Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers

Gabriel Zwart[1,*], Byron C. Crump[2], Miranda P. Kamst-van Agterveld[1], Ferry Hagen[1], Suk-Kyun Han[3]

[1]Centre for Limnology, Netherlands Institute of Ecology, Royal Netherlands Academy of Arts and Sciences, Rijksstraatweg 6, 3631 AC Nieuwersluis, The Netherlands

[2]The Ecosystem Center, Marine Biological Laboratory, 7 MBL Street, Woods Hole, Massachusetts 02543, USA

[3]University of Dankook, Laboratory of Microbial Ecology, Anseo San 29, South Korea

ABSTRACT: In order to identify patterns in bacterial community composition in freshwater habitats, we analyzed the available database of 16S rDNA sequences from freshwater plankton, including 24 new sequences from Parker River (Massachusetts, USA), 42 from Lake Soyang (South Korea) and 148 from Lake IJssel (The Netherlands). At this point, combined diversity studies using random cloning have deposited 689 bacterial and 75 plastid 16S rDNA sequences from the water column of rivers and lakes in North America, Europe and Asia. Systematic comparisons with the global database showed that the majority of the bacterial sequences were most closely related to other freshwater clones or isolates, while relatively few were closest to sequences recovered from soils or marine habitats. This habitat-specific clustering suggests that the clustered 16S rDNA sequences represent species or groups of species that are indigenous to freshwater. We have discerned 34 phylogenetic clusters of closely related sequences that are either restricted to freshwater or dominated by freshwater sequences. Of these clusters, 23 contained no cultivated organisms. These putative freshwater clusters were found among the alpha-, beta- and gamma-*Proteobacteria*, the *Cytophaga-Flavobacterium-Bacteroides* group, the *Cyanobacteria*, the *Actinobacteria*, the *Verrucomicrobia*, the green non-sulfur bacteria and candidate division OP10. This study shows that rivers and lakes have a specific planktonic bacterial community distinct from bacteria in neighboring environments such as soil and sediments. It also points out that these planktonic bacteria are distributed in diverse freshwater ecosystems around the world.

KEY WORDS: Microbial diversity · Ribosomal RNA gene · Freshwater · Habitat · Polymerase chain reaction · Phylogeny · Nucleotide sequence database

INTRODUCTION

Only 6 years ago, our knowledge of bacteria that naturally occur in freshwater ecosystems was restricted to organisms that could be grown in culture. However, it was known that most bacteria in the environment could not be cultivated (Buck 1979, reviewed in Amann et al. 1995, Hugenholtz et al. 1998). Therefore, with the exception of taxa from the division *Cyanobac-teria*, which are relatively amenable to cultivation and morphological differentiation, we had very limited insight into which bacteria are common in freshwater ecosystems. In fact, on the basis of cultivation techniques Rheinheimer (1980) concluded that bacteria found in groundwater, spring water and streams also occur in soils, and that there was no clear separation between soil bacteria and aquatic bacteria. These observations call into question the existence of a unique freshwater bacterial flora.

The advent of molecular techniques and especially the polymerase chain reaction (PCR; Saiki et al. 1985)

---

*E-mail: zwart@cl.nioo.knaw.nl

has made it possible to obtain information on microbial community composition directly, without cultivation (Giovannoni et al. 1990). An environmental sample can be inventoried for taxa present by direct nucleic acid isolation, followed by amplification of particular marker genes and analysis of the sequence of base pairs. The most widely used marker gene is the small subunit rRNA gene (16S rDNA), and the recent application of molecular techniques in a variety of habitats has produced a large set of sequences from this gene. This growing database has taught us that the diversity of the microbial world is much larger than we were able to estimate before the use of molecular techniques (Pace 1997, Hugenholtz et al. 1998). However, a clear view of the species or groups of species that we can expect in particular environments is still lacking. This is due in part to the great diversity of bacteria. In addition, an overview is lacking due to the focused approach followed in many molecular diversity studies. While most studies compare retrieved sequences from the sites under study to the global database (e.g. large studies by Glöckner et al. 2000 and Urbach et al. 2001), a comprehensive comparison of datasets from similar sites has not been presented. We argue that combining the sequence data from separate studies in similar habitats would reveal patterns in bacterial distribution. From such patterns we can learn about diversity, global distribution, habitat specificity and niche differentiation, and possibly obtain indications as to physiology and function of as yet uncultivated bacteria.

In a limited comparison of freshwater studies, Zwart et al. (1998) previously observed remarkable similarities among 16S rDNA sequences recovered from 3 different lake sites. Since then several new datasets have become available and at present, when we take a 16S rDNA sequence from a particular freshwater body and compare it to the global database, it is not unusual to find several sequences from other freshwater systems to be its closest relatives. This observation provides initial evidence of a unique freshwater microbial flora and demonstrates the need to obtain an overview of freshwater bacterial diversity.

In this study we present an overview and analysis of available bacterial 16S rDNA sequences recovered from freshwater lakes and rivers. We include 24 new 16S rDNA sequences from Parker River (Massachusetts, USA), 42 new sequences from Lake Soyang (South Korea) and 148 new sequences from Lake IJssel (The Netherlands). Comparison with the databases shows that the majority of freshwater sequences are most similar to other sequences recovered from this habitat. We will show that many of these sequences occur in tight phylogenetic groupings dominated by freshwater bacteria.

MATERIALS AND METHODS

16S rDNA sequences from Parker River, Lake IJssel and Lake Soyang. Parker River is a small woodland river with relatively high carbon content from forest runoff (Hopkinson et al. 1998). For DNA isolation, a subsurface sample (500 ml) was taken at a dam located immediately upstream of the estuary, in December 1999. The sample was filtered (0.2 µm) to collect a total community sample. The methods for DNA isolation from the filter, PCR amplification using general bacteria-specific primers (Table 1), and cloning and sequencing of nearly full-length 16S rDNA were as

Table 1. Bacterial diversity studies using community DNA, direct polymerase chain reaction (PCR) amplification and random cloning of 16S rRNA genes in freshwater lakes or rivers

Site	State/country	Coordinates	Study	Primers used for cloning (*Escherichia coli* positions) Forward	Reverse
Toolik Lake	Alaska, USA	68° N, 149° W	Bahr et al. (1996)	515–533	1492–1510
Adirondack Mountain lakes	New York, USA	43° N, 74° W	Hiorns et al. (1997)	8–27	1492–1512
Lake Loosdrecht	Netherlands	52° N, 5° E	Zwart et al. (1998)	8–27	1492–1509
Lake Baikal	Russia	52° N, 107° E	Semenova et al. (1998)	50–68	519–536
Lake Baikal	Russia	52° N, 107° E	Glöckner et al. (2000)	514–533	1385–1402
Lake Cadagno	Switzerland	46° N, 8° E	Bosshard et al. (2000)	8–27	1492–1510
Lake Gossenkölle	Austria	47° N, 11° E	Glöckner et al. (2000)	8–27	1492–1510
Lake Fuchskuhle	Germany	53° N, 13° E	Glöckner et al. (2000)	8–27	1492–1510
Crater Lake	Oregon, USA	43° N, 122° W	Urbach et al. (2001)	8–27	1518–1537
Lake Soyang	South Korea	38° N, 128° E	This paper	8–27	1492–1510
Lake IJssel	The Netherlands	52° N, 5° E	This paper	9–27	1492–1509
Columbia River	Washington, USA	46° N, 123° W	Crump et al. (1999)	8–27	1492–1510
Parker River	Massachusetts, USA	43° N, 71° W	This paper	8–27	1492–1510

described previously (Crump et al. 1999). Nearly full-length 16S rDNA sequences were determined of 24 randomly selected clones (designated PRD for Parker River dam) and deposited at Genbank/EMBL/DDBJ under accession numbers AF289149 to AF289172.

Lake Soyang is the largest reservoir in Korea. It is a dendritic freshwater lake with a length of 60 km, mean width of 0.5 km and maximum depth of 100 m. Samples were taken at different depths near the dam (37° 57′ N, 127° 49′ E) using a Niskin sampler on November 1, 1997, and transferred to a dark bottle at 4°C, and approximately 1 l of each sample was filtered through a Sterivex GV (0.22 μm) filter (Millipore). Filters were incubated with 1.5 ml of lysis buffer (10 mM Tris·Cl, 100 mM EDTA, 15 mg ml^{-1} lysozyme, pH 8.0) at 37°C for 60 min. Then the mixture was incubated with Proteinase K (final concentration 50 μg ml^{-1}) at 55°C for 30 min, after which sodium dodecyl sulfate was added (1% w/v final concentration), and 3 cycles of freezing (–70°C) and thawing (65°C) were performed. After extractions with phenol-chloroform-isoamyl alcohol (25:24:1) and chloroform-isoamyl alcohol (24:1), the DNA was precipitated in isopropanol at –20°C. Bacterial 16S rRNA genes were PCR amplified in a 50 μl reaction mixture containing 500 ng template DNA, 10 mM Tris·HCl pH 8.3, 40 mM KCl, 50 μg ml^{-1} bovine serum albumin (BSA), 1.5 mM MgCl2, deoxynucleoside triphosphates (2.5 mM each), 100 pmol of each primer (27F [5′-AGAGTTTGATCMTGGCTCAG-3′] and 1492R [5′-GGYTACCTTGTTACGACTT-3′], adapted from Lane 1991) and 2.5 U *Taq* polymerase. After an initial incubation at 94°C for 4 min, 30 cycles were run of 94°C for 1 min, 55°C for 1 min and 72°C for 2 min, followed by a final incubation at 72°C for 10 min. After excision of appropriate bands from a 0.8% agarose gel and purification using a Geneclean III kit (Bio101), the PCR products were cloned into pGEM-T vector (Promega). From 331 clones originating from different depths in the pelagic, 42 were selected for sequencing (lengths 252 to 531, accession numbers AF107496 to AF107537) on the basis of *Msp*I and *Alu*I restriction patterns. Clone designations were SY1 retrieved from surface sample, SY2 from 10 m depth, SY3 from 20 m, SY4 from 50 m, and SY5 from 80 m. Sequences retrieved from water column-sediment interphase (SY6) and sediment (SE) were not included in basic local alignment search tool (BLAST) analyses (see below).

Lake IJssel is one of western Europe's largest lakes (surface area 113 200 ha, average depth 5 m; Berger & Sweers 1988, Gons et al. 1998). It was closed off from the sea in 1932 and has been freshwater since 1937. The lake receives more than 70% of its water from the River IJssel, in addition to water from the surrounding, mainly agricultural land. For DNA isolation integrated depth samples were taken at 3 different locations in the lake (52° 40′ N, 5° 32′ E, near the village of Urk; 52° 50′ N, 5° 13′ E, near Medemblik; 52° 57′ N, 5° 19′ E, near Stavoren). DNA isolation, PCR amplification with general bacterial primers (Table 1), cloning of nearly full-length 16S rDNA and sequencing were performed according to the methods described earlier (Zwart et al. 1998). Sequences from Lake IJssel with designations Urk0, Sta0 and Med0 were from samples taken at stations Urk, Medemblik and Stavoren, respectively, on 13 September 1995. From these 3 samples, 91 clones were obtained containing the 16S rRNA insert, and 21 of these were selected for sequencing on the basis of denaturing gradient gel electrophoresis (DGGE) band position. Twenty clones were of bacterial origin (sequenced lengths 527 to 927, accession numbers AJ416156 to AJ416175) and 1 was from a chloroplast (accession number AJ427423). Sequences with designations Sta1 to Sta6 (lengths 331 to 1021) were from a sample taken at station Stavoren on 29 July 1997, with the digit 1 to 6 indicating the respective culture dishes used for plating transformed *Escherichia coli*. From this Stavoren sample, 132 clones were selected for sequencing on the basis of *Msp*I and *Hin*P1I restriction patterns from 259 screened clones containing inserts of the appropriate size. Of these 132 sequences, 105 were bacterial sequences (accession numbers AJ416176 to AJ416280), 22 were chloroplast sequences (accession numbers AJ427424 to AJ427445), and 5 were apparently artifacts and were discarded (4 were chimeric as identified through phylogenetic analysis of partial sequences and non-complementarity of helix regions, and 1 contained a sequence with closest match to the cloning vector).

Datasets included in analysis. This analysis was restricted to studies that obtained 16S rDNA sequences from the water column of lakes or rivers through direct amplification, using general bacterial primers, from DNA and random cloning. At the time of our analysis (July 2001), 9 bacterial diversity studies (performed at 12 sites) met these criteria (Tables 1 & 2). The datasets of 16S rDNA sequences from these studies as deposited in the EMBL/Genbank/DDBJ databases are relatively small, comprising a total of 689 bacterial 16S rDNA sequences and 75 plastid sequences. The Adirondack Mountain Lakes study included 7 lakes in the same geographical region, but, for brevity, will be mentioned as 1 site in this paper. A study relating recovered 16S rDNA sequences to chemical differences among these lakes was published by Methé & Zehr (1999). The Columbia River study included sequences recovered from the estuary and coastal ocean near the river outlet, as well as sequences from a freshwater site in the lower reaches of the river, but only the freshwater sequences were

Table 2. General characteristics of the freshwater bodies studied

Lakes:	Area (ha)	Max. depth (m)	Mean depth (m)	Type	Trophic status	Ice cover	Watershed type	Comments
Toolik lake	150	25	7	Arctic lake	Oligotrophic	$\sim$9 mo yr^{-1}	Arctic tundra	
Adirondack Mountain lakes	2–68	4–23		Forest lakes	Oligotrophic Mesotrophic	$\sim$1 mo yr^{-1}	Forest	7 lakes, pH ranging from 5–7
Lake Loosdrecht	980	3	2	Peat lake	Eutrotrophic	Rarely	Agricultural/urban	
Lake Baikal	3150000	1620	730	Foothill lake	Oligotrophic	4 mo yr^{-1}	Forest/mountain steppe	
Lake Cadagno	260	21	9	Meromictic alpine lake	Oligotrophic	7 mo yr^{-1}	Mountainous	Sampled at oxic/ anoxic interface
Lake Gossenkölle	2	10	5	Mountain lake	Oligotrophic	8 mo yr^{-1}	Mountainous	
Lake Fuchskuhle	2	6	4	Forest/bog lake	Mesotrophic-Acidotrophic	Rarely	Forest/bog	
Crater Lake	53200	589		Crater lake	Oligotrophic	Rarely	Sparsely vegetated crater walls	Sampled at 0, 120 and 500 m depth
Lake Soyang	2600	118	34	Water reservoir		No	Mountainous/forest	Sampled at 0, 10, 20 and 50 m
Lake IJssel	113600	6	5	Lagoon	Eutrophic	Rarely	Agricultural/urban	
Rivers:	Discharge (m³ s^{-1})			Type	Trophic status	Ice cover	Watershed type	
Columbia River	5500			River lower reach	Oligotrophic	No	Forest/agricultural/urban	
Parker River	1			Small river	Mesotrophic	Partial $\sim$2 mo yr^{-1}	Forest/urban	

included in the dataset analyzed (see below). Two separate random cloning studies were performed on Lake Baikal, and sequences from both were included in the analysis (Table 1).

The methods used to create 16S rDNA clone libraries were similar in the 9 bacterial diversity studies, but not identical. Most studies collected a total plankton sample from the water column. However, in the Toolik Lake study, water samples were prefiltered using a 1 µm pore size filter, thus excluding bacteria living in colonies or attached to particles. In the Columbia River study, particle-attached bacteria were separated from free-living bacteria using a 3 µm pore size filter and both fractions were analyzed. DNA isolation methods differed slightly among studies and, in addition, there are differences in the primer sets used for cloning (Table 1).

BLAST analyses. The program BLAST (Altschul et al. 1997) uses a fast heuristic searching algorithm to identify nucleotide stretches of high similarity to a query sequence. The program produces a list of best matching sequences with similarity scores. We used version BLASTN 2.2.1 from the National Center for Biotechnology Information (NCBI) web pages (http://www.ncbi.nlm.nih.gov/BLAST/) to examine the EMBL/Genbank/DDBJ databases for sequences with high similarity to the 764 16S rDNA sequences from freshwater.

Phylogenetic analyses. Alignments were made for each major division of the *Bacteria* using the freshwater sequences as well as the first 5 to 10 best matching sequences from BLAST analyses. Identical sequences were combined and unalignable segments of the sequences were excluded from further analyses. All presented sequence similarities are percentages of uncorrected nucleotide identities between 2 compared sequences calculated on the basis of alignments. Gaps were not taken into account.

Phylogenetic analyses were accomplished with the PAUP 4.0b4a for Macintosh program (Swofford 1999). Substitution models for estimation of distance matrices were chosen using likelihood ratio tests (LRT) (Sullivan & Swofford 1997, Huelsenbeck & Crandall 1997, McArthur & Koop 1999) calculated with the program Modeltest Version 3.06 (Posada & Crandall 1988). Distance matrices were estimated using these models (Table 3)

Table 3. Regions of the 16S rDNA gene (*Escherichia coli* numbering system) and models of DNA substitution used for each tree presented in Fig. 1. ef: assume equal base frequencies; G: estimate parameter of the gamma distribution; I: estimate proportion of invariable sites; SYM: Zharkikh (1994); TIM: Rodríguez et al. (1990); TrN: Tamura & Nei (1993)

Phylogenetic tree	*E. coli* region	Models of DNA substitution
A. *Proteobacteria*-α	136–468	SYM+I+G
B. *Proteobacteria*-β	134–523	TrN+I+G
C. *Proteobacteria*-β	134–523	TrNef+I+G
D. *Proteobacteria*-γ	134–527	TrNef+I+G
E. *Verrucomicrobia*	134–216, 252–425	TrNef+I+G
F. *Actinobacteria*	134–438	TIMef+I+G
G. Green non-sulfur bacteria	134–216, 252–438	TrN+G
H. CFB group	134–384	SYM+I+G
I. Candidate division OP10	134–548	TrN+I+G

under maximum likelihood criteria. Minimum evolution trees were determined using 3 iterations of tree and parameter estimations, with the last iteration including 100 random-addition replicates. Bootstrapping of the data used both distance and parsimony estimations using the same models and parameters as heuristic searches on 100 replicates, with 1 random-addition replicate per bootstrap replicate, under minimum evolution criteria. The computation time for *Actinobacteria* sequences bootstrap analysis using parsimony was necessarily reduced to a manageable length by omitting some taxa and by limiting the number of trees retained in each replicate to 100 000. Multiple analyses, using different regions of the 16S rRNA gene, were conducted to determine the phylogenetic position of partial sequences that did not include the region used to produce the trees presented in this paper.

RESULTS AND DISCUSSION

Represented bacterial divisions

BLAST analyses supported identification of 75 of the 764 sequences as chloroplast sequences and 689 sequences as of bacterial origin. Chloroplast sequences were not analyzed in this study.

The major bacterial divisions represented in most or all freshwater sites are the *Proteobacteria* (alpha- and beta-subdivisions), the *Cytophaga-Flavobacterium-Bacteroides* (CFB) group, the *Actinobacteria* and the *Verrucomicrobia* (Table 4). Except for the *Verrucomicrobia*, the widespread occurrence of organisms from these divisions in freshwater is in agreement with the findings of *in situ* hybridization studies, which used probes specific for these groups (Glöckner et al. 1999).

The ubiquity of members of the *Verrucomicrobia* division has not yet been shown in hybridization studies due to a lack of probes specific for this phylogenetic division.

The distribution of sequence types among bacterial divisions in these freshwater studies clearly distinguishes the bacterial communities from communities in marine and soil habitats. As was first pointed out by Methé et al. (1998), the ubiquity of *Proteobacteria* of the β-subdivision in freshwater is in sharp contrast to the relative absence of this subdivision in the open oceans. Divisions that are well represented in clone libraries from soils, including the Gram-positive bacteria with low G+C content, the *Acidobacterium*/*Holophaga* division and the *Planctomycetes* (e.g. Borneman et al. 1996, Barns et al. 1999, Felske et al. 1998, Nüsslein & Tiedje 1998, McCaig et al. 1999) appear to be poorly represented in the water column of freshwater ecosystems.

Typical freshwater bacteria

Preliminary examinations of 16S rDNA sequences from our own and other freshwater studies showed that the most closely related sequences frequently originated from another freshwater study. To explore this finding more systematically we investigated all 689 bacterial 16S rRNA sequences from the available freshwater datasets for similarities to database sequences using the program BLAST. The best matching sequence for each freshwater sequence not originating from the same site was tabulated and an overview is shown in Table 5. A total of 373 of the 689 examined bacterial freshwater sequences (54%) showed highest similarity to a sequence recovered from the water column of another freshwater body. Of these 373 sequences, 260 had their closest relatives among the investigated 689 sequences, demonstrating that there were many similarities among the 12 study sites; 39 were most closely related to clones from other freshwater studies, among which are enclosure studies of Riemann & Winding (2001) and Šimek et al. (2001), and 74 of the 373 were most closely related to cultivated strains isolated from freshwater plankton (Table 5). The number of best matching sequences in the database that originate from freshwater should be evaluated against the proportion of sequences from different habitats present in the database. However, the EMBL/Genbank/DDBJ databases are not set up for habitat queries, and very frequently information on sample origin is missing from the entries. Nevertheless, the number of sequences deposited from studies in soils and marine environments is probably greater

Table 4. Distribution of clones from freshwater studies among bacterial divisions. CFB: *Cytophaga-Flavobacterium-Bacteroides* group

Study site	Proteobacteria-α	Proteobacteria-β	Proteobacteria-γ	Proteobacteria-δ-ε	CFB	Actinobacteria	Gram-pos. low G+C	Cyanobacteria	Verrucomicrobia	Planctomycetes	Green non-sulfur	Holophaga	OP10	Other	Unknown bacterial	Total bacterial	Plastids	Total
Toolik Lake	2	4														6		6
Adirondack Mt Lakes	17	19	8		16	17			6			1				84	8	92
Lake Loosdrecht	1	2			5			11	2						1	22	1	23
Lake Baikal	21	11	7	5	4	24		7	2	3		5				89	3	92
Lake Cadagno	1		4	4	6	1	2		1		2			1	2	24	2	26
Lake Gossenkölle	5	5	1		16	6						2		2	3	40		40
Lake Fuchskuhle	7	23	1	3	14	10			5							63		63
Crater Lake	12	14	2	1	9	37			18	12	7		12	2	1	127	33	160
Lake IJssel	21	10	2	2	17	18		35	2	9				4	5	125	23	148
Lake Soyang	4	3		4	4	6	1	3	12	1	1				3	42		42
Columbia River	5	12	5		3	10			6				1		2	44	4	48
Parker River	1	11			5	1								2	3	23	1	24
Total	97	114	30	19	99	130	3	56	54	25	10	8	13	11	20	689	75	764

than the number of such studies in freshwater (Table 6). Therefore, the high proportion of sequences in the freshwater studies that find their best match in other freshwater ecosystems is remarkable. Clearly, bacteria can be transported from one habitat to another, and therefore the finding of an organism in a particular environment does not necessarily mean that it grows and thrives in that environment. However, the high number of sequences from unrelated freshwater sites clustering together phylogenetically, combined with the relatively low number of closely related sequences from other habitat types, cannot be explained by transport or coincidence, but rather suggests that the clustered 16S rDNA sequences represent species or groups of species that are indigenous to freshwater.

Freshwater clusters

To further investigate the clustering of bacterial 16S rDNA sequences, we conducted phylogenetic analyses. Many of the tested sequences were highly similar or identical to their closest matching database relatives. However, other freshwater sequences were only distantly related to their closest relatives, sometimes with a similarity as low as 85%. Pairs of sequences at such distances may group together in phylogenetic analyses due to the lack of related taxa. However, these loose groupings have little meaning and will likely not hold together when new related sequences become available. Therefore, for the purpose of this paper we defined a putative freshwater cluster as a monophyletic group of 16S rDNA sequences, containing at least 2 sequences that are at least 95% identical and originating from the pelagic of at least 2 different freshwater sites.

We sought such clusters by examining the phylogenetic position of the 373 freshwater sequences that were most similar to other freshwater sequences in BLAST analyses. Starting with these closely related freshwater sequences, we added increasingly more distant taxa irrespective of origin to investigate the contours of the phylogenetic clusters and performed BLAST analyses on all cluster members to search for closely related taxa that were not yet included. One problem in this analysis was that many sequences were partial and there was no overlap among some. To mediate this problem we performed separate phylogenetic analyses in 2 regions: approximately positions 100 to 400 and positions 600 to 900 (*Escherichia coli* numbering). The region 100 to 400 generally was most variable, allowing better differentiation of clusters, and was represented in the majority of sequences. Therefore, the clusters were mainly based on analysis of the

Table 5. Overview of origins of closest database sequences (highest basic local alignment search tool [BLAST] scores, but not originating from the same site) to the bacterial 16S rDNA sequences from the 9 freshwater studies

Origin	Isolate	Clone	Unknown	Total (%)
Water column of lakes, rivers, ponds, ditches and wells	74	299	0	373 (54)
Activated sludge, bioreactors	7	24	1	32 (5)
Cooling coils, distilled water, trickling filter, mine drainage, geothermal spring, microbial mats, lake ice	16	13	0	29 (4)
Aquifers, ground water	2	12	0	14 (2)
Freshwater sediments and flooded soils	7	21	0	28 (4)
Soil (not flooded)	9	77	1	87 (13)
Marine	8	22	0	30 (4)
Brackish	5	9	0	14 (2)
Other	22	10	1	33 (5)
Unknown	29	13	7	49 (7)
Total	179	500	10	689 (100)

100 to 400 region and sequences that did not span this region (from studies in Toolik Lake, Adirondack Mountain Lakes and the Lake Baikal study by Glöckner et al. 2000) were included if they were very closely related (>98%) in the 600 to 900 region to cluster members that overlapped both regions. Analysis of the available sequences that covered nearly the entire 16S rRNA gene for the region 100 to 1300 did not contradict phylogenetic position or coherence of clusters.

Table 7 lists 34 bacterial clusters that met our criteria and were stable in most analyses. The phylogenetic trees in Fig. 1 show the position of these clusters in the major divisions of the *Bacteria*. The clusters were named after a cultivated species that is clearly within the cluster or, if no cultivated species is included, after a clone from the cluster for which the longest sequence is available. A number of clusters overlap with clusters or groups described by Hiorns et al. (1997), Glöckner et al. (2000) or Urbach et al. (2001). However, frequently the clusters described by these authors are broader than those described here. For that reason and because we wanted to use a uniform naming system, we frequently deviated from earlier designations. Five clusters are not shown in Fig. 1. Among these are 4 cyanobacterial clusters containing isolates of which the phylogeny has been thoroughly described (Honda et al. 1999, Robertson et al. 2001). The fifth cluster not shown is cluster CL500-15, which contains 3 closely related sequences from deep layers in Lake Soyang (SY5-55) and Crater Lake (CL500-15 and CL500-100). They have 98% similarity in the overlapping region 31 to 306 (*Escherichia coli* numbering). The sequence CL500-15 was tentatively classified as a planctomycete (Urbach et al. 2001). However, it is divergent from all other planctomycetes, making its position within this division uncertain.

Fourteen of the clusters thus far consist of sequences from 2 sites only (Table 7). Twenty clusters contain sequences from 3 sites or more. Among these, 6 clusters are represented in clone libraries from at least 6 of the 12 listed sites (LD12, Fig. 1A; *Polynucleobacter necessarius*, LD28 and *Rhodoferax* sp. BAL47, Fig. 1B,C; and ACK-m1 and Sta2-30, Fig. 1F), indicating that bacteria from these phylogenetic groupings are widely distributed and thrive in diverse freshwater systems.

Based on the present database, 14 clusters contain only sequences from the water column of rivers, lakes and ponds (Table 7). Twelve clusters contain, in addition to freshwater plankton sequences, sequences from habitats that are in contact with freshwater, but no marine or soil sequences. These habitats in contact with

Table 6. Database search strategy and resulting hits. Search of Genbank using Entrez (Schuler et al. 1996) at http://www.ncbi.nlm.nih.gov/entrez/ including 'all fields' and date limit set to contain all entries until 1 August 2001. The search is indicative since information on the origin of sequences is not standardized, and frequently absent, in the database entries

Search keywords	Number of hits
16S	51 265
16S AND (freshwater OR lake OR river OR pond)	2856
16S AND marine	6611
16S AND soil	4482
16S AND (clone OR uncultured)	17 714
16S AND (clone OR uncultured) AND (freshwater OR lake OR river OR pond)	1641
16S AND (clone OR uncultured) AND marine	3695
16S AND (clone OR uncultured) AND soil	3366

Table 7. Overview of putative freshwater clusters. Freshwater contact means sequence from estuaries, sludge, flooded soils or rice paddies. Similarities within clusters are minimum sequence similarities between all pairs of sequences within the cluster, for *Escherichia coli* region 100–400

Bacterial cluster name	Division	Toolik	Adirondack	Loosdrecht	Baikal	Cadagno	Gossenkölle	Fuchskuhle	Crater	Soyang	IJssel	Columbia River	Parker River	Other freshwater	Isolates	Only freshwater	Freshwater contact	Includes marine	Includes soil	Includes clinical	Similarity within cluster (%)
LD12	α-*Proteob.*	X	X	X	X				X	X	X	X				*					>99
Brevundimonas intermedia	α-*Proteob.*				X					X				X	*		*	*			>99
CR-FL11	α-*Proteob.*									X	X					*					>99
GOBB3-C201	α-*Proteob.*								X					X	*		*				>99
Novosphingobium subarctica	α-*Proteob.*						X	X							*		*		*		>96
Polynucleobacter necessarius	β-*Proteob.*	X	X	X	X			X			X	X	X		*		*				>93
LD28	β-*Proteob.*	X	X	X	X					X	X					*					>97
GKS98	β-*Proteob.*						X	X	X	X						*					>98
Ralstonia pickettii	β-*Proteob.*	X			X					X					*		*		*	*	>95
Rhodoferax sp. Bal47	β-*Proteob.*				X		X	X	X	X	X	X			*		*	*			>95
GKS16	β-*Proteob.*								X	X						*					>99
Methylobacter psychrophilus	γ-*Proteob.*					X			X	X					*		*				>99
CL120-10	*Verruco.*								X	X						*					>97
CL0-14	*Verruco.*								X	X	X					*					>98
FukuN18	*Verruco.*		X					X		X		X					*		*		>93
Sta2-35	*Verruco.*								X	X						*					>99
LD19	*Verruco.*			X						X						*					>99
ACK-m1	*Actinob.*		X	X[a]	X				X	X	X	X	X				*	*			>91
STA2-30	*Actinob.*		X		X	X			X	X	X	X				*					>99
MED0-06	*Actinob.*				X				X	X	X	X					*		*		>94
URK0-14	*Actinob.*								X	X	X					*					>98
CL500-29	*Actinob.*				X											*					>96
LD2	CFB			X								X				*					>99
FukuN47	CFB							X	X	X						*					>93
PRD01a001B	CFB		X		X								X	X		*					>96
CL500-6	CFB								X	X						*					>97
GKS2-216	CFB		X				X		X								*		*		>94
Synechococcus 6b	*Cyanob.*			X	X					X				X	*		*		*		>95
Planktothrix agardhii	*Cyanob.*			X										X	*		*				>99
Aphanizomenon flos aquae	*Cyanob.*			X						X	X			X	*		*				>96
Microcystis	*Cyanob.*			X						X				X	*		*				>94
CL500-11	GNS								X	X						*					>98
CLO-84	OP10								X			X				*					>95
CL500-15	Planct[b]								X	X						*					>97[c]

[a]Sequence from DNA reamplified from denaturing gradient gel electrophoresis (DGGE) gel (AJ007869); [b]The divergent sequence CL500-15 has been allocated to the *Planctomycetes* (Urbach et al. 2001); [c]Similarity calculated for *E. coli* positions 31 to 306. GNS: green non-sulfur bacteria

Fig. 1. Minimum evolution trees showing the phylogenetic positions of bacterial freshwater clusters among the (A) α-*Proteobacteria*, (B,C) β-*Proteobacteria*, (D) γ-*Proteobacteria*, (E) *Verrucomicrobia*, (F) *Actinobacteria*, (G) green non-sulfur bacteria, (H) *Cytophaga-Flavobacterium-Bacteroides* (CFB) group and I candidate division OP10. Outgroup taxa for each tree (not shown) included *Escherichia coli*, *Rhodoferax fermentans*, *Bacillus subtilis*, *Leuconostoc lactis*, *Pirellula staleyi*, *Verrucomicrobium spinosum*, *Desulfovibrio desulfuricans* and *Agrococcus jenensis*. The bold-printed taxa are from sequences obtained in 1 of the 9 freshwater bacterial diversity studies of Table 1. Clone designations: Arc, Toolik Lake; ACK, Adirondack Mountain Lakes; Baikal, Lake Baikal; CL, Crater Lake; CR-FL, Columbia River (free-living fraction); CR-PA Columbia River (particle-attached fraction); CRE-FL, Columbia River estuary (free-living fraction); CRE-PA, Columbia River estuary (particle-attached fraction); CRO, Columbia River coastal outlet; Fuku, Lake Fuchskuhle; GKS, Lake Gossenkölle; LCK, Lake Cadagno; LD, Lake Loosdrecht; PRD, Parker River; SY, Lake Soyang; Sta, Med and Urk, Lake IJssel. Bold brackets indicate putative freshwater clusters. Cluster designations in parentheses indicate broader group designations from Hiorns et al. (1997), Glöckner et al. (2000) and Urbach et al. (2001) except for alphaV, betaIV, and gammaI, which exactly match clusters LD12, LD28 and *Methylobacter psychrophilus*, respectively. Origins of sequences provided in parentheses are as follows: FW, freshwater; sed., sediment; sludge, activated sludge; unk., unknown. Bootstrap values at nodes were calculated using distance (before slash) and parsimony (after slash). An asterisk indicates bootstrap values less than 50. Analyses were done using sequence stretches of 300 to 500 bases in the first third (5′) of the 16S rDNA gene

freshwater include estuaries, flooded soil, activated sludge and lake ice. From these habitats we can expect intensive exchange with the freshwater water column.

Four clusters contain clones from marine ecosystems in addition to freshwater sequences. The clusters *Rhodoferax* Bal47 (Fig. 1C) and ACK-m1 (Fig. 1F)

contain many freshwater sequences, as well as sequences of estuarine and coastal marine origin, but no sequences from the open ocean. The cluster *Brevundimonas intermedia* (Fig. 1A) contains a single sequence from the coast of one of the Palau islands, as well as sequences from aquifers and several plank-

A. α-Proteobacteria

B. β-Proteobacteria (tree 1)

C. β-Proteobacteria (tree 2)
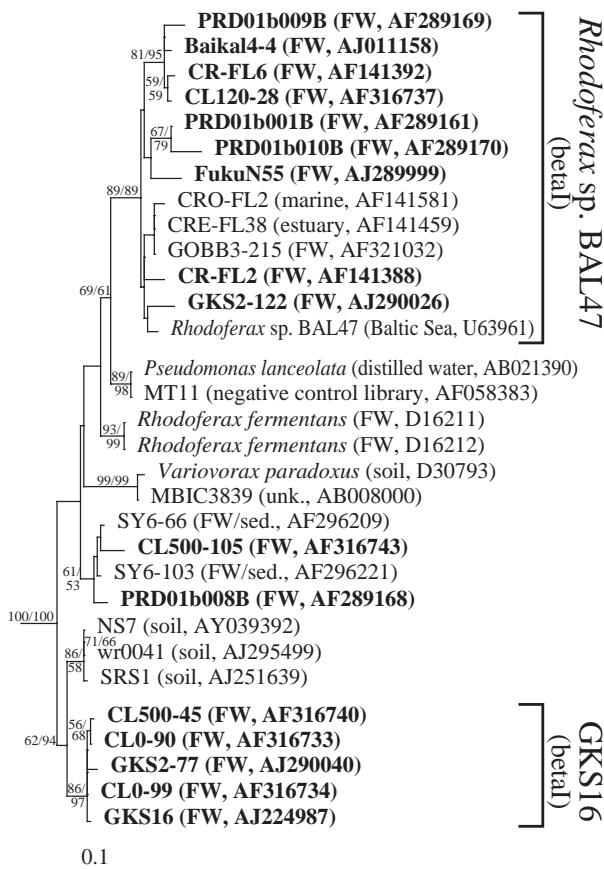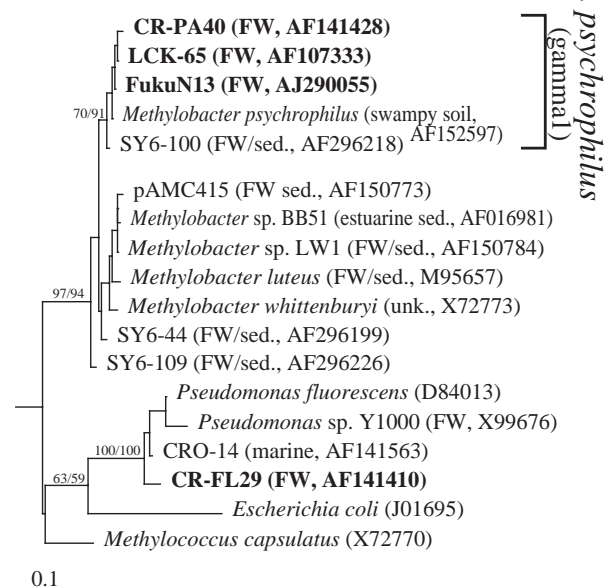

Rhodoferax sp. BAL47 (beta)

PRD01b009B (FW, AF289169)
81/95 — **Baikal4-4 (FW, AJ011158)**
59/59 — **CR-FL6 (FW, AF141392)**
— **CL120-28 (FW, AF316737)**
67/79 — **PRD01b001B (FW, AF289161)**
— **PRD01b010B (FW, AF289170)**
89/89 — **FukuN55 (FW, AJ289999)**
— CRO-FL2 (marine, AF141581)
— CRE-FL38 (estuary, AF141459)
— GOBB3-215 (FW, AF321032)
69/61 — **CR-FL2 (FW, AF141388)**
— **GKS2-122 (FW, AJ290026)**
— *Rhodoferax* sp. BAL47 (Baltic Sea, U63961)

89/98 — *Pseudomonas lanceolata* (distilled water, AB021390)
— MT11 (negative control library, AF058383)
93/99 — *Rhodoferax fermentans* (FW, D16211)
— *Rhodoferax fermentans* (FW, D16212)
99/99 — *Variovorax paradoxus* (soil, D30793)
— MBIC3839 (unk., AB008000)
— SY6-66 (FW/sed., AF296209)
— **CL500-105 (FW, AF316743)**
61/53 — SY6-103 (FW/sed., AF296221)
— **PRD01b008B (FW, AF289168)**

100/100 — NS7 (soil, AY039392)
71/66 — wr0041 (soil, AJ295499)
86/58 — SRS1 (soil, AJ251639)

GKS16 (beta)

56/68 — **CL500-45 (FW, AF316740)**
62/94 — **CL0-90 (FW, AF316733)**
— **GKS2-77 (FW, AJ290040)**
86/97 — **CL0-99 (FW, AF316734)**
— **GKS16 (FW, AJ224987)**

0.1

D. γ-Proteobacteria

M. psychrophilus (gamma)

CR-PA40 (FW, AF141428)
LCK-65 (FW, AF107333)
FukuN13 (FW, AJ290055)
70/91 — *Methylobacter psychrophilus* (swampy soil, AF152597)
— SY6-100 (FW/sed., AF296218)

— pAMC415 (FW sed., AF150773)
— *Methylobacter* sp. BB51 (estuarine sed., AF016981)
— *Methylobacter* sp. LW1 (FW/sed., AF150784)
97/94 — *Methylobacter luteus* (FW/sed., M95657)
— *Methylobacter whittenburyi* (unk., X72773)
— SY6-44 (FW/sed., AF296199)
— SY6-109 (FW/sed., AF296226)

— *Pseudomonas fluorescens* (D84013)
— *Pseudomonas* sp. Y1000 (FW, X99676)
100/100 — CRO-14 (marine, AF141563)
63/59 — **CR-FL29 (FW, AF141410)**
— *Escherichia coli* (J01695)
— *Methylococcus capsulatus* (X72770)

0.1

Fig. 1 (continued)

E. Verrucomicrobiales

FukuN18   LD19   Sta2-35

FukuN106 (FW, AJ289983)
— saf3-020 (soil, AF078286)
SY2-10 (FW, AF107505)
FukuN18 (FW, AJ289992)
CR-FL5 (FW, AF141391)
SY1-18 (FW, AF107499)

100/100 — **LD19 (FW, AF009974)**
— **Sta4-51 (FW, AJ416177)**

87/70 — 99/99 — DA101 (soil, Y07576)
— EA25 (soil, U51864)

100/100 — **CL120-84 (FW, AF316729)**
— **Sta2-35 (FW, AJ416176)**

97/94 — *Prosthecobacter dejongeii* (FW, U60012)
100/96 — *Prosthecobacter fusiformis* (FW, U60015)
— *Verrucomicrobium spinosum* (FW, X90515)

CL120-10 (CL120-10)

— **CR-FL27 (FW, AF141408)**
— **CL120-23 (FW, AF316721)**
— **CL120-135 (FW, AF316727)**
93/98 — **CL120-10 (FW, AF316720)**
— **CR-FL25 (FW, AF141406)**

97/98 — CL0-14 (CL120-10)

98/100 — CRE-FL31 (estuary, AF141455)
— **CL500-5 (FW, AF316728)**
— **CL0-14 (FW, AF316714)**
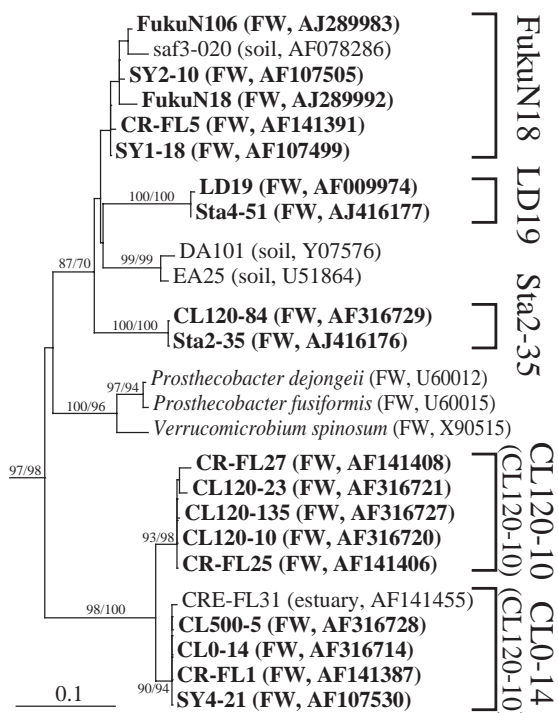— **CR-FL1 (FW, AF141387)**
90/94 — **SY4-21 (FW, AF107530)**

0.1

tonic freshwater sequences. The genus *Synechococcus* is paraphyletic, occurring in at least 3 relatedness groups (Turner 1997, Honda et al. 1999). The *Synechococcus* 6b cluster was designated thus by Robertson et al. (2001), and contains cloned sequences and isolates from coastal and open ocean sites, as well as from freshwater sites.

The *Ralstonia pickettii* cluster (Fig. 1B) contains sequences retrieved through cloning from the Columbia River and its estuary, from 2 lakes and from flooded soil. In addition, it contains isolates from the denitrifier genus *Ralstonia*, which have been obtained from plant, soil and clinical specimens. Interestingly, it also contains sequences that have been obtained from the stomach of the 'iceman', which had been immersed in glacial meltwater for a prolonged time (Cano et al. 2000).
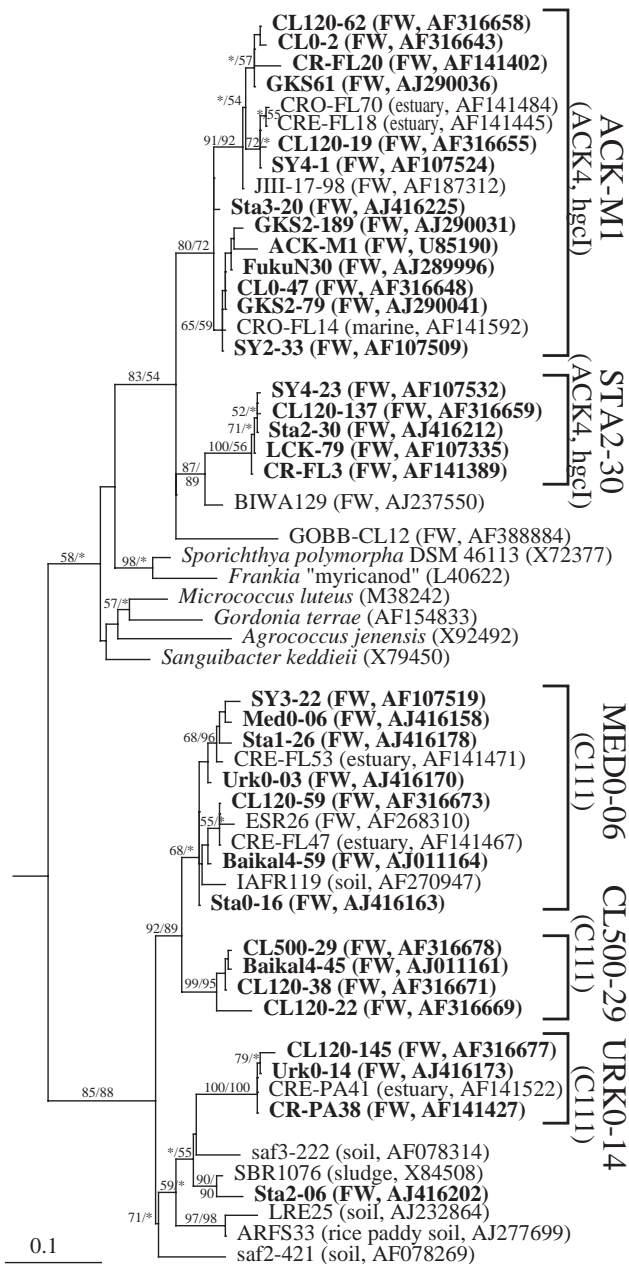
The clusters FukuN18 (Fig. 1E) and Med0-06 each contain a single sequence from soil in addition to several pelagic freshwater sequences from different sites. The cluster *Novosphingobium subarctica* (Fig. 1A) is diffuse and contains freshwater sequences, as well as soil and sludge sequences. This cluster and cluster GOBB3-C201 belong to the *Sphingomonadaceae*.

At present 11 of 34 clusters described here contain cultivated species (Table 7). This provides entry to possible function of the cluster members, although it is known that small differences at the 16S rDNA level may hide considerable differences at the genome level and phenotype. Treatment of possible physiology is outside of the scope of this paper.

F. Actinobacteria

G. Green Non-Sulfur

Fig. 1 (continued)
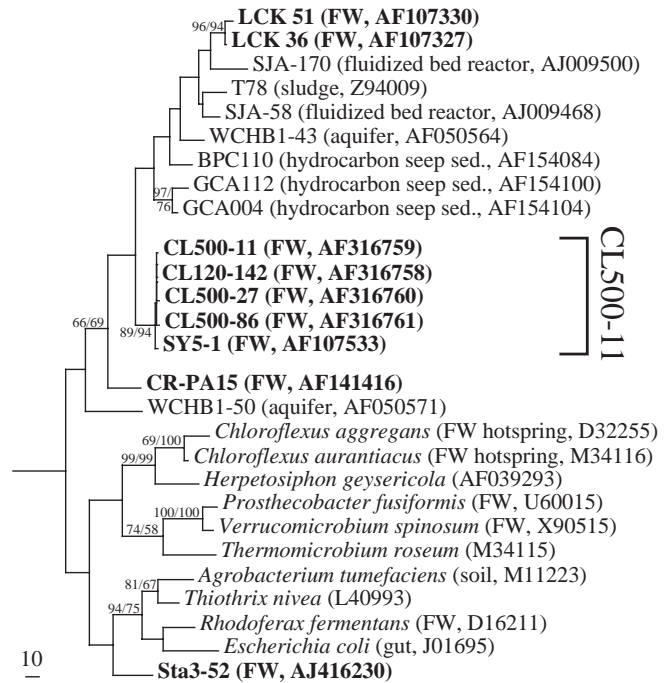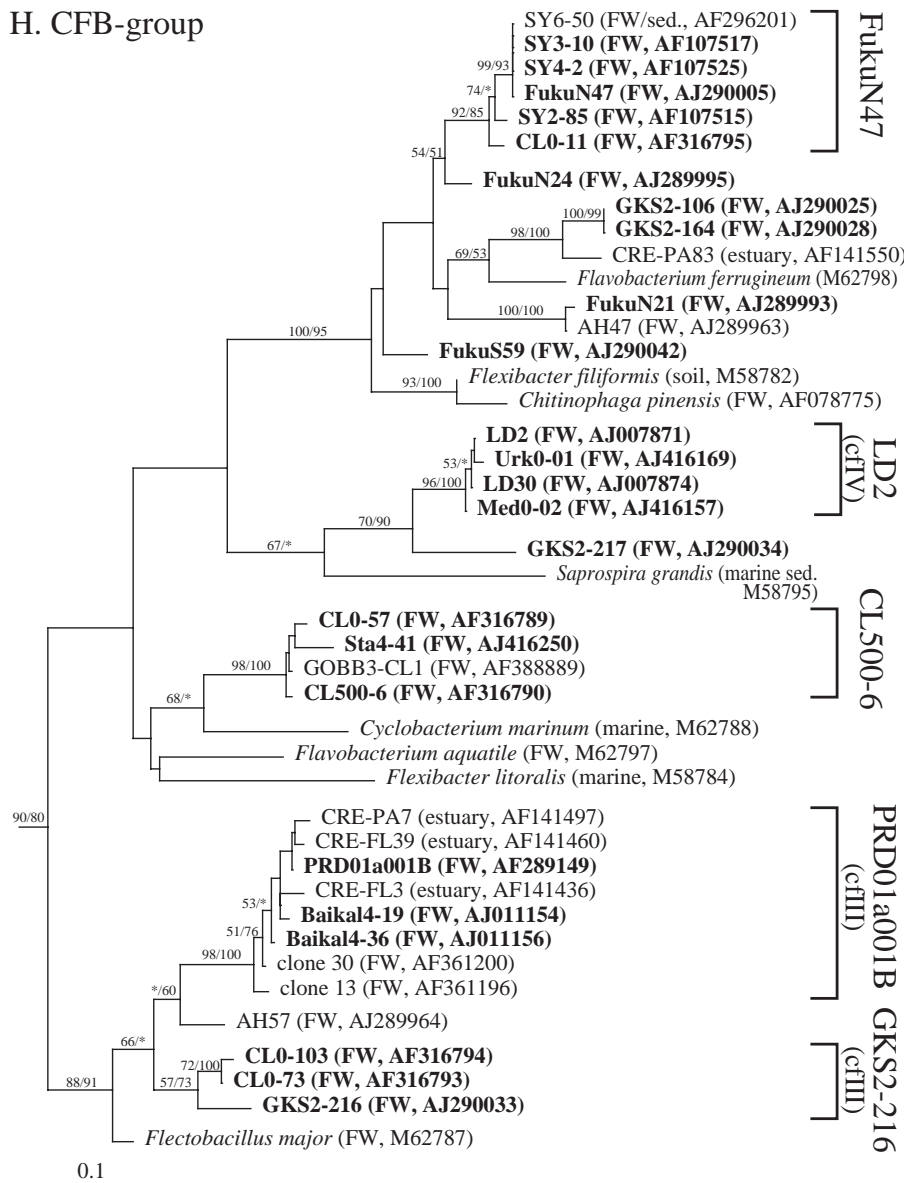
ized *Actinobacteria*. Within these large groups we distinguish 5 well-defined freshwater clusters supported by high bootstrap values. The first large group was designated ACK-4 by Hiorns et al. (1997) and hgcI by Glöckner et al. (2000). It contains clusters ACK-m1 and Sta2-30. Probe hybridization studies, directed at the larger ACK-4/hgcI group, have recently presented compelling evidence for high abundance of this group in a number of studied lakes (Glöckner et al. 2000, Urbach et al. 2001). Within-group similarity of the ACK-4/hgcI group was 83% (calculated for positions 124 to 376, *Escherichia coli* numbering). The broadness of this group and the clear split within the group justifies recognition of clusters ACK-m1 and Sta2-30, which will likely be found to have different physiological and ecological properties. Sta2-30 is a tight cluster with 99% internal similarity. The within-cluster similarity of ACK-m1 is relatively low (91%), which reflects the possibility of yet further subdivision. The second large *Actinobacteria* group was designated C111 by Urbach et al. (2001). A probe for this group did not give strong rRNA hybridization signals in Crater Lake (Urbach et al. 2001). Within this larger group C111 (similarity 85%) we distinguish freshwater clusters Urk0-14 (similarity 98%), CL500-29 (96%) and Med0-06 (94%).

At this point the indicated freshwater clusters are putative. The remarkable clustering of closely related 16S rDNA sequences from freshwater among the huge number of sequences in the global database is

The finding that 23 of 34 clusters contain no cultivated species demonstrates the usefulness of PCR-based techniques in uncovering microbial diversity in freshwater. Three bacterial divisions deserve special attention: the *Actinobacteria*, the *Verrucomicrobia* and the CFB group. Each of these divisions contributes many taxa from freshwater to the clusters described in this paper, but so far no bacteria have been cultivated of any of these clusters.

The *Actinobacteria* contain 2 large monophyletic groups that are dominated by freshwater sequences (Fig. 1F). Both groups are distant from any character-

H. CFB-group

SY6-50 (FW/sed., AF296201)
SY3-10 (FW, AF107517)
SY4-2 (FW, AF107525)
FukuN47 (FW, AJ290005)
SY2-85 (FW, AF107515)
CL0-11 (FW, AF316795)
99/93
74/*
92/85
54/51

FukuN47

FukuN24 (FW, AJ289995)

GKS2-106 (FW, AJ290025)
GKS2-164 (FW, AJ290028)
CRE-PA83 (estuary, AF141550)
Flavobacterium ferrugineum (M62798)
100/99
98/100
69/53

FukuN21 (FW, AJ289993)
AH47 (FW, AJ289963)
100/100

FukuS59 (FW, AJ290042)
Flexibacter filiformis (soil, M58782)
Chitinophaga pinensis (FW, AF078775)
93/100
100/95

LD2 (FW, AJ007871)
Urk0-01 (FW, AJ416169)
LD30 (FW, AJ007874)
Med0-02 (FW, AJ416157)
53/*
96/100

LD2 (cfIV)

70/90

GKS2-217 (FW, AJ290034)
Saprospira grandis (marine sed. M58795)
67/*

CL0-57 (FW, AF316789)
Sta4-41 (FW, AJ416250)
GOBB3-CL1 (FW, AF388889)
CL500-6 (FW, AF316790)
98/100

CL500-6

68/*

Cyclobacterium marinum (marine, M62788)
Flavobacterium aquatile (FW, M62797)
Flexibacter litoralis (marine, M58784)

90/80

CRE-PA7 (estuary, AF141497)
CRE-FL39 (estuary, AF141460)
PRD01a001B (FW, AF289149)
CRE-FL3 (estuary, AF141436)
Baikal4-19 (FW, AJ011154)
Baikal4-36 (FW, AJ011156)
clone 30 (FW, AF361200)
clone 13 (FW, AF361196)
53/*
51/76
98/100

PRD01a001B (cfIII)

AH57 (FW, AJ289964)
*/60

CL0-103 (FW, AF316794)
CL0-73 (FW, AF316793)
72/100
57/73

GKS2-216 (FW, AJ290033)

GKS2-216 (cfIII)

66/*

88/91

Flectobacillus major (FW, M62787)

0.1

I. OP10

CL500-48 (FW, AF316757)
CL0-84 (FW, AF316751)
CL500-18 (FW, AF316755)
CL0-1 (FW, AF316746)
CRE-PA63 (estuary, AF141535)
CR-PA13 (FW, AF141415)
CL120-121 (FW, AF316754)
65/*
81/97
100/98

CL0-84 (CL0-1)

96/99

WR848 (soil, AJ292921)
WR864 (soil, AJ292920)
100/100

SJA-22 (AJ009456)

OPB80 (FW hotspring, AF027089)
OPB90 (FW hotspring, AF027090)
100/100

SJA-121 (AJ009490)
SJA-171 (AJ009501)

77/92

72/54

OPB50 (FW hotspring, AF027092)

0.1

Fig. 1 (continued)

evidence that a variety of indigenous freshwater bacteria exist. However, for each individual cluster, further research is required to establish the habitat range. For mixed clusters like the *Ralstonia pickettii* cluster or the *Novosphingobium subarctica* cluster, it is feasible that the bacteria can grow in a broad range of habitats. It is also possible that different members of such clusters have different natural habitats. Alternatively, for each of the indicated clusters, it is possible that sequences have been picked up as contaminants, either through natural dispersal processes or as laboratory artifacts.

Similarities between freshwater ecosystems and evenness of freshwater bacterial communities

Clearly, the number of 16S rDNA sequences retrieved from each of the freshwater study sites is relatively small and none can be regarded as a 'complete' inventory. However, in spite of these small sample sizes, each of the studies contains many sequences that have close relatives in other freshwater ecosystems. Urbach et al. (2001) emphasized the unique character of Crater Lake's microbial community and discussed in depth the potential selective factors influencing bacterioplankton community composition. Among these factors is the absence of riverine influx, which eliminates the transport of terrestrial microorganisms and terrestrial organic matter and nutrients. Interestingly, if we disregard redundancies within the Crater Lake clone library, a large percentage of the Crater Lake sequences (65%) match best with sequences from other freshwater systems, most often with more than 95% similarity. The Crater Lake dataset shares 19 relatively tight clusters with one or more other freshwater ecosystems (Table 7).
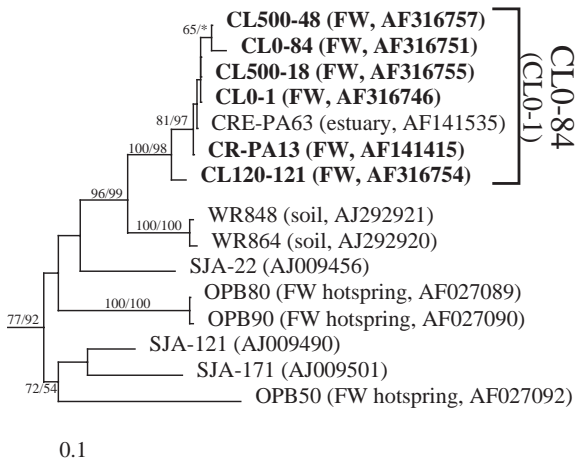
Apparently the chance of picking up a sequence type from freshwater that has already been found in another freshwater study is relatively high. This may in part be explained by biases in DNA extraction and PCR amplification. However, similar biases apply to molecular studies in soil, in which low redundancy is found in and between studies (Borneman et al. 1996, Felske et al. 1998, McCaig et al. 1999, Dunbar et al. 1999). Therefore, other explanations are needed. The relatively high redundancy in the combined freshwater studies indicates 2 important aspects of freshwater microbial communities. First, in spite of physical, chemical and biological differences, freshwater ecosystems share critical driving forces that select bacteria common to many diverse freshwater habitats. Second, freshwater bacterial communities have a relatively low evenness as compared to communities in soil. That is, a relatively small number of bacterial groups seem to dominate the freshwater plankton.

Evolutionary separation of saline and freshwater species

Bacteria and archaea that grow in the oceans and freshwater frequently show a clear evolutionary separation. This is apparent from the identification of phylogenetic clusters that are restricted to either freshwater or marine habitats, and also from the prominence of the beta-subdivision of the *Proteobacteria* in freshwater versus its near absence from the oceans. It suggests that the successful transition of microorganisms between low and high salinity environments is a rare event. Probably this transition requires multiple adaptations to maintain effective osmotic regulation and protein conformation.

A number of prokaryotic groups are restricted to aquatic habitats and show a bifurcation into a freshwater lineage and a marine lineage. These groups include the SAR11/LD12 group and cultivated strains from the Caulobacter group described by Stahl et al. (1992). Members of these groups may perform similar functions in their respective habitats, which would make them good candidates for investigation of salinity adaptation mechanisms.

There are also 4 examples of clusters that include both marine and freshwater organisms (*Brevundimonas intermedia* cluster, *Rhodoferax* sp. BAL47 cluster, ACK-m1 cluster and *Synechococcus* 6b), suggesting that transitions have occurred more frequently in these groups or that some groups can better tolerate a wide range of salt concentrations. Advances in cultivation techniques and genome sequencing will probably be required before we can further resolve the physiological constraints and evolutionary history of these clusters of microorganisms.

Conclusions

The results presented here show that the comparison of separate molecular studies can yield insights that cannot be generated from any of these studies alone. The combined data identify common bacteria across various freshwater environments. From this, it appears that microorganisms are not selected accidentally but that strong drivers structure the freshwater plankton assemblage (Nold & Zwart 1998). The water bodies studied differ in climate, dimensions, trophic status, pH, dominating phototrophs, fish stock and many other parameters, and it is very likely that they harbor microorganisms specific for their particular conditions. However, apparently these water bodies share important traits that select the species or groups of species in the identified clusters.

This paper has taken a generalized view of freshwater habitats, not considering the diversity between them. No doubt, further accumulation of sequence data will allow future differentiation of clusters to different freshwater habitats and niches within them. The identification of phylogenetic clusters of bacteria that are common in freshwater in general provides a starting point for probe design and cultivation efforts, and thus should propel directed research to ecology and functions in freshwater microbial communities.

LITERATURE CITED

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman D (1997) Gapped BLAST and PSI-blast: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microb Rev 59:143–169

Bahr M, Hobbie JE, Sogin ML (1996) Bacterial diversity in an Arctic lake: a freshwater SAR11 cluster. Aquat Microb Ecol 11:271–277

Barns SM, Takala SL, Kuske CR (1999) Wide distribution and diversity of members of the bacterial kingdom acidobacterium in the environment. Appl Environ Microbiol 65: 1731–1737

Berger C, Sweers HE (1988) The IJsselmeer and its phytoplankton—with special attention to the suitability of the

lake as a habitat for *Oscillatoria agardhii* Gom. J Plankton Res 10:579–599

Borneman JP, Skroch W, O'Sullivan KM, Palus JA, Rumjanek NG, Jansen JL, Nienhuis J, Triplett EW (1996) Molecular microbial diversity of an agricultural soil in Wisconsin. Appl Environ Microbiol 62:1935–1943

Bosshard PP, Santini Y, Grüter D, Stettler R, Bachofen R (2000) Bacterial diversity and community composition in the chemocline of the meromictic alpine Lake Cadagno as revealed by 16S rDNA analysis. FEMS Microbiol Ecol 31: 173–182

Buck JD (1979) The plate count in aquatic microbiology. In: Costerton JW, Colwell RR (eds) Native aquatic bacteria: enumeration, activity, and ecology (ASTM STP 695). American Society for Testing and Materials, West Conshohocken, p 19–28

Cano RJ, Tiefenbrunner F, Ubaldi M, Del Cueto C and 5 others (2000) Sequence analysis of bacterial DNA in the colon and stomach of the Tyrolean Iceman. Am J Phys Anthropol 112:297–309

Crump BC, Armbrust EV, Baross JA (1999) Phylogenetic analysis of particle-attached and free-living bacterial communities in the Columbia River, its estuary, and the adjacent coastal ocean. Appl Environ Microbiol 65:3192–3204

Dunbar J, Takala S, Barns SM, Davis JA, Kuske CR (1999) Levels of bacterial community diversity in four arid soils compared by cultivation and 16S rRNA gene cloning. Appl Environ Microbiol 65:1662–1669

Felske A, Wolterink A, Vanlis R, Akkermans ADL (1998) Phylogeny of the main bacterial 16S rRNA sequences in Drentse A grassland soils (the Netherlands). Appl Environ Microbiol 64:871–879

Giovannoni SJ, Britschgi TB, Moyer CL, Field KG (1990) Genetic diversity in Sargasso Sea bacterioplankton. Nature 345:60–63

Glöckner FO, Fuchs BM, Amann R (1999) Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence in situ hybridization. Appl Environ Microbiol 65:3721–3726

Glöckner FO, Zaichikov E, Belkova N, Denissova L, Pernthaler J, Pernthaler A, Amann R (2000) Comparative 16S rRNA analysis of lake bacterioplankton reveals globally distributed phylogenetic clusters including an abundant group of actinobacteria. Appl Environ Microbiol 66: 5053–5065

Gons HJ, Ebert J, Kromkamp J (1998) Optical teledetection of the vertical attenuation coefficient for downward quantum irradiance of photosynthetically available radiation in turbid inland waters. Aquat Ecol 31:299–311

Hiorns WD, Methé BA, Nierzwickibauer SA, Zehr JP (1997) Bacterial diversity in Adirondack mountain lakes as revealed by 16S rRNA gene sequences. Appl Environ Microbiol 63:2957–2960

Honda D, Yokota A, Sugiyama J (1999) Detection of seven major evolutionary lineages in cyanobacteria based on the 16S rRNA gene sequence analysis with new sequences of five marine Synechococcus strains. J Mol Evol 48:723–739

Hopkinson C, Buffam I, Hobbie J, Vallino J and 11 others (1998) Terrestrial inputs of organic matter to coastal ecosystems: an intercomparison of chemical characteristics and bioavailability. Biogeochemistry 43:211–234

Huelsenbeck JP, Crandall KA (1997) Phylogeny estimation and hypothesis testing using maximum likelihood. Annu Rev Ecol Syst 28:437–466

Hugenholtz P, Goebel BM, Pace NR (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol 180:4765–4774

Lane DJ (1991) 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M (eds) Nucleic acid techniques in bacterial systematics. Wiley, New York, p 115–175

McArthur AG, Koop BF (1999) Partial 28S rDNA sequences and the antiquity of hydrothermal vent endemic gastropods. Mol Phylogenet Evol 13:255–274

McCaig AE, Glover LA, Prosser JI (1999) Molecular analysis of bacterial community structure and diversity in unimproved and improved upland grass pastures. Appl Environ Microbiol 65:1721–1730

Methé BA, Zehr JP (1999) Diversity of bacterial communities in Adirondack lakes: do species assemblages reflect lake water chemistry? Hydrobiologia 401:77–96

Methé BA, Hiorns WD, Zehr JP (1998) Contrasts between marine and freshwater bacterial community composition: analyses of communities in Lake George and six other Adirondack lakes. Limnol Oceanogr 43:368–374

Nold SC, Zwart G (1998) Patterns and governing forces in aquatic microbial communities. Aquat Ecol 32:17–35

Nüsslein K, Tiedje JM (1998) Characterization of the dominant and rare members of a young Hawaiian soil bacterial community with small-subunit ribosomal DNA amplified from DNA fractionated on the basis of its guanine and cytosine composition. Appl Environ Microbiol 64: 1283–1289

Pace NR (1997) A molecular view of microbial diversity and the biosphere. Science 276:734–740

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818

Rheinheimer G (1980) Aquatic microbiology, 2nd edn. Wiley, New York

Riemann L, Winding A (2001) Community dynamics of free-living and particle-associated bacterial assemblages during a freshwater phytoplankton bloom. Microb Ecol 42: 274–285

Robertson BR, Tezuka N, Watanabe MM (2001) Phylogenetic analyses of Synechococcus strains (cyanobacteria) using sequences of 16S rDNA and part of the phycocyanin operon reveal multiple evolutionary lines and reflect phycobilin content. Int J Syst Evol Microbiol 51: 861–871

Rodríguez F, Oliver JF, Marín A, Medina JR (1990) The general stochastic model of nucleotide substitution. J Theor Biol 142:485–501

Saiki R, Scharf S, Faloona F, Mullis K, Horn C, Erlich H, Arnheim N (1985) Enzymatic amplification of β-globin genomic sequences and restriction site analyses of sickle cell anemia. Science 230:1350–1354

Schuler GD, Epstein JA, Ohkawa H, Kans JA (1996) Entrez: molecular biology database and retrieval system (1996). In: Doolittle RF (ed) Computer methods for macromolecular sequence analysis. Methods Enzymol 266:141–162

Semenova EA, Kuznedelov KD (1998) A study of the biodiversity of Baikal picoplankton by comparative analysis of 16S rRNA gene 5'-terminal regions. Mol Biol 32:754–760

Šimek K, Pernthaler J, Weinbauer MG, Hornák K, Dolan JR, Nedoma J, Masín M, Amann R (2001) Changes in bacterial community composition and dynamics and viral mortality rates associated with enhanced flagellate grazing in a mesoeutrophic reservoir. Appl Environ Microbiol 67: 2723–2733

Stahl DA, Key R, Flesher B, Smit J (1992) The phylogeny of marine and freshwater Caulobacters reflects their habitat. J Bacteriol 174:2193–2198

Sullivan J, Swofford DL (1997) Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. J Mamm Evol 4:77–86

Swofford DL (1999) PAUP* Phylogenetic analysis using parsimony (*and other methods), version 4. Sinauer Associates, Sunderland

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10: 512–526

Turner S (1997) Molecular systematics of oxygenic photosynthetic bacteria. In: Bhattacharya D (ed) Origins of algae and their plastids. Springer, Berlin, p 13–52

Urbach E, Vergin KL, Young L, Morse A, Larson GL, Giovan-

noni SJ (2001) Unusual bacterioplankton community structure in ultra-oligotrophic Crater Lake. Limnol Oceanogr 46:557–572

Zharkikh A (1994) Estimation of evolutionary distances between nucleotide sequences. J Mol Evol 39:315–329

Zwart G, Hiorns WD, Methé BA, Van Agterveld MP, Huismans R, Nold SC, Zehr JP, Laanbroek HJ (1998) Nearly identical 16S rRNA sequences recovered from lakes in North America and Europe indicate the existence of clades of globally distributed freshwater bacteria. Syst Appl Microbiol 21:546–556