System. Appl. Microbiol. 21, 546–556 (1998)

SYSTEMATIC AND APPLIED MICROBIOLOGY

Nearly Identical 16S rRNA Sequences Recovered from Lakes in North America and Europe Indicate the Existence of Clades of Globally Distributed Freshwater Bacteria*

GABRIEL ZWART[1], WILLIAM D. HIORNS[2], BARBARA A. METHÉ[2], MIRANDA P. VAN AGTERVELD[1], RAYMOND HUISMANS[1], STEPHEN C. NOLD[1]**, JONATHAN P. ZEHR[2], and HENDRIKUS J. LAANBROEK[1]

[1]Netherlands Institute of Ecology, Centre for Limnology, Nieuwersluis, The Netherlands
[2]Darrin Fresh Water Institute, Rensselaer Polytechnic Institute, Troy, New York, USA

Received August 17, 1998

Summary

We compared bacterial 16S ribosomal RNA gene sequences recovered from Lake Loosdrecht, the Netherlands, to reported sequences from lakes in Alaska and New York State. In each of the three lake systems, which differ in pH and trophic state, some sequence types were found without related sequences (sequence identity <90%) in the data sets from the other two systems. Two sequences in the *Actinomycetes* and *Verrucomicrobia* radiations were more closely related to sequences from the New York lakes data set than to any other sequence in the global databases. However, the most striking similarities were found in the subdivisions alpha and beta of the *Proteobacteria*. In these subdivisions three different clusters of highly related bacteria were identified (97–100% sequence identity) that were represented in all three lake regions. The clusters contained no members other than freshwater bacteria. One cluster falls within a monophyletic aquatic supergroup that apparently diverged early in evolution into an exclusive freshwater cluster and an exclusive marine cluster, the so-called SAR11 cluster. The detection of these three bacterial clades in lakes distinguished by geographic distance as well as physical and chemical diversity suggests that these organisms are dispersed globally and that they possess unique functional capabilities enabling successful competition in a wide range of freshwater environments.

Key words: Global Distribution – Freshwater Bacteria – Ribosomal RNA – SAR11 Cluster – *Proteobacteria* – Dispersal – Bacterioplankton – Community analysis – Lakes

Introduction

During the last decade researchers investigating microbial diversity in the environment have increasingly made use of molecular biological techniques (PACE, 1997). By directly extracting microbial gene sequences from environmental samples, naturally occurring microorganisms can be identified, or at least classified, with much greater taxonomic resolution than microscopic examination allows and without the bias introduced by prior cultivation (BROCK, 1987; WARD et al., 1990; SUZUKI et al., 1997). The small subunit ribosomal RNA gene (referred to as the 16S rRNA gene in *Bacteria* and *Archaea*) has been the target of many environmental surveys of microbial diversity. As a result, the number of small subunit ribosomal RNA sequences in the databases that are derived from the environment is rapidly growing. This growing store of information allows the comparison of different ecosystems, and is starting to reveal patterns in the distribution of microorganisms (NOLD and ZWART, 1998).

Molecular surveys of bacterial diversity have primarily been performed in pelagic marine and soil habitats. Interestingly, many of the reported marine sequences group into a few phylogenetically coherent clusters. Members of three important proteobacterial clusters, designated the SAR11, SAR83, and SAR86, clusters, as well as the

*Publication 2445 of the Netherlands Institute of Ecology, Centre for Limnology, Nieuwersluis, The Netherlands

**Present address: Center for Microbial Ecology, Michigan State University, Michigan, USA

Abbreviations: DGGE, denaturing gradient gel electrophoresis

cyanobacterial cluster SAR6/SAR7 have been found both in the Pacific Ocean and in the Atlantic Ocean (GIOVANNONI et al., 1990; SCHMIDT et al., 1991; DELONG et al., 1993; FUHRMANN et al., 1993; MULLINS et al., 1995; RAPPÉ et al., 1997). Therefore, these clusters of related bacteria appear to be globally distributed. However, one could argue that through connective currents the worlds oceans in fact form a single large water body rather than a set of separated basins, resulting in a homogeneous distribution of marine microorganisms.

The present report concerns geographically separated freshwater ecosystems. To date, only two studies have been published in which 16S rRNA clone libraries of freshwater bacterioplankton were created. These studies explored bacterial diversity in Toolik lake in Alaska (BAHR et al., 1996) and in a group of lakes in the Adirondack mountains in New York State (HIORNS et al., 1997; METHÉ et al., 1998). With the aim to seek patterns in bacterial community composition and gain understanding of freshwater bacterial distribution, we compared these two published data sets with a clone library we created from the bacterioplankton of Lake Loosdrecht in the Netherlands. Several physical and chemical parameters differentiate these lakes. The North American lake systems are both low in nutrients and primary productivity (oligotrophic) (BAHR et al., 1996; O'BRIEN et al., 1992, METHÉ et al., 1998) whereas the Dutch lake is rich in nutrients and highly productive (eutrophic) (VAN LIERE and JANSE, 1992). Furthermore, the lakes differ in average pH; the Adirondack mountain lakes are slightly acidic (pH 5 to 6.5) while Toolik lake and Lake Loosdrecht are slightly alkaline (pH 7.5 to 8.5). The combined bacterial rDNA sequence data sets from these lakes contain sequences from the divisions of *Cyanobacteria,* gram-positive bacteria, *Cytophagales, Verrucomicrobia, Fibrobacter* and relatives and the *Proteobacteria.* Here we report on those divisions that are represented in the data sets from two or all three lake systems. Interesting similarities between lake systems are found in the *Verrucomicrobia* division and in the *Actinomycetes* subdivision of the gram-positive bacteria. However, most strikingly, despite the geographic distance and diversity separating these ecosystems, we present evidence for the existence of three groups of very closely related bacteria in the proteobacterial subdivisions alpha and beta that inhabit all three lake systems. Furthermore, one of these proteobacterial freshwater groups has a phylogenetically related counterpart in the marine environment, the widely distributed SAR11 cluster.

Materials and Methods

Description of lakes: Lake Loosdrecht (52° 11' N, 5° 4' E) is a shallow lake (depth 2.5 m maximum, 1.9 m average) with a surface area of approximately 980 ha, which is almost constantly mixed through the action of wind. The water temperature varies between 3 °C in winter to 24 °C in the summer. Total phosphorus fluctuates between 50 and 150 µg l^{-1} and chlorophyll-*a* fluctuates between 110 and 150 µg l^{-1}. The lake is constantly turbid with resuspended detritus and filamentous cyanobacteria (Secchi disk depth 0.35–0.55 m). Bacterial numbers approximate 1×10^7 ml^{-1}. Toolik lake (68° 38' N, 149° 36' W) has a surface area of 149 ha, a depth of 25 m maximum and 7 m average, and is covered with ice approximately 9 months per year (O'BRIEN et al., 1992, BAHR et al., 1996). The lake exhibits summer stratification and the epilimnetic temperature ranges between 12 °C and 18 °C in July. Total phosphorus averages 7 µg l^{-1} and photic zone chlorophyll-*a* averages around 1.3 µg l^{-1}. Summer Secchi disk depth readings are between 4 and 7 m and bacterial numbers approximate 0.5×10^6 ml^{-1}. The Adirondack mountain lake study was performed in seven lakes in the southwest region of the Adirondacks (43° 41' to 43° 49' N, 74° 29' to 75° 03' W) varying in maximum depth from 5 to 23 m with surface areas between 2.4 and 68 ha (HIORNS et al., 1997, METHÉ et al., 1998). Total phosphorus concentrations vary between lakes from 2 to 14 µg l^{-1}, chlorophyll-*a* ranges from 1.1 to 2.4 µg l^{-1}, Secchi disk depths range from 2.6 to 5.9 m for the different lakes and bacterial numbers are around 1×10^6 ml^{-1} for all lakes. For all lakes described oxygen is at or near saturation levels at the sampled depths (subsurface or epilimnetic).

Sampling and DNA isolation: Water samples from Lake Loosdrecht were collected immediately below the surface in sterile bottles, from a jetty on the north-east bank of the lake and stored in the dark at 4 °C. Within two hours after sampling, 50 ml of each sample was filtered over a polycarbonate membrane filter (0.2 µm cut-off, 2.5 cm diameter, BA83 Schleicher and Schuell, Dassel, Germany). The filter was cut in two with a sterile scalpel and each half was stored in a microcentrifuge tube at –80 °C until further processing. To lyse the cells, 0.5 g of zirconium beads (0.1 mm diameter), 0.5 ml TE buffer (10 mM Tris, pH 7.6, 1 mM EDTA) and 0.5 ml buffered phenol (pH 7 to 8) were added to the tubes containing the filter and the tubes were vigorously shaken (5,000 rpm) on a Mini Beadbeater (Biospec Products, Bartlesville, OK, USA) for two min with intermittent cooling on ice. The tubes were then centrifuged for 5 min at 10,000×*g* and the upper (aqueous) phase was collected and extracted twice with phenol-chloroform-isoamylalcohol (25:24:1). The DNA was then precipitated by adding one tenth volume of 3 M sodium acetate (pH 5) and two volumes of 96% (v/v) ethanol and centrifuging for 30 min at 14,000×*g*. Subsequently, the DNA was dissolved in water and purified on a Wizard column (Promega, Madison, WI, USA) according to the manufacturer's recommendations.

Clone library construction and sequence determination: To generate near full-length 16S rDNA clones, the extracted DNA was subjected to PCR amplification with the following primers: F27 (5'-AGAGTTTGATCMTGGCTCAG-3', LANE et al., 1991) which is specific for most *Bacteria* and R1492 (5'-GRTACCTTGTTACGACTT-3') which is specific for most *Bacteria* and *Archaea*. Numbering refers to the *Escherichia coli* 16S rRNA gene position corresponding to the 3' end of the primers. PCR amplification was performed using a PE480 thermocycler (Perkin-Elmer, Foster City, CA, USA) in a 50 µl reaction mixture containing approximately 100 ng of purified DNA, 10 mM Tris/HCl pH 8.3, 50 mM KCl, 0.01% w/v gelatin, 200 µM of each deoxynucleotide, 1.5 mM $MgCl_2$, 2.5 units of Taq polymerase (Boehringer Mannheim, Mannheim, Germany) and 0.5 µM of each primer. The temperature and cycling conditions were as follows. First, a preincubation at 94 °C for 5 min; then 25 cycles of 94 °C for 1 min, 55 °C for 1 min and 72 °C for 1 min, and a final incubation at 72 °C for 5 min. The PCR products were excised from an agarose gel (1.0%) and purified with QIAquick spin columns (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The product was ligated into pGEM-T vector (Promega) and transfected through heat-shock into Epicurian Coli XL-1-Blue MRF' supercom-

petent *E. coli* cells (Stratagene, La Jolla, CA, USA). The presence of inserts was determined by performing a PCR directly on a sample from white (positive) bacterial colonies, using primers flanking the cloning site on the vector. Plasmids were purified from successfully transfected clones using the High Pure Plasmid isolation kit of Boehringer Mannheim.

To determine the sequences of plasmid inserts, cycle-sequence reactions were performed using Thermosequenase (Amersham, Little Chalfont; UK) according to the manufacturer's instructions. Fragment separation, detection and base calling was performed using a Vistra DNA Sequencer 725 (Amersham). The sequences were determined in two directions with two vector-specific primers flanking the cloning site: M13 (–21) forward (5'-TGTAAAACGACGGCCAGT-3') and M13 (–26) reverse (5'-GAAACAGCTATGACCATG-3') which were labelled with Texas Red, and five primers binding at sites in the 16S rDNA: F357 (5'-CCTACGGGAGGCAGCAG-3'), R1053 (5'-AGCTGACGACAGCCATGC-3') and R1221 (5'-CATTGTAGCACGTGTGTAGCC-3') which were Texas Red-labeled, and F797plus (5'-GCGTTCTTCATCGTTGCGAG-CAAACRGGATTAGATACCC-3') and R518plus (5'-GCGTTCTTCATCGTTGCGAG-ATTACCGCGGCTGCTGG-3'). The latter two unlabelled primers (in combination with R1492 and F27, respectively) were used to generate PCR products from the plasmids which subsequently were subjected to cycle sequencing. The Texas Red-labelled primer used for sequencing of these PCR products (Stef1Tex 5'-GCGTTCTTCATCGTTGCGAG-3') binds specifically to the site introduced in the PCR reaction through the extensions at the 5' end of F797plus and R518plus. Database accession numbers of reported nucleotide sequences are shown in Table 1.

DGGE profiling: The DNA extracted from the water samples was amplified for analysis by denaturing gradient gel electrophoresis (DGGE) (FISHER and LERMAN, 1979) using the PCR procedure essentially as described by MUYZER et al. (1993). The PCR primers were F357GC (5'-CGCCCGCCGCGCCCCGCGCCCGGCCCGCCGCCCCCGCCCCCCTACGGGAGGCAGCAG-3'), which contains a GC-rich ‚clamp' and is specific for most *Bacteria*, and R518 (5'-ATTACCGCGGCTGCTGG-3') which is specific for most *Bacteria, Archaea* and *Eukarya*. The temperature cycling conditions were as follows. After a preincubation at 94 °C for 5 min, a total of 25 cycles were performed of 94 °C for 1 min, T_A for 1 min and 72 °C for 1 min. In the first twenty cycles, T_A decreased by 1 °C, stepwise, each two cycles, from 65 °C in the first cycle to 56 °C in the twentieth. In the last five cycles T_A was 55 °C. Cycling was followed by 5 min incubation at 72 °C. PCR buffers were as described above. For analysis of cloned sequences by DGGE, the same approximately 200 bp region of the 16S rDNA was amplified from the plasmids using the primers F357GC and R518. Reaction conditions and thermocycling were as described above.

DGGE was performed essentially as described by MUYZER et al. (1993). Briefly, similarly sized PCR products were separated on a 1.5 mm thick, vertical gel containing 8% (w/v) polyacrylamide (37.5:1 acrylamide:bisacrylamide) and a linear gradient of the denaturants urea and formamide, increasing from 30% at the top of the gel to 60% at the bottom. Here, 100% denaturant is defined as 7 M urea and 40% v/v formamide.

Electrophoresis was performed in a buffer containing 40 mM Tris, 40 mM acetic acid, 1 mM EDTA, pH 7.6 (0.5× TAE) for 16 hours at 75 V. The gel was stained for 1 h in 0.5×TAE containing 0.5 μg ml^{-1} ethidium bromide followed by 20 min destaining in 0.5× TAE. An image of the gel was recorded with a CCD camera system (The Imager, Appligene, Illkirch, France). The image as a whole was electronically manipulated solely to improve the contrast using Aldus Photostyler 2.0 (Aldus Corporation, Seattle, WA, USA).

Sequencing of DNA from DGGE bands: A small block of gel from the middle of the target band was excised from the DGGE gel with a surgical knife. The block was placed into a 2 ml screw-cap tube and 0.5 g of beads and 0.5 ml TE was added. To elute the DNA from the gel, the tube was shaken with a Mini Beadbeater for two min at 5,000 rpm and incubated overnight at 4 °C. To determine the migration characteristics of the eluated DNA, 5 μl of the supernatant was subjected to PCR with the primers F357GC and R518 as described above and analysed by DGGE. For sequence determination, the eluted DNA (5 μl of supernatant) was amplified with the primers F357 (no GC-clamp) and R518plus, and AmpliTaq Gold (Perkin-Elmer) using the PCR conditions described above. The amplification product (50 μl) was purified by electrophoresis (2% agarose, 0.5×TAE) and extraction from an excised gel fragment using the QIAquick Gel Extraction kit (Qiagen). Five μl out of the 30 μl spin column eluate was subjected to sequence reactions with the primer Stef1Tex as described above.

Phylogenetic analyses: The 16S rDNA sequences recovered from Lake Loosdrecht water and the published sequences from Toolik lake and the Adirondack mountain lakes were screened against sequences in the Ribosomal Database Project (MAIDAK et al., 1997) using the program Similarity Rank (http://rdpgopher.life.uiuc.edu/RDP/commands/rank.html), and against Genbank/EMBL/DDBJ sequences using the program Blast (ALTSCHUL et al., 1990) (via http://www.ncbi.nlm.nih.gov/BLAST/). Then, the recovered sequences and the sequences with highest similarity as obtained by the above procedures were aligned to reference sequences from all described bacterial phyla obtained from the 16S rRNA database (VAN DE PEER et al., 1997). This alignment was made on the basis of comparison of secondary structural elements in the ribosomal RNA using the Dedicated Comparative Sequence Editor (DCSE) (DE RIJK et al., 1993). Sequence identity values were computed on the basis of this alignment using the program package TREECON (VAN DE PEER et al., 1994). No corrections were made for multiple substitutions and gaps were not taken into account.

Phylogenetic distance trees were constructed with the program package TREECON in which evolutionary distances were corrected for multiple substitutions according to the algorithm of JUKES and CANTOR (1969). Gaps were taken into account as single mutational events regardless of their length. The resulting distance matrix was used to construct an evolutionary tree using the neighbour-joining method (SAITOU & NEI, 1987). The consistency of the tree branches was assessed by bootstrap analysis from resampled data.

In addition we performed maximum likelihood and parsimony analyses using the test versions 4.0d61 and 4.0d63 for DOS of the program PAUP* developed by DAVID L. SWOFFORD (Laboratory of Molecular Systematics, Smithsonian Institution, Washington DC, USA).

Results

For all three lake systems a data set of bacterial 16S rRNA gene sequences has been generated from DNA obtained from the water column. For Lake Loosdrecht we extracted total community DNA from a water sample collected in January 1996. Amplification and cloning of nearly complete bacterial 16S rRNA genes from this DNA resulted in a library of 79 clones. Among these clones, 23 different sequence types were discriminated through electrophoretic separation by DGGE. Eleven sequence types were of cyanobacterial origin, 5 were from

Table 1. Bacterial 16S rDNA sequences from Lake Loosdrecht.

clone	(length)	EMBL acc. no	division	identity to closest database sequence[a]	
LD1	1432	AJ007870	*Cytophagales*	<90%	
LD2	991	AJ007871	*Cytophagales*	90%	*Haliscomenobacter hydrossis* (M58790)
LD3	1006	AJ007872	*Cytophagakes*	92%	*Flavobacterium thalpophilum* (M58779)
LD21	1444	AJ007873	*Cytophagales*	94%	granular sludge clone MUG23 (AB011315)
LD30	990	AJ007874	*Cytophagales*	90%	*Haliscomenobacter hydrossis* (M58790)
LD10	1461	AJ007875	*proteobacteria*	<90%	
LD12	1408	Z99997	*Proteobacteria* (α)	>99%	Conc22, Conc34, ACK-M20, ACK-DH2, ACD-DE33[b]
LD17	1454	Z99998	*Proteobacteria* (β)	>98%	Conc31, ACK-C4
LD28	1463	Z99999	*Proteobacteria* (β)	>99%	Conc33, ACK-C30
LD19	1440	AF009974	*Verrucomicrobia*	<90%	
LD29	1482	AF009975	*Verrucomicrobia*	93%	ACK-DE41
LD4	649	AJ006279	*Cyanobacteria*	96%	*Oscillatoria limnetica* str. MR1 (AJ007908)
LD5	653	AJ007865	*Cyanobacteria*	92%	*Prochlorothrix hollandica* (AJ007907)
LD6	649	AJ006280	*Cyanobacteria*	>99%	*Oscillatoria limnetica* str. MR1 (AJ007908)
LD7	1409	AJ007864	*Cyanobacteria*	93%	*Prochlorothrix hollandica* (AJ007907)
LD8	1409	AJ006281	*Cyanobacteria*	99%	*Aphanizomenon flos-aquae* (Z82809)
LD9	1405	AJ006282	*Cyanobacteria*	98%	*Synechococcus* sp. PCC6307 (AF001477)
LD15	653	AJ006283	*Cyanobacteria*	91%	*Prochlorothrix hollandica* (AJ007907)
LD16	1406	AJ007866	*Cyanobacteria*	>99%	*Prochlorothrix hollandica* (AJ007907)
LD18	654	AJ006284	*Cyanobacteria*	>99%	*Oscillatoria agardhii* (X84811)
LD22	651	AJ006285	*Cyanobacteria*	>99%	*Prochlorothrix hollandica* (AJ007907)
LD25	653	AJ006286	*Cyanobacteria*	91%	*Prochlorothrix hollandica* (AJ007907)
LD27	1379	AJ007876	chloroplasts	97%	Ocean clone OCS54 (AF001657)
excband8[c]	162	AJ007867	*Actinomycetes*	99%	*Mycobacterium* sp. (U46146)
excband14[c]	147	AJ007869	*Actinomycetes*	94%	ACK-M21 (ACK-4 group)

[a]comparison to sequences from EMBL/Genbank/DDBJ databases as of may 1998.
[b]prefixes ‚Conc' refer to cloned sequences from Toolik Lake (BAHR et al., 1996), prefixes ‚ACK' refer to cloned sequences from the Adirondack Mountain Lakes (HIORNS et al., 1997).
[c]sequence amplified from DNA extracted from excised DGGE band.

the division *Cytophagales*, 4 were from the division *Proteobacteria*, 2 were from the division *Verrucomicrobia* (published previously, ZWART et al., 1998) and one was from a chloroplast. In addition, two bands in the DGGE pattern that did not correspond to any of the clones from the library were excised from the gel. Sequence analysis of DNA amplified from these bands indicated that the DNA originated from species of the *Actinomycetes* subdivision of the gram-positive bacteria. The phylogenetic affiliation and closest database relatives of each of the sequence types from Lake Loosdrecht are shown in table 1.

The published clone library from the Adirondack mountain lakes was constructed in a similar manner from total community DNA and contains 108 16S rDNA Sequences of which 16 belong to the *Cytophagales*, 59 belong to the *Proteobacteria*, 17 belong to the *Actinomycetes*, 6 belong to the *Verrucomicrobia*, 1 is related to the genus *Acidobacterium* (*Fibrobacter* division) and 9 originate from chloroplasts or cyanelles (HIORNS et al., 1997, METHÉ et al., 1998). The published Toolik lake data set partly consists of sequences from cultivated bacteria and partly of bacterial 16S rRNA sequences that were obtained ‚directly', through PCR amplification of bacterioplankton DNA (BAHR et al., 1996). The bacterioplankton for this ‚direct' retrieval from Toolik lake was prefiltered with 1 µm cutoff. We will only include the ‚directly' obtained sequences from the Toolik lake data set for comparison to the other two lake systems' data sets. These sequences all belong to the *Proteobacteria* (n = 13).

In the following paragraphs we will compare sequences from the lake systems data sets in those divisions that have sequences from two or all three lake systems. To be able to compare as many sequences from the databases as possible we have performed phylogenetic analyses using the gene segment in which most of these sequences overlap. The limited length of overlap does not allow analysis of early evolutionary events but does show phylogenetic grouping.

Verrucomicrobia

Fig. 1 shows a phylogenetic tree containing most of the sequences belonging to the *Verrucomicrobia* that are currently in the EMBL/Genbank/DDBJ databases and have overlap in the 16S rRNA Region 607 to 884 (*E. coli* numbering). None of the environmental sequences in this bacterial division are closely related to sequences obtained from cultured bacteria (<85% identity). The six sequences from the Adirondack data set are found in four different branches of the tree, with ACK-M10 being the most divergent. The closest relatives in the EMBL/Genbank/DDBJ databases to Loosdrecht sequence LD29 are Adirondack sequences ACK-DE36 and ACK-DE41

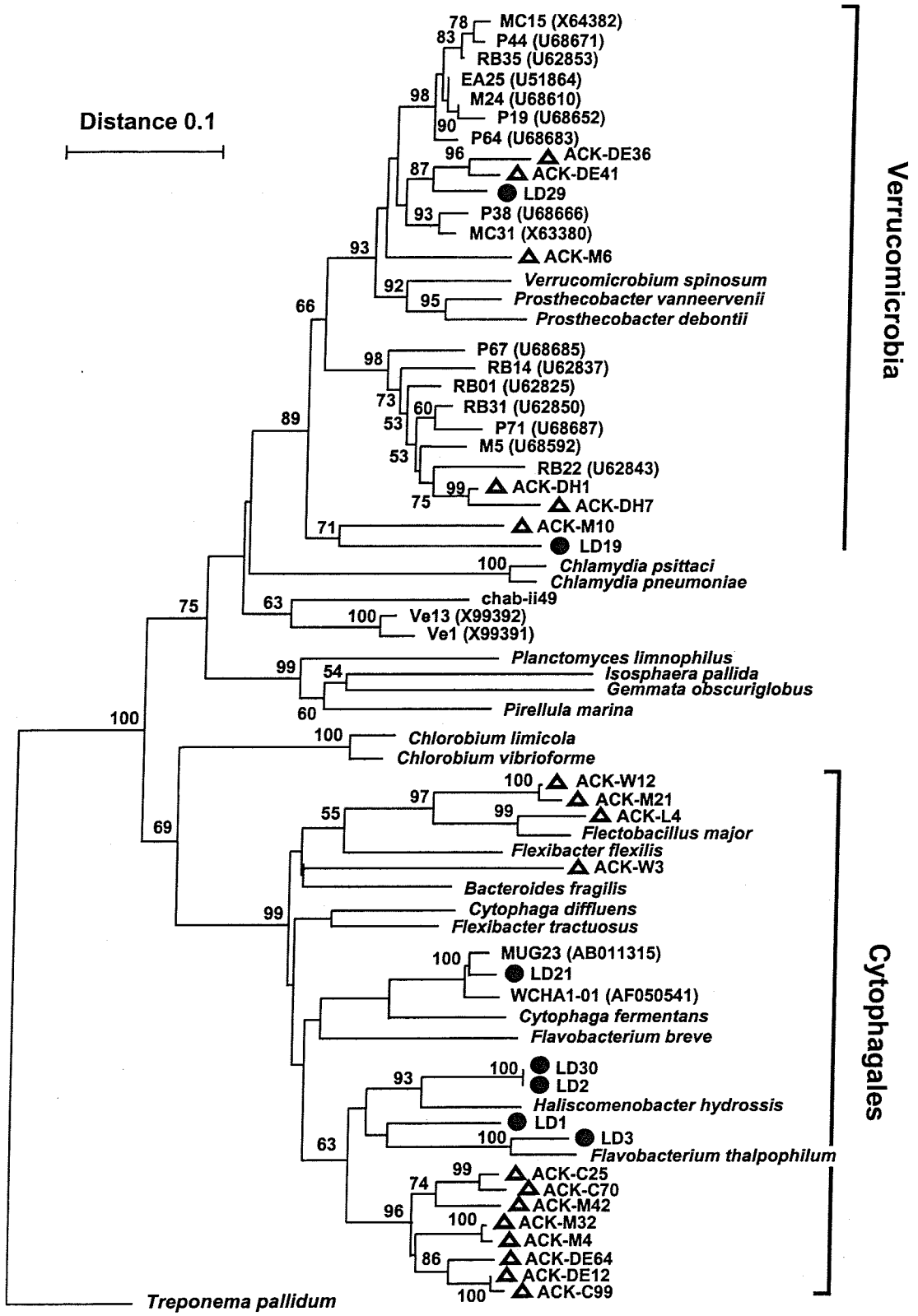

Fig. 1. Distance tree showing the phylogenetic position in the divisions *Verrucomicrobia* and *Cytophagales* of 16S rDNA clones obtained from Adirondack mountain lakes, New York State (HIORNS et al., 1997, METHÉ et al., 1998) (△), and Lake Loosdrecht, the Netherlands (●), in comparison to sequences from the nearest cultivated species and other environmental clones (accession numbers between brackets). The tree was constructed using 16S rRNA nucleotides 607 to 884 (*E. coli* numbering) and *Treponema pallidum* as outgroup. The scale bars correspond to 0.1 changes per nucleotide position. The numbers at the nodes indicate the percentage of occurrence of the cluster at the right of the nodes (if higher than 50%) among 500 bootstrapped replicates. Clone chab-ii49 has been retrieved from the Mediterranean Sea (H. SCHAEFER, unpublished data).

with 91 and 93% sequence identity, respectively. The sequences LD19 and ACK-M10 group together in this tree due to a lack of closer sequences in the databases but have an identity as low as 80%, illustrating that our view of verrucomicrobial diversity is far from complete.

Cytophagales

Twelve Adirondack sequences of sufficient length in the *Cytophagales* division are found along 4 different branches in the tree (Fig. 1). Eight of these sequences are

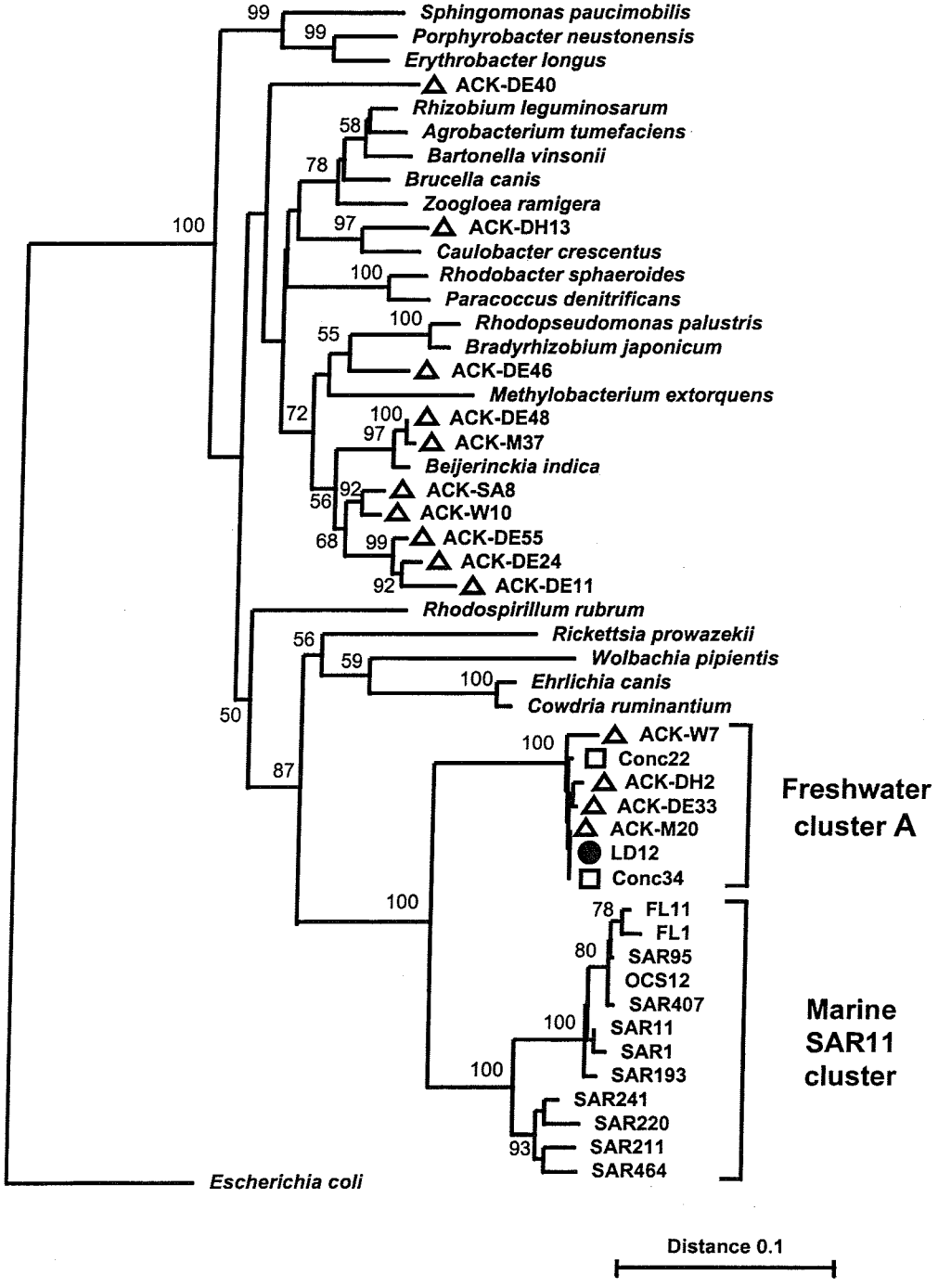

Fig. 2. Distance tree showing the phylogenetic position in the proteobacterial subdivision alpha of 16S rDNA clones obtained from Toolik lake (□), the Adirondack mountain lakes (△), and Lake Loosdrecht (●), in comparison to sequences from the nearest cultivated species and other environmental clones. The tree was constructed from 500 bootstrapped replicates using nucleotides 607–929 and *E. coli* as outgroup. Bootstrap values and scale bar as in Fig. 1.

found in one monophyletic group (designated the subdivision II ACK group by HIORNS et al., 1997) with a minimum within-group-identity of 89%. This group has no close relatives that have been cultivated. The five Loosdrecht sequences in the *Cytophagales* division occupy four different branches (LD2 and LD30 have 99% identity), and show no phylogenetic association to any of the Adirondack sequences. LD21 is related to two bacterial sequences from other freshwater sources: clone MUG23 which was obtained from a granular sludge reactor for wastewater treatment in Japan (SEKUGUCHI et al., 1998) and clone WCHA1-01 (acc. no. AF050541) which was recovered from a fuel-contaminated aquifer in Michigan, USA (NORMAN R PACE, personal communication).

Actinomycetes

Sixteen of 17 Adirondack sequences in the *Actinomycetes* division are found in a single monophyletic group (designated ACK-4 group by HIORNS et al., 1997) with a minimum within group identity of 94% (not shown). The nearest cultivated species have identies of

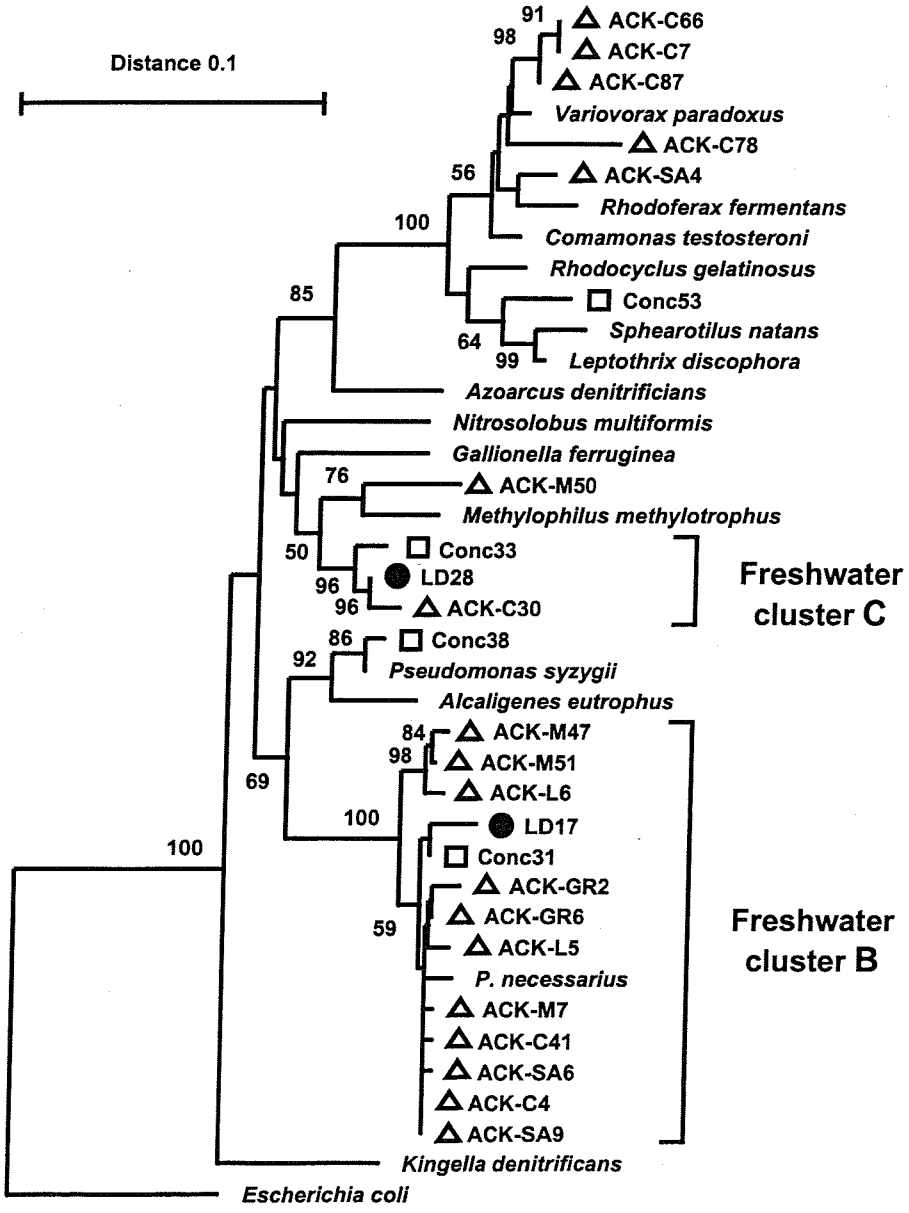

Fig. 3. Distance tree showing the phylogenetic position in the proteobacterial subdivision beta of 16S rDNA clones obtained from Toolik lake (□), the Adirondack mountain lakes (△) and Lake Loosdrecht (●), in comparison to sequences from the nearest cultivated species and other environmental clones. The tree was constructed from 500 bootstrapped replicates using nucleotides 607–929 and *E. coli* as outgroup. Bootstrap values and scale bar as in Fig. 1.

approximately 90% to this group. The sequence information from the excised DGGE bands of amplified Loosdrecht DNA is limited (approximately 150 nucleotides). Nevertheless, it is interesting that one of these sequences (excband14) is most closely related to the ACK-4 group (Table 1). The other sequence which was retrieved from an excised DGGE band (excband8) is closely related to sequences from species of the genus *Mycobacterium*.

Proteobacteria

The highest similarities between the three lake systems are found in the proteobacterial division. We identified three phylogenetic clusters in this division (designated A, B and C in Figs. 2 and 3) which each contain sequences from all three lake systems. Within each of these clusters sequence identities exceed 97% for all pairs of sequences and 99% for many pairs. High bootstrap values (>95%) separate the identified clusters from other related sequences in distance analyses using approximately 300 bp (Fig. 2 and 3). The coherence of all clusters and complete separation of other database sequences was confirmed with equally high bootstrap values in distance and parsimony analyses using approximately 800 nucleotides (*E. coli* numbering 515–1313) of the 16S rRNA gene (not shown).

Furthermore, comparison of the detected sequences to the more than 6,000 bacterial 16S rRNA sequences known to date showed that the three clusters exclusively contain sequences from freshwater environments.

Alpha subdivision

Freshwater cluster A, including sequence type LD12 from Lake Loosdrecht, belongs to the alpha subdivision of the *Proteobacteria* (Fig. 2). The 16S rDNA sequences within this cluster exhibit similarities between 99.0%–100%. As reported by BAHR et al. (1996) for the cloned sequences Conc22 and Conc34, the closest relatives to freshwater cluster A occur in the marine SAR11 cluster, at an average similarity of 86.6% between the two clusters. The similarity of the freshwater cluster A to the closest cultivated species *(Cowdria ruminantium)* is approximately 83%.

The freshwater cluster A and the marine SAR11 cluster form a monophyletic aquatic supergroup (Fig. 2), supported by bootstrap values of 100% in the maximum parsimony and distance matrix analyses. In addition, at ten nucleotide positions in the 16S rRNA, characters were found that were shared by all presently known members of this SAR11 supergroup (marine and freshwater) but were different in most (>98%) other known members of the alpha subdivision of the *Proteobacteria* (*E. coli* positions 543–544 (AC), 602 (U), 636 (A), 771 (A), 776 (U), 783:799 (G:C) and 1363 (G)). Thus, the two clusters have a common ancestor, which appears to be more recent than the ancestor to the entire alpha subdivision of the *Proteobacteria*. Since the supergroup contains only sequences recovered from aquatic habitats, the common ancestor most likely was an aquatic microorganism.

Bacteria containing the LD12 sequence type from cluster A appear to be present throughout annual cycles in Lake Loosdrecht. DGGE analysis of amplified 16S rDNA showed that all 22 water samples collected between May 1995 and November 1997 exhibited a relatively intense band at the same vertical position in the gel as clone LD12. Gel material of the ‚LD12' band (▷) was excised from four of these DGGE patterns (Fig. 4). The nucleotide sequence obtained by directly sequencing reamplified DNA from this band was identical to the LD12 sequence in all four cases, confirming the persistent presence of bacteria with this sequence type in the lake water.

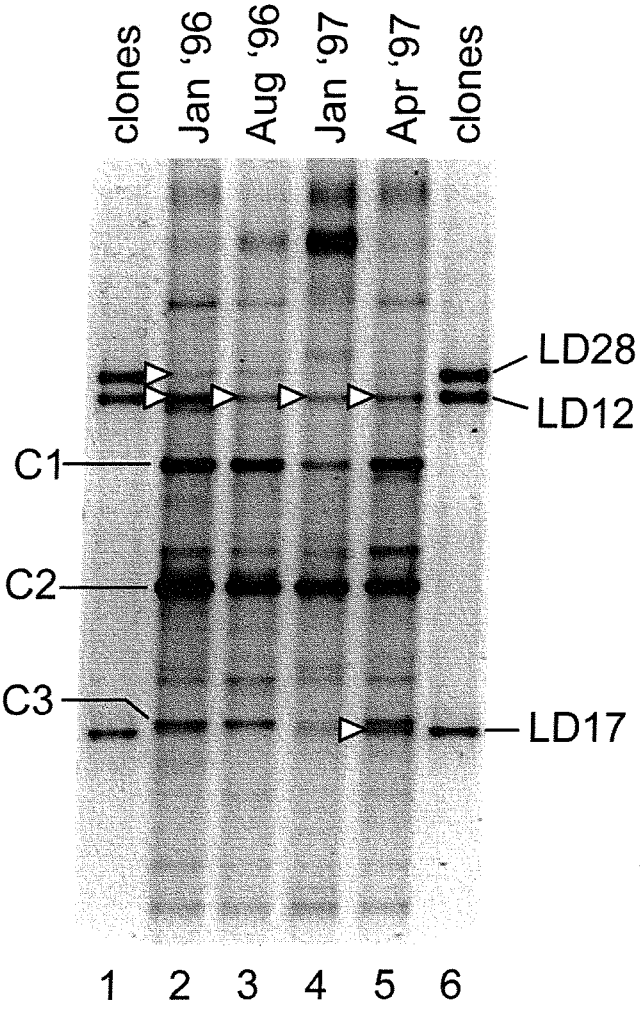

Fig. 4. DGGE analysis of clones LD12, LD17 and LD28 (lanes 1 and 6) and water samples from Lake Loosdrecht at different time points (lanes 2 to 5). Bands marked with arrowheads (▷) were excised to confirm if the nucleotide sequences of the DNA in these bands matched the comigrating clones (see text). In all cases the reamplification products migrated to the same position in DGGE as the bands from which they originated (not shown). Direct sequences recovered from bands marked C1, C2 and C3 were of cyanobacterial origin.

Beta subdivision

Freshwater clusters B and C belong to the beta subdivision of the *Proteobacteria* (Fig. 3). Cluster B has within-group similarities of 97.3–99.7%. In addition to sequence types from the three lakes, cluster B also contains a sequence obtained from borehole groundwater in Sweden (X91466, PEDERSEN et al., 1996, not shown) and the sequence from *Polynucleobacter necessarius*, an endosymbiont of a freshwater ciliate (HECKMAN, 1975; SPRINGER et al., 1996). Within cluster C the sequence identities are 99.0 to 100%. The closest cultivated species *(Methylophilus methyloptrophus)* has a maximum similarity of 96.6% to this cluster.

Bands in the DGGE pattern corresponding to the LD17 sequence type of cluster B were most intense in the April and May samples from Lake Loosdrecht and were absent from most of the samples in autumn and winter. Direct sequencing of reamplified DNA from the ‚LD17' band (▷) in the April 1997 pattern (Fig. 4) revealed a nucleotide sequence identical to sequence type LD17, confirming the seasonal fluctuation in abundance of the corresponding bacteria. Bands in the DGGE patterns at the same vertical position as sequence type LD28 of cluster C displayed low intensity and the presence of the corresponding sequence could not be confirmed in excised gel material (▷), suggesting that sequence type LD28 was present at low abundance among the amplified sequences.

Uncertain affiliation

Phylogenetic analyses (distance, parsimony, maximum likelihood) position Loosdrecht clone LD10 consistently within the proteobacterial division. However, the closest relatives in the current database have identities as low as 80% to this sequence, and positioning within any of the proteobacterial subdivisions is uncertain. The sequence of LD10 is fully compatible with known secondary structures for the 16S rRNA but displays a number of unique features among which an exceptionally long helix no. 11 (*E. coli* numbering 198–219).

Discussion

The strategies for constructing the data sets for the three lake systems were different in a number of ways, including differences in DNA isolation methods and the amplification primers used. Moreover, the Toolik lake study did not sample total community DNA. By prefiltering the water samples with a 1 µm filter prior to DNA extraction (BAHR et al., 1996), the filamentous, irregularly shaped, large or aggregated bacteria could have been selected against, whereas the Adirondack and Loosdrecht studies collected all microorganisms from pelagic samples retained by a 0.2 µm filter. Considering these methodological differences, the limited size of the data sets and the physicochemical diversity of the lakes, it is not surprising to find differences in composition of the clone libraries. Thus, we can not determine whether the lack of overlap in the *Cytophagales* division, although represented both in the Adirondack and the Loosdrecht data sets, is due to physico-chemical differences between the lake systems, to incomplete dispersal or to incomplete sampling. For the cyanobacteria, however, the large differences in representation of this group in the sequence data sets of the lakes are consistent with microscopic examinations of phytoplankton. The large proportion of cyanobacterial sequences in the Loosdrecht databases reflects the predominance of cyanobacteria in the lake, which can be attributed to the availability of high nutrient concentrations in combination with favourable light conditions, temperature and pH (VAN LIERE, 1992). The absence of cyanobacterial sequences from the North American data sets agrees with the insignificant occurrence of cyanobacteria in phytoplankton of both Toolik lake (O'BRIEN, 1992) and the Adirondack Mountain Lakes, which may partly be explained by the low nutrient concentrations in these lakes.

Therefore, differences in the three data sets are not unexpected and will partly be caused by the physico-chemical and climatic differences between the lakes. However, it is remarkable to observe the great similarities in the proteobacterial radiation.

The discovery of bacteria with nearly identical 16S rRNA sequences occurring in three isolated lake regions at large geographic distances suggests that bacteria or clades of bacteria with global distribution exist in freshwater habitats. Although the bacterial species concept and its boundaries are not firmly established (STACKEBRANDT and GOEBEL, 1994; PALYS et al., 1997), the high similarities of the ribosomal RNA gene sequences and the restriction to freshwater of all cluster members suggets that the corresponding bacteria are related at the species level. However, the confirmation of species identity must await the isolation of the bacteria in pure culture.

Cultivation may also reveal phenotypic differences between cluster members, related to the differences in trophic status and pH of the freshwater environments in which the organisms were detected. Such phenotypic differences may occur between subsets of individuals within a species (BEGON et al., 1990; see also FERRIS et al., 1997 and FIELD et al., 1997).

In the marine habitat the worldwide occurrence of bacterial clusters may be attributed to homogenous mixing between ocean basins, but such homogenisation is not possible between the studied lakes. Evidently, land masses and the salt water body between the lakes have not prevented dispersal of the described freshwater bacteria, although, as yet it is unclear how frequently such dispersal events take place. Considering the slow rate of change of the ribosomal RNA genes (OCHMAN and WILSON, 1987), the dispersal to the different continents may have occurred millions of years ago.

Successful colonisation of a habitat depends not only on bacterial dispersion but also on the ability of the bacteria to grow under the provided environmental conditions. Although many 16S rRNA sequences have been

generated from different environments, including marine and soil sites, the monophyletic clusters we describe contain only freshwater sequences. It appears that the freshwater habitat selects very particular phylogenetic clusters. This selection is not a local phenomenon acting under specific conditions but appears to be effective in chemically diverse freshwater habitats. The bacteria in the different clusters apparently possess unique properties that allow them to compete successfully for ecological niches common to these diverse habitats.

In this respect the ecological functions performed by bacteria from the freshwater cluster A and the marine SAR11 cluster are especially intriguing. Although there is a complete phylogenetic separation between the marine and freshwater members within the supergroup containing both clusters, they are ancestrally related, strictly aquatic, and are phylogenetically distinct from other known bacterial groups. Members of both clusters are widely distributed and readily detected in their respective environments. Combined, these observations suggest that members of the two clusters, although adapted to specific environmental conditions, may perform the same important ecological functions in aquatic environments.

Acknowledgements

We thank Dr. E. J. VAN HANNEN for computational support and helpful discussions.

References

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W., LIPMAN, D. J.: Basic local alignment search tool. J. Mol. Biol. 215, 403–410 (1990)

BAHR, M., HOBBIE, J. E., SOGIN, M. L.: Bacterial diversity in an arctic lake: a freshwater SAR11 cluster. Aquat. Microb. Ecol. 11, 271–277 (1996)

BEGON, M., HARPER, J. L., TOWNSEND, C. R.: Ecology: Individuals, Populations, and Communities. Blackwell Scientific Publications, Boston, MA, USA 1990

BROCK, T. D.: The study of microorganisms *in situ*: progress and problems, pp. 1–17. In: Symp. Soc. Gen. Microbiol. 41 (M. FLETCHER, T. R. GRAY, J. G. JONES, eds.) Cambridge, UK, Cambridge University Press 1987

DE RIJK, P., DE WACHTER, R.: DCSE v.2.54, an interactive tool fot sequence alignment and secondary structure research. Comput. Applic. Biosci. 9, 735–740 (1993)

DELONG, E. F., FRANKS, D. G., ALLDREDGE, A. L.: Phylogenetic diversity of aggregate-attached vs free-living marine bacterial assemblages. Limnol. Oceanogr. 38, 924–934 (1993)

FERRIS, M. J., WARD, D. M.: Seasonal distributions of dominant 16S rRNA-defined populations in a hot spring microbial mat examined by denaturing gradient gel electrophoresis. Appl. Environ. Microbiol. 63, 1375–1381 (1997)

FIELD, K. G., GORDON, D., WRIGHT, T., RAPPÉ, M. S., URBACH, E., VERGIN, K., GIOVANNONI, S. J.: Diversity and depth-specific distribution of SAR11 cluster rRNA genes from marine planktonic bacteria. Appl. Environ. Microbiol. 63, 63–70 (1997)

FISHER, S. G., LERMAN, L. S.: Length-independent separation of DNA restriction fragments in two-dimensional gel electrophoresis. Cell 16, 191–200 (1979)

FUHRMAN, J. A., MCCALLUM, K., DAVIS, A. A.: Phylogenetic diversity of subsurface marine mcirobial communities from the Atlantic and Pacific Oceans. Appl. Environ. Microbiol. 59, 1294–1302 (1993)

GIOVANNONI, S. J., BRITSCHGI, T. B., MOYER, C. L., FIELD, K. G.: Genetic diversity in Sargasso Sea bacterioplankton. Nature 344, 60–63 (1990)

HASEGAWA, M., KISHINO, H., YANO, T.: Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 21, 160–174 (1985)

HECKMANN, K.: *Omikron*, ein essentieller Endosymbiont von *Euplotes aediculatus*. J. Protozool. 22, 97–104 (1975)

HEDLUND, B. P., GOSINK, J. J., STALEY, J. T.: Verrucomicrobia div. nov., a new division of the Bacteria containing three new species of *Prosthecobacter*. Anton. Leeuwenhoek Int. J. Gen. Microbiol. 72, 29–38 (1997)

HIORNS, W. D., METHÉ, B, A., NIERZWICKIBAUER, S. A., ZEHR, J. P.: Bacterial diversity in Adirondack mountain lakes as revealed by 16S rRNA gene sequences. Appl. Environ. Microbiol. 63, 2957–2960 (1997)

JUKES, T. H., CANTOR, C. R.: Evolution of protein molecules, pp. 21–132. In: Mammalian protein metabolism, vol. 3 (H. N. MUNRO, ed.) New York, NY, USA, Academic Press 1969

LANE, D. J.: 16S/23S rRNA sequencing, pp. 115–175. In: Nucleic acid techniques in bacterial systematics (E. STACKEBRANDT, M. GOODFELLOW, eds.) Chichester, UK, Wiley 1991

MAIDAK, B. L., OLSEN, G. J., LARSEN, N., OVERBEEK, R., MCCAUGHEY, M. J., WOESE, C. R.: The RDP (Ribosomal Databases Project). Nucleic Acids Res. 25, 109–111 (1997)

MULLINS, T. D., BRITSCHGI, T. B., KREST, R. L., GIOVANNONI, S. J.: Genetic comparisons reveal the same unknown bacterial lineages in Atlantic and Pacific bacterioplankton communities. Limnol. Oceanogr. 40, 148–158 (1995)

MUYZER, G., DE WAAL E. C,; UITTERLINDEN, A. G.: Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl. Environ. Microbiol. 59, 695–700 (1993)

NOLD, S. C., ZWART, G.: Patterns and governing forces in aquatic microbial communities. Aquatic Ecology 32, 17–35 (1998)

O'BRIEN, W. J., HERSHEY, A. E., HOBBIE, J. E. HULLAR, M. A., KIPPHUT, G. W., MILLER, M. C., MOLLER, B., VESTAL, J. R.: Control mechanisms of arctic lake ecosystems: a limnocorral experiment. Hydrobiologia 240, 143–188 (1992)

OCHMAN, H., WILSON A. C.: Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. J. Mol. Evol. 26, 74–86 (1987)

PACE, N. R.: A molecular view of microbial diversity and the biosphere. Science 276, 734–740 (1997)

PALYS, T., NAKAMURA, L. K., COHAN, F. M.: Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. Int. J. Syst. Bact. 47, 1145–1156 (1997)

PEDERSEN, K., ARLINGER, J., EKENDAHL, S., HALLBECK, L.: 16S rRNA gene diversity of attached and unattached bacteria in boreholes along the access tunnel to the Aspö hard rock laboratory, Sweden. FEMS Microbiol. Ecol. 19, 249–262 (1996)

RAPPÉ, M. S., KEMP, P. F., GIOVANNONI, S. J.: Phylogenetic diversity of marine coastal picoplankton 16S rRNA genes cloned from the continental shelf off Cape Hatteras, North Carolina. Limnol. Oceanogr. 42, 811–826 (1997)

SAITOU, N., NEI, M.: The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425 (1987)

SCHMIDT, T. M., DELONG, E. F., PACE, N. R.: Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. J. Bacteriol. 173, 4371–4378 (1991)

SEKIGUCHI, Y., KAMAGATA, Y., SYUTSUBO, K., OHASHI, A., HARADA, H., NAKAMURA, K.: Phylogenetic diversity of mesophilic and thermophilic granular sludges determined by 16S rRNA gene analysis. Microbiology (UK) 144, 2655–2665 (1998)

SPRINGER, N., AMANN, R., LUDWIG, W., SCHLEIFER, K.-H., SCHMIDT, H.: *Polynucleobacter necessarius,* an obligate bacterial endosymbiont of the hypotrichous ciliate *Euplotes aediculatus*, is a member of the betasubclass of proteobacteria. FEMS Microbiol. Lett. 135, 333–336 (1996)

STACKEBRANDT, E., GOEBEL, B. M.: A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int. J. Syst. Bacteriol. 44, 846–849 (1994)

SUZUKI, M. T., RAPPÉ, M. S., HAIMBERGER, Z. W., WINFIELD, H., ADAIR, N., STROBEL, J., GIOVANNONI, S. J.: Bacterial diversity among small-subunit rRNA gene clones and cellular isolates from the same seawater sample. Appl. Environ. Microbiol. 63, 983–989 (1997)

VAN DE PEER, Y., DE WACHTER, R.: TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. Comput. Applic. Biosci. 10, 569–570 (1994)

VAN DE PEER, Y., JANSEN, J., DE RIJK, P., DE WACHTER, R.: Database on the structure of small ribosomal subunit RNA. Nucleic Acids Res. 24, 111–116 (1997)

VAN LIERE, L., JANSE, J. H.: Restoration and resilience to recovery of the Lake Loosdrecht ecosystem in relation to phosphorus flow. Hydrobiologia 233, 95–104 (1992)

WARD, D. M., WELLER, R., BATESON, M. M.: 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. Nature 345, 63–65 (1990)

ZWART, G., HUISMANS, R., VAN AGTERVELD, M. P., VAN DE PEER, Y., DE RIJK, P., EENHOORN, H., MUYZER, G., VAN HANNEN, E. J., LAANBROEK, H. J.: Divergent members of the bacterial division *Verrucomicrobiales* in a temperate freshwater lake. FEMS Microbiol. Ecol. 25, 159–169 (1998)

Corresponding author:
Dr. GABRIEL ZWART, Netherlands Institute of Ecology, Centre for Limnology, Rijksstraatwweg 6, 3631 AC Nieuwersluis, The Netherlands
Phone: +31 294 239315; Fax: +31 294 232224;
e-mail: zwart@cl.nioo.knaw.nl