et al.*, 2006a) contain catalase and peroxidase

genes. The activity and specificity of methanogen populations for the water column is an interesting topic for future research.

Differences in the structure of diversity

Samples from the deeper MA and RL stand out from the Trout area bog lakes, because they contain microdiversity i.e. an abundance of clone sequences differing only in a few base positions. In contrast to the Adelaide area bog lakes, in particular NSB and TB harbor more genetic diversity between clones and few closely related sequences. Based on these data we hypothesize that the isolated and undisturbed water column in the deeper lakes promotes microdiversification resulting from neutral mutations, whereas competition between genotypes may be driving diversification in the dimictic bog lakes of the Trout area. The degree of microdiversity may then be a function of the disturbance frequency (Connell, 1978), which is driven by the depth of the bog lake (Wetzel, 2001), when more shallow lakes experience complete mixing more regularly than deeper dimictic lakes such as RL. Other depth-associated factors such as productivity, habitat volume, microbial community composition and diversity and overall population size may contribute to the correlation between depth and genetic diversity identified here. In contrast, the unique population structure of SSB with one dominant genotype may result from a recent selective sweep, and recolonization by a small population, or overall smaller population. The intriguing correlation between average nucleotide identity within each lake versus the maximum depth of the sampling site in Figure 7 encourages further research to test the hypothesis that with increasing depth, the average nucleotide identity increases because of more extensive microdiversification in contrast with more divergent methanogen sequences.

Conclusions

Methanogenic archaea were detected in the anaerobic hypolimnion of humic bog lakes. Spatial structure exists among lake populations, which may have resulted from migration barriers or selection allowing populations within each lake to diverge independently. Each lake harbored a unique set of abundant sequence types that were present at all sampling times. Fall and spring turnover of the dimictic lakes did not change the composition of the methanogen populations. Based on our data, we hypothesize that differences in the distribution of diversity within populations from meromictic lakes and shallower dimictic lakes may be driven by differences in mixing regime (for example, the frequency and strength of disturbance) or other depth-associated factors.

Acknowledgements

We thank KD McMahon, A Shade, C Caceres and H Cadillo-Quiroz for critical feedback, and discussion of methods and results, B Dalsing and D Mattis for assistance with DNA extractions and sequencing, A Shade and students and staff at UW Madison Trout Lake Station for assistance with sampling and P Burke, N Held, M Reno and three anonymous reviewers for helpful comments on the article. Funding was provided by NSF via the North Temperate Lakes Microbial Observatory (MCB-0702653) and the North Temperate Lakes Long Term Ecological Research Site (NTL-LTER) (NSF DEB-0217533 and DEB-0822700).

References

- Achtman M, Wagner M. (2008). Microbial diversity and the genetic nature of microbial species. *Nat Rev Microbiol* **6**: 431–440.
- Banning N, Brock F, Fry JC, Parkes RJ, Hornibrook ERC, Weightman AJ. (2005). Investigation of the methanogen population structure and activity in a brackish lake sediment. *Environ Microbiol* **7**: 947–960.
- Bräuer S, Cadillo-Quiroz H, Yashiro E, Yavitt JB, Zinder SH. (2006a). Isolation of a novel acidiphilic methanogen from an acidic peat bog. *Nature* **442**: 192–194.
- Bräuer S, Yashiro E, Ueno NG, Yavitt JB, Zinder SH. (2006b). Characterization of acid-tolerant H₂/CO₂-utilizing methanogenic enrichment cultures from an acidic peat bog in New York State. *FEMS Microbiol Ecol* **57**: 206–216.
- Cadillo-Quiroz H, Brauer S, Yashiro E, Sun C, Yavitt J, Zinder S. (2006). Vertical profiles of methanogenesis and methanogens in two contrasting acidic peatlands in Central New York State, USA. *Environ Microbiol* **8**: 1428–1440.
- Cadillo-Quiroz H, Yashiro E, Yavitt JB, Zinder SH. (2008). Characterization of the archaeal community in a minerotrophic fen and terminal restriction fragment length polymorphism-directed isolation of a novel hydrogenotrophic methanogen. *Appl Environ Microbiol* **74**: 2059–2068.
- Chantratita N, Wuthiekanun V, Limmathurotsakul D, Vesaratchavest M, Thanwisai A, Amornchai P *et al*. (2008). Genetic diversity and microevolution of *Burkholderia pseudomallei* in the environment. *Plos Neglected Tropical Diseases* **2**: e182.
- Clayton L, Attig J, Mickelson D, Johnson M, Syverson K. (2002). Glaciation of Wisconsin. Wisconsin Geological and Natural History Survey, 3rd edn. Educational Series 36.
- Collins G, Woods A, McHugh S, Carton MW, O'Flaherty V. (2003). Microbial community structure and methanogenic activity during start-up of psychrophilic anaerobic digesters treating synthetic industrial wastewaters. *FEMS Microbiol Ecol* **46**: 159–170.
- Connell JH. (1978). Diversity in tropical rain forests and coral reefs—high diversity of trees and corals is maintained only in a non-equilibrium state. *Science* **199**: 1302–1310.
- Curtis TP, Sloan WT. (2004). Prokaryotic diversity and its limits: microbial community structure in nature and implications for microbial ecology. *Curr Opin Microbiol* **7**: 221–226.
- Dhillon A, Lever M, Lloyd KG, Albert DB, Sogin ML, Teske A. (2005). Methanogen diversity evidenced by molecular characterization of methyl coenzyme M reductase A (McrA) genes in hydrothermal sediments of the guaymas basin. *Appl Environ Microbiol* **71**: 4592–4601.
- Earl J, Hall G, Pickup RW, Ritchie DA, Edwards C. (2003). Analysis of methanogen diversity in a hypereutrophic lake using PCR-RFLP analysis of Mcr sequences. *Microbiol Ecol* **46**: 270–278.
- Eiler A, Bertilsson S. (2004). Composition of freshwater bacterial communities associated with cyanobacterial blooms in four Swedish lakes. *Environ Microbiol* **6**: 1228–1243.
- Eiler A, Bertilsson S. (2007). Flavobacteria blooms in four eutrophic lakes: linking population dynamics of freshwater bacterioplankton to resource availability. *Appl Environ Microbiol* **73**: 3511–3518.
- Erkel C, Kube M, Reinhardt R, Liesack W. (2006). Genome of rice cluster I archaea—the key methane producers in the rice rhizosphere. *Science* **313**: 370–372.
- Ermiler U, Grabarse W, Shima S, Goubeaud M, Thauer RK. (1997). Crystal structure of methyl coenzyme M reductase: the key enzyme of biological methane formation. *Science* **278**: 1457–1462.
- Excoffier L, Laval G, Schneider S. (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**: 47–50.
- Fraser C, Alm EJ, Polz MF, Spratt BG, Hanage WP. (2009). The bacterial species challenge: making sense of genetic and ecological diversity. *Science* **323**: 741–746.
- Galand PE, Saarnio S, Fritze H, Yrjala K. (2002). Depth related diversity of methanogen archaea in Finnish oligotrophic fen. *FEMS Microbiol Ecol* **42**: 441–449.
- Grundmann GL, Debouzie D. (2000). Geostatistical analysis of the distribution of NH₄⁺ and NO₂⁻ oxidizing bacteria and serotypes at the millimeter scale along a soil transect. *FEMS Microbiol Ecol* **34**: 57–62.
- Hales BA, Edwards C, Ritchie DA, Hall G, Pickup RW, Saunders JR. (1996). Isolation and identification of methanogen-specific DNA from blanket bog peat by PCR amplification and sequence analysis. *Appl Environ Microbiol* **62**: 668–675.
- Huber T, Faulkner G, Hugenholtz P. (2004). Bellerophon; a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* **20**: 2317–2319.
- Hudson RR, Slatkin M, Maddison WP. (1992). Estimation of levels of gene flow from DNA-sequence data. *Genetics* **132**: 583–589.
- Hunt DE, David LA, Gevers D, Preheim SP, Alm EJ, Polz MF. (2008). Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* **320**: 1081–1085.
- Jones SE, Chiu CY, Kratz TK, Wu JT, Shade A, McMahon KD. (2008). Typhoons initiate predictable change in aquatic bacterial communities. *Limnol Oceanography* **53**: 1319–1326.
- Jones SE, McMahon KD. (2009). Species-sorting may explain an apparent minimal effect of immigration on freshwater bacterial community dynamics. *Environ Microbiol* **11**: 905–913.
- Juday C, Birge E. (1941). Hydrography and morphometry of some Northeastern Wisconsin lakes. *Trans Wisconsin Acad Sci Arts and Lett* **33**: 21–72.

- Juottonen H, Galand PE, Yrjala K. (2006). Detection of methanogenic archaea in peat: comparison of PCR primers targeting the *McrA* gene. *Res Microbiol* **157**: 914–921.
- Kent AD, Jones SE, Yannarell AC, Graham JM, Lauster GH, Kratz TK *et al.* (2004). Annual patterns in bacterio-plankton community variability in a humic lake. *Microbiol Ecol* **48**: 550–560.
- Klepac-Ceraj V, Bahr M, Crump BC, Teske AP, Hobbie JE, Polz MF. (2004). High overall diversity and dominance of microdiverse relationships in salt marsh sulphate-reducing bacteria. *Environ Microbiol* **6**: 686–698.
- Kumar S, Nei M, Dudley J, Tamura K. (2008). MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings Bioinformatics* **9**: 299–306.
- Leclerc M, Delbes C, Moletta R, Godon JJ. (2001). Single strand conformation polymorphism monitoring of 16S rDNA archaea during start-up of an anaerobic digester. *FEMS Microbiol Ecol* **34**: 213–220.
- Leclerc M, Delgenes JP, Godon JJ. (2004). Diversity of the archaeal community in 44 anaerobic digesters as determined by single strand conformation polymorphism analysis and 16S rDNA sequencing. *Environ Microbiol* **6**: 809–819.
- Lehours AC, Bardot C, Thenot A, Debroas D, Fonty G. (2005). Anaerobic microbial communities in Lake Pavin, a unique meromictic lake in France. *Appl Environ Microbiol* **71**: 7389–7400.
- Lehours AC, Evans P, Bardot C, Joblin K, Gerard F. (2007). Phylogenetic diversity of archaea and bacteria in the anoxic zone of a meromictic lake (Lake Pavin, France). *Appl Environ Microbiol* **73**: 2016–2019.
- Lozupone C, Hamady M, Knight R. (2006). UniFrac—an online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics* **7**: 371.
- Ludwig W, Strunk O, Westram R *et al.* (2004). ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363–1371.
- Lueders T, Chin KJ, Conrad R, Friedrich M. (2001). Molecular analyses of methyl-coenzyme M reductase alpha-subunit (*McrA*) genes in rice field soil and enrichment cultures reveal the methanogenic phenotype of a novel archaeal lineage. *Environ Microbiol* **3**: 194–204.
- Lueders T, Friedrich M. (2000). Archaeal population dynamics during sequential reduction processes in rice field soil. *Appl Environ Microbiol* **66**: 2732–2742.
- Luton PE, Wayne JM, Sharp RJ, Riley PW. (2002). The *McrA* gene as an alternative to 16S rRNA in the phylogenetic analysis of methanogen populations in landfill. *Microbiol-Sgm* **148**: 3521–3530.
- Maeder DL, Anderson I, Brettin TS, Bruce DC, Gilna P, Han CS *et al.* (2006). The *Methanosarcina barkeri* genome: comparative analysis with *Methanosarcina acetivorans* and *Methanosarcina mazei* reveals extensive rearrangement within methanosarcinal genomes. *J Bacteriol* **188**: 7922–7931.
- Middelboe M, Hagstrom A, Blackburn N, Sinn B, Fischer U, Borch NH *et al.* (2001). Effects of bacteriophages on the population dynamics of four strains of pelagic marine bacteria. *Microbiol Ecol* **42**: 395–406.
- Nelson CE. (2009). Phenology of high-elevation pelagic bacteria: the roles of meteorologic variability, catchment inputs and thermal stratification in structuring communities. *ISMEJ* **3**: 13–30.
- Paradis E, Claude J, Strimmer K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**: 289–290.
- Pender S, Toomey M, Carton M, Eardly D, Patching JW, Colleran E *et al.* (2004). Long-term effects of operating temperature and sulphate addition on the methanogenic community structure of anaerobic hybrid reactors. *Water Res* **38**: 619–630.
- R development core team (2009). R: a language and environment for statistical computing. R Foundation for Statistical Computing <http://www.R-project.org>.
- Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA. *et al.* (2003). Genome divergence in two prochlorococcus ecotypes reflects oceanic niche differentiation. *Nature* **424**: 1042–1047.
- Schloss PD, Handelsman J. (2005). Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* **71**: 1501–1506.
- Shade A, Jones SE, McMahon KD. (2008). The influence of habitat heterogeneity on freshwater bacterial community composition and dynamics. *Environ Microbiol* **10**: 1057–1067.
- Shade A, Kent AD, Jones SE, Newton RJ, Triplett EW, McMahon KD. (2007). Interannual dynamics and phenology of bacterial communities in a Eutrophic lake. *Limnol Oceanography* **52**: 487–494.
- Wetzel RG. (2001). *Limnology: Lake and River Ecosystems*. Academic Press: London, UK, 850.
- Whitaker RJ, Grogan DW, Taylor JW. (2003). Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* **301**: 976–978.
- Winfrey MR, Zeikus JG. (1979). Microbial methanogenesis and acetate metabolism in a meromictic lake. *Appl Environ Microbiol* **37**: 213–221.
- Wright S. (1951). The genetical structure of populations. *Ann Eugenics* **15**: 323–354.
- Wu QL, Hahn MW. (2006). Differences in structure and dynamics of polynucleobacter communities in a temperate and a subtropical lake, revealed at three phylogenetic levels. *FEMS Microbiol Ecol* **57**: 67–79.
- Yannarell AC, Triplett EW. (2005). Geographic and environmental sources of variation in lake bacterial community composition. *Appl Environ Microbiol* **71**: 227–239.
- Zumstein E, Moletta R, Godon JJ. (2000). Examination of two years of community dynamics in an anaerobic bioreactor using fluorescence polymerase chain reaction (PCR) single-strand conformation polymorphism analysis. *Environ Microbiol* **2**: 69–78.
- Zwart G, Crump BC, Agterveld MPKV, Hagen F, Han SK. (2002). Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquatic Microbiol Ecol* **28**: 141–155.
- Zwickl DJ. (2008). GARLI (genetic algorithm for rapid likelihood inference) https://www.nescent.org/wg_garli/Main_Page.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)