J Mol Evol (2000) 51:587–599
DOI: 10.1007/s002390010122

JOURNAL OF MOLECULAR EVOLUTION

Horizontal Transfer of Archaeal Genes into the Deinococcaceae: Detection by Molecular and Computer-Based Approaches

Lorraine Olendzenski,[1] Lei Liu,[1,2,*] Olga Zhaxybayeva,[1] Ryan Murphey,[1,‡] Dong-Guk Shin,[2] J. Peter Gogarten[1]

[1] Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT 06269, USA
[2] Department of Computer Sciences and Engineering, University of Connecticut, Storrs, CT 06269, USA

Received: 3 April 2000 / Accepted: 14 August 2000

Abstract. Members of the Deinococcaceae (e.g., *Thermus, Meiothermus, Deinococcus*) contain A/V-ATPases typically found in Archaea or Eukaryotes which were probably acquired by horizontal gene transfer. Two methods were used to quantify the extent to which archaeal or eukaryotic genes have been acquired by this lineage. Screening of a *Meiothermus ruber* library with probes made against *Thermoplasma acidophilum* DNA yielded a number of clones which hybridized more strongly than background. One of these contained the prolyl tRNA synthetase (RS) gene. Phylogenetic analysis shows the *M. ruber* and *D. radiodurans* prolyl RS to be more closely related to archaeal and eukaryal forms of this gene than to the typical bacterial type. Using a bioinformatics approach, putative open reading frames (ORFs) from the prerelease version of the *D. radiodurans* genome were screened for genes more closely related to archaeal or eukaryotic genes. Putative ORFs were searched against representative genomes from each of the three domains using automated BLAST. ORFs showing the highest matches against archaeal and eukaryotic genes were collected and ranked. Among the top-ranked hits were the A/V-ATPase catalytic and noncatalytic subunits and the prolyl RS genes. Using phylogenetic methods, ORFs were analyzed and trees assessed for evidence of horizontal gene transfer. Of the 45 genes examined, 20 showed topologies in which *D. radiodurans* homologues clearly group with eukaryotic or archaeal homologues, and 17 additional trees were found to show probable evidence of horizontal gene transfer. Compared to the total number of ORFs in the genome, those that can be identified as having been acquired from Archaea or Eukaryotes are relatively few (approximately 1%), suggesting that interdomain transfer is rare.

Key words: Horizontal gene transfer — BLAST comparison — Genomes — *Deinococcus* — *Thermus* — *Meiothermus* — A/V-ATPase — Prolyl-tRNA synthetase — Biotin carboxylase — Enolase — Bioinformatics

Introduction

The Deinococcaceae are classified as Bacteria based on ribosomal RNA sequence, cell wall, and lipid composition (Woese 1987; Hensel et al. 1986). In 16S rRNA phylogenies, this group represents a lineage that branches off before the radiation of the majority of bacterial lineages, and in many phylogenetic reconstructions the Deinococcaceae form a clade together with the green nonsulfur bacteria (e.g., Maidak et al. 1999). Genera in the Deinococcaceae include the radiation-resistant *Deinococcus*, thermophilic *Thermus* (Wlliams and Sharp 1995), and mesophilic *Meiothermus* [formerly *Thermus*

* *Present address:* W.M. Keck Center for Comparative and Functional Genomics, University of Illinois at Urbana–Champaign, Urbana, IL 61801 USA
‡ *Present address:* CuraGen Corporation, 555 Long Wharf Drive, New Haven, CT 06511, USA
Correspondence to: J. Peter Gogarten; *e-mail:* gogarten@uconnvm.uconn.edu

(Nobre et al. 1996)]. While many characteristics clearly identify the Deinococcaceae as bacteria, *Thermus thermophilus* does not have an F-ATPase like other bacteria, but a complete A/V-ATPase operon (Yokoyama et al. 1990; Tsutsumi et al. 1991). Typically, A-ATPases are found only in Archaea, while V-ATPases are found only in Eukaryotes. Both major subunits (A and B) of the *Thermus* ATPase are archaeal in character (Gogarten et al. 1992) and their presence in these Bacteria can be explained by horizontal gene transfer from an achaeon (Hilario and Gogarten, 1993).

Distribution of ATPases Among the Deinococcaceae. Complete sequencing of the *D. radiodurans* genome (White et al. 1999) reveals that this organism also contains an A/V-ATPase operon. In bootstrapped distance analyses of the amino acid sequences of the catalytic and noncatalytic ATPase subunits, *T. thermophilus* and *D. radiodurans* branch off in between the A-ATPase and the V-ATPase sequences (Fig. 1). For this reason, we refer to this ATPase as an A/V-ATPase. Basic local alignment search tool (BLAST) searches using F-ATPase subunits as query sequences against the *D. radiodurans* genome return only the homologous A/V-ATPase subunits, indicating that there is no F-ATPase encoded by this genome. The same conclusion was reached for *T. thermophilus* HB 8 and HB 24 using Southern blot analyses and PCR with redundant primers on genomic DNA (K.H. Schleifer, University of Munich, personal communication; Gogarten, unpublished). Among the genera *Thermus* and *Meiothermus,* biochemical assays and N-terminal amino acid sequencing show that *T. aquaticus* YT-1, *T. thermophilus* HB27, and *M. chliarophilus* contain an A/V-type ATPase. However, at least two *Thermus* species, *T. scotoductus* and *T. filiformis,* appear to possess bacterial type F-ATPases (Radax et al. 1998).

Before the discovery of an A/V-ATPase in *T. thermophilus,* it was assumed that all Bacteria had F-ATPases. However, new data from whole-genome sequencing has revealed A-ATPases in the spirochetes *Borrelia burgdorferi* (Fraser et al. 1997) and *Treponema pallidum* (Fraser et al. 1998), as well as in *Chlamydia trachomatis* (Stephens et al. 1998), *Chlamydophila pneumoniae* [National Center for Biotechnology Information (NCBI) accession No. AAD18241] and *Chlamydia pneumoniae* (Kalman et al. 1999). *Enterococcus hirae* contains both a typical F-type ATPase and a sodium-pumping A/V-type ATPase (Takase et al. 1994). An A/V-type ATPase has also been detected in *Clostridium fervidis* but the complete sequence is not available (Höner zu Bentrup et al. 1997). The reverse has also been found: a sequence encoding an F-ATPase was obtained from the archaeon *Methanosarcina barkeri* (Sumi et al. 1992, 1997).

Within the Deinococcaceae, the ATPase operon is not the only evidence of horizontal transfer of nonbacterial type genes. Using malate dehydrogenase (mdh) and the paralogous lactate dehydrogenase sequences to root a phylogenetic tree, *Thermus flavus* was shown to group with eukaryotic cytosolic mdh sequences (Iwabe et al. 1989). Analysis of an expanded data set (Olendzenski and Gogarten, unpublished data) groups the *Deinococcus* sequence with the *Thermus* sequence, with both most closely related to the eukaryotic cytosolic forms. This suggests a horizontal transfer into the Deinococcaceae from a eukaryote or an organism close to the ancestor of the eukaryotic nucleocytoplasmic component.

Among prokaryotes, a number of other genes yield phylogenies that differ from the typical or accepted phylogeny of three domains. These markers place archaeal sequences within the Bacteria, suggesting a transfer of a number of genes from the Bacteria to the Archaea (Gogarten et al. 1996; Olendzenski and Gogarten 1999). Computer analysis of archaeal and bacterial genomes also reveals episodes of horizontal gene transfer (e.g., Ribeiro and Golding 1998). An unexpectedly large fraction of the *Methanococcus jannaschii* gene products, 44%, showed a higher similarity to bacterial proteins than eukaryotic ones (Koonin et al. 1997), while the genome of *Thermotoga maritima,* a bacterium, has been interpreted to contain a large number of archaeal genes [(Nelson et al. 1999); however, see Logsdon and Faguy (1999) for a different interpretation].

Using two complimentary approaches, we have searched for other genes in the Deinococcaceae which may have been transferred from the archaeal/eukaryotic lineage. Using a molecular biology approach, a mixed population of random probes made from the genomic DNA of the archaeon *Thermoplasma acidophilum* was used to screen a genomic DNA library from *Meiothermus* (*Thermus*) *ruber,* a member of the Deinococcaceae which branches between *Thermus* and *Deinococcus* (Williams and Sharp 1995) and forms a sister taxon to the thermophilic *Thermus* species (Maidak et al. 1999). Using a bioinformatics approach, we compared open reading frames (ORFs) generated from the *Deinococcus* genome data available at the time against completed and annotated genomes from each of the three domains, *Methanococcus jannaschii, Saccharomyces cerevisiae, Escherichia coli, Bacillus subtilis,* and *Aquifex aeolicus,* and selected those genes that were more similar to archaeal/eukaryotic sequences than to their bacterial homologues. Gene trees of these candidate ORFs were then analyzed for evidence of horizontal gene transfer (HGT).

Materials and Methods

Library Construction and Screening. Genomic DNA from *Meiothermus ruber* (ATCC 35948) was purified on a cesium chloride gradient. Purified DNA was partially digested with *Sau*3AI (Gibco BRL) and size fractionated on a sucrose gradient. Then 1- to 3-kb fragments (2-kb average size) were ligated to *Bam*HI-predigested, CIAP-treated ZAP Express vector arms (Stratagene). Recombinant phages were packaged

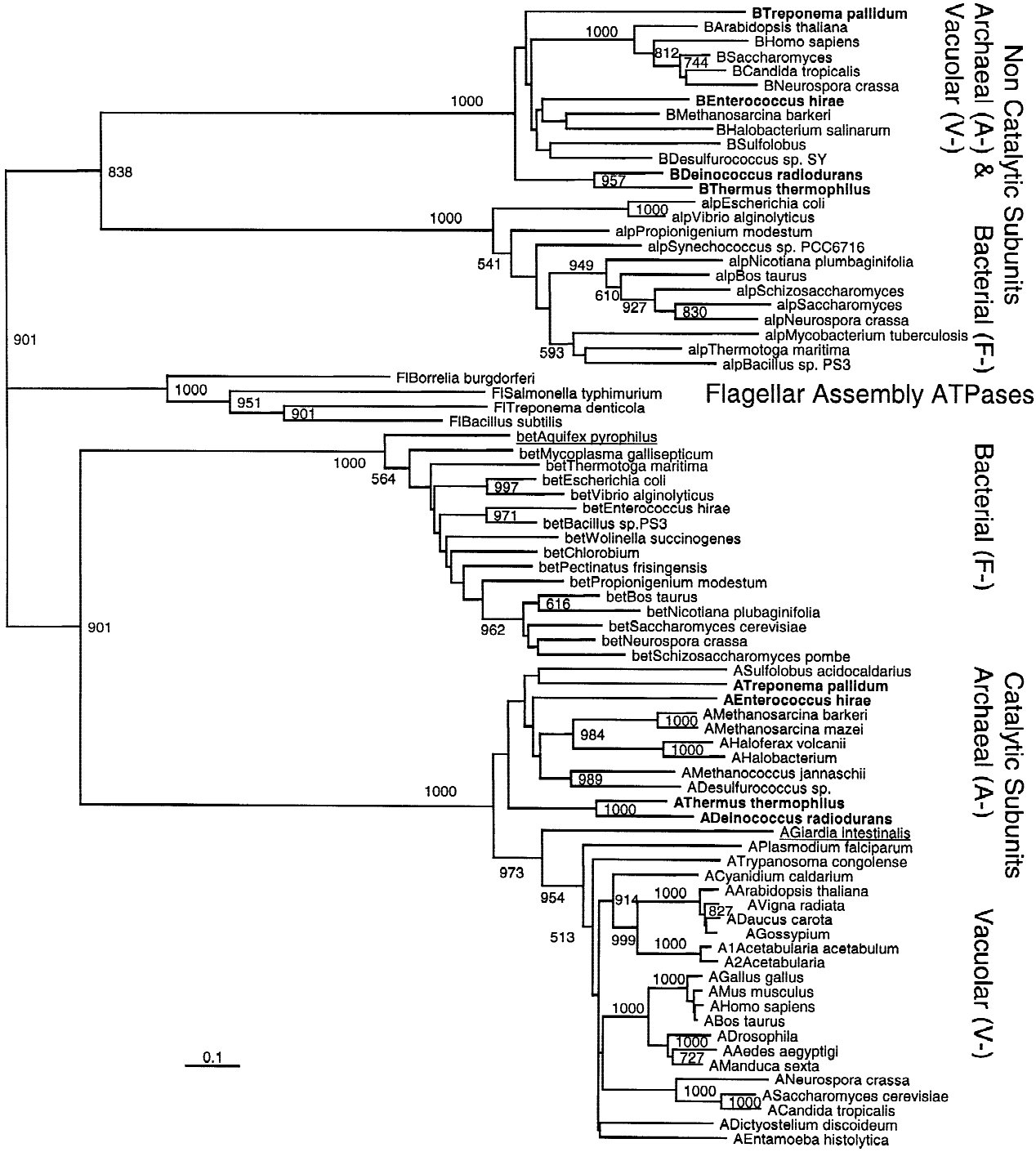

Fig. 1. Distance tree based on amino acid alignment of catalytic and noncatalytic subunits of proton-pumping ATPases from Bacteria, Archaea, and Eukaryotes. The catalytic and noncatalytic subunits are the products of an ancient gene duplication; the tree can be rooted by using one of the duplicated pair (paralogue) as an outgroup (Gogarten et al. 1989). The genera in *boldface* are bacterial species (including *Thermus* and *Deinococcus*) which have acquired an archaeal-type ATPase. The *underlined* genera represent the deepest branches in the bacterial and eukaryotic lineages. B, A/V-ATPase B subunit; alp, F-ATPase α subunit; F1, flagellar assembly ATPase subunit; bet, F-ATPase β subunit; A, A/V-ATPase A subunit.

using Gigapack II packaging extracts (Stratagene) to yield a library with $>10^7$ primary transformands. The primary library was plated on XL1Blue-MRF′ and 20 random clones were picked for sequencing. The library was then amplified to a titer of 10^9 pfu/ml. A digoxigenin-labeled probe was made against cesium chloride gradient-purified *Thermoplasma acidophilum* genomic DNA using random nucleotide octamers as primers and primer extension with Klenow fragment DNA polymerase following the manufacturer's instructions (Boehringer Mannheim). The probe was hybridized overnight (65–68°C) to plaques which had been transferred to nylon membranes (Amersham) and washed in 2× SSC, 0.1% SDS twice for 5 min at room temperature and in 0.5× SSC, 0.1% SDS twice for 15 min at 63°C. Clones showing an

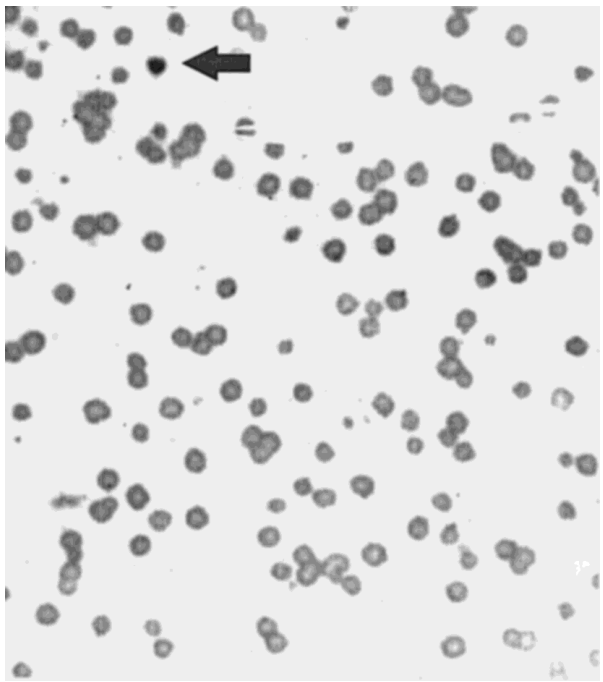

Fig. 2. Portion of primary plaque lift from *Meiothermus ruber* library hybridized to digoxygenin-labeled probe made against purified *Thermoplasma acidophilum* genomic DNA. Clones showing increased hybridization against background (*arrow*) were sequenced.

increased level of hybridization against background were picked for sequencing (Fig. 2).

Sequencing Clones from the Library. Nine hybridized clones and 20 randomly picked clones were excised using ExAssist helper phage to obtain double-stranded pBK-CMV phagemid for sequencing. Excisions were performed using the standard single-clone excision protocol in the ZAP Express Predigested Vector Kit (Stratagene). Plasmids were prepared using standard alkaline lysis procedures (Sambrook et al. 1989). DNA was quantified on a DyNA Quant 200 fluorimeter (Hoefer). Sequencing reactions were performed using standard annealing protocol for double-stranded DNA with the Cy5 Autoread Sequencing Kit (Pharmacia) and Cy5-labeled T3 and T7 primers. Sequence reactions were run on an ALF automated sequencer (Pharmacia) using standard acrylamide gel and running procedures. Sequence tags (STs) were searched against the nonredundant databank at the NCBI using BLASTX.

Screening the Library for A/V-ATPase and Sequencing. Primers NP2 and NP3 (Starke and Gogarten 1993) were used to generate an approximately 160-bp PCR product from the A subunit of the A/V-ATPase of *Meiothermus ruber*. PCR reactions were performed using a hot start and a 42°C annealing temperature. The resulting product was labeled using PCR with digoxigenin-labeled dNTPs (Boehringer Mannheim). The probe (20 ng/ml) was hybridized overnight (65–68°C) to plaques which had been transferred to two nylon membranes (Amersham) and washed in 2× SSC, 0.1% SDS twice for 5 min at room temperature and in 0.5× SSC, 0.1% SDS twice for 15 min at 63°C. Positive plaques were picked and stored in SM buffer and screened using PCR as described above with primers NP2 and NP3 and 1 μl of SM buffer in which plaques had been stored as template. Plaques showing the presence of the 160-bp band were purified through two more rounds of hybridization, followed each time by a PCR reaction to confirm the presence of the desired insert. The resulting two clones were excised as above. Plasmids were prepared using alkaline lysis (Sambrook et al. 1989) followed by PEG precipitation (0.4 *M* NaCl and 6.5% PEG 8000, final concentration). DNA was quantified on a DyNA Quant 200 fluorimeter (Hoefer). Three hundred nanograms of plasmid DNA were used for sequencing reaction. Sequencing reactions were performed using half reactions (4 μl) of ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Mix (Perkin–Elmer) and were run on an ABI PRISM 377 automated sequencer (Perkin–Elmer).

Generation of Potential ORFs from Deinococcus *Data.* The *Deinococcus* genome data used were in preliminary form, as uncorrected data without annotation or analysis and assembled into contigs (http://www.tigr.org). An ORF generating program was created which translated and searched each contig in all six possible reading frames for stop codons and recorded the number of nucleotides between two adjacent stop codons. If the length was greater than 300 bp, the sequence was reported as a potential ORF and assigned an ID. ID designations were generated automatically with the following format: gdr_xxx_ORF*y.z*, where gdr_xxx is the contig number, *y* is an integer between 0 and 5 representing the six different frames, and *z* is an integer representing the number of an ORF in a particular contig and frame. This rather permissive approach to defining putative ORFs was chosen because it guarantees that most of the protein coding regions were included in putative ORFs even when the gene was disrupted by an artificial frame shift due to a sequencing error. The many putative ORFs generated by this approach that do not correspond to actual ORFs will not negatively impact the performed analyses because these incorrectly identified ORFs do not result in significant matches with respect to the reference genomes. In three-dimensional plots, these artificial ORFs are located in a cloud closest to the origin (see Figs. 4 and 5).

Automation of BLAST to Process all ORFs in a Batch. A local version of the BLAST program was installed on a Sun workstation by downloading the executable file of gapped-blast (BLAST 2.0) from the NCBI (http://www.ncbi.nlm.nih.gov). The BLAST program was wrapped in a Java RMI server which would spawn a BLAST process upon a request from the client program. Outputs returned from the BLAST program were parsed by the server program. Gapped BLAST was run for each ORF against amino acid sequences of ORFs from a single reference genome, using BLASTP with the BLOSUM62 matrix with filtering for low-complexity regions. For each BLAST run, only the top hit was reported. The ID of the query sequence, the ID of the matching sequence, the high score pair (HSP), and expected (*E*) values for each comparison were retained, and these data stored in a tab delimited file. If there was no hit, HSP and *E* were assigned values of 0 and 10.0, respectively. These data were then loaded into an Oracle database. Each potential *D. radiodurans* ORF was compared against each of five individual reference genomes: *Saccharomyces cerevisiae* (Goffeau et al. 1997), *Methanococcus jannaschii* (Bult et al. 1996), *Escherichia coli* (Blattner et al. 1997), *Bacillus subtilis* (Kunst et al. 1997), and *Aquifex aeolicus* (Deckert et al. 1998). For each microorganismal reference genome, files containing the complete list of translated ORFs were downloaded from the following sites: *M. jannaschii*, ftp://ftp.tigr.org/pub/data/m_jannaschii/; *S. cerevisiae*, ftp://genome-ftp.stanford.edu/pub/yeast; *E. coli*, ftp://ftp.genome.wisc.edu/pub/sequence/; *A. aeolicus*, ftp://ncbi.nlm.nih.gov/genbank/genomes/bacteria/Aquae/; and *B. subtilis*, ftp://ftp.pasteur.fr/pub/GenomeDB/SubtiList/FlatFiles/. Output data from each comparison were stored in one table.

Ranking and Visualization of the Comparison Results. To visualize the overall comparison, HSP and (−log*E*) values for BLAST searches obtained against the *S. cerevisiae*, *E. coli*, and *M. jannaschii* genomes were plotted on three-dimensional graphs (see Figs. 4 and 5). Additionally, the following difference was calculated for each putative *Deinococcus* ORF: δ_{ai} = maximum($S_{y.ai}$, $S_{m.ai}$) − Maximum($S_{e.ai}$, $S_{b.ai}$, $S_{a.ai}$), where ai is the ORF id, and S_y, S_m, S_e, S_b, and S_a are the set of HSP scores obtained by BLAST against the *S. cerevisiae*, *M. jan-*

naschii, E. coli, B. subtilis, and *A. aeolicus* genomes, respectively. All ORFs were ranked in a descendent order according to δ_{ai}. Larger δ_{ai} values indicate ORFs which are more similar to eukaryotic and archaeal genes than to bacterial ones. A rotatable three-dimensional plot of the HSP values obtained in BLAST searches against the *S. cerevisiae, M. jannaschii,* and *E. coli* genomes was created using a Java program. Points residing on or close to the line equidistant from the three axes were identified and used for phylogenetic analysis. Sequences were included if they had scores of greater than 50 for each of the three coordinates and fell into a cone surrounding the $x = y = z$ axis that was defined by all points whose distance from the $x = y = z$ line was less than or equal to one-tenth the distance of the point from the origin of the coordinate system.

Phylogenetic Analysis of the Candidate ORFs. Phylogenetic trees of each of the ORFs of the 45 top-ranked δ_{ai} values were constructed. Additionally, trees were calculated for those ORFs which fell on or close to the line equidistant from all three axes in the plot of HSP scores (see above). Using gapped BLASTP, each of the selected ORFs was compared to the nonredundant protein database from the NCBI. Putative orthologues were chosen from the obtained matches to represent all three domains. Sequences were aligned using ClustalX 1.64 (Thompson et al. 1997). Using neighbor joining (Saitou and Nei 1987) as implemented in ClustalX and 100 or 1000 bootstrapped replicates, consensus trees were calculated excluding all positions with gaps and correcting for multiple hits (see documentation, ClustalX 1.64). Those data sets which were not excessively large were also analyzed using PUZZLE 4.0 (Strimmer and von Haeseler 1996), a quartet puzzling method, with correction for among-site rate variation (ASRV) using eight gamma rate categories. Bootstrapped neighbor-joining trees of ORFs whose HSP values were equidistant from *M. jannaschii, E. coli,* and *S. cerevisiae* were also calculated using Phylowin (Galtier et al. 1996) with gaps excluded and using either a PAM matrix or a Poisson correction. Trees were evaluated visually for evidence of horizontal transfer from the Archaea or Eukaryotes into *D. radiodurans.*

Results

Library Screening. Screening of approximately 10,000 plaques from the *Meiothermus ruber* library with random probes made against *T. acidophilum* yielded five plaques which showed increased hybridization. Some of these were screened through a second round of hybridization and picked for sequencing. Attempts were made to sequence 20 random clones and 9 screened clones with both T3 and T7 primers. Clones which failed to yield sequences of more than 175 bp were not searched on the NCBI website. Of 20 STs obtained from randomly picked clones and searched against the nonredundant database at NCBI, 11 showed no match or matches which were nonsignificant ($E \geqslant 10^{-4}$). Of nine significant matches, seven showed the highest matches to eubacterial sequences, while two showed the highest matches to eukaryotic homologues. No distinctly archaeal sequences were found in the randomly picked clones. Of eight STs obtained from strongly hybridizing plaques, four showed no match or no significant match, while two showed the highest matches to eubacterial sequences. Two STs showed the highest matches to eukaryotic sequences. One of these had closest similarity to eucaryal and archaeal prolyl tRNA synthetases. Bootstrapped distance and quartet puzzling analyses of the full-length data set of prolyl aminoacyl tRNA synthetase (RS) genes, including the *D. radiodurans* sequence, with and without the short *M. ruber* prolyl RS fragment show that the prolyl RS genes form two distantly related groups supported by high bootstrap and quartet puzzling values (Fig. 3). The same groups were highly supported in bootstrapped parsimony analysis as implemented in Paup*4.0 [(Swofford 2000); scoring gaps as missing data and using TBR branch swapping and three starting trees obtained by random addition in the analysis of each bootstrapped sample]. The majority of Bacteria are found in one group, while the Eukaryotes and Archaea are found in the other. The prolyl tRNA synthetases of *Mycoplasma, Borrelia, Deinococcus,* and the *Meiothermus* fragment (not shown) group with the eukaryotic homologues. The two yeast homologues which group with the majority of Bacteria are probably of mitochondrial origin (Stehlin et al. 1998).

Screening with the *T. acidophilum* whole-genome probe did not result in the identification of a V/A-ATPase containing clone. However, screening the *Meiothermus ruber* genomic library with the probe specific for a 160-bp portion of the A/V-ATPase retrieved several clones. Sequence data from one of these showed the highest similarity to A-subunits of A/V-type vacuolar ATPases.

Computer-Based Search of the Deinococcus *Genome.* Using the ORF generating program, we generated 15826 potential ORFs from the prerelease data set of the *Deinococcus* genome (http://www.tigr.org). A plot of *E* values obtained for each ORF when analyzed with BLAST against the *S. cerevisiae, E. coli,* and *M. jannaschii* genomes is shown in Fig. 4. Arrows indicate ORFs which are more similar to eukaryotic or archaeal homologues.

ORFs were ranked according to the difference in BLAST score between the bacterial genomes and either the yeast or the *M. janaschii* genome. The following formula was used for ranking: $[\max(HSP_{Eukaryotes}, HSP_{Archaea}) - \max(HSP_{Bacteria})]$. The 45 highest-scoring ORFs are listed in Table 1. For each ORF, the HSP value against each reference genome is given, with the difference calculated as described above. Assigned functions are based on the GenBank assignment of the highest-scoring match for each ORF. Phylogenetic trees for each ORF were evaluated for evidence of HGT. The results are tallied in Table 1 in the column labeled HGT. Reconstructed phylogenies including bootstrap and quartet puzzling support values are available at http://carrot.mcb.uconn.edu/hgtpaper/. The most easily interpreted cases of horizontal gene transfer were those topologies which showed Bacteria, Archaea, and Eukaryotes as distinct groups, but with few bacterial representatives including *D. radiodurans* grouping with the Eukaryotes or Archaea (Table 1, footnote 1). This was seen for the top three entries in the table, the A and B subunits of the

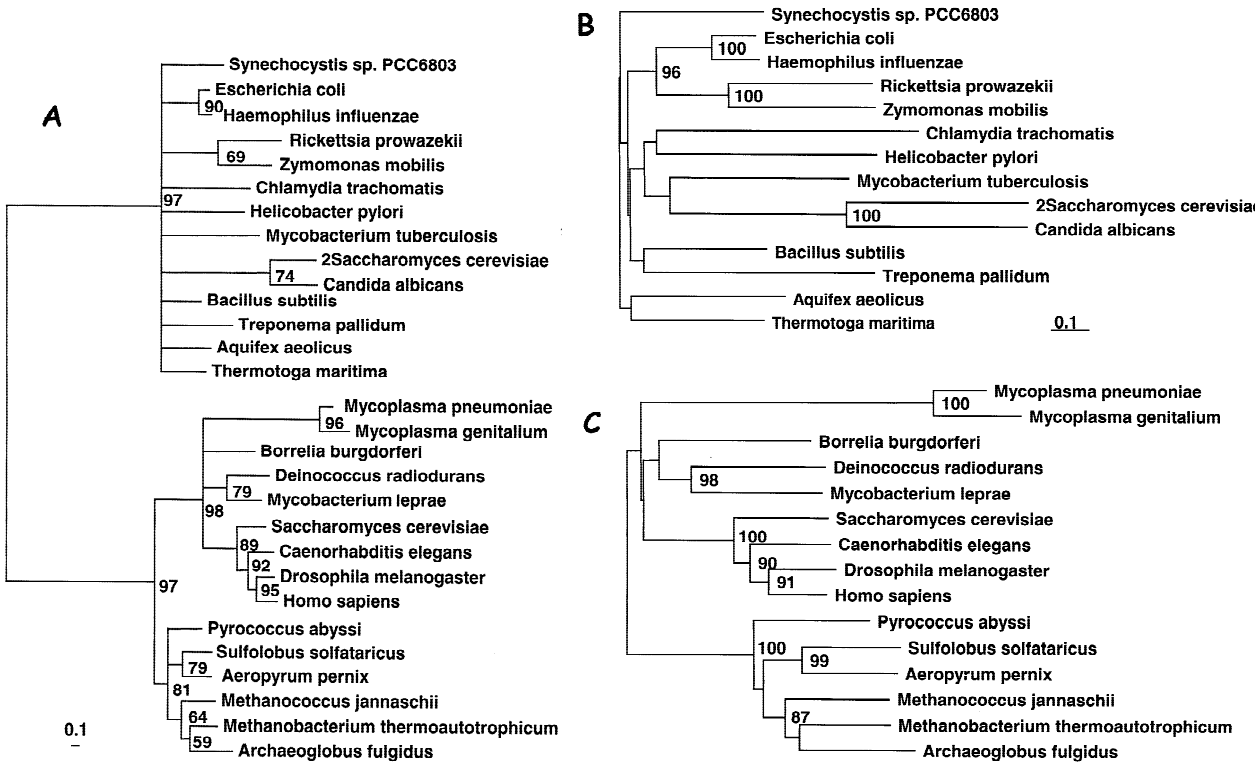

Fig. 3. A Quartet puzzling analysis of prolyl-tRNA synthetase (prolyl RS) amino acid sequences. Sequences were aligned using the profile alignment option of ClustalX (Thompson et al. 1997) with default parameters. Numbers reflect quartet puzzling support values for each node. Nodes with values of less than 50% have been collapsed. Analysis was corrected for among-site rate variation using the gamma distribution (shape parameter $\alpha = 1.97$). All sequences were obtained from GenBank except those of *Deinococcus radiodurans* and *Treponema pallidum*, which were obtained through BLAST search from TIGR, and *Sulfolobus solfataricus*, which was kindly provided by Mark Regan and Christoph Sensen, Institute for Marine Biosciences, Dalhousie University, Halifax, Nova Scotia, Canada. The prolyl RS from *Meiothermus ruber* (not shown) groups with the *D. radiodurans* sequence. The spirochete *Borrelia burgdorferi* and the two *Mycoplasma* species also have the archaeal/eukaryotic form of this gene. The yeast homologues, which group among the Bacteria (*2Saccharomyces* and *Candida*) are assumed to be of mitochondrial origin. **B, C** Neighbor-joining analysis of each type of prolyl tRNA synthetase [bacterial (B), archaeal/eukaryotic (C)], showing branching order. Trees were calculated using Phylo_win (Galtier et al. 1996) with global gap removal and the PAM matrix for calculation of pairwise distances. One hundred bootstrapped replicates were calculated; bootstrap values greater than 85% are indicated on the tree.

ATPase (Fig. 1) and the prolyl RS (Fig. 3), which had already been detected by other techniques as probable examples of genes horizontally transferred from Archaea or Eukaryotes into the Deinococcaceae. Of the 45 ORFs examined, 20 showed topologies in which *Deinococcus* homologues clearly grouped with eukaryotic or archaeal homologues (designated +), 17 ORFs were interpreted to show probable evidence of HGT [designated (+); see Table 1 footnotes and Discussion], and 3 were found to show no evidence of transfer (designated −) in that *Deinococcus* grouped with the bacterial homologues, as in a "conventional" phylogeny. The five remaining ORFs could not be interpreted due to lack of data (NED), the presence of numerous paralogues in the databank making identification of orthologues difficult, or high divergence of the *Deinococcus* sequence, which may have caused long-branch artifacts during phylogenetic reconstruction (designated ?).

Thirty-three high-scoring ORFs were found to lie close to the line equidistant from the *E. coli*, *M. jannaschii*, and *S. cerevisiae* axes when HSP values were plotted (Fig. 5). The first 17 of these are listed in Table 2. BLAST searches using these sequences revealed that they all had numerous homologues in the data bank, indicating that they were highly conserved and easily recognizable among a variety of organisms. However, contrary to expectation, the obtained molecular phylogenies did not reveal indications for frequent recent HGT between Archaea and Bacteria. In most cases, the Archaea formed only one or two monophyletic groups. Eukaryotes sometimes were interspersed among bacterial sequences probably representing contributions of genes from mitochondria or plastids.

Discussion

Both *D. radiodurans* and *M. ruber* contain genes which are more closely related to archaeal and eukaryotic genes than they are to the bacterial homologues present in the three bacterial reference genomes used in this study. The

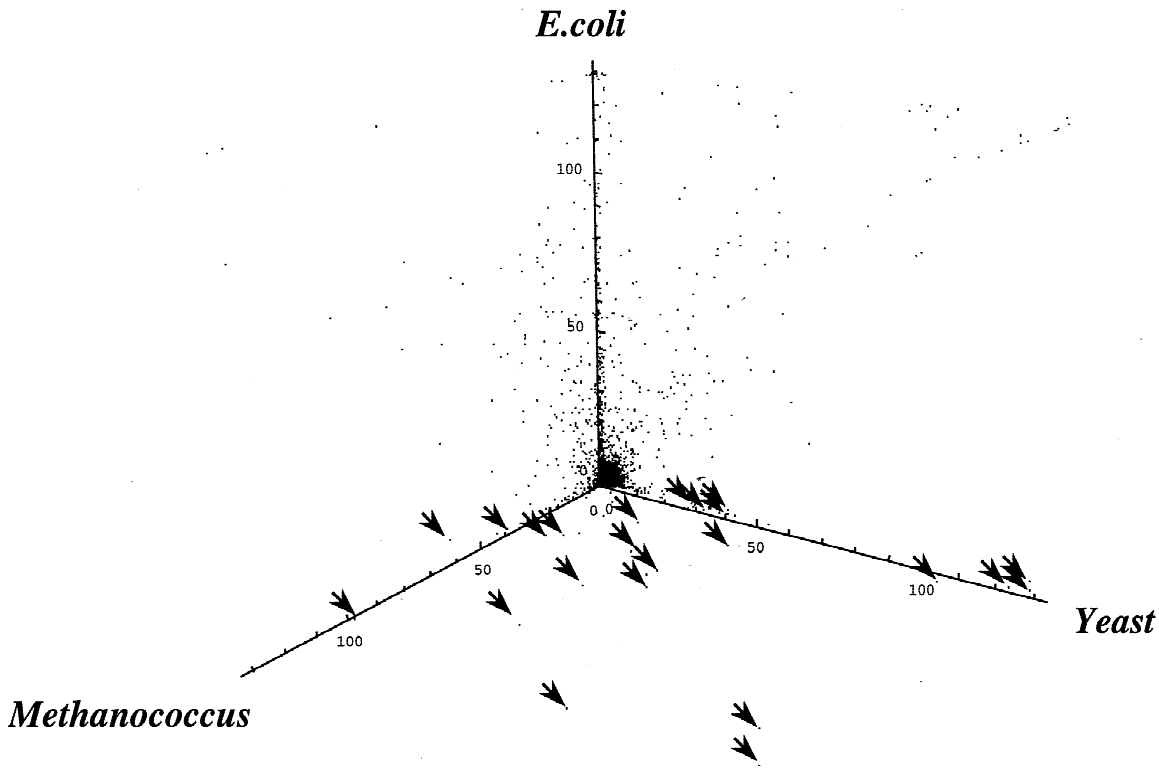

Fig. 4. Plot of *E* values obtained from BLAST analysis of each ORF against the genomes of *Saccharomyces cerevisiae* (yeast), *Escherichia coli*, and *Methanococcus jannaschii*. Each point (*x*, *y*, *z*) on the three-dimensional graph represents one putative ORF from the *Deinococcus radiodurans* genome. The coordinates are as follows: *x* is the −log*E* value from BLAST against the yeast genome; *y* is the −log*E* value from BLAST against the *M. jannaschii* genome; *z* is the −log*E* value from BLAST against the *E. coli* genome. *Arrowheads* indicate points which are more similar to yeast and/or *Methanococcus* genes than to *E. coli* genes based on the *E* value. These represent candiate genes horizontally transferred from Archaea or Eukaryotes to the bacterial domain.

reference genomes were chosen to provide a good representation of the bacterial domain. *Aquifex aeolicus* represents a putatively deep-branching lineage, whereas *B. subtilis* and *E. coli* represent the Gram positives and Gram negatives, respectively. While the outcome of the initial screening is dependent on the choice of reference genomes, the further phylogenetic analyses of the candidate genes are based on all homologues present in the nonredundant protein database at the NCBI. Both the wet lab and the in silico approach detected genes of archaeal/eukaryal character in the Deinococcaceae; however, our experiments suggest that the proportion of these types of genes is not large compared to the number of genes in the complete genome. In the case of *Meiothermus*, no distinctly archaeal sequences were found in the STs from randomly picked clones. Although library screening detected a *Thermus* clone containing a prolyl RS that is significantly more similar to eukaryotic and archaeal sequences than to eubacterial ones, screening with the *Thermoplasma* probe did not detect A/V-ATPase encoding sequences. Since data were acquired only for up to 500 bp from each end of the 1- to 3-kb inserts, it is possible that archaeal/eukaryal genes are located in the interior of some of these clones. The method of screening was successful, however, in that it was able to differentiate a hybridized clone that contained a sequence similar to an archaeal type from random clones.

Partial sequencing of a clone retrieved from the *M. ruber* library using a V/A-ATPase A-subunit specific probe did confirm that *M. ruber* contains an A/V-type ATPase. It is possible that a common ancestor of the *Deinococcus*, *Meiothermus*, and *Thermus* lineages acquired this gene via HGT and then lost its original F-ATPase. However, *T. scotoductus* and *T. filiformis* were reported to contain an F-ATPase (Radax et al. 1998). The presence of an A/V-ATPase in the deeper-branching *M. ruber* and *D. radiodurans* indicates that the presence of an A/V-type ATPase is a primitive trait (pleisiomorphy) for the Deinococcaceae. Could *T. scotoductus* and *T. filiformis* have acquired the F-ATPase later in another transfer event? Or was the F-ATPase originally present in the Deinococcaceae maintained in the common ancestor and then selectively lost multiple times in the descendents of the lineage? More sequence data are needed from *T. scotoductus*, *T. filiformis*, and other members of the Deinococcaceae to trace the likely order of events.

The computer-based screening of the *Deinococcus* ORFs yielded the A/V-ATPase subunits and the prolyl RS as the top-scoring hits, confirming our previous data that these genes were likely acquired by HGT. The pres-

Table 1. ORFs ranked according to the maximum difference in high-scoring pairs [max (HSP Eukaryotes, HSP Archaea) − max (HSP Bacteria)]

ORF_ID	*S.c.*	*M.j.*	*E.c.*	*B.s.*	*A.a.*	Difference	HGT	Function
66_ORF1.17	497	561	131	144	125	417	+	ATPase B subunit[1]
66_ORF0.13	290	467	90	91	97	370	+	ATPase A subunit[1]
80_ORF3.7	328	244	75	65	69	253	+	Prolyl-tRNA synthetase (proS)[1]
12_ORF1.14	380	200	122	134	142	238	(+)	Homocitrate synthase[2]
240_ORF4.3	33	227	32	36	31	191	−	Cobyric acid synthase[3]
36_ORF2.6	170	224	30	33	32	191	+	Glycyl-tRNA synthetase[1,4]
95_ORF4.11	228	27	43	36	56	172	(+)	Metalloprotease[5]
125_ORF3.8	218	70	62	59	46	156	NED	Hypothetical protein[6]
19_ORF3.57	282	33	92	127	27	155	(+)	*O*-Acetyl-L-homoserine sulfhydrylase
23_ORF4.32	31	186	28	37	27	149	(−)	Cobyrinic acid *a,c*-diamide synthase[3,7]
66_ORF2.12	141	185	32	28	40	145	+	ATPase A subunit[1,8]
31_ORF4.25	220	28	75	68	76	144	+	NADPH adrenodoxin/oxedored.[9]
114_ORF1.2	28	168	27	25	27	141	+	Hypothetical protein[10]
39_ORF4.40	123	148	34	34	27	114	+	Cleavage and polyad. factor[3,11]
121_ORF0.7	29	143	29	30	31	112	+	2-Phosphoglycerate kinase[10]
21_ORF4.16	32	153	28	39	41	112	(+)	Hypothetical protein[3,12]
18_ORF3.4	272	38	170	49	50	102	−	Threonine synthase[13]
95_ORF3.11	125	25	31	29	29	94	(+)	Zinc metalloprotease[5,14]
119_ORF5.16	41	171	79	68	62	92	+	NADH oxidase, water-forming (nox)[15]
85_ORF1.8	224	290	123	198	110	92	+	Aspartyl-tRNA synthetase[1]
86_ORF5.7	27	116	28	27	29	87	(+)	Hypothetical protein MJ1477[3,7]
104_ORF0.3	246	243	102	161	116	85	+	Isoleucyl-tRNA synthetase[1]
104_ORF2.2	308	266	166	231	207	77	+	Isoleucyl-tRNA synthetase (ileS)[1,8]
66_ORF0.17	61	102	27	27	28	74	+	ATPase D subunit[1]
122_ORF4.1	78	100	27	27	27	73	(+)	Cleavage and polyad. factor[3,8,11]
209_ORF3.4	135	191	120	87	67	71	+	Arginyl-tRNA synthetase[16]
54_ORF0.15	188	238	151	166	167	71	(+)	3-Isopropylmalate dehydrogenase[1,17]
218_ORF0.5	27	106	30	40	26	66	(+)	Conserved hypothetical protein[18]
6_ORF5.20	29	94	29	27	24	65	(+)	Hypothetical protein[10]
19_ORF4.49	101	25	39	38	27	62	?	Homoserine acetyltransferase (met2)[5,19]
171_ORF4.5	41	136	75	51	74	61	?	Potassium channel[20]
189_ORF2.5	0	135	27	75	25	60	(+)	Hypothetical protein[3,7,20]
129_ORF4.3	29	90	31	27	29	59	?	Hypothetical protein[21]
44_ORF2.20	116	37	38	58	25	58	+	L-Kynurenine hydrolase[22]
125_ORF4.4	114	44	58	48	32	56	+	Hypothetical protein[23]
240_ORF5.1	0	83	0	28	24	55	−	Cobalamin biosynthesis protein B[3]
50_ORF4.4	128	26	29	73	33	55	(+)	Aqualysin precursor (aa 1 to 513)[24]
282_ORF3.1	79	25	27	25	27	52	+	Zinc metalloprotease[5,8]
7_ORF4.17	80	25	0	29	30	50	(+)	Conserved hypothetical protein[25]
5_ORF2.46	139	24	26	92	45	47	(+)	Aqualysin precursor (aa 1 to 513)[8,24]
159_ORF3.2	0	75	29	27	26	46	(+)	Conserved hypothetical protein[10]
70_ORF2.13	27	75	29	27	27	46	(+)	Hypothetical protein[3,7]

ence of an archaeal/eukaryal-type prolyl RS sequence in *M. ruber* and *D. radiodurans* becomes even more intriguing when correlated with the fact that, in the crenarcheote *Sulfolobus sulfataricus*, the next gene downstream from the A-ATPase operon is an archaeal-type prolyl RS (personal communication, M. Ragan, Institute of Marine Sciences, and D. Faguy, Dalhousie University, Halifax, NS). This finding suggests that both the A-ATPase operon and the prolyl RS gene could have been transferred together in a single event that took place before the split between *Thermus* and *Deinococcus.* However, in the genome of *D. radiodurans* the ATPase operon and the prolyl RS are separated. Furthermore, the spirochete *Borrelia* also contains both the A/V-ATPase and the archaeal-type prolyl RS; however, *Treponema*, a close relative of *Borrelia*, contains an A/V-ATPase but a bacterial prolyl RS. At present the resolution of the molecular phylogenies is insufficient to decide with confidence whether the A/V-ATPase and prolyl RS in Deinococcaceae and Spirochetes originated from the same donor or whether these represent independent interdomain transfers.

Some amino acyl tRNA synthetases have been recognized as genes that have been transferred among domains frequently (Nagel and Doolittle 1995; Doolittle and Handy 1998; Wolf et al. 1999; Tumbula et al. 1999). In *Deinococcus,* we find archaeal/eukaryal representatives of the prolyl RS, glycyl RS, isoleucyl RS, arginyl RS, and aspartyl RS. Prior to this, the aspartyl RS were thought to follow a "standard phylogeny" (Doolittle and Handy 1998). The *Deinococcus* sequence is most closely related to a number of archaeal homologues, suggesting that they were acquired by *Deinococcus* via HGT from an archeaon. However, horizontal transfer might not be

Table 1. Continued

ORF_ID	*S.c.*	*M.j.*	*E.c.*	*B.s.*	*A.a.*	Difference	HGT	Function
164_ORF4.3	201	27	120	159	99	42	+	2,4-Dienoyl-CoA reductase[26]
39_ORF5.22	132	178	117	136	116	42	+	Glucose-1-phosphate thymidylyltransferase[27]
66_ORF2.10	31	77	36	28	28	41	+	ATPase C subunit[12,28]

Note. For each ORF, the maximum HSP value obtained for homologues from each reference genome is listed. In the column labeled HGT, each phylogeny is scored for evidence of horizontal gene transfer from Archaea or Eukarya to *Deinococcus.* +, trees whose topology clearly indicates transfer; (+), probable evidence of HGT (see footnotes below and Discussion); −, no evidence of transfer (*Deinococcus* groups with Bacteria); NED, not enough data for interpretation; ?, analysis not done because numerous paralogues make identification of orthologues difficult or high divergence of the *Deinococcus* sequence may have caused long-branch artifacts. Assigned functions are based on the GenBank assignment of the highest-scoring match for each ORF. *S.c., Saccharomyces cerevisiae; M.j., Methanococcus jannaschii; E.c., Escherichia coli; B.s., Bacillus subtilis; A.a., Aquifex aeolicus.*

[1] Archaeal/eukaryal type clearly different from typical bacterial homologues. Some bacteria including *Deinococcus* possess the archaeal/eukaryal type of enzyme.

[2] Few eukaryotic sequences avaialble. *Deinococcus* clusters with homologues from *Saccharomyces* (mt and cyt) and are clearly separate from typical bacterial homologues.

[3] Closest homologues are bacterial; a homologue is absent in bacterial reference genomes screened (*E. coli, B. subtilus,* and *A. aeolicus*).

[4] Two major groups, an archaeal/eukaryal group and a bacterial group, exist. Alignment between the two is questionable.

[5] Only a few homologues in databank. All are eukaryal or bacterial.

[6] Only five significant matches in databank, with a disjunct taxonomic distribution (*Mycobacterium, Saccharomyces,* and *Chlamydia*).

[7] Only a few archaeal and bacterial homologues in databank; the disjunct taxonomic distribution could be explained by HGT.

[8] Duplicate entries appear, probably due to a frameshift error in preliminary data.

[9] Only archaeal and eukaryotic homologues detected in database.

[10] Only archaeal homologues detected in database.

[11] Except for three bacterial homologues, all other significant hits (23) are to Archaea and Eukaryotes.

[12] Except for bacterial homologues in two species, all other significant hits are to Archaea.

[13] Primarily bacterial homologues, with some eukaryotic representatives. Although the high score is to the yeast homologue, phylogenetic reconstruction shows it to group with Bacteria.

[14] Closest homologues are bacterial (spirochetes and chlamydias). All other homologues are eukaryal.

[15] Does not group with typical bacterial homologues in phylogenetic reconstruction; *Deinococcus,* along with *Staphylococcus* and *Borrelia,* group with archaeal/eukaryal homologues.

[16] Closest homologue in phylogenetic reconstruction is archaeal. Eukaryotes group with bacterial homologues.

[17] Archaeal and deinococcal enzymes form a paralogous group.

[18] Except for one additional bacterial homologue, all other significant hits are to Archaea.

[19] *Deinococcus* sequence very divergent from other data bank entries.

[20] Not enough information retained in the sequences to reconstruct phylogeny reliably.

[21] Multiple paralogues present in *Deinococcus.* Phylogenetic relationships cannot be resolved.

[22] *Deinococcus* sequence groups with *Streptomyces* homologue as sister group to Eukaryotes; no detectable homologue in *Escherichia, Methanococcus,* or *Aquifex.*

[23] Sequence groups with eukaryotic homologues and a few other bacterial sequences. This cluster is clearly separate from the other bacterial and archaeal homologues.

[24] Many paralogues in data bank. Few bacterial matches, and the majority to fungal sequences. The disjunct distribution might be explained by HGT.

[25] Few homologues in data bank. All are deinococcal (two) or fungal. The disjunct distribution might be explained by HGT.

[26] Highest-scoring homologues are fungal. Numerous paralogues in data bank.

[27] Typical bacterial homologues form a clade distinct from the archaeal clade. The archaeal type is found in *Deinococcus* and a few other bacteria.

[28] Closest homologue is *Thermus.*

the complete explanation of the aspartyl RS data set, as some Bacteria and many Eukaryotes contain two aspartyl RS genes. Those in Bacteria seem to be of two highly divergent types explained by either acquisition by HGT or an ancient duplication followed by loss in some lineages, while those in the Eukaryotes can be explained by a more recent gene duplication.

In some cases, the computer-based screening method implemented here gave negative results. The top-ranked sequences in Table 1, which showed no evidence of HGT (i.e., their closest homologues were bacterial, as expected), were those that were not represented in any of the bacterial reference genomes. For this reason, they obtained large HSP score differences during our screening and appeared among the top-ranked hits. Additionally, we certainly underestimate the number of candidate genes. We would not detect horizontally transferred genes that were not highly conserved or were absent in the reference genomes chosen. The different A/V-ATPase subunits provide a good illustration of positives and false negatives encountered in homologue detection. The A/V-ATPases are multisubunit complexes consisting of at least seven different subunits (Bowman and Bowman 1996). The highly conserved A and B subunits (Gogarten 1994) are ranked at the top in Table 1. Only two additional subunits of the A/V-ATPases are included among the high-scoring ORFs (subunits D and C). However, the different A/V-ATPase subunits are highly coevolved and coadapted. Therefore, it is likely that the whole ATPase encoding operon, and not just four of eight subunits, was transferred. In the *Deinococcus* genome, the ORFs neighboring the genes encoding the A/V-ATPase subunits A, B, C, and D also encode A/V-ATPase subunits; however, these subunits are so little conserved that they are not placed high in Table 1. The following gives the name of the other identified V/A-ATPase subunits, the species, and the *E* value for the homologue most similar to the *Deinococcus* ORFs (out-

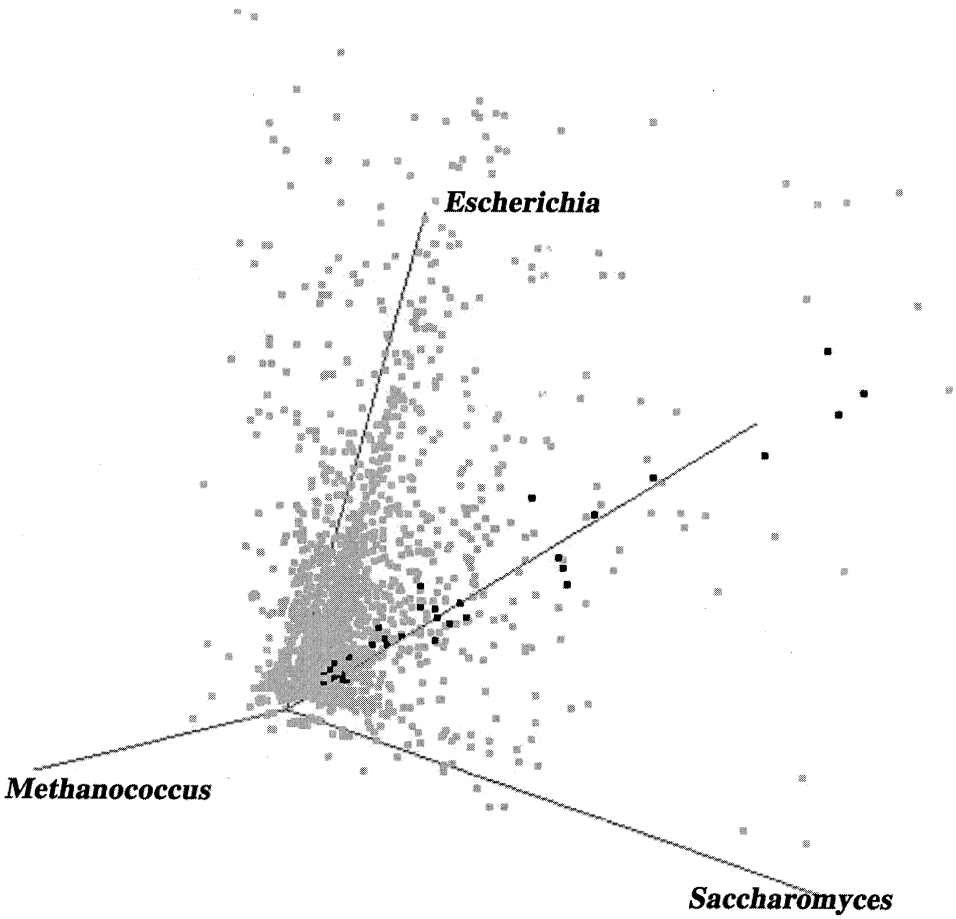

Fig. 5. Three-dimensional plot of high scores obtained in BLAST search against the *Saccharomyces cerevisiae*, *Methanococcus jannaschii*, and *Escherichia coli* genomes. *Black squares* represent ORFs close to a line which is equally distant from all three axes.

Table 2. Putative frequently transferred ORFs

ORF_ID	*S.c.*	*M.j.*	*E.c.*	*B.s.*	*A.a.*	Function
gdr_49_ORF2.1	381	431	482	527	508	Enolase (2-phosphoglycerate dehydratase)
gdr_164_ORF3.2	377	466	451	497	451	Biotin carboxylase
gdr_121_ORF0.9	404	457	427	484	474	Acetolactate synthase
gdr_5_ORF0.3	425	417	376	445	462	Tryptophan synthase β chain
gdr_118_ORF3.15	251	305	317	322	344	Carbamoylphosphate synthetase small subunit
gdr_12_ORF4.6	230	266	271	319	318	Succinyl-CoA synthetase β chain
gdr_170_ORF1.15	224	216	269	268	204	ATP-dependent RNA helicase dead homologue
gdr_191_ORF2.25	277	263	236	280	298	Biotin carboxylase
gdr_22_ORF0.11	217	252	221	293	268	Histidinol dehydrogenase
gdr_136_ORF3.8	239	264	213	452	360	Methionyl-tRNA synthetase
gdr_83_ORF2.9	150	133	162	211	133	Thioredoxin reductase
gdr_43_ORF2.11	161	169	161	326	333	Chorismate synthase
gdr_204_ORF0.1	133	168	147	169	177	Anthranilate synthase component i
gdr_46_ORF0.9	145	146	147	158	120	Nucleoside diphosphate kinase
gdr_165_ORF2.3	155	152	144	182	154	Aspartate carbamoyltransferase
gdr_225_ORF5.1	122	128	141	130	139	Anthranilate synthase
gdr_63_ORF2.15	153	163	141	135	150	Transitional endoplasmic reticulum ATPase

Note. For each ORF, the maximum HSP value obtained for homologues from each reference genome is listed. Abbreviations are as in Table 1, Note.

side the Deinococcaceae; gapped BLAST search of the nonredundant data bank using BLASTP with default options at the NCBI): subunit E, *Haloferax volcani*, 0.026; subunit C (the last entry in Table 1), *Methanococcus jannaschii*, $3 \cdot 10^{-13}$; subunit F, *Pyrococcus horikoshii*, 0.12; and subunit I, *Archaeoglobus fulgidus*, $5 \cdot 10^{-23}$. The match of *D. radiodurans* subunit I to the archaeal genome used as reference, *M. jannaschii*, has an *E* value

of only $8 \cdot 10^{-11}$. The putative proteolipid, although part of the operon, is not included in this list because it does not show any significant similarities in a normal BLAST search. In a pairwise comparison using BLAST, the *E* value for comparing the proteolipid from *D. radiodurans* to *M. jannaschii* is 6.1. This analysis of the *Deinococcus* V/A-ATPase operon clearly illustrates that the chosen approach detects only HGTs of highly conserved molecules.

We have tried to determine whether the top-ranked ORFs obtained in our screen represent genes that were likely acquired by the Deinococcaceae via HGT by interpreting phylogenetic analyses. If the closest homologues to the *Deinococcus* sequences were clearly archaeal and/or eukaryal, with few other closely related bacterial homologues, these genes were scored as positive for HGT. Taking a suggestion from W. Ford Doolittle, we also considered genes that have a very disjunct distribution (i.e., are present in only a few taxa from two domains) to represent cases where these genes may have been transferred between the two domains. (See below for a discussion of alternative explanations.)

In evaluating the trees obtained, we have interpreted patterns of extremely disjunct distribution in a small number of species as evidence of HGT. Trees showing a traditional three-domain topology, with Archaea as the sister group to Eukaryotes and the majority of Bacteria in a single group, are also interpreted to show HGT when a few bacterial species branch with Archaea or Eukaryotes. Other interpretations of these data are possible, however [e.g., evolution in a common ancestor followed by excessive independent gene loss, unrecognized paralogy, and extreme differences in substitution rates (Philippe and Forterre 1999; Gogarten et al. 1996)]. For example, the observation that some Bacteria contain A/V-type ATPases can be explained by the presence of two types of molecules (eukaryotic/archaeal and bacterial) in the common ancestor followed by independent loss of the archaeal form in the majority of Bacteria and loss of the bacterial form in virtually all of the Archaea (Tsutsumi et al. 1991; Forterre et al. 1993). The many independent and convergent losses of A/V-ATPases and F-ATPases that are required to explain this distribution lead us to reject the hypothesis that both the A/V- and the F-ATPase were already present in the last common ancestor. The presence of A/V-ATPases in *Thermus, Deinococcus,* and the few other Bacteria in which they have been found can be explained more simply by horizontal transfer of genes into the Deinococcaceae lineage (Hilario and Gogarten 1993; Olendzenski and Gogarten 1999). Similarly, trees showing distribution of genes in a few, distantly related species could have arisen by the evolution of the molecule in a common ancestor, followed by many instances of independent loss, but are more simply explained by HGT from one organism where the molecule evolved into a few distantly related species. For each postulated interdomain horizontal transfer considered individually, an alternative explanation based on unrecognized gene duplication and parallel gene losses cannot be completely ruled out; however, if one applies this explanation to the many interdomain horizontal transfers that we have detected in the *Deinococcus radiodurans* genome, and the many additional interdomain transfers detected by others (for a review see Doolittle 1999), one would be forced to postulate a multitude of parallel gene losses and an extremely complex last common ancestor that would have contained all possible biochemical pathways, many of them in duplicate form (Hilario and Gogarten 1993; Doolittle 2000).

Open reading frames that are candidates for genes frequently transferred among the three domains would be expected to have HSP values of nearly the same magnitude compared to a Bacterial, Archaeal, or Eukaryotic reference genome, i.e., a phylogenetic signature should be absent for these ORFs. If these transfers occurred recently, the HSP values should be high. Using this rationale we selected those ORFs that, in a three-dimensional plot of HSP values, fell within a cone surrounding the $x = y = z$ line (Fig. 5). The ORFs within this cone were ranked according to their distance from the origin. In contrast to our expectation, close inspection of the phylogenies obtained for the 17 most conserved ORFs (Table 2) did not support our hypothesis (i.e., that these genes had been frequently and recently transferred between the domains). If genes were being exchanged frequently between Archaea and Bacteria, we would expect to see these sequences interspersed with each other in phylogenetic analyses. This is not the case. For most of these data sets, if one considers only the putative orthologous sequences, the Archaea form one or two monophyletic groups. While many of these data sets are compatible with one or two ancient horizontal transfer events from the Bacteria to the Archaea, or vice versa, none of these data sets reveals more recent or more frequent transfers. Many of the calculated phylogenies do support acquisition of bacterial forms by eukaryotes, but these might represent transfers associated with the endosymbiosis that gave rise to mitochondria or plastids. We believe, instead, that these ORFs represent genes that are very highly conserved and that have not diverged greatly during the course of evolution.

The techniques implemented here were able to extract conserved genes from the *Deinococcus radiodurans* genome that have undergone interdomain horizontal transfer. The use of phylogenetic analysis beyond BLAST comparisons was critical in verifying candidate horizontally transferred genes (Logsdon and Faguy 1999). While it remains difficult to pinpoint these transfer events on the tree of life, the recipient of the transfer event that provided the A/V-ATPase and the prolyl RS to the Deinococcaceae appears to have lived before the divergence of *Thermus* and *Deinococcus.* The prolyl RS phy-

logeny suggests that the donor was a eukaryote or an organism at the base of the eukaryotes. While several interdomain HGT transfer events could be detected, it seems that these transfers were infrequent and limited to relatively few genes compared to the whole genome.

At first glance this finding corroborates traditional systematic concepts of prokaryotic evolution. If only a few genes were horizontally transferred between divergent species, the remainder of the genes appears to be vertically inherited. Taking the Deinococcaceae as an example, HGT could be regarded as an important but rare exception and a natural bacterial classification reflecting shared ancestry and based on the majority of vertically inherited genes would emerge after weeding out the odd horizontal transfers. However, this optimistic conclusion is premature. Our finding is also compatible with the extreme assumption that the top-level taxonomic categories (i.e., the domains) are being exclusively determined by horizontal transfer frequency, i.e., Bacteria contain mainly bacterial genes because they more frequently exchange genes with other Bacteria than with Archaea. Our analysis does not address the relative importance of horizontal versus vertical inheritance in determining the top-level taxonomic categories; it only affirms the validity of two prokaryotic domains as useful and appropriate categories regardless of the mode by which they are maintained and came into existence. To decide to what extent prokaryotic taxonomic units might reflect HGT frequencies or vertical inheritance, analyses like the one described here need to be repeated and the results compared for interkingdom, interfamily, and intergenera HGTs.

Acknowledgments. We thank Parin Chaivisuthangkura, Scott McNamara, Marisa Merlo, and Andre King for help in screening the *Meiothermus ruber* library. We also thank W. Ford Doolittle for stimulating discussions regarding the extent of horizontal gene transfer in microbial evolution. This work was supported through the NASA Exobiology program.

References

Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of Escherichia coli K-12. Science 277:1453–1474

Bowman BJ, Bowman EJ (1996) Mitochondrial and vacuolar ATPases. In: Brambl R, Marzluf GA (eds) The Mycota III, biochemistry and molecular biology. Springer-Verlag, Berlin, pp 57–83

Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, Kerlavage AR, Dougherty BA, Tomb JF, Adams MD, Reich CI, Overbeek R, Kirkness EF, Weinstock KG, Merrick JM, Glodek A, Scott JL, Geoghagen NSM, Venter JC (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii.* Science 273:1058–1073

Deckert G, Warren PV, Gaasterland T, Young WG, Lenox AL, Graham DE, Overbeek R, Snead MA, Keller M, Aujay M, Huber R, Feldman RA, Short JM, Olsen GJ, Swanson RV (1998) The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus.* Nature 392:353–358

Doolittle RF, Handy J (1998) Evolutionary anomalies among the aminoacyl-tRNA synthetases. Curr Opin Genet Dev 8:630–636

Doolittle WF (1999) Phylogenetic classification and the universal tree. Science 284:2124–2129

Doolittle WF (2000) The nature of the universal ancestor and the evolution of the proteome. Curr Opin Struct Biol 10:355–358

Forterre P, Benachenhou-Lafha N, Confalonieri F, Duguet M, Elie C, Labedan B (1993) The nature of the last universal ancestor and the root of the tree of life, still open questions. BioSystems 28:15–32

Fraser CM, Casjens S, Huang WM, et al. (1997) Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi.* Nature 390:580–586

Fraser CM, Norris SJ, Weinstock GM, et al. (1998) Complete genome sequence of *Treponema pallidum* the syphilis spirochete. Science 281:375–388

Galtier N, Gouy M, Gautier C (1996) SeaView and Phylo_win, two graphic tools for sequence alignment and molecular phylogeny. Comput Applic Biosci 12:543–548

Goffeau A, Aert R, Agostini-Carbone ML, et al. (1997) The yeast genome directory. Nature 387(Suppl):1–105

Gogarten JP (1994) Which is the most conserved group of proteins? Homology—orthology, paralogy, xenology and the fusion of independent lineages. J Mol Evol 39:541–543

Gogarten JP, Kibak H, Dittrich P, Taiz L, Bowman EJ, Bowman BJ, Manolson MF, Poole RJ, Date T, Oshima T, Konishi J, Denda K, Yoshida M (1989) Evolution of the vacuolar H^+-ATPase: Implications for the origin of eukaryotes. Proc Natl Acad Sci USA 86: 6661–6665

Gogarten JP, Starke T, Kibak H, Fichmann J, Taiz L (1992) Evolution and isoforms of V-ATPase subunits. J Exp Biol 172:137–147

Gogarten JP, Hilario E, Olendzenski L (1996) Gene duplications and horizontal gene transfer during early evolution. In: Roberts DMcL, Sharp P, Alderson G, Collins M (eds) Evolution of microbial life, Society for General Microbiology 54. Cambridge University Press, Cambridge, pp 267–292

Hensel R, Demharter W, Kandler O, Kroppenstedt M, Stackebrandt E (1986) Chemotaxonomic and molecular-genetic studies of the genus *Thermus:* Evidence for a phylogenetic relationship of *Thermus aquaticus* and *Thermus ruber* to the genus *Deinococcus.* Int J Syst Bact 36:444–453

Hilario E, Gogarten JP (1993) Horizontal transfer of ATPase genes—The tree of life becomes a net of life. BioSystems 31:111–119

Höner zu Bentrup K, Ubbink-kok T, Lolkem JS, Konings WN (1997) An Na^+-pumping V1VO-ATPase complex in the thermophilic bacterium *Clostridium fervidus.* J Bacteriol 179:1274–1279

Iwabe N, Kuna KI, Hasegawa M, Osawa S, Miyata T (1989) Evolutionary relationship opf archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc Natl Acad Sci USA 86:9355–9359

Kalman S, Mitchell W, Marathe R, Lammel C, Fan J, Hyman RW, Olinger L, Grimwood J, Davis RW, Stephens RS (1999) Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis.* Nature Genet 21:385–389

Koonin EV, Mushegian AR, Galperin MY, Walker DR (1997) Comparison of archaeal and bacterial genomes: Computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea. Mol Microbiol 25:619–637

Kunst F, Ogasawara N, Moszer I, et al. (1997) The complete genome sequence of the gram-positive bacterium *Bacillus subtilis.* Nature 390:249–256

Logsdon JM, Faguy DM (1999) *Thermotoga* heats up lateral gene transfer. Curr Biol 9(19):R747–R751

Maidak BL, Cole JR, Parker CT Jr, Garrity GM, Larsen N, Li B, Lilburn TG, McCaughey MJ, Olsen GJ, Overbeek R, Pramanik S,

Schmidt TM, Tiedje JM, Woese CR (1999) A new version of the RDP (Ribosomal Database Project). Nucleic Acids Res 27:171–173

Nagel GM, Doolittle RF (1995) Phylogenetic analysis of the aminoacyl-tRNA synthetases. J Mol Evol 40:487–498

Nelson KE, Clayton RA, Gill SR, et al. (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima.* Nature 399:323–329

Nobre MF, Truper HG, DaCosta MS (1996) Transfer of *Thermus ruber* (Longinova et al. 1984), *Thermus silvanus* (Tenreiro et al. 1995) and *Thermus chliarophilus* (Tenreiro et al. 1995) to *Meiothermus* gen. nov. as *Meiothermus silvanus* comb.nov., and *Meiothermus chliarophilus* comb. nov., respectively, and emendation of the genus *Thermus.* Int J Syst Bact 46:604–606

Olendzenski L, Gogarten JP (1999) Gene transfer in early evolution. In: Seckbach J (ed) Enigmatic microbes and life in extreme environments. Kluwer Academic, Dordrecht, pp 15–27

Philippe H, Forterre P (1999) The rooting of the universal tree of life is not reliable. J Mol Evol 49:509–523

Radax C, Sigurdsson O, Greggvidsson GO, Aichinger N, Gruber C, Kristjansson JK, Stan-Lotter H (1998) F- and V-ATPases in the genus *Thermus* and related species. Syst Appl Microbiol 21:12–22

Ribeiro S, Golding BG (1998) The mosaic nature of the eukaryotic nucleus. Mol Biol Evol 15:779–788

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

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Starke T, Gogarten JP (1993) A conserved intron in the V-ATPase A subunit genes of plants and algae. FEBS Lett 315:252–258

Stehlin C, Burke B, Yang F, Lilu H, Shiba K, Musier-Forsyth K (1998) Species-specific differences in the operational RNA code for aminoacylation of tRNApro. Biochemistry 37:8605–8613

Stephens RS, Kalman S, Lammel CJ, Fan J, Marathe R, Aravind L, Mitchell WP, Olinger L, Tatusov RL, Zhao Q, Koonin EV, Davis RW (1998) Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis.* Science 282:754–759

Strimmer K, von Haeseler A (1996) Quartet puzzling: A quartet maximum-liklihood method for reconstructing tree toplogies. Mol Biol Evol 13:964–969

Sumi M, Sato MH, Denda K, Date T, Yoshida M (1992) A DNA fragment homologous to F1-ATPase beta subunit was amplified from genomic DNA of *Methanosarcina barkeri:* Indication of an archaebacterial F-type ATPase. FEBS Lett 314:207–210

Sumi M, Yohda M, Koga Y, Yoshida M (1997) F0F1-ATPase genes from an archaebacterium, *Methanosarcina barkeri.* Biochem Biophys Res Commun 241:427–433

Swofford DL (2000) Paup*4.0 Beta4a. Sinauer Associates, Sunderland, MA

Takase K, Kakinuma S, Yamato I, Konishi K, Igarashi K, Kakinuma Y (1994) Sequencing and characterization of the ntp gene cluster for vacuolar-type Na(+)-translocating ATPase of *Enterococcus hirae.* J Biol Chem 269:11037–11044

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876–4882

Tsutsumi S, Denda K, Yokoyama K, Oshima T, Date T, Yoshida M (1991) Molecular cloning of genes encoding major two subunits of a eubacterial V-Type ATPase from *Thermus thermophilus.* Biochim Biophys Acta 1098:13–20

Tumbula D, Vothknecht UC, Kim HS, Ibba M, Min B, Li T, Pelaschier J, Stathopoulos C, Becker H, Soll D (1999) Archaeal aminoacyl-tRNA synthesis: Diversity replaces dogma. Genetics 152:1269–1276

White O, Eisen JA, Heidelberg JF, et al. (1999) Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. Science 286:1571–1577

Williams R, Sharp R (1995) The taxonomy and identification of *Thermus.* In: Williams R, Sharp R (eds) Thermus species. Biotechnology handbooks 9. Plenum Press, New York, pp 1–42

Woese CR (1987) Bacterial evolution. Microbiol Rev 51:221–271

Wolf YI, Aravind L, Grishin NV, Koonin EV (1991) Evolution of aminoacyl-tRNA synthetases—Analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. Genome Res 9:689–710

Yokoyama K, Oshima T, Yoshida M (1990) *Thermus thermophilus* membrane-associated ATPase; Indication of a eubacterial V-type ATPase. J Biol Chem 265:21946–21950