Helen C. Rees · William D. Grant · Brian E. Jones
Shaun Heaphy

Diversity of Kenyan soda lake alkaliphiles assessed by molecular methods

Abstract DNA was extracted from water and sediment samples taken from soda lakes of the Kenyan-Tanzanian Rift Valley. DNA was also extracted from microbial enrichment cultures of sediment samples. 16S rRNA genes were amplified by the polymerase chain reaction and microbial diversity was studied using denaturing gradient gel electrophoresis (DGGE) of 16S rDNA amplicons. Cloning and sequencing of single DGGE bands showed that they usually contained mixed amplicons. Several of the amplicon sequences had high identities, up to 99%, with 16S rRNA genes of organisms previously isolated from soda lakes, while others were only distantly related, with identities as low as 82%. Phylogenetic analysis of the sequenced amplicons indicated that sequences were related to the haloarchaeal, *Bacillus*/*Clostridium*, *Rhodobacterium*/*Thioalcalovibrio*/ *Methylobacter*, and *Cytophaga*/*Flavobacterium*/ *Bacteroides* (CFB) groups and the enterobacteria/*Aeromonas*/*Vibrio* part of the γ3 subdivision of the Proteobacteria.

Keywords 16S rDNA · Alkaliphile · Denaturing gradient gel electrophoresis · Phylogenetics · Soda lake

H. C. Rees (✉) · W. D. Grant · S. Heaphy
Department of Microbiology and Immunology,
University of Leicester, University Road,
Leicester, LE1 9HN, UK
E-mail: H.C.Rees@Sheffield.ac.uk
Tel.: +44-114-2225745
Fax: +44-114-2225700

B. E. Jones
Genencor International BV, Archimedesweg 30,
2333 CN Leiden, The Netherlands

Present address: H. C. Rees
Department of Civil and Structural Engineering,
University of Sheffield, Mappin Street, Sheffield, S1 3JD, UK

Introduction

Soda lakes and soda deserts are the most stable naturally occurring alkaline environments on earth, where pH values of 10 and above are common. These environments are characterized by large amounts of sodium carbonate, or complexes of this salt, formed by evaporative concentration (Grant and Tindall 1980). Other salts, especially sodium chloride, also concentrate, leading to the formation of alkaline saline lakes. The Kenyan-Tanzanian Rift Valley contains a number of lakes of this type, which range from 5% w/v to 35% w/v (saturation) salts and have pH values of 8.5 to >11.5.

The soda lake microbial community contains alkaliphilic representatives of all the major trophic groups of bacteria and archaea. Between these groups, there is cycling of carbon, sulfur, and nitrogen under aerobic and anaerobic conditions present in the lakes. Cyanobacteria, notably *Arthrospira platensis*, and *Cyanospira rippkae*, are responsible for photosynthetic primary production in dilute lakes. There is also an unquantified contribution to primary productivity made by anoxygenic phototrophic bacteria of the genus *Ectothiorhodospira* (Jones et al. 1998). In hypersaline lakes cyanobacteria and anoxygenic phototrophs from the genus *Halorhodospira* and also *Rhodobaca bogoriensis* (Milford et al. 2000) may be responsible for primary productivity.

A high rate of primary production (10 g carbon m^{-1} day^{-1}) (Melack and Kilham 1974) supports chemoorganotrophic populations in both aerobic and anaerobic environments. The aerobic environment contains prokaryotic groups of considerable phylogenetic diversity, including haloalkaliphilic archaea (Grant et al. 1999). Chemoorganotrophic populations are biochemically very active, hydrolyzing many different polymers and producing sugars and amino acids. These may be used as substrates for the fermentation of simple compounds by anaerobic fermentors. Fatty acids produced by

anaerobes may be consumed by other groups such as the acetogenic bacteria, including *Natroniella acetigena*, *Thermosyntropha lipolytica* (Svetlichnyi et al. 1996), and *Tindallia magadiensis* (Zavarzin et al. 1999).

Organic material degraded by anaerobic digestion produces substrates for methanogens such as *Methanosalus zhilinaeae* isolated from Lake Magadi (Zhilina and Zavarzin 1994). The methane produced is oxidized by methane-oxidizers, methanotrophs, assigned to the *Methylobacter* genus, although a recently isolated methane oxidizer, AMO1, is most closely related to *Methylmicrobium pelagium* (Sorokin et al. 2000).

The sulfur cycle in these lakes utilizes sulfur and sulfate presumably generated by *Ectothiorhodospira* and *Halorhodospiria* sp. (the link between the carbon and sulfur cycles), and also aerobic sulfur-oxidizers. Sulfur-oxidizing bacteria belong to two groups, those similar to the non-lithotrophic *Halomonas deleya* and those assigned to the novel genus *Thioalcalovibrio* (Sorokin et al. 2001a, 2001b). Sulfate-reducing bacteria then complete the cycle; *Desulphonatronovibrio hydrogenovorans* and *Desulphonatrum lacustre* (Zhilina et al. 1997; Pikuta et al. 1998) have been isolated from Siberian soda lakes.

The nitrogen cycle in these lakes involves the production of ammonia by fermentative anaerobes such as *Tindallia magadii* (Kevbrin et al. 1998). Ammonia is utilized by methanotrophs and nitrifiers, producing nitrate. Nitrate, in turn, is utilized by the chemoorganotrophs, creating a link between the nitrogen and carbon cycles.

Much of this diversity has been discovered by traditional culturing and taxonomic procedures. More recent studies have looked productively at the 16S rRNA genes of isolates (Duckworth et al. 1996; Jones et al. 1994), and later restricted studies of whole populations using 16S rDNA clone libraries were performed (Grant et al. 1999), the latter revealing novel archaeal phylotypes. However, these methods are time consuming, and more rapid fingerprinting techniques have been developed. One of the most useful methods is denaturing gradient gel electrophoresis (DGGE), developed for population studies by Muyzer et al. (1993). An advantage of this technique is that DNA can be recovered from the gels to allow sequencing and identification of the population present. This makes DGGE a powerful tool for the study of the phylogenetic diversity of environmental samples.

It is evident that the full extent of microbial diversity, and therefore trophic relations, in the Kenyan-Tanzanian soda lakes is not yet known. Total microbial diversity can best be assessed by determining 16S rRNA gene sequences. This has and will continue to reveal previously undetected and unknown phylogenetic groups, some of which may be exclusive to the soda lake environment.

The research presented here has used the techniques of DGGE and 16S rRNA gene sequencing to assess microbial diversity in five East African soda lakes.

Materials and methods

Sample collection

Samples were collected from five soda lakes of the Kenyan-Tanzanian Rift Valley in January 1999. Lake Magadi, a saturated hypersaline alkaline lake situated at 1°43′-2°00′S and 36°13′-36°18′E, and four more dilute lakes, namely, Crater Lake (Lake Sonachi) (0°49′S and 36°16′E), Lake Elmenteita (0°25′S and 36°15′E), Lake Nakuru (0°23′S and 36°05′E), and Lake Bogoria (0°20′N and 36°15′E) were sampled. Water and soil/sediment samples were collected into sterile "whirlpaks" (Astell Scientific, Sidcup, UK) by pouring into the bags or by using a sterile tongue depressor to scoop up soil as described previously (Duckworth et al. 1996). The samples were refrigerated after 10 days, upon return to Leicester.

Preparation of media and enrichment cultures

Media for the enrichment of specific groups of organisms were based on a modified Horikoshi medium (mHK)—consisting, in grams per liter, of 10 g glucose, 1 g bacto-yeast extract, 0.2 g K_2HPO_4 made up to 500 ml with distilled water, and 10 g Na_2CO_3 and 40 g NaCl—made up to 500 ml with distilled water. The two solutions were mixed after autoclaving. This medium, mHK, was used as a general growth medium. Two other media were also used: mCMC, with glucose replaced with 10 g carboxymethylcellulose and mOO, with glucose replaced with 10 ml olive oil. These media were used to select for the growth of cellulase and lipase/esterase-producing organisms, respectively.

Two hundred milliliters of sterile growth media was inoculated with 1 ml of the appropriate sample as follows: mCMC and mHK were inoculated with 1 ml of a mixed water and sediment sample from Lake Nakuru and mOO was inoculated with 1 ml of a mixed water and sediment sample from Lake Elmenteita. The enrichment cultures were grown for 3–5 days at 37°C before DNA extraction was carried out.

DNA extraction and PCR amplification of 16S rRNA genes

Environmental mud, water, or sediments from each individual lake were first combined to form a mixed sample for DNA extraction for each of the five lakes. DNA was extracted using the method of Chen and Kuo (1993) with the addition of a prior lysozyme treatment with 50 µl lysozyme solution (50 mg ml^{-1} in 10 mM Tris-HCl pH 8.0, 1 mM EDTA) for 30 min at 37°C in order to degrade Gram-positive cell walls. DNA from the lakes was labeled B (Lake Bogoria), C (Crater lake), E (Lake Elmenteita), M (Lake Magadi), or N (Lake Nakuru). DNA was extracted from enrichment cultures using the method of Pitcher et al. (1989) as this gave a better yield, with around 50 µg DNA per ml starting volume. DNA from enrichments was labeled mHK (from the modified Horikoshi medium), mCMC (from the modified carboxymethylcellulose medium), or mOO (from the modified olive oil medium).

Archaeal and bacterial 16S rDNA genes were amplified in a Perkin Elmer PCR Thermocycler using "touchdown" PCR (Don et al. 1991). Briefly, this comprises an initial melting of the template DNA at 96°C for 5 min, 2 cycles of 1 min at 94°C, 1 min at 66°C, and 1 min at 72°C; the annealing temperature was then dropped by 1°C for every two cycles until the actual annealing temperature of 56°C was reached, where 10 cycles were performed with a final extension of 10 min at 72°C. The reaction mix contained the following components: 32.75 µl nanopure H_2O, 5 µl 10x reaction buffer, 4 µl of a 10 mM dNTP mix, 6 µl $MgCl_2$, 1 µl forward primer, 1 µl reverse primer, 1 µl DNA template, and 0.25 µl "Herculase" polymerase (Stratagene, La Jolla, Calif.). Two rounds of PCR were performed. In the first round the whole 16S rDNA gene

was amplified using 27Fa (*Escherichia coli* positions 8–27), 5′-TC(CT)GGTTGATCCTG(GC)CGG-3′, and rP1 (*E. coli* positions 1512–1492), 5′-ACGG(TCA)TACCTTGTTACGACTT-3′ for Archaea and 27Fb, 5′-AGAGTTTGATCCTGGCTCAG-3′, and rP1 for Bacteria (brackets represent two- or threefold redundancy). In the second round a 576-bp fragment of the 16S rDNA gene was amplified from the first-round PCR product using the primers 27Fa-GC consisting of primer 27Fa with a 40-bp GC clamp attached to the 5′ end of the primer as follows: 5′-CGCCCGCCGCGCGCGGCGGGCGGGGCGGGGGCACG GGGGG-3′, rD2 (*E. coli* positions 536–519), and 5′-G(TA)AT-TACCGCGC(GT)GCTG-3′ for Archaea, and 27Fb-GC and rD2 for Bacteria.

Denaturing gradient gel electrophoresis

PCR products were purified using Qiagen quick-spin columns (Qiagen Ltd., Crawley, UK) according to the manufacturer's instructions. The PCR products were then analyzed by DGGE on a urea-formamide gradient of 30–36% w/v denaturant for bacterial products and 20–50% w/v denaturant for archaeal products in 6% v/v acrylamide gels. The gels were run for 5 h at a constant voltage of 200 V and a constant temperature of 60°C under conditions described by Muyzer et al. (1993). Gels were stained with ethidium bromide in 0.5x TAE (stock solution 50× TAE: 242 g Tris base, 57.1 ml glacial acetic acid, 100 ml 0.5 M EDTA, pH 8.0) for 20 min and then destained in 0.5× TAE buffer for a further 20 min to remove any unbound ethidium bromide. The bands were visualized using an ultra violet transilluminator.

Gel images were taken with a Kodak DC290 zoom digital camera, and a computer-based analysis was performed using Kodak Electrophoresis Documentation and Analysis System (EDAS) 290 software. All bands in a gel lane were marked; the positions of those bands exceeding the background Gaussian absorbance were compared between lanes on the same gel. Any bands migrating to the same position in the gel were presumed to be potentially identical amplicons.

Cloning and sequencing of DGGE bands

DNA was recovered from the DGGE polyacrylamide gels by excising the bands under ultraviolet light and transferring into 40 μl Tris-HCl (pH8.0) for 48 h to allow the DNA to elute from the gel. Eluted samples were amplified using touchdown PCR as described previously, except that 10 μl DNA template was used. The resulting amplicons were cloned into pPCR-Script AmpSK (+) cloning vector (Stratagene) according to the manufacturer's instructions.

For each of the excised bands that were cloned into the pPCR-Script AmpSK (+) cloning vector, two different amplicons were sequenced. Sequencing was carried out at the University of Leicester's Protein and Nucleic Acid Chemistry Laboratory using the Perkin Elmer "BigDye" terminator chemistry and the model 377 ABI automated DNA sequencer. Sequencing was direct from the plasmid using M13 primers, the sequence of which occurred in the plasmid's multiple cloning site and surrounded the inserted amplicons: M13F 5′-GTAAAACGACGGCCAGT-3′ and M13R 5′-GGAAACAGCTATGACCATG-3′.

Phylogenetic analysis

Sequence data was edited using the Applied Biosystems multisequence editor SeqEd version 1.0.3 program before analysis with the FASTA database query program, available on the World Wide Web (http://bioweb.pasteur.fr/seqanal/interfaces/fasta.html). The edited sequences were entered into the GeneDoc program to allow alignment of the sequenced amplicons to sequences from closely related organisms found using the FASTA database search. All phylogenetic trees were constructed using the TREECON for Windows package (Van de Peer and De Wachter 1993). Bootstrap analysis (100 replications) was applied.

Nucleotide sequence accession numbers

The sequences determined in this study have been deposited in the EMBL database under accession numbers AJ517850–AJ517908.

Results

DGGE analysis of bacterial amplicons

A visual inspection of the bacterial DGGE gel did not show any striking differences among the environmental DNA samples (Fig. 1A). In every case the bacterial amplicons migrated to two major positions on the gels, with only a few bands lying outside of these. The first cluster was found at 31.5–32% denaturant and the second at 34–35% denaturant (Fig. 1A). However, the digital analysis illustrated more subtle differences (Rees 2002). Different DNA samples contained different numbers of bands, some with a unique mobility. DNA samples prepared from enrichments had a band profile different from the samples that they were derived from. For instance, the mCMC (carboxymethylcellulose enrichment) derived from a Lake Nakuru sample (N) had only one band in common (Fig. 1A). Enrichments also resulted in the appearance of new bands on the DGGE gels, indicating that the environmental samples were more complex than the DGGE analysis suggests.

DGGE analysis of archaeal amplicons

Unlike the bacterial amplicons discussed above, visual inspection of the archaeal amplicons on the DGGE gel showed a striking difference in the banding patterns among the samples. The archaeal amplicons did not form a few clusters of bands, as was the case with the bacterial amplicons (Fig. 1B, C). The 20–50% gel (Fig. 1B) contained more bands than the 30–36% gel (Fig. 1C), which were also better separated. Different environmental DNA samples showed a similar number of bands. DNA prepared from the enrichment cultures resulted in the appearance of new bands on the DGGE gels with different amplicon patterns. This again indicated a more diverse community structure than the initial DGGE results of the environmental samples would suggest.

16S rRNA gene sequencing results

The bacterial and archaeal bands chosen for sequencing were taken from the full gradient range of the DGGE gels. DNA was recovered from the gels and reamplified. Direct sequencing of these products resulted in mixed
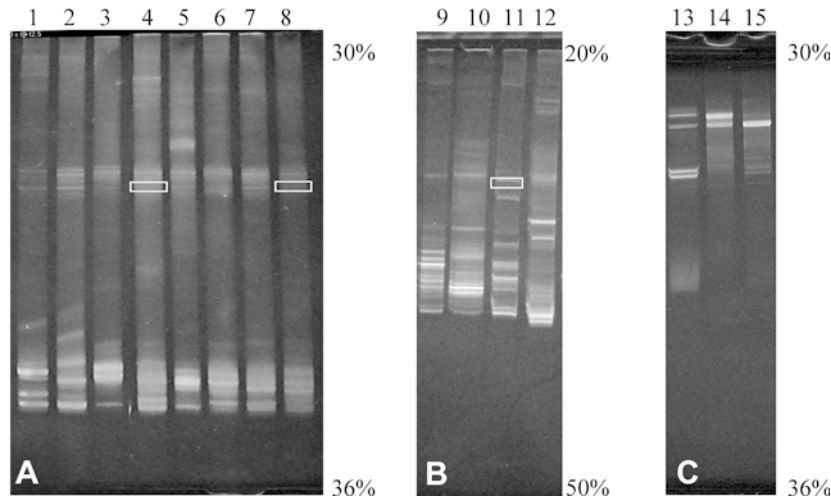
Fig. 1 A 30–36% DGGE gel showing bacterial 570 bp 16S rDNA banding patterns for lake and enrichment samples as follows: 1, modified Horikoshi enrichment; 2, modified carboxymethylcellulose enrichment; 3, modified olive oil enrichment; 4, Lake Bogoria; 5, Crater Lake; 6, Lake Elmenteita; 7, Lake Magadi; and 8, Lake Nakuru. *White boxes* mark bacterial band 34. **B** 20–50% DGGE gel showing archaeal 570 bp 16S rDNA banding patterns for lake and enrichment samples as follows: 9, modified Horikoshi enrichment; 10, modified carboxymethylcellulose enrichment; 11, modified olive oil enrichment; and 12, Lake Elmenteita. *White box* marks archaeal band 13. **C** 30–36% DGGE gel showing archaeal 570 bp 16S rDNA banding patterns for lake samples as follows: 13, Lake Bogoria; 14, Lake Elmenteita; and 15, Lake Magadi

sequences, suggesting that the bands contained different amplicons. Accordingly, the PCR products were cloned and analyzed by restriction fragment length polymorphism (RFLP). Two clones with different RFLP patterns from each band were subjected to DNA sequence determination. None of the sequenced amplicons (60) showed 100% identity to database entries.

Sequencing did show that the two amplicons from each individual band were usually different. We observed this in 80% of cases. In a few instances the two amplicons from the same band, despite being different, would share highest identity to the same bacterial or archaeal sequence. For instance, the archaeal band 13 (Fig. 1B, boxed in white) from the mOO enrichment sample contained at least two amplicons with highest identity to *Haloferax mediterranei* at 93.7% and 96.6%, respectively.

Bands that migrated to the same place on the gel did not always contain the same DNA species. In one example, the bacterial gel band 34 (Fig. 1A, boxed in white) appeared in both the Lake Bogoria (B) and Lake Nakuru (N) samples, but the four amplicon sequences that we determined had highest identity to three differing species. From Lake Bogoria they were *Natronobacterium pharaonis* (a haloarchaeal sequence) and "*Clostridium alcalibutyricum*." Both the Lake Nakuru clones had highest identity to an uncultured member of the *Bacteroides*.

As indicated above, we found bacterial amplicons with identity to archaeal sequences and vice versa. This

is probably due to mispriming during the PCR (Sommer and Tautz 1989).

Phylogenetic analysis

DNA sequence analysis allowed the determination of a section of the 16S rRNA gene in the region between positions 27 and 400 (*E. coli* numbering). Phylogenetic analysis of these amplicon sequences revealed a range of identities to several groups of bacteria and archaea (Figs. 2A–F). A total of 56 clones were sequenced, of which 43 could be placed into trees. Of the 13 clones not placed into trees, 9 had too low an identity to allow for sensible alignment and therefore were removed, and 4 had sequences of less than 320 bp and therefore were not included.

Fourteen archaeal-related amplicons (Fig. 2A) were related to the genera *Haloferax*, *Haloarcula*, and *Natronobacterium*. Many of the amplicons were closely related to *Natronobacterium pharaonis*, with 98.05–99.53% identity. On further inspection of these amplicons, all except mCMCa18.6 and Ea10.1 were found to share greater than 98% identity to each other.

Four archaeal primer amplicons were found to be related to the enterobacteria/*Aeromonas*/*Vibrio* part of the γ3 subdivision of the Proteobacteria (Fig. 2B). They had between 90% and 95% identity to a γ-proteobacterium and isolates from Lakes Bogoria, Elmenteita, and Nakuru found during previous studies on the alkaline soda lakes (Duckworth et al. 1996). These amplicons are again presumably due to mispriming events.

A further three archaeal primer amplicon sequences were related to members of the *Thioalcalovibrio* group (Fig. 2C). These sequences had 89–97% identity to an uncultured γ-proteobacterium and also relatively close identity, 89% and 95%, to *Thioalcalovibrio* species. *Thioalcalovibrio* have been isolated from alkaline soda lakes in Kenya and from the hypersaline alkaline Mono Lake in California (Sorokin et al. 2000, 2001a, 2001b). Also in this tree is a bacterial amplicon whose sequence has 87.9% identity to an uncultured

Fig. 2A–F Phylogenetic trees showing the relationship among 16S rRNA gene sequences from Kenyan soda lakes obtained in this study. The trees were constructed using the Jukes and Cantor (1969) evolutionary distance matrix and a neighbor-joining tree topology. The values indicate the percentage of occurrence in 100 bootstrapped trees, and the *scale bar* represents 0.02 nucleotide substitution. Where organisms have been deposited, the culture collection numbers are indicated. Type species are denoted by [T]. **A** Rooted phylogenetic tree of the archaeal-related amplicons (*bold*). **B** Rooted phylogenetic tree of the enterobacteria/ *Aeromonas*/*Vibrio*-related amplicons. **C** Rooted phylogenetic tree of the *Thioalkalovibrio*/ *Methylobacter*-related amplicons. **D** Rooted phylogenetic tree of the *Bacillus*/*Clostridium*-related amplicons. **E** Rooted phylogenetic tree of the *Rhodobacterium*-related amplicons. **F** Rooted phylogenetic tree of the *Cytophaga*/*Flavobacterium*/ *Bacteriodetes*-related amplicons

A

0.02 substitutions per site

B

0.02 substitutions per site
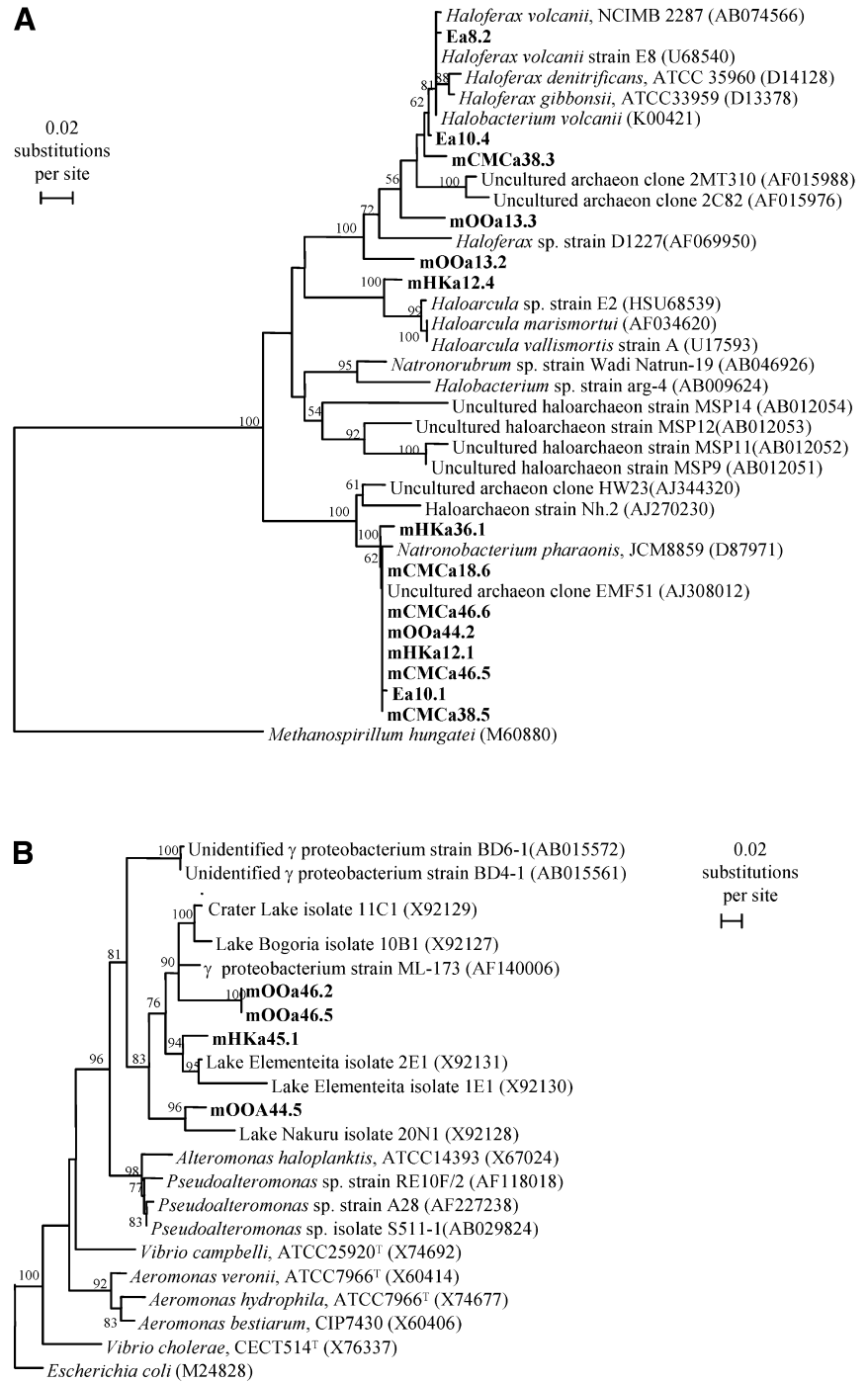
γ-Proteobacteria and lower identity, approximately 87%, to the *Thioalcalovibrio* species present. Many of the organisms present in these two branches of the trees have been isolated from soda lake environments, making the clone sequences plausible members of the soda lake community.

Several of the bacterial amplicon sequences had high identity to examples in the *Bacillus*/*Clostridium* group (Fig. 2D). Although none share highest identity to isolates from the Kenyan soda lakes, these are present within both parts of the tree, confirming that they are

also plausible members of the soda lake community. One of the amplicon sequences, Cb30.2, has only a very low identity to all of the sequences shown (81.42% identity to *Bacillus halodurans*). The low identity may mean that this particular sequence is a representative of an unknown genus.

Seven sequences (Fig. 2E) had between 91% and 96% identity to *Rhodobacter* spp. and *Rhodobaca bogoriensis*. This tree also has low bootstrap values in several places, which indicates some uncertainties in the tree structure. However, again these amplicons are

68

Fig. 2A–F (Contd.)

C

0.02
substitutions
per site

```
63 ┌ Uncultured γ proteobacterium clone ML1218J-16 (AF452605)
88 ├ Uncultured γ proteobacterium clone ML516-J12(AF452611)
57 └ Uncultured γ proteobacterium clone ML110M-3(AF452604)
65 ┌ Ea5.6
82 ├ 83 Ea5.5
63 │ Ea8.3
   └ Bacterium strain ARH1 (AF151432)
84 ┌ Thioalcalovibrio versutus strain AL2ᵀ (AF126546)
88 ├ Thioalcalovibrio nitratus strain ALJ12ᵀ (AF126547)
66 ├ Uncultured γ proteobacterium clone ML623J-18 (AF507820)
   └ Thioalcalovibrio jannaschii strain ALM2ᵀ (AF329083)
87 ┌ Phototrophic bacterium DMS2111(X93478)
   └ Thioalcalovibrio denitrificans (AF126545)
77 100 ┌ Methylobacter sp. BB5.1 (AF016981)
100 ├ Methylobacter luteus ACM3304 (X72772)
    └ Methylobacter whittenburyi ACM3310 (X727738)
100 72 ┌ Unidentified γ proteobacterium strain BD3-1 (AB015547)
86 ├ Unidentified γ proteobacterium strain BD3-6 (AB015548)
99 ├ Unidntified γ proteobacterium strain BD5-16 (AB015571)
   └ Unidentified γ proteobacterium strain BD6-6 (AB015576)
85 69 ┌ Eb26.4
   └ Uncultured γ proteobacterium clone DSS15 (AJ401384)
94 100 ┌ Marinobacter hydrocarbonoclasticus, ATCC27132ᵀ (AB021372)
    ├ Marinobacter hydrocarbonoclasticus (Y16735)
    └ Marinobacter aquaeolei strain VT8(AJ000726)
Escherichia coli (M24828)
```

D

0.02
substitutions
per site

```
53 ┌ Bacterial sp. Lake Nakuru isolate 18N1 (X92159)
   └ Bacterial sp. Lake Bogoria isolate 66B4 (X92158)
88 ┌ Bacillu sp. strain 8-1 (AB043842)
72 ├ Bacillus pseudofirmus, DSM8715 (X76439)
   ├ Bacillus sp. strain S2 (AB043860)
   └ Bacillus pseudofirmus strain FTU (AF406790)
84 ┌ mHKb35.4
100 ┌ mOOb28.5
   └ mOOb28.1
   Bacillus horikoshii, DSM8719 (X76443)
   Bacillus subtilis, NCDO1769 (X60646)
99 ┌ Bacillus alcalophilus, DSM485ᵀ (X76436)
99 ┌ Bacillus halodurans (AP001507)
   └ Bacillus sp., DSM8718 (X76442)
69 97 ┌ Halobacillus litoralis strain SL-4 (X94558)
100 ├ Sporosarcina halophila, NCIMB9251 (X62174)
   ├ Virgibacillus sp. strain YIM-kkny16 (AY121439)
100 ┌ Bacillus horti strain K13(D87035)
   └ Bacillus sp. strain 5215(AF078813)
89 63 ┌ Cb61.1
   ├ Bacillus clarkii, DSM8720 (X76444)
82 ├ Bacterial sp. Lake Nakuru isolate 17N1 (X92168)
   ├ Halophilic bacterium MBIC3303 (AB015022)
   ├ Bacterial sp. Lake Natron isolate 105NT4 (X92166)
78 ┌ Bacillus agaradhaerans, DSM8721 (X76445)
   └ Bacterial sp. Lake Elementeita isolate E31 (X92162)
61 100 ┌ Bacterial sp. Lake Natron isolate 103NT4 (X92163)
   └ Bacterial sp. Lake Elementeita isolate WE1 (X92164)
Bacillus thermodenitrificans strain NS (AF114424)
100 Cb30.2
95 ┌ Bb34.6
   └ Clostridium alcalibutyricum isolate E2SE1-C (AJ271454)
100 100 ┌ Mb69.1
   └ Nb1.1
63 65 ┌ Clostridium elementeitii isolate E23E1-B (AJ271453)
93 ├ Tindallia sp. strain APO (AF373919)
100 ┌ Clostridium aminovorans isolate B7FT-A (AJ271455)
70 ├ Tindallia magadii, DSM10318 (Y15626)
   └ Clostridium bogorii isolate B8NSI-C (AJ271457)
Escherichia coli (M24828)
```

plausible members of the soda lake community since *Rhodobaca bogoriensis,* which was isolated from Lake Bogoria, one of the Kenyan soda lakes (Milford et al. 2000), appears to be a close relative.

The remaining six amplicon sequences showed high identity to members of the CFB group (Fig. 2F). Within this group were sequences of uncultured *Bacteriodes,* and *Cytophaga,* which have recently been isolated from the hypersaline alkaline Mono Lake in California and from

hypersaline lakes in Hawaii (Humayoun et al. unpublished results; Donachie et al. unpublished results). As yet cytophagas have not been isolated from the Kenyan soda lakes, but the identity (≤ 91%) of the bacterial amplicon sequences to these uncultured organisms suggests the presence of these types within the East African lakes.

The results of the phylogenetic tree construction therefore indicate that the majority of the clones

Fig. 2A–F (Contd.)

E

0.02
substitutions
per site
⊢ ⊣

```
        ┌ Rhodobacter sp. strain TCRI 3 (AB017796)
      63│
      99├ Rhodobacter capsulatus ATCC11166 (D16428)
        └ Rhodobacter sp. strain TCRI 2 (AB017798)
      74┌─ Rhodobacter veldkampii ATCC35703 (D16421)
      87├─ Rhodobacter adriaticum, DSM2781 (D16418)
        └─ Rhodobacter sulfidophilum, DSM1374 (D16423)
   Bb12.3
        ┌─ Rhodobacter azotoformans strain KA25(D70846)
        ├─ Rhodobacter sphaeroides (X53854)
    100┌ Rhodobacter sp. strain TCRI 14(AB017799)
        └ Rhodobacter blastica, ATCC33485 (D16429)
   Bb12.1
   Mb35.1 Unidentified alpha proteobacterium strain
             JTB359 (AB015247)
                          Cb61.2
   100    100 Eb57.6
              Mb35.6
        ─ Rhodobaca bogoriensis strain LBB1(AF248638)
   mCMCb35.1
        ──── Bacillus subtilis, NCDO1769 (X60646)
```

F

0.02
substitutions
per site
⊢ ⊣

```
                                   100│ Nb34.5
                                57 │  Nb34.1
                                  ┌── Uncultured Bacteriodetes clone ML635J-56 (AF507862)
                              61│64└── Uncultured Bacteriodetes clone ML635J-40 (AF507859)
                                  100┌── Uncultured Bacteriodetes clone ML635J-44 (AF507864)
                            98│        └── Uncultured Bacteriodetes clone ML635J-8 (AF507857)
                        90│       └── Uncultured Bacteriodetes clone ML635J-1 (AF507860)
                      88│   └── Cytophaga sp. strain JTB250 (AB015264)
                        │   ┌── Cytophaga sp. strain JTB248 (AB015263)
                      67│ 100┌ Cytophaga sp. strain AA8-3 (AB017048)
                        │ 84└ Cytophaga sp. strain AA8-2 (AB017047)
                        │    └── Cytophaga sp. strain BD2-17 (AB015545)
                     100│  59  100┌ Cytophaga sp. strain JTB244 (AB015262)
                        │          └ Psychroserpens burtonensis, ACAM188 (U62913)
                        │       100┌ Uncultured bacterium clone MK22 (AF087075)
                   100│    100│    ├ Uncultured bacterium clone MK21 (AF087074)
                        │          └ Uncultured bacterium clone MK04 (AF087071)
                   100│        └── Cytophaga sp. strain JTB251 (AB015265)
                        │    53┌── Uncultured CFB group bacterium strain FL13A07 (AF446287)
                        │  53│  Mb69.6
                        │        └── Uncultured bacterium clone PHOS-HE45 (AF314423)
                        │ ── Uncultured CFB group bacterium clone ML110M-11 (AF452597)
                   90│  Nb63.2
                   88│ ── Uncultured Cytophagales clone LA7-B21N (AF513957)
                   83│ Cb30.1
                   97│ ── Uncultured Bacteriodetes clone ML602J-37 (AF507870)
                    │  74┌ Uncultured CFB group bacterium clone ML110J-4 (AF452593)
                        └ Uncultured CFB group bacterium clone ML110J-29 (AF452591)
           ──── Escherichia coli (M24828)
```

---

investigated appear to be plausible members of the soda lake community.

Discussion

The phylogenetic diversity of the Kenyan soda lakes has until very recently been limited to the study of culti-vatable microorganisms using conventional taxonomic procedures, including the analysis of 16S rRNA genes (Jones et al. 1994; Duckworth et al. 1996, reviewed in Jones et al. 1998). The Duckworth study found organisms with identity to organisms in the γ 3 division of the Proteobacteria, including *Halomonas/Deleya*, *Aeromonas/Vibrio*, and *Pseudomonas,* while other groups formed distinct groups without any close affinity to known taxa. Gram-positive isolates included examples from both the high G + C and low G + C divisions. Two of the high G + C isolates identified with the known taxon *Dietzia*, whereas others were loosely associated with *Arthrobacter* and *Terrabacter*. Low G + C isolates were associated with members of the *Bacillus* spectrum, forming at least two distinct phylogenetic groups. Archaeal types isolated from the hypersaline Lake Magadi included haloalkaliphilic organisms related to *Natronobacterium* and *Natronococcus* spp., which were originally isolated and described from this site (Tindall et al. 1984). Other species that have been isolated from the Kenyan soda lakes include *Thioalcalovibrio* and *Thioalkalimicrobium* spp., *Methylomicobium* spp.

(Sorokin et al. 2001a, 2001b, 2002), and the anoxygenic phototroph *Rhodobaca bogoriensis* from Lake Bogoria (Milford et al. 2000). Only one study involving direct recovery of phylotypes from DNA extracted from one of these lakes, Lake Magadi, has been reported prior to this work. Sequencing 16S rRNA gene clone libraries derived from DNA extracted on site from Lake Magadi brines identified novel archaeal phylotypes that were not closely related to any of the haloalkaliphilic archaea currently in culture (Grant et al. 1999).

We used DGGE to profile 16S rDNA amplicons recovered from five lake samples without cultivation. Material from the same sites was enriched to investigate the effect on the population present. Cultivation in all three cases studied resulted in a different banding pattern of 16S rDNA amplicons. Cultivation appeared to change the community structure but did not greatly alter the number of bands observed.

At first these results suggested that the lakes and enrichment samples were not as diverse as expected. The 8 samples studied each contained between 11 and 19 bacterial bands that migrated to 71 unique positions and 11–17 archaeal bands that migrated to 48 unique positions. This would be consistent with about 100 species being present, which would seem to be a large underestimate. It is generally thought that most microorganisms cannot be cultured, and a search of the literature has indicated that something like this number of species have already been cultured from the lakes. This apparent discrepancy seems to be accounted for by the limits of resolution of DGGE, at least in our hands. Bands did not correspond to single amplicons. Bands with identical mobility contained different amplicon sequences. Almost all the bands contained multiple amplicons when sequenced. We did not determine how many amplicons were actually present due to resource limitations, but many/most bands certainly contained more than two amplicons, as judged by RFLP analysis. As DGGE appeared to have a limited resolution, we carried out a terminal restriction fragment length polymorphism (t-RFLP) study of the same DNA samples (Rees 2002). Results indicated that each sample contained between 25 and 96 different amplicons, itself likely to be an underestimate.

Other workers have noted the limited capacity of DGGE band resolution. For example, when studying microbial mats and bacterial biofilms, the presence of only 10 distinguishable bands was found (Muyzer et al. 1993). In another study of bacterioplankton samples (Lindstrom 1998), between 6 and 15 bands were found.

Comparison of the 60 amplicon sequences with the database entries showed identities of between 82% and 99%. In the majority of cases, close identity to groups of organisms already known to be present in the Rift Valley soda lake environment was found, i.e., *Bacillus/Clostridium*, enterobacteria/*Aeromonas/Vibrio, Thioalcalovibrio/Methylomicobium/Rhodobacterium*, and haloarchaeal groups (Fig. 2A–F). The results shown in Fig. 2F also clearly indicate the presence of organisms related to the *Cytophaga*/Flavobacter/Bacteriodes group of microorganisms. No evidence of these has previously been found in these lakes, although they have been detected in similar lakes in California and Hawaii (sequences related to *Cytophaga*/Flavobacter/Bacteriodes have been deposited in the database, accession numbers AF452590–AF452599 and AF513955–AF513959, although these have not yet been formally published). This is an important finding, since members of this group are involved in nutrient recycling from complex carbohydrates.

Finally, it is noteworthy that none of the amplicons sequenced had 100% identity to sequences in the database. The clones that had sequence identities of over 98% to a known organism may represent the same species. Those sequences that share an identity between 88% and 98% are usually considered to be part of the same genus (Stackebrandt and Goebel 1994). On this basis, tempered with the fact that only 500 bp of a 1.5 kbp gene was sequenced, most of the PCR amplicons described here probably represent new members of known genera. In a few cases where the identity was less than 88%, the rDNA could represent new genera that may be specific to soda lakes.

References

Chen WP, Kuo TT (1993) A simple and rapid method for the preparation of gram-negative bacterial genomic DNA. Nucleic Acids Res 21:2260–2260

Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS (1991) Touchdown PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res 19:4008–4008

Duckworth AW, Grant WD, Jones BE, van Steenburgen R (1996) Phylogenetic diversity of soda lake alkaliphiles. FEMS Microbiol Ecol 9:181–191

Grant WD, Tindall BJ (1980) Isolation of alkaliphilic bacteria. In: Herbert RA, Codd GA (eds) Microbial growth and survival in extremes of environment. Academic Press, London, pp 27–36

Grant S, Grant WD, Jones BE, Kato C, Li L (1999) Novel archaeal phylotypes from an East African alkaline saltern. Extremophiles 3:139–145

Jones BE, Grant WD, Collins NC, Mwatha WE (1994) Alkaliphiles: diversity and identification. In: Priest FG, Ramos-Cormenzana A, Tindall BJ (eds) Bacterial diversity and systematics. Plenum Press, New York, pp 195–230

Jones BE, Grant WD, Duckworth AW, Owenson GG (1998) Microbial diversity of soda lakes. Extremophiles 2:191–200

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic Press, New York, pp 21–132

Kevbrin VV, Zhilina TN, Rainey FA, Zavarzin GA (1998) *Tindallia magadii* gen. nov., sp. nov.: an alkaliphilic anaerobic ammonifier from soda lake deposits. Curr Microbiol 37:94–100

Lindstrom ES (1998) Bacterioplankton community composition in a boreal forest lake. FEMS Microbiol Ecol 27:163–174

Melack JM, Kilham P (1974) Photosynthetic rates of phytoplankton in tropical African soda lakes. Hydrobiologia 81:71–85

Milford AD, Achenback LA, Jung DO, Madigan MT (2000) *Rhodobaca bogoriensis* gen. nov. and sp. nov., an alkaliphilic purple nonsulfur bacterium from African Rift Valley soda lakes. Arch Microbiol 174:18–27

Muyzer G, de Waal EC, Uitterlinden AG (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified gene coding for 16S rRNA. Appl Environ Microbiol 59:695–700

Pikuta EV, Lysenko AM, Zhilina TN (1998) Distribution of *Desulfonatronovibrio hydrogenovorans* in soda lakes of Tuva. Microbiology 66:216–221

Pitcher DG, Saunders NA, Owen RJ (1989) Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. Lett Appl Microbiol 8:151–156

Rees HC (2002) Environmental gene screening. PhD thesis, University of Leicester, Leicester, UK

Sommer R, Tautz D (1989) Minimal homology requirements for PCR primers. Nucleic Acids Res 17:6749

Sorokin DY, Jones BE, Kuenen JG (2000) An obligate methylotrophic, methane-oxidizing *Methylomicrobium* species from a highly alkaline environment. Extremophiles 4:145–155

Sorokin DY, Lysenko AM, Mityushina LL, Tourova TP, Jones BE, Rainey FA, Robertson LA, Kuenen GJ (2001a) *Thioalkalimicrobium aerophilum* gen. nov., sp nov and *Thioalkalimicrobium sibericum* sp nov., and *Thioalkalivibrio versutus* gen. nov., sp. nov., *Thioalkalivibrio nitratis* sp nov and *Thioalkalivibrio denitrificans* sp nov., novel obligately alkaliphilic and obligately chemolithoautotrophic sulfur- oxidizing bacteria from soda lakes. Intl J Syst Evol Microbiol 51:565–580

Sorokin DY, Kuenen GJ, Jetten MSM (2001b) Denitrification at extremely high pH values by the alkaliphilic, obligately chemolithoautotrophic, sulphur-oxidising bacterium *Thioalkalivibrio denitrificans* strain ALJD. Arch Microbiol 175:94–101

Sorokin DY, Gorlenko VM, Tourova TP, Tsapin A, Nealson KH, Kuenen GJ (2002) *Thioalkalimicrobium cyclicum* sp nov and *Thioalkalivibrio jannaschii* sp nov., novel species of haloalkaliphilic, obligately chemolithoautotrophic sulfur-oxidizing bacteria from hypersaline alkaline Mono Lake (California). Int J Syst Evol Microbiol 52:913–920

Stackebrandt E, Gobel UB (1994) Taxonomic note: a place for DNA-DNA reassociation and 16SrRNA sequence analysis in the present species definition in bacteriology. Int J Syst Bacteriol 44:846–849

Svetlitshnyi V, Rainey F, Weigel J (1996) *Thermosyntropha lipolytica* gen. nov., sp. nov., a lypolytic, anaerobic, alkalitolerant, thermophilic bacterium utilising short- and long-chain fatty acids in syntrophic coculture with a methanogenic archaeum. Int J Syst Bacteriol 46:1131–1137

Tindall BJ, Ross HNM, Grant WD (1984) *Natronobacterium* gen. nov. and *Natronococcus* gen. nov., two new genera of haloalkaliphilic archaebacteria. Syst Appl Microbiol 5:41–57

Van de Peer Y, De Wachter R (1993) TREECON: A software package for the construction and drawing of evolutionary trees. Comput Appl Biosci 9:177–182

Zavarzin GA, Zhilina TN, Kevbrin VV (1999) The alkaliphilic microbial community and its functional diversity. Microbiology 68:503–521

Zhilina TN, Zavarzin GA (1994) Alkaliphilic anaerobic community at Ph-10. Curr Microbiol 29:109–112

Zhilina TN, Zavarzin GA, Rainey FA, Pikuta EN, Osipov GA, Kostrikina NA (1997) *Desulfonatronovibrio hydyogenovorans* gen nov, sp nov, an alkaliphilic, sulfate-reducing bacterium. Int J Syst Bacteriol 47:144–149