

# Molecular Phylogenetic Analysis of Archaeal Intron-Containing Genes Coding for rRNA Obtained from a Deep-Subsurface Geothermal Water Pool

KEN TAKAI\* AND KOKI HORIKOSHI

*Deep-sea Microorganisms Research Group, Japan Marine Science & Technology Center, Yokosuka 237-0061, Japan*

Received 9 July 1999/Accepted 8 September 1999

**Molecular phylogenetic analysis of a naturally occurring microbial community in a deep-subsurface geothermal environment indicated that the phylogenetic diversity of the microbial population in the environment was extremely limited and that only hyperthermophilic archaeal members closely related to *Pyrobaculum* were present. All archaeal ribosomal DNA sequences contained intron-like sequences, some of which had open reading frames with repeated homing-endonuclease motifs. The sequence similarity analysis and the phylogenetic analysis of these homing endonucleases suggested the possible phylogenetic relationship among archaeal rRNA-encoded homing endonucleases.**

Since sizable populations of viable microorganisms were found in terrestrial and ocean subsurface environments, there has been increasing interest in microbial communities and diversities in deep-subsurface environments (3, 7, 13, 17, 21, 27–29, 34). Subsurface microorganisms are often viewed as important because they could play a significant role in subsurface geochemical processes, and some of these strains may have novel metabolic properties potentially useful for industrial processes, bioremediation, or biotechnology (10, 11). Hyperthermophilic or thermophilic microorganisms are likely major and important members of deep-subsurface microbial communities, considering that the temperature of subsurface environments increases with increasing depth. In fact, a number of thermophiles and hyperthermophiles within the domains of *Bacteria* and *Archaea* have been isolated from deep continental or sea oil reservoirs (14, 21, 26, 27). However, thermophilic microbial diversity in deep-subsurface environments other than oil reservoirs is poorly understood.

Here, we sought to determine the microbial diversity in a hot-subsurface biosphere. Deep-subsurface geothermal water pools are found in active volcanic areas and are often tapped by geothermal electric power plants. As the first step in comprehending thermophilic microbial diversity in such environments, molecular phylogenetic analysis based on small-subunit rRNA gene (SSU rDNA) sequencing was undertaken. In this study, a sample from a deep-subsurface geothermal water pool was obtained 1,500 m down in a production well of the Hachoubaru geothermal plant in Oita Prefecture, Japan. Approximately 100 liters of effluent hot water (96°C, but over 250°C in situ at a depth of 1,500 m) was collected from immediately beyond the end of the production well. The chemical composition of the water was described by Hirowatari et al. (15). Sixty liters of water was immediately filtered by 0.22- $\mu$ m-pore-size 47-mm-diameter cellulose acetate filters (Advantec, Tokyo, Japan), and the microbial particles were collected on the filters. The filters were washed with NET buffer (50 mM Tris-HCl [pH 8.0], 150 mM NaCl, 100 mM EDTA) twice. Nucleic acids were extracted from the microbial particles on the filters and puri-

fied according to the method of Takai and Sako (31). Approximately 10 ng of DNA was recovered from 60 liters of the water sample. Based on this DNA yield (0.17 pg/ml) and assuming an average cellular DNA content of 2 fg (2), the calculated microbial population density was  $0.8 \times 10^2$  cells/ml. This was in good agreement with the population density determined by epifluorescence microscopy direct count (approximately  $10^2$  cells/ml) with 4',6-diamidino-2-phenylindole (DAPI). No experimental contaminant (32) was detected throughout the procedure of DNA extraction and purification.

SSU rDNAs were amplified by PCR using LA *Taq* polymerase with GC buffer (TaKaRa, Kyoto, Japan). Reaction mixtures in which the concentration of each oligonucleotide primer was 0.1  $\mu$ M and that of DNA template was 0.1 ng/ $\mu$ l were prepared. Thermal cycling was performed with the GeneAmp PCR system 9600 (Perkin-Elmer, Foster City, Calif.), and the conditions were as follows: denaturation at 96°C for 20 s, annealing at 50°C for 45 s, and extension at 72°C for 120 s for a total of 30 cycles. The oligonucleotide primers Arch21F (12) and 1492R (19) were used for archaeal rDNA, and Uni515F and Uni1408R (30, 31) were used for all microbial rDNAs. As usual, PCR products approximately 1.5 and 0.9 kb in size are expected with these archaeon-specific primers and universal primers, respectively. However, products 2.9 and 2.4 kb in size, respectively, were obtained from the reactions. These rDNA products of unusual length have often been reported in characterization of hyperthermophilic archaea such as *Pyrobaculum*, *Thermoproteus*, and *Aeropyrum* (8, 16, 25) that are isolated from various hot-water environments. In view of the occurrence of introns in rDNA of hyperthermophilic archaea, the cloning and sequencing of these PCR products were carried out as described by Takai and Horikoshi (30).

The partial rDNA sequences (400 bp corresponding to positions 536 to 935 of *Escherichia coli* rDNA numbering) were analyzed with the similarity\_rank and align\_sequence from the Ribosomal Database Project (20) and the gapped-BLAST search algorithm (1, 6) to examine the microbial and archaeal populations in a deep-subsurface geothermal water pool. Sixty-eight and fifty-two clones were analyzed from the universal and archaeal primer PCR libraries, respectively. Sequence similarity analysis among the partially sequenced clones indicated that two distinct clone types (pHGPU6 and pHGPU21, 37 and 31 clones of 68 clones, respectively) were recovered from the

\* Corresponding author. Present address: Pacific Northwest National Laboratory, P.O. Box 999, Mail stop P7-50, Richland, WA 99352. Phone: (509) 373-3386. Fax: (509) 376-1321. E-mail: Ken.Takai@pnl.gov.

universal primer PCR library and that these clone types almost completely matched two distinct clone types (pHGPA1 and pHGPA13, 24 and 28 clones of 52 clones, respectively) observed in the archaeal primer PCR library. Low sequence similarity (below 65%) between the partial sequences of pHGPA1 and pHGPA13 was found. Gapped-BLAST analysis of the partial sequences of pHGPA1 and pHGPA13 revealed coexistence of highly and poorly conserved regions in their sequences. When analysis was performed on these regions, the highly conserved regions of pHGPA1 and pHGPA13 were found to be quite similar to the rDNA sequences of *Pyrobaculum* species (over 99.0%). In contrast, the poorly conserved regions had certain similarities to the sequences of rRNA introns (033-II and 061-II) of *Thermoproteus* sp. strains IC-033 and IC-061 (16). These results suggested that the clones in the universal and archaeal primer PCR libraries represented the microbial rDNAs containing intron-like sequences and that all of the rDNA sequences recovered were derived from hyperthermophilic archaeal members. In order to characterize the gene structure of rDNA clones, almost complete sequences of the representative types of rDNA clones (2,930 bp for pHGPA1 and 2,909 bp for pHGPA13) were determined.

Based on the multiple alignments with the rDNA sequences of *Pyrobaculum* members, the insertion sites of intron-like sequences were roughly assumed, and then the secondary structures around exon-intron junction sites were manually constructed according to the convention proposed by Thompson and Daniels (33) and Lykke-Andersen et al. (23). As a result, three (pHGPA1-a, -b, and -c) and four (pHGPA13-a, -b, -c, and -d) possible introns were found in the rDNA sequences of pHGPA1 and pHGPA13, respectively. All the introns except for the second intron of pHGPA13 (pHGPA13-b) had a secondary structure specific to archaeal rRNA introns, having a core structure consisting of a bulge-helix-bulge structure, a long stable stem, and a terminal loop. In terms of their core structures, these putative introns were quite similar to the rRNA introns of *Thermoproteus* sp. strains IC-033 and IC-061 (16). These results strongly suggested that the unusually long rDNA clones obtained from the deep-subsurface geothermal pool contained several intervening sequences that were closely related to archaeal rRNA introns.

Several putative introns having relatively long terminal inserts (pHGPA1-a, pHGPA1-c, pHGPA13-b, and pHGPA13-d) were found to contain a single open reading frame (ORF) each. The ORFs encoded proteins consisting of 234, 69, 160, and 108 residues in the case of pHGPA1-a, pHGPA1-c, pHGPA13-b, and pHGPA13-d, respectively. These ORFs were found to contain amino acid sequences similar to LAGLI-DADG-like stretches, which are considered to be motifs shared by intron- and intein-encoded homing endonucleases (24). In the archaeal rRNA intron-encoded homing endonucleases found so far, the LAGLI-DADG-like stretches appear to be repeated with an interval of about 60 to 90 amino acid residues (8, 9, 16, 18, 22, 25). It seems likely, therefore, that the repeated LAGLI-DADG-like stretch is a feature of a proper ORF incurring no lethal mutation or frameshift. Although the products of the ORFs in pHGPA1-a and pHGPA13-b had the repeated LAGLI-DADG-like stretches with intervals of 88 and 73 residues, respectively, pHGPA1-c and pHGPA13-d products had one LAGLI-DADG-like stretch each. A second LAGLI-DADG-like stretch was found in the products of other reading frames in the case of pHGPA1-c and pHGPA13-d. When the amino acid sequences encoded by the ORFs in pHGPA1-c and pHGPA13-d were modified by a single nucleotide deletion at the nucleotide position 206 of pHGPA1-c and in the nucleotide position 224 of

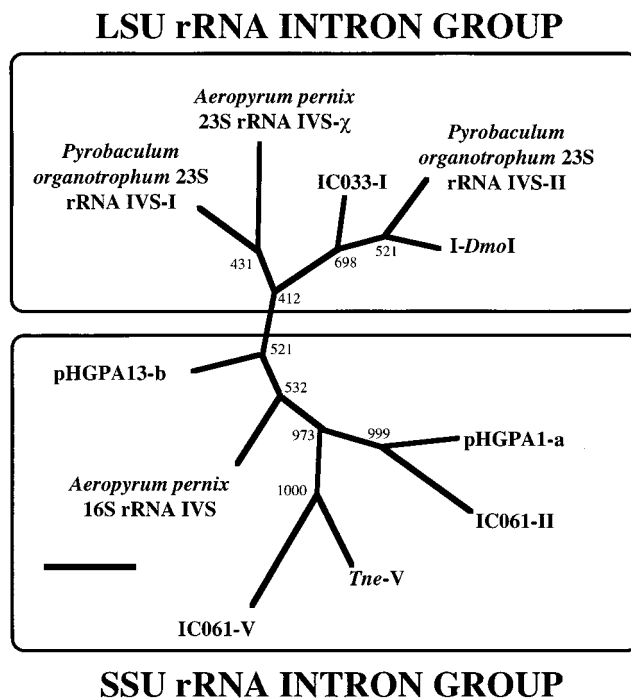


FIG. 1. Unrooted phylogenetic tree of archaeal rRNA intron-encoded homing endonucleases. The tree was inferred by applying the neighbor-joining method to 81 homologous positions of the amino acid sequence. Each number shows the bootstrap value on the branching (1,000 replicates). The scale bar indicates 0.1 substitution per residue. The amino acid sequences in this figure are based on the nucleotide sequences of archaeal rRNA introns under the following GenBank accession numbers: *Aeropyrum pernix* 16S and 23S rRNA IVSs, AB008745; IC 033-I, AB009616; IC 061-II and -V, AB009617; *TneV*, AB009618; *I-DmoI*, X03263; pHGPA1-a, AB027539; and pHGPA13-b, AB027540. The amino acid sequences of *Pyrobaculum organotrophum* 23S rRNA IVS-I and -II are from GenPept, under accession no. JC1382 and JC1383, respectively. LSU, large subunit.

pHGPA13-d, the highest sequence similarity was obtained with other archaeal homing endonucleases. These results indicated that both ORFs of pHGPA1-c and pHGPA13-d incurred mutations that shifted the reading frames, resulting in smaller size and loss of the second LAGLI-DADG-like stretch. The putative junction sites and lengths of intron-like sequences and the initiation and terminal sites and amino acid sequences of ORF products were described in the DDBJ database under accession no. AB027539 for pHGPA1 and AB027540 for pHGPA13.

Although an increasing number of homing-endonuclease-like proteins have been found to be encoded in archaeal rRNA genes, the amino acid sequence-based relatedness and the phylogenetic relationship among such homing-endonuclease-like proteins are little understood. The amino acid sequences of archaeal rRNA intron-encoded proteins, including those of pHGPA1-a and pHGPA13-b, were aligned to focus on the repeated LAGLI-DADG-like stretch. Except for the ORF in the *Pyrobaculum aerophilum* 16S rRNA intron, all of the amino acid sequences reported so far were potentially alignable. Although no identical amino acid residue was found in the multiple alignment, 16 similar amino acid residues were identified in the LAGLI-DADG-like stretches and in other parts. The result indicated that most of the archaeal rRNA intron-encoded homing endonucleases were related to each other on the primary structure level. The phylogenetic analysis also supported the relationship among these archaeal rRNA intron-encoded homing endonucleases (Fig. 1). Although the general

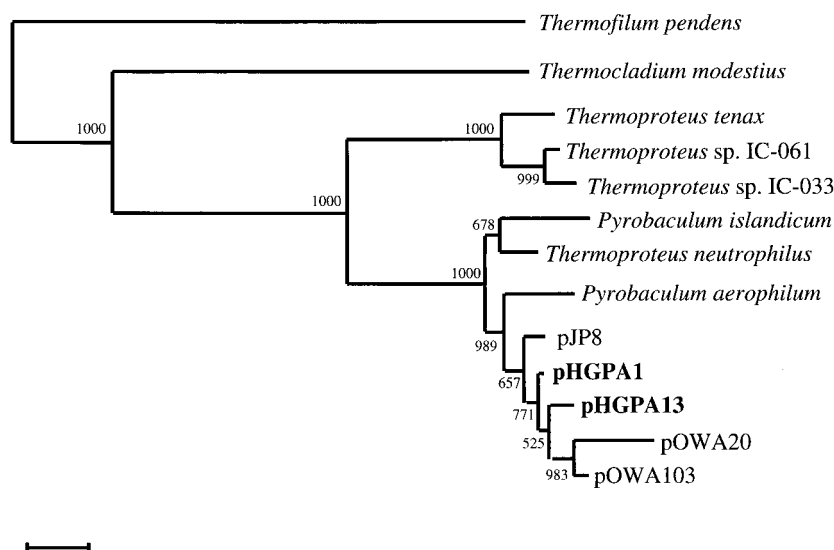


FIG. 2. Phylogenetic tree of cultivated strains and rDNA clones within *Thermoproteales* members. The tree was inferred by a neighbor-joining analysis of 414 homologous positions of the rDNA sequence in the case of each organism or clone. rDNA clones were derived from a shallow marine hydrothermal vent (pOWA) (31) and from sediment in a Yellowstone National Park hot spring (pJP) (5). Bold letters indicate rDNA clones obtained from a deep-subsurface geothermal environment. The scale bar represents 0.02 nucleotide substitution per sequence position. The number of bootstrap resamplings out of 1,000 is indicated. The rDNA sequences in this figure are from GenBank under the following accession numbers: *Thermofilum pendens*, X14835; *Thermocladium modestius*, AB005296; *Thermoproteus tenax*, M35966; *Thermoproteus* sp. strain IC-061, AB009617; *Thermoproteus* sp. strain IC-033, AB009616; *Thermoproteus neutrophilus*, AB009618; *Pyrobaculum islandicum*, L07511; *P. aerophilum*, L07510; pJP8, L25309; pOWA20, AB007314; pOWA103, AB007311; pHGPA1, AB027539; and pHGPA13, AB027540.

tree topology was not significantly supported by bootstrap examination, it revealed the presence of two distinct lineages: one group contained most of the SSU rRNA intron-encoded endonucleases, and the other group consisted of the large subunit rRNA intron-encoded entities and several SSU rRNA intron types (Fig. 1).

Using the exon regions of sequences, the phylogenetic analysis of rDNA clones obtained from a deep-subsurface geothermal environment was carried out by the neighbor-joining and maximum-likelihood methods with the ODEN software package (version 1.1; National Institute of Genetics, Mishima, Japan) and the PHYLIP package (version 3.5; obtained from J. Felsenstein, University of Washington, Seattle). Phylogenetic analysis by both methods resulted in trees with similar topologies (Fig. 2). As indicated by the extent of sequence similarity, both archaeal rDNA clones were closely related to the members of *Thermoproteales*, especially to *Pyrobaculum* species (Fig. 2). However, both rDNA clones had closer relationships with rDNA clones pJP8, obtained from the sediment of a Yellowstone National Park hot spring, and pOWA20 and pOWA103, obtained from the vent water of a shallow marine hydrothermal vent, than with the cultivated *Pyrobaculum* species (4, 5, 31). This implied that uncultivated and unidentified *Thermoproteales* members were widely distributed in terrestrial, shallow marine, and deep-subsurface hot-water environments.

The presence of a thermophilic microbial population was evident upon analysis of DNA recovered from a deep-subsurface geothermal vent water pool. All of the microbial rDNA sequences obtained from the water sample from an environment with a temperature over 250°C represented archaeal rDNA sequences closely related to those of hyperthermophilic *Thermoproteales* members. As determined by microscopic observation, all of these microbial cells were straight rods with an average length of 4 to 8  $\mu\text{m}$  and a width of about 0.5 to 0.8  $\mu\text{m}$ , which corresponded well with the morphological characteristics of *Thermoproteales* members. The chemical composition of

the hot water from the production wells was frequently analyzed by the water quality department of the power plant, and it was proved that the hot water and vapor were derived from the deep geothermal water pool (15). In addition, no rDNA sequences of mesophilic microorganisms were recovered from the sample, and no experimental contamination was detected. These results excluded possible contamination from upper layers of sediments and other contamination throughout the experimental procedure. Hence, the archaeal rDNA clone population might represent the true microbial population in the deep-subsurface geothermal environment.

To date, a great phylogenetic diversity of archaea has been found in a variety of hot-water environments, such as a Yellowstone National Park hot spring (4, 5), a shallow marine hydrothermal vent (31), and various deep-sea hydrothermal vents (30). Lower microbial population density ( $10^2$  cells/ml) and less phylogenetic diversity were observed in a deep-subsurface geothermal water pool than in other hot-water environments. Two phylotypes of archaeal members found in the deep-subsurface water pool were closely related to *Pyrobaculum* species, which are known as the most hyperthermophilic archaeal members isolated from geothermal freshwater systems (28). The extraordinarily high temperature and the confined hydrological system of the deep-subsurface geothermal pool are likely to allow only the most hyperthermophilic archaeal members and to preclude an increase in deep-subsurface microbial diversity. In addition, the geochemical condition of the deep-subsurface hot-water environment might have great impact on the deep-subsurface microbial populations. Further molecular phylogenetic analyses will be helpful in comprehending the magnitude, composition, and diversity of the microbial communities in deep-subsurface environments. The isolation and cultivation method should also help in obtaining further insight into the thermophilic microbial communities in the deep-subsurface environments. These subjects are the focus of future work.



**Nucleotide sequence accession numbers.** The sequences of the 2,930-bp and the 2,909-bp rDNA clones (pHGPA1 and pHGPA13, respectively) determined in this study have been deposited in DDBJ under accession no. AB027539 and AB027540.

We thank Wayne R. Bellamy for editing English usage in the manuscript.

This work was supported in part by a domestic research fellowship provided by the Japan Science and Technology Corporation.

#### REFERENCES

- Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389–3402.
- Bakken, L. R., and R. A. Olsen. 1989. DNA content of soil bacteria of different cell size. *Soil Biol. Biochem.* **21**:789–793.
- Bale, S. J., K. Goodman, P. A. Rochelle, J. R. Marchesi, J. C. Fry, A. J. Weightman, and R. J. Parkes. 1997. *Desulfovibrio profundus* sp. nov., a novel barophilic sulfate-reducing bacterium from deep sediment layers in the Japan Sea. *Int. J. Syst. Bacteriol.* **47**:515–521.
- Barns, S. M., C. F. Delwiche, J. D. Palmer, and N. R. Pace. 1996. Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc. Natl. Acad. Sci. USA* **93**:9188–9193.
- Barns, S. M., R. E. Fundyga, M. W. Jeffries, and N. R. Pace. 1994. Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proc. Natl. Acad. Sci. USA* **91**:1609–1613.
- Benson, D. A., M. S. Boguski, D. J. Lipman, J. Ostell, and B. F. F. Ouellette. 1998. GenBank. *Nucleic Acids Res.* **26**:1–7.
- Boone, D. R., Y. Liu, Z.-J. Zhao, D. L. Balkwill, G. R. Drake, T. O. Stevens, and H. C. Aldrich. 1995. *Bacillus infernus* sp. nov., an Fe(III)- and Mn(IV)-reducing anaerobe from the deep terrestrial subsurface. *Int. J. Syst. Bacteriol.* **45**:441–448.
- Burggraf, S., N. Larsen, C. R. Woese, and K. O. Stetter. 1993. An intron within the 16S ribosomal RNA gene of the archaeon *Pyrobaculum aerophilum*. *Proc. Natl. Acad. Sci. USA* **90**:2547–2550.
- Dalgaard, J. Z., and R. A. Garrett. 1992. Protein-coding introns from the 23S rRNA-encoding gene form stable circles in the hyperthermophilic archaeon *Pyrobaculum organotrophum*. *Gene* **121**:103–110.
- Fredrickson, J. K. 1992. DOE explores subsurface biosphere. *ASM News* **58**:183.
- Fredrickson, J. K., F. J. Brockman, D. J. Workman, S. W. Li, and T. O. Stevens. 1991. Isolation and characterization of a subsurface bacterium capable of growth on toluene, naphthalene, and other aromatic compounds. *Appl. Environ. Microbiol.* **57**:796–803.
- Fuhrman, J. A., K. McCallum, and A. A. Davis. 1992. Novel major archaeobacterial group from marine plankton. *Nature* **356**:148–149.
- Gold, T. 1992. The deep, hot biosphere. *Proc. Natl. Acad. Sci. USA* **89**:6045–6049.
- Greene, A. C., B. K. C. Patel, and A. J. Sheehy. 1997. *Deferribacter thermophilus* gen. nov., sp. nov., a novel thermophilic manganese- and iron-reducing bacterium isolated from a petroleum reservoir. *Int. J. Syst. Bacteriol.* **47**:505–509.
- Hirowatari, K., E. Nishiyama, and A. Tanaka. 1981. Research in chemical processing of water in geothermal electric power plants. *Proc. Res. Kyusyu Electric Power Co.* **56**:15–27. (In Japanese.)
- Itoh, T., K. Suzuki, and T. Nakase. 1998. Occurrence of introns in the 16S rRNA genes of members of the genus *Thermoproteus*. *Arch. Microbiol.* **170**:155–161.
- Kieft, T. L., J. K. Fredrickson, T. C. Onstott, Y. A. Gorby, H. M. Kostandiarithes, T. J. Bailey, D. W. Kennedy, S. W. Li, A. E. Plymale, C. M. Spadoni, and M. S. Gray. 1999. Dissimilatory reduction of Fe(III) and other electron acceptors by a *Thermus* isolate. *Appl. Environ. Microbiol.* **65**:1214–1221.
- Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequence. *J. Mol. Evol.* **16**:111–120.
- Kjems, J., and R. A. Garrett. 1985. An intron in the 23S ribosomal RNA gene of the archaeobacterium *Desulfofurococcus mobilis*. *Nature* **318**:675–677.
- Lane, D. J. 1985. 16S/23S sequencing, p. 115–176. In E. Stackebrandt and M. Goodfellow (ed.), *Nucleic acid techniques in bacterial systematics*. John Wiley & Sons, New York, N.Y.
- Larsen, N., G. J. Olsen, B. L. Maidak, M. J. McCaughey, R. Overbeek, T. J. Macke, T. L. Marsh, and C. R. Woese. 1993. The Ribosomal Database Project. *Nucleic Acids Res.* **21**:3021–3023.
- L'Haridon, S., A.-L. Reysenbach, P. Glatat, D. Prieur, and C. Jeanthon. 1995. Hot subterranean biosphere in a continental oil reservoir. *Nature* **377**:223–224.
- Lykke-Andersen, J., C. Aagaard, M. Semionkov, and G. A. Garrett. 1997. Archaeal introns: splicing, intercellular mobility and evolution. *Trends Biochem. Sci.* **22**:326–331.
- Lykke-Andersen, J., and R. A. Garrett. 1994. Structural characteristics of the stable RNA introns of archaeal hyperthermophiles and their splicing junctions. *J. Mol. Biol.* **243**:846–855.
- Mueller, J. E., M. Bryk, N. Loizos, and M. Belfort. 1993. Homing endonucleases, p. 111–143. In S. M. Linn, R. S. Lloyd, and R. J. Roberts (ed.), *Nucleases*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Nomura, N., Y. Sako, and A. Uchida. 1998. Molecular characterization and postsplicing fate of three introns within the single rRNA operon of the hyperthermophilic archaeon *Aeropyrum pernix* K1. *J. Bacteriol.* **180**:3635–3643.
- Ravot, G., M. Magot, M.-L. Fardeau, B. K. C. Patel, G. Prensier, A. Egan, J.-L. Garcia, and B. Ollivier. 1995. *Thermotoga elfii* sp. nov., a novel thermophilic bacterium from an African oil-producing well. *Int. J. Syst. Bacteriol.* **45**:308–314.
- Stetter, K. O., R. Huber, E. Blöchl, M. Kurr, R. D. Eden, M. Fielder, H. Cash, and I. Vance. 1993. Hyperthermophilic archaea are thriving in deep North Sea and Alaskan oil reservoirs. *Nature* **365**:743–745.
- Stevens, T. O., and J. P. McKinley. 1995. Lithoautotrophic microbial ecosystems in deep basalt aquifers. *Science* **270**:450–454.
- Takai, K., and K. Horikoshi. 1999. Genetic diversity of archaea in deep-sea hydrothermal vent environments. *Genetics* **152**:1285–1297.
- Takai, K., and Y. Sako. 1999. A molecular view of archaeal diversity in marine and terrestrial hot water environments. *FEMS Microbiol. Ecol.* **28**:177–188.
- Tanner, M. A., B. M. Goebel, M. A. Dojka, and N. R. Pace. 1998. Specific ribosomal DNA sequences from diverse environmental settings correlate with experimental contaminants. *Appl. Environ. Microbiol.* **64**:3110–3113.
- Thompson, L. D., and C. J. Daniels. 1988. A tRNA<sup>TP</sup> intron endonuclease from *Halobacterium volcanii*: unique substrate recognition properties. *J. Biol. Chem.* **263**:17951–17959.
- Whitman, W. B., D. C. Coleman, and W. J. Wiebe. 1998. Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci. USA* **95**:6578–6583.