

Takashi Itoh · Norimichi Nomura · Yoshihiko Sako

Distribution of 16S rRNA introns among the family Thermoproteaceae and their evolutionary implications

Received: 15 May 2002 / Accepted: 25 December 2002 / Published online: 8 March 2003
© Springer-Verlag 2003

Abstract Novel 16S rRNA introns were detected in four new strains within the family Thermoproteaceae. *Pyrobaculum oguniense* TE7^T and *Thermoproteus* sp. IC-062 housed introns of 32 and 665–668 bp after positions 1205 and 1213 (*Escherichia coli* numbering system), respectively. *Caldivirga maquilingsensis* IC-167^T had two introns of 37 and 140 bp after positions 901 and 908, respectively. *Vulcanisaeta distributa* IC-065 had a 691-bp intron after position 1391. All the introns larger than 650 bp encoded the LAGLI-DADG type proteins. The intron-encoded proteins of *P. oguniense* TE7^T and *Thermoproteus* sp. IC-062 are cognate with the proteins encoded by introns inserted at the same position in other *Pyrobaculum*/*Thermoproteus* strains and phylotypes. The intron-encoded protein of *V. distributa* IC-065 is partially related to that of a *Pyrobaculum* phylotype. A large-scale deletion in the second intron of *Caldivirga maquilingsensis* IC-167^T is suspected. Based on these newly found introns and hitherto known 16S rRNA introns, the evolutionary movements of the 16S rRNA introns and the encoded LAGLI-DADG type proteins are discussed.

Keywords Archaea · Hyperthermophile · Intron · LAGLI-DADG-like motif · Thermoproteaceae · 16S rRNA gene

Introduction

Within the domain Archaea, introns are detected in the genes of tRNAs, rRNAs, and proteins homologous to the eukaryotic centromere-binding factor 5 (Lykke-Andersen et al. 1997; Watanabe et al. 2002). The tRNA introns are distributed widely over the phyla Euryarchaeota and Crenarchaeota, while the introns in the protein-coding genes are found in the genome-sequenced Crenarchaeota strains (Watanabe et al. 2002). Meanwhile, the rRNA introns are so far confined to the 16S rRNA and 23S rRNA genes of the two crenarchaeotic orders Thermoproteales and Desulfurococcales (Burggraf et al. 1993; Dalgaard and Garrett 1992; Itoh et al. 1998; Kjems and Garrett 1985, 1991; Nomura et al. 1998). Furthermore, the rRNA introns occur sporadically even in these two crenarchaeotic orders.

The archaeal rRNA introns are composed of core structures and terminal loops (Lykke-Andersen and Garrett 1994). The former comprise bulge-helix-bulge structures at which the intron is spliced by RNA endonuclease. The lengths of the terminal loops are variable, ranging from several bases to more than 600 bases. The large intronic terminal loops usually carry open reading frames (ORFs) containing the LAGLI-DADG-like motif, which is one of the four conserved amino acid sequence motifs of homing endonucleases (reviewed by Belfort and Roberts 1997; Chevalier and Stoddard 2001), although some of the frames seem to have undergone frame-shift mutations (Itoh et al. 1998; Takai and Horikoshi 1999). Until now, several intron-encoded proteins have been shown to cleave the intronless alleles at the vicinities of the respective intron-insertion sites (Dalgaard et al. 1993; Lykke-Andersen and Garrett 1994; Morinaga et al. 2000). Hyperthermophilic archaea usually possess single 16S rRNA-23S rRNA operons; therefore, the presence of the “homing” endonuclease sequence in the archaeal rRNA introns may confer an infectious nature on the rRNA introns per se. Actually, the 23S rRNA intron

Communicated by K. Horikoshi

T. Itoh (✉)
Japan Collection of Microorganisms,
RIKEN (The Institute of Physical and Chemical Research),
2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan
E-mail: ito@jcm.riken.go.jp
Tel.: +81-48-4679560
Fax: +81-48-4624860

N. Nomura · Y. Sako
Division of Applied Biosciences,
Graduate School of Agriculture,
Kyoto University, Kyoto 606-8502, Japan

of *Desulfurococcus mobilis* has been shown to invade the intronless allele of *Sulfolobus acidocaldarius* (Aagaard et al. 1995). Conversely, the rRNA introns can be lost, possibly by recombination with the cDNA copy of the spliced mRNA (Belfort and Perlman 1995). The possibility of losing the ancestral rRNA introns among *Thermoproteus* strains has been suggested (Itoh et al. 1998). Thus, gain and loss of the rRNA introns are probably common events among these hyperthermophilic crenarchaeotes. Nevertheless, a paucity of the identified rRNA introns has hampered our understanding of the evolution and population dynamics of rRNA introns and the corresponding intron-encoded proteins in the natural environment.

Recently, with an increase of 16S rRNA sequences reported, new 16S rRNA introns have been found in several culturable strains and phylotypes of the family Thermoproteaceae (Itoh et al. 1999, 2002; Sako et al. 2001; Takai and Horikoshi 1999). In this paper, we compare several newly identified 16S rRNA introns and the hitherto known 16S rRNA introns and discuss the possible evolutionary movements of the 16S rRNA introns and the encoded LAGLI-DADG proteins.

Materials and methods

Strains

The 16S rRNA introns newly detected in four strains of the family Thermoproteaceae were sequenced in this study. *Thermoproteus* sp. IC-062 was isolated from hot spring water collected at Souzan-Onsen, Kanagawa, Japan, where another 16S rRNA intron-containing *Thermoproteus* sp. IC-061 was isolated (Itoh et al. 1998). *Pyrobaculum oguniense* TE7^T (JCM 10595^T) was isolated from hot spring effluent at Oguni-cho, Kumamoto, Japan (Sako et al. 2001). Strain IC-065 (JCM 11215), a strain of the new species *Vulcanisaeta distributa* designated within the novel genus *Vulcanisaeta* (Itoh et al. 2002) in the family Thermoproteaceae, was isolated from solfataric soil at Ohwakudani, Kanagawa, Japan. *Caldivirga maquilingensis* IC-167^T (JCM 10307^T) was an isolate from hot spring water at Mt. Maquiling, Laguna, the Philippines (Itoh et al. 1999).

Detection of 16S rRNA introns, sequencing, and phylogenetic analysis

Genomic DNA was isolated and purified by the method of Lauerer et al. (1986) or Tamaoka (1994). The 16S rRNA gene was amplified by PCR with the PCR primers described by Itoh et al. (1998) (for *Thermoproteus* sp. IC-062), Itoh et al. (1999) (for *C. maquilingensis* IC-167^T), Itoh et al. (2002) (for *V. distributa* IC-065), and Sako et al. (2001) (for *P. oguniense* TE7^T). Sequences of the 16S rRNA genes and cDNAs of the 16S rRNA transcripts were determined as described previously (Itoh et al. 1998; Sako et al. 2001). The phylogenetic analysis was performed with the Clustal X program (Thompson et al. 1997) and the dot matrix analyses were performed with GeneWorks (IntelliGenetics Inc.). The phylogenetic tree was reconstructed by the neighbor-joining method of Saitou and Nei (1987) and was estimated by bootstrap sampling (Felsenstein 1985).

Results and discussion

Phylogenetic relationships of the family Thermoproteaceae

A phylogenetic tree of the family Thermoproteaceae derived from the 16S rRNA exon sequences is shown in Fig. 1. Strain IC-062 was identified as *Thermoproteus* sp. by cell shape (rod-shaped), growth temperature (up to 95 °C), DNA base composition (56.5 mol% G + C), and a 16S rRNA exon sequence identical to that of strain IC-061.

Members of the genus *Pyrobaculum*, together with *Thermoproteus neutrophilus* JCM 9278^T and the phylotypes pHGPA1, pHGPA13, and pBA2, formed a coherent clade with more than 98.1% sequence similarities to each other and were closely related to the *Thermoproteus* strains ($\geq 96.5\%$ sequence similarities). Strains of

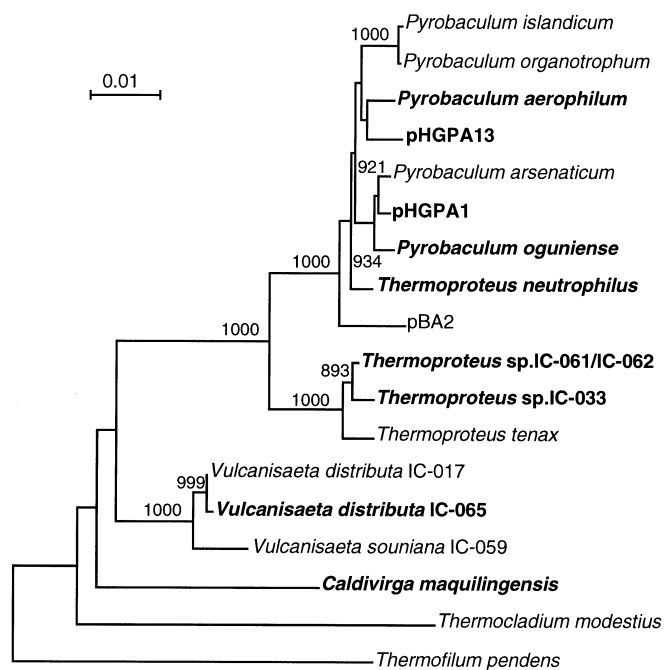


Fig. 1 A phylogenetic tree of the family Thermoproteaceae derived from the 16S rRNA exon sequences. Strains and phylotypes having the 16S rRNA introns are shown in *bold* letters. The numbers indicate bootstrap values of 1,000 trials (only values greater than 800 are shown). The bar represents 0.01 substitutions per nucleotide position. The sequences cited are as follows: *Caldivirga maquilingensis* IC-167^T, AB013926; *Pyrobaculum aerophilum* im2^T, L07510; *Pyrobaculum arsenaticum* PZ6^T, AJ277124; *Pyrobaculum islandicum* geo3^T, L07511; *Pyrobaculum oguniense* TE7^T, AB029339; *Pyrobaculum organotrophum* JCM 9190^T, AB063647; *Thermocodium modestius* IC-125^T, AB005296; *Thermofilum pendens* Hvv3^T, X14835; *Thermoproteus neutrophilus* JCM 9278^T, AB009616; *Thermoproteus tenax*, M35966; *Thermoproteus* sp. IC-033, AB009616; *Thermoproteus* sp. IC-061, AB009617; *Thermoproteus* sp. IC-062, AB081846; *Vulcanisaeta distributa* IC-017^T, AB063630; *Vulcanisaeta distributa* IC-065, AB063639; *Vulcanisaeta souniana* IC-059^T, AB063646; pBA2, AF176346; pHGPA1, AB027539; pHGPA13, AB027540

the genera *Vulcanisaeta* and *Caldivirga* were positioned in separate lineages.

Detection of 16S rRNA introns and its core structure

After PCR amplification of the 16S rRNA genes from genomic DNAs, agarose gel electrophoresis of the reaction mixtures revealed that the amplified DNAs of *Pyrobaculum oguniense* TE7^T, *Thermoproteus* sp. IC-062, *Vulcanisaeta distributa* IC-065, and *Caldivirga maquilingsense* IC-167^T were larger than the normal 16S rRNA genes of Thermoproteaceae strains. Sequence analysis showed that these amplified 16S rRNA genes contained intervening sequences as shown in Table 1. All the inserted sequences possessed the putative intron core structures that exist in all archaeal rRNA introns so far discovered (Lykke-Andersen and Garrett 1994; Itoh et al. 1998; Nomura et al. 1998; Takai and Horikoshi 1999). Moreover, for *P. oguniense* TE7^T, *V. distributa* IC-065, and *C. maquilingsensis* IC-167^T (IC-062 was not examined), the inserted sequences were absent in the cDNAs of the corresponding 16S rRNA transcripts. Thus, the intervening sequences were identified as introns.

Introns found in the genera *Thermoproteus* and *Pyrobaculum*

P. oguniense TE7^T and *Thermoproteus* sp. IC-062 had two 16S rRNA introns, after positions 1205 and 1213, that have been detected in several other strains and phylotypes of the genera *Thermoproteus* and *Pyrobaculum*, as shown in Table 1. The introns after position 1205 range from 32 to 34 bases in length, and the sequences of introns 061-IV, 062-IV, Pog-IV, pHGPA1-b, and pHGPA13-c are identical or almost identical (only

one base difference exists in pHGPA1-b) (see Table 1 for nomenclature of the introns).

The introns after position 1213 range from 662 to 688 bases in length, and the whole intron sequences can be aligned. Among these large introns, 061-V, 062-V, and Tne-V possess ORFs containing the two LAGLI-DADG motifs and occupying almost the whole region of the terminal insert. Homologous ORFs that are apparently shortened by putative occurrence of insertion and deletion in the nucleotide sequences are found in the remaining introns (033-V, Pog-V, pHGPA1-c, and pHGPA13-d). By comparing the nucleotide sequences altogether, reconstruction of the encoded proteins in the four introns (i.e., 033-V, Pog-V, pHGPA1-c, and pHGPA13-d) is theoretically possible. As shown in Fig. 2, the evolutionary relationship based on 190 amino acid positions of the proteins (including the reconstructed proteins) encoded by the 16S rRNA introns inserted after position 1213 agreed well with the phylogenetic tree based on the 16S rRNA exons, with the exception of the Tne-V-encoded protein. The Tne-V-encoded protein is distantly related to the remaining proteins. These findings may indicate that the common ancestor of the genera *Pyrobaculum* and *Thermoproteus* had the LAGLI-DADG ORFs in the introns after position 1213 and that the Tne-encoded protein had substituted the original protein. This interpretation, however, needs to be confirmed by identifying the cognate proteins from more new isolates, particularly strains living in the vicinity of Iceland where *T. neutrophilus* was isolated.

Among the members having introns after position 1213, *Thermoproteus* spp. IC-033 and IC-061 as well as pHGPA1 possess another 16S rRNA intron after position 781 (i.e., 033-II, 061-II, and pHGPA1-a, respectively), which encodes another LAGLI-DADG protein (the ORF of 033-II seems to have been derived from frame-shift mutations; Itoh et al. 1998). Interestingly,

Table 1 Occurrence of 16S rRNA introns in strains and phylotypes within the family Thermoproteaceae. Positions of introns 062-V, Vdi, Cma-I, and Cma-II are deduced from the cleavage sites of the RNA endonucleases (Kjems and Garrett 1991)

Strain or phylotype	Intron name (position in 16S rRNA gene, length in nucleotide)	Reference
<i>Pyrobaculum aerophilum</i> im2 ^T	Pae (373, 713)	Burggraf et al. 1993
<i>Pyrobaculum oguniense</i> TE7 ^T	Pog-IV (1205, 32), Pog-V (1213, 665)	This study
pHGPA1	pHGPA1-a (781, 763), pHGPA1-b (1205, 32), pHGPA1-c (1213, 665)	Takai and Horikoshi 1999
pHGPA13	pHGPA13-a (781, 96), pHGPA13-b1 ^a (901, 34), pHGPA13-b2 ^a (908, 614), pHGPA13-c (1205, 32), pHGPA13-d (1213, 665)	Takai and Horikoshi 1999
<i>Thermoproteus neutrophilus</i> JCM 9278 ^T	Tne-IV (1205, 34), Tne-V (1213, 663)	Itoh et al. 1998
<i>Thermoproteus</i> sp. IC-033	033-I (548, 627), 033-II (781, 762), 033-III (1092, 636), 033-IV (1205, 33), 033-V (1213, 682)	Itoh et al. 1998
<i>Thermoproteus</i> sp. IC-061	061-II (781, 764), 061-IV (1205, 32), 061-V (1213, 688)	Itoh et al. 1998
<i>Thermoproteus</i> sp. IC-062	062-IV (1205, 32), 062-V (1213, 688)	This study
<i>Vulcanisaeta distributa</i> IC-065	Vdi (1391, 691)	This study
<i>Caldivirga maquilingsensis</i> IC-167 ^T	Cma-I (901, 37), Cma-II (908, 140)	This study

^a The two introns pHGPA13-b1 and pHGPA13-b2 are not identified in Takai and Horikoshi 1999

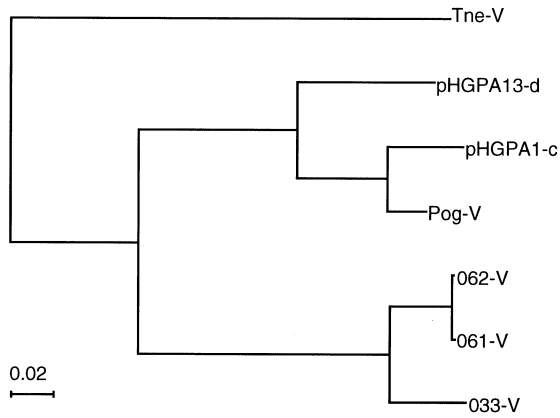


Fig. 2 A phylogenetic tree of the LAGLI-DADG proteins encoded by the 16S rRNA introns inserted after position 1213 (in *Escherichia coli* numbering system). Reconstructed amino acid sequences were used for the putative encoded proteins of 033-V, Pog-V, pHGPA1-c, and pHGPA13-d (see text). The bar represents 0.02 substitutions per amino acid position. See Table 1 for nomenclature of the introns

the evolutionary distances of the two encoded proteins of *Thermoproteus* sp. IC-061 and pHGPA 1 are quite similar: the amino acid sequence similarities between introns 061-II and pHGPA1-a and between introns 061-V and pHGPA1-c were 31% and 28%, respectively. These introns might have evolved at similar evolutionary rates in the 16S rRNA genes.

Unlike strain IC-061, strain IC-062 lacks the intron after position 781 in the 16S rDNA. Both strains were isolated from the same sampling site and have identical 16S rRNA exon sequences. Moreover, a close relative of these strains, *Thermoproteus* sp. IC-033, also possesses an intron after position 781 (Itoh et al. 1998). These facts may suggest that strain IC-062 had lost the intron after position 781.

Introns found in *C. maquilingensis* IC-167^T

In the 16S rDNA of *C. maquilingensis* IC-167^T, two introns, Cma-I and Cma-II, exist in close proximity (Fig. 3). The insertion sites of the two introns appear to be the same as those of pHGPA13-b1 and pHGPA13-b2, respectively. Moreover, the Cma-II intron has the same insertion site as the 16S rRNA intron of *Aeropyrum pernix* K1^T (ApeI α) (Nomura et al. 1998). The bulge-helix-bulge structures and the long stems, particularly the regions adjacent to the bulge-helix-bulge structures, of the two introns of *C. maquilingensis* IC-167^T show high similarities with the counterpart introns of pHGPA13. However, the long stem of Cma-II differs markedly from that of ApeI α . The putative terminal loop of Cma-II (98 bp) is apparently shorter than that of pHGPA13-b2 (571 bp) and ApeI α (653 bp). Within the putative terminal insert of Cma-II, CT-rich stretches resembling an archaeal transcriptional terminator were detected. Such CT-rich sequences are often found near the stop codons

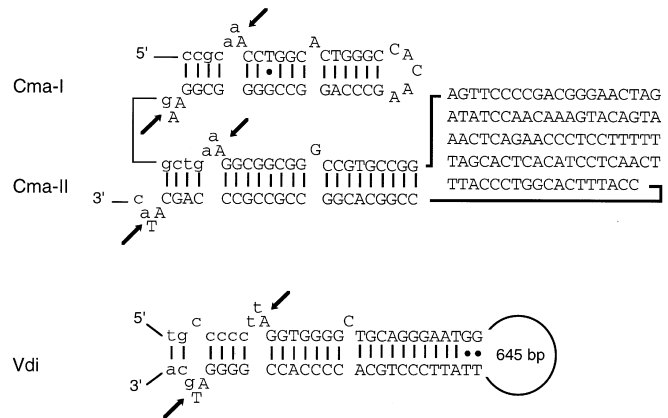


Fig. 3 Putative structures of the intron cores of Cma-I, Cma-II (upper), and Vdi (lower). Exon sequences are shown in lowercase letters. Arrows denote the putative cleavage sites

of the archaeal rRNA intron ORFs (Itoh et al. 1998). Furthermore, DNA–DNA dot matrix analysis revealed a certain degree of similarity between the terminal insert of Cma-II and the downstream region of pHGPA13-b2 encoding the LAGLI-DADG protein (e.g., the stretch from the 37th to 78th nucleotides of Cma-II and that from the 534th to 575th nucleotides of pHGPA13-b share 69% identity without gaps). This fact implies that the terminal loop sequence of Cma-II may be a remnant of the nucleotide sequence encoding a protein that shared a common trait with the pHGPA13-b2-encoded protein. No significant homology was found between the terminal inserts of Cma-II (or pHGPA13-b2) and ApeI α . At the moment, it is not clear whether these proteins originated from the same ancestral protein.

Intron found in *V. distributa* IC-065

The 16S rRNA intron (Vdi) found in *V. distributa* strain IC-065 intervenes at a hitherto unknown intron-insertion position. The intron has a typical core consisting of a bulge-helix-bulge structure and a long stable stem, as shown in Fig. 3. The long terminal insert contains a single ORF, corresponding to 203 amino acid residues, which spans almost the entire insert. The G + C content of the insert is 35.09 mol%, which is significantly lower than the 16S rRNA gene exon. The putative encoded protein had two LAGLI-DADG-like stretches (LMATGVALEG and VLRWAFTLEG). Moreover, AT-rich and CT-rich sequences are found 23–30 bp upstream of the putative start codon of the ORF and around the stop codon of the ORF, respectively. The features described above are consistent with most of the archaeal rRNA introns containing large ORFs. However, protein–protein dot matrix analysis showed no significant similarities between the Vdi-encoded protein and other LAGLI-DADG proteins, except for the protein sequence encoded by 061-II and pHGPA1-a downstream of the second LAGLI-DADG motifs (e.g.,

the 149th to 172nd amino acid stretch of the Vdi-encoded protein and the 166th to 189th amino acid stretch of the 061-II- and pHGPA1-a-encoded proteins are strictly or strongly conserved, according to the definition of Thompson et al. 1994, in 50% of the residues). This fact may suggest that the Vdi-encoded LAGLI-DADG protein (or those encoded by 061-II and pHGPA1-a) is chimeric in origin.

In the present study, comparison of the 16S rRNA introns and the intron-encoded LAGLI-DADG proteins of strains within the family Thermoproteaceae permits a glimpse of the evolutionary movements of these introns. The introns acquired in the rRNA genes could be coevolved with the rRNA exons, or they could be eliminated from the genes or undergo mutations. Some introns could propagate in another host organism. Thus, the evolutionary or population dynamics of the rRNA introns in the natural environment should be estimated statistically with sizable examples. Direct detection and sequencing of the rRNA intron-containing genes from geographically different geothermal environments, as suggested by Takai and Horikoshi (1999), could be the next strategy for this purpose.

References

- Aagaard C, Dalgaard JZ, Garrett RA (1995) Intercellular mobility and homing of an archaeal rDNA intron confers a selective advantage over intron⁻ cells of *Sulfolobus acidocaldarius*. Proc Natl Acad Sci USA 92:12285–12289
- Belfort M, Perlman PS (1995) Mechanisms of intron mobility. J Biol Chem 270:30237–30240
- Belfort M, Roberts RJ (1997) Homing endonucleases: keeping the house in order. Nucleic Acid Res 25:3379–3388
- Burggraf S, Larsen N, Woese CR, Stetter KO (1993) An intron within the 16S ribosomal RNA gene of the archaeon *Pyrobaculum aerophilum*. Proc Natl Acad Sci USA 90:2457–2550
- Chevalier BS, Stoddard BL (2001) Homing endonucleases: structure and functional insight into the catalysts of intron/intein mobility. Nucleic Acids Res 29:3757–3774
- Dalgaard JZ, Garrett RA (1992) Protein-coding introns from the 23S rRNA encoding gene form stable circles in the hyperthermophilic archaeon *Pyrobaculum organotrophum*. Gene 121:103–110
- Dalgaard JZ, Garrett RA, Belfort M (1993) A site specific endonuclease encoded by a typical archaeal intron. Proc Natl Acad Sci USA 90:5414–5417
- Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791
- Itoh T, Suzuki K, Nakase T (1998) Occurrence of introns in the 16S rRNA genes of members of the genus *Thermoproteus*. Arch Microbiol 170:155–161
- Itoh T, Suzuki K, Sanchez PC, Nakase T (1999) *Caldivirga maquilingensis* gen. nov., sp. nov., a new genus of rod-shaped crenarchaeote isolated from a hot spring in the Philippines. Int J Syst Bacteriol 49:1157–1163
- Itoh T, Suzuki K, Nakase T (2002) *Vulcanisaeta distributa* gen. nov., sp. nov. and *Vulcanisaeta soumiana* sp. nov., hyperthermophilic, rod-shaped crenarchaeotes isolated from hot springs in Japan. Int J Syst Evol Microbiol 52:1097–1104
- Kjems J, Garrett RA (1985) An intron in the 23S ribosomal RNA gene of the archaeobacterium *Desulfurococcus mobilis*. Nature 318:675–677
- Kjems J, Garrett RA (1991) Ribosomal RNA introns in archaea and evidence for RNA conformational changes associated with splicing. Proc Natl Acad Sci USA 88:439–443
- Lauerer G, Kristjansson JK, Langworthy TA, König H, Stetter KO (1986) *Methanothermus sociabilis* sp. nov., a second species within the *Methanothermaceae* growing at 97 °C. Syst Appl Microbiol 8:100–105
- Lykke-Andersen J, Garrett RA (1994) Structural characteristics of the stable RNA introns of archaeal hyperthermophiles and their splicing junctions. J Mol Biol 243:846–855
- Lykke-Andersen J, Aagaard C, Semionenko M, Garrett RA (1997) Archaeal introns: splicing, intercellular mobility and evolution. Trends Biochem Sci 22:326–331
- Morinaga Y, Nomura N, Sako Y, Uchida A (2000) Substrate recognition of homing endonuclease I-Ape I in a hyperthermophilic archaeon (Abstract). Third International Congress on Extremophiles, Hamburg, Germany, p 212
- Nomura N, Sako Y, Uchida A (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
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425
- Sako Y, Nunoura T, Uchida A (2001) *Pyrobaculum oguniense* sp. nov., a novel facultative aerobic and hyperthermophilic archaeon growing at up to 97 °C. Int J Syst Evol Microbiol 51: 303–309
- Takai K, Horikoshi K (1999) Molecular phylogenetic analysis of archaeal intron-containing genes coding for rRNA obtained from a deep-subsurface geothermal water pool. Appl Environ Microbiol 65:5586–5589
- Tamoka J (1994) Determination of DNA base composition. In: Goodfellow M, O'Donnell AG (eds) Chemical methods in prokaryotic systematics. Wiley, Chichester, pp 463–470
- Thompson JD, Higgins DG, Gibson TJ (1994) Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882
- Watanabe Y, Yokobori S, Inaba T, Yamagishi A, Oshima T, Kawarabayashi Y, Kikuchi H, Kita K (2002) Introns in protein-coding genes in Archaea. FEBS Lett 510:27–30