JOURNAL OF BACTERIOLOGY, Oct. 1997, p. 6360–6366 Vol. 179, No. 20
0021-9193/97/$04.00+0

The Presence of a *dnaK* (HSP70) Multigene Family in Members of the Orders *Planctomycetales* and *Verrucomicrobiales*

NAOMI WARD-RAINEY,* FRED A. RAINEY,† AND ERKO STACKEBRANDT

DSMZ—German Collection of Microorganisms and Cell Cultures, D-38124 Braunschweig, Germany

Received 6 March 1997/Accepted 10 July 1997

Sequences of the *dnaK* gene, coding for the 70-kDa heat shock protein (HSP70), were determined for six members of the order *Planctomycetales*, including representatives of three genera, and for the only cultivated member of the order *Verrucomicrobiales*, *Verrucomicrobium spinosum*. A fragment of the *dnaK* gene was amplified from these strains by PCR with oligonucleotide primers targeting regions of the *dnaK* gene that are conserved at the amino acid level, and the resulting PCR products were cloned into a plasmid vector. Sequence analysis of the cloned *dnaK* fragments revealed the presence of two different types of *dnaK* sequence in one of the planctomycete strains, *Planctomyces maris*, and in *V. spinosum*. Only one type of *dnaK* sequence was found for each of the remaining strains. Phylogenetic analysis of the partial sequence data suggested that the majority of planctomycete strains, including one of the *Planctomyces maris* sequences, form a coherent phylogenetic group branching adjacent to other main lines of descent within the domain *Bacteria*, as has been shown previously by 16S rRNA sequence analysis. One of the two *V. spinosum dnaK* sequences also appears to constitute a separate lineage within the gram-negative bacteria. Each of the remaining sequences from *P. maris* and *V. spinosum*, together with the single sequence obtained from *Planctomyces limnophilus*, appeared to be unrelated to the other planctomycete sequences and to occupy a position distant from that of other gram-negative bacteria. The phylogenetic diversity of *dnaK* sequences exhibited by *P. maris* and *V. spinosum* was comparable to that found in *Synechococcus* sp. strain PCC7942 and *Escherichia coli*, the only other prokaryotes for which a *dnaK* multigene family has been demonstrated.

The 70-kDa heat shock protein (HSP70) is one of a set of proteins referred to as stress or heat shock proteins (HSPs), which are synthesized in greater amounts under conditions of stress, including elevated temperature, hypoxia, and exposure to ethanol, and are thought to protect the cell from the toxic effects of these stresses (21). HSP70 also constitutes a major cellular protein under normal growth conditions and is essential for normal growth (6, 10). In eukaryotes, HSP70 is part of a multigene family, some members of which are present in different cellular compartments or expressed under different physiological conditions (18, 42, 46). The gene encoding a protein related to HSP70 in the domain *Bacteria* is called *dnaK* (2). The only prokaryotes for which a *dnaK* multigene family have been previously demonstrated are the cyanobacterium *Synechococcus* sp. strain PCC7942 (25, 26) and *Escherichia coli* (32).

HSP70 is one of a number of proteins which are both ubiquitous and highly conserved in their amino acid sequence and therefore are useful for measuring phylogenetic relationships among members of the *Bacteria* and between *Bacteria* and members of the other two domains (4, 7, 14, 16). In most cases, the phylogenies obtained from conserved protein sequences are compared against those obtained from 16S rRNA and rDNA, which can be considered the benchmark molecule for the reconstruction of phylogenetic relationships among members of the *Bacteria* (27, 36, 47). Some protein sequences, for example ATPase β-subunits and elongation factor EF-Tu (22), produce trees that correlate well with those obtained from small-subunit rRNA data, supporting the phylogenetic distinctness of the domains *Bacteria* and *Archaea* (47, 48) and recovering similar phylogenies within the domain *Bacteria*. Some of these proteins allow better resolution of phylogenetic relationships within groups of closely related members of the *Bacteria* than can be obtained with 16S rRNA, for example, phylogenetic analysis of DNA gyrase subunit B sequences for members of the genera *Pseudomonas* (49) and *Acinetobacter* (50). Another category of proteins produce phylogenetic trees that appear to contradict the three-domain classification system based on rRNA sequence analysis of Woese et al. (48) and the common ancestry of the domains *Archaea* and *Eucarya* based on protein sequence data from duplicated gene families (11, 19). These proteins, including HSP70 (13, 14, 16, 17), glutamine synthetase (5, 33, 40), and glutamate dehydrogenase (3, 33), indicate a specific relationship between the *Eucarya* and the gram-negative bacteria, on one hand, and between the *Archaea* and the gram-positive bacteria, on the other hand, rather than between the *Archaea* and the *Eucarya*. The HSP70 tree topology is supported by shared HSP70 sequence signatures characteristic for certain groupings (13).

Within the domain *Bacteria*, *dnaK* sequences are available for relatively few lineages, with no sequences from representatives of the orders *Planctomycetales* and *Verrucomicrobiales*. The phylogenetic distinctness of members of the orders *Planctomycetales* (31, 34, 35, 43) and *Verrucomicrobiales* (44) and the phylogenetic diversity within these orders (43, 44) justify their inclusion in attempts to reconstruct phylogenetic relationships among members of the domain *Bacteria* and between the *Bacteria*, *Archaea*, and *Eucarya*, using HSP70 sequences. To allow comparative phylogenetic analyses with a more representative selection of phylogenetic diversity within the domain *Bacteria*, we obtained *dnaK* sequence data from six members of the order *Planctomycetales* and from *Verrucomicrobium spinosum.*

* Corresponding author. Present address: Department of Microbiology, Life Sciences Building, Louisiana State University, Baton Rouge, LA 70803. Phone: (504) 334-2127. Fax: (504) 388-2597.

† Present address: Department of Microbiology, Louisiana State University, Baton Rouge, LA 70803.

TABLE 1. Bacterial strains for which HSP70 sequence data were obtained

Species or strain	Strain name	Medium	Reference for medium	EMBL accession no.
Planctomyces limnophilus	DSM 3776^T	PYGV	37	Y14789
Planctomyces maris	DSM 8797^T	Difco marine agar		Y14790, Y14796
Pirellula marina	DSM 3645^T	M14	30	Y14795
Pirellula staleyi	DSM 6068^T	M14	30	Y14794
Strain 140		M13	30	Y14793
Verrucomicrobium spinosum	DSM 4136^T	M13	30	Y14791, Y14792

MATERIALS AND METHODS

Bacterial strains and culture conditions. Strains for which sequence data were obtained are listed in Table 1. The type strains of the validly described planctomycete species were obtained from the German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany. *V. spinosum* DSM 4136^T and strain 140 were provided by Heinz Schlesner, University of Kiel, Kiel, Germany. The strains were cultivated on solid media and incubated aerobically at 28°C for 6 days. The culture media used are listed in Table 1.

Isolation and purification of genomic DNA. Genomic DNA was isolated and purified as described previously (28).

PCR-mediated amplification of *dnaK*. *dnaK* was amplified by PCR in a model 480 apparatus (Perkin-Elmer, Foster City, Calif.) with the degenerate primers and the thermal profile described by Tilly et al. (41). The PCR mixture contained 1× PCR buffer (Boehringer, Mannheim, Germany), 200 μM each deoxynucleoside triphosphate (Boehringer), 50 to 100 ng of genomic DNA, and 5 μg of each amplification primer. The final volume of the PCR mixture was adjusted to 100 μl by the addition of distilled H_2O (dH_2O), and the reaction mixture was overlaid with 80 μl of sterile mineral oil.

Cloning of PCR-amplified *dnaK* fragments. PCR products were purified and concentrated with the Prep-A-Gene kit (Bio-Rad, Hercules, Calif.) as specified by the manufacturer. DNA was eluted in 25 μl of dH_2O. Purified PCR products were cloned into the plasmid vector pCR-Script (Stratagene, La Jolla, Calif.), and the ligation mixture was transformed into Epicurian Coli XL1-Blue MRF′ Kan supercompetent cells (Stratagene), as directed by the manufacturer. Plasmid DNA was obtained from recombinant transformant colonies by lysis of a small amount of cell material resuspended in 100 μl of dH_2O and heated to 98°C for 8 min in a thermal cycler. The lysate was centrifuged to remove bacterial debris, and 1 μl of the supernatant was used in a PCR to amplify the cloned inserts. The PCR conditions were as described for *dnaK* above, with the exception that the primers used were M13-20 and M13rev (5′GTAAAACGACGGCCAGT3′ and 5′GGAAACAGCTATGACCATG3′, respectively; 0.5 μg of each primer), which bind to sequences flanking the multiple-cloning site in the pCR-Script vector (39), and the following thermal profile was used: 28 cycles of 1 min of primer annealing at 52°C, 2 min of extension at 72°C, and 1 min of denaturation at 94°C. A final extension step of 5 min at 72°C was performed. Reamplified clone inserts were purified with the Prep-A-Gene kit (Bio-Rad) as described above, and elution was performed in 50 μl of dH_2O, 0.5 to 1.0 μl of which was used in each sequencing reaction.

Sequencing of cloned *dnaK* fragments. Sequencing reactions were performed with the PRISM ReadyReaction DyeDeoxy Terminator cycle-sequencing kit with AmpliTaq FS (Applied Biosystems, Foster City, Calif.) and a model 9600 Perkin-Elmer Cetus thermal cycler, as specified in the protocol and with the thermal profile recommended by Applied Biosystems. The clone inserts were sequenced with the M13-20 and M13rev primers flanking the pCR-Script cloning site. Sequencing reaction products were purified as recommended by Applied Biosystems and electrophoresed on a 6% (wt/vol) polyacrylamide sequencing gel for 12 h with an Applied Biosystems model 373A automated DNA sequencer.

Data analysis. Translated *dnaK* or HSP70 sequences were aligned manually with the ae2 sequence editor (24) against a database containing the following reference sequences (in alphabetical order [the strain is indicated where available; the accession number in parentheses refers to SWISSPROT unless indicated otherwise]): *Bacillus megaterium* (P05646), *Bacillus subtilis* 168/MB11 (P17820), *Borrelia burgdorferi* CA12 (P28608), *Brucella ovis* (Q05981), *Caulobacter crescentus* (P20442), *Chlamydia pneumoniae* TWAR (P27542), *Chlamydia trachomatis* serovar L2 (P17821), *Clostridium acetobutylicum* DSM 1731 (P30721), *Clostridium perfringens* (P26823), *Entamoeba histolytica* (cytoplasmic) (EMBL M84652), *Erysipelothrix rhusiopathiae* E1-6P (Q05647), *Escherichia coli* (P04475), *Giardia lamblia* (cytoplasmic) Portland1/ATCC 2088 (EMBL U04874), *Giardia lamblia* (endoplasmic reticulum) Portland1/ATCC 2088 (EMBL U04875), *Halobacterium cutirubrum* (P42372), *Haloarcula marismortui* (Q01100), *Homo sapiens* (cytoplasmic) (P08107), *Lactococcus lactis* MG1363 (P42368), *Leishmania major* (mitochondrial) WR300 P14834, *Methanosarcina mazei* S6 (P27094), *Mycobacterium leprae* (P19993), *Mycobacterium paratuberculosis* (Q00488), *Mycobacterium tuberculosis* Erdmann (P32723), *Mycoplasma genitalium* ATCC 33530/G-37 (P47547), *Pisum sativum* (chloroplast) (L03299), *Porphyra umbilica* (chloroplast) Avonport (P30723), *Saccharomyces cerevisiae* (cytoplasmic) S288C (P10591), *Saccharomyces cerevisiae* (endoplasmic reticulum) (P16474), *Saccharomyces cerevisiae* (mitochondrial) D273-10B (P12398), *Staphylococcus aureus* 912 (P45554), *Streptomyces coelicolor* A3(2) (Q05558), *Streptomyces griseus* (EMBL D14499), *Synechococcus* sp. strain PCC7942 (EMBL D28550, D28551, and D29968), *Thermoplasma acidophilum* clone 3 (EMBL L35529), and *Zea mays* (cytoplasmic) (P11143).

The dendrograms presented or described in this paper were constructed with the treeing algorithms contained in the PHYLIP package (9). SEQBOOT was used to generate 100 data sets from the sequence data by the bootstrapping approach (8). Similarities were calculated with PROTDIST and the PAM120 matrix. NEIGHBOR was used for the construction of neighbor-joining (29) trees from the sequence similarity values, and a consensus neighbor-joining tree was obtained with CONSENSE.

RESULTS

PCR-mediated amplification of *dnaK*. A fragment of the *dnaK* gene was successfully amplified by PCR from six planctomycete strains and *V. spinosum*. The amplified fragment was approximately 650 nucleotides in length and corresponded to the sequence encoding amino acid positions 148 through 364 of *Escherichia coli* HSP70 (2).

Direct sequencing of *dnaK* PCR products. Partial *dnaK* sequence data were obtained for six planctomycete strains and *V. spinosum*. The number of clones sequenced for each strain were as follows: *Planctomyces limnophilus*, 15; *Planctomyces maris*, 10; *Planctomyces brasiliensis*, 11; *Pirellula marina*, 10; *Pirellula staleyi*, 9; strain 140, 3; *V. spinosum*, 9.

The deduced amino acid sequences determined in this study showed sequence similarity to other currently available HSP70 sequences. Two different *dnaK* sequences, encoding different HSP70 proteins, were obtained for each of *P. maris* (clones 1 and 8) and *V. spinosum* (clones 1 and 4). Only one type of *dnaK* sequence was detected for each of the remaining strains. Multiple *dnaK* genes have been previously found in only two prokaryotes, the cyanobacterium *Synechococcus* sp. strain PCC7942, for which three different *dnaK* sequences were found (25, 26), and *Escherichia coli*, for which a gene bearing sequence similarity to *dnaK* was described in 1994 (32). The multiple *dnaK* sequences within the strains investigated in this study were not closely related to each other. For example, the *P. maris* clone 8 sequence shared only 41.7% amino acid identity with the *P. maris* clone 1 sequence. This value is significantly lower than the level of amino acid identity between *P. maris* clone 8 and other planctomycetes (70.7 to 75.7%) and between clone 8 and other representatives from the domains *Bacteria* (59.8 to 69.1%). The two *dnaK* sequences from *V. spinosum* (clones 1 and 4) shared 50.6% amino acid identity. This is comparable to the 49.6% amino acid identity shared by the two most unrelated *dnaK* sequences from *Synechococcus* sp. strain PCC7942 (25, 26).

Phylogenetic analyses. To determine whether the partial *dnaK* sequences obtained in this study could be used with confidence for phylogenetic analysis, two data sets were used. Data set 1 contained the translated partial *dnaK* sequences determined in this study and translated reference *dnaK* sequences from representatives of the main lines of descent

```
E.coli           AQRQATKDAGRIAGLEVKRIINEPTAAALAYGLDKG---TGNRTIAVYDL 26
P.limnophilus    EQRRATIAAGQMAGLKVERIVNEPTAAAIAYGLHEA---DSQKTAVIIDL 21
1P.maris         HQRNATKQAGELAGLNVRRIINEPTAAALTYGFHDR---QAEKKLIVIDL 21
4V.spinosum      AQRNATKLAGEQAGLTVERILAEPTAAALAYGLDKL---EEHRRIAVYDL 23
1V.spinosum      SQRNATKAAGEIAGLTVRRIINEPTAASLAYGLDK----KKDEKIAVYDL 21
Strain140        AQRQATKDAGQIAGLEVARIINEPTAAALAYGLDK----KKDESIIVFDL 18
Pi.staleyi       AQRQATKDAGQIAGLEVARIINEPTAAALAYGLDK----KKNEKIVVFDL 19
Pi.marina        AQRQATKDAGQIAGLEVARIINEPTAAALAYGLGK----QKSEKIAVFDL 20
8P.maris         AQRQATKDAGQISGLEVSRIINEPTAAALAYGLEK----KSDEKIVVFDF 16
S.cerevisiae(m)  SQRQATKDAGQIVGLNVLRVVNEPTAAALAYGLEKS----DSKVVAVFDL 17
C.trachomatis    SQRASTKDAGRIAGLDVKRIIPEPTAAALAYGIDKE----GDKKIAVFDL 18
1PCC7942         AQRQATKDAGAIAGLEVLRIVNEPTAAALSYGLDKL---HENSRILVFDL 19
2PCC7942         SQRQATKDAGKIAGLEVLRIINEPTAAALAYGLDK----KSNERILVFNL 19
3PCC7942         SQRQATRDAGRIAGLEVKRILNEPTAASLAYGLDR----RDNQTILVFDL 21
M.tuberculosis   AQRQATKDAGQIAGLNVLRIVNEPTAAAPGYGLDKG---EKEQRILVFDL 20
H.marismortui    RQRQATKDAGKIAGFEVERIVNEPTAAAMAYGLDDE----SDQTVLVYDL 19
B.subtilis       AERQATKDAGKIAGLEVERIINEPTAAALAYGLDKT---DEDQTILVYDL 20
Me.mazei         SQRQATKDAGAIAGLEVLRIINEPTAASLAYGLDKG---DIDQKILVYDL 18
Z.mays(C)        SQRQATKDAGVIAGLNVMRIINEPTAAAIAYGLDKKATSSGEKNVLIFDL 21

E.coli           GGGTFDISIIEI-----DEVDGEKTFEVLATNGDTHLGGEDFDSRLINYL 42
P.limnophilus    GGGTFDVSI--V-------EMFEGVLEIRASAGEIFLGGEDFTDACVSQI 35
1P.maris         GGGTFDVTA--M-------EVFEGTLEIISTAGESMLGGEDFTDRILAKV 37
4V.spinosum      GGGTFDVSV--L-------EMRDGVFQVLSTAGDTQLGGDDIDRSLAEWI 37
1V.spinosum      GGGTFDISVLEI---------GDGVFEVLATDGDTHLGGDDWDNKLIQWI 34
Strain140        GGGTFDVSVLEVADSGDEEQESR-VFQVISTSGDTHLGGDDFDEALIHYV 32
Pi.staleyi       GGGTFDVSILEVAG-ADSAET--KVFEVISTNGDTHLGGDDFDEALIHFV 35
Pi.marina        GGGTFDISILEVSPPEGEEEGERTVFEVISTNGDTHLGGDDFDEELIHYV 35
8P.maris         GGGTFDVSVLEV---------GDEVIETLSTNGDGHLGGDDFDEELINHI 29
S.cerevisiae(m)  GGGTFDISILDI---------DNGVFEVKSTNGDTHLGGEDFDIYLLREI 31
C.trachomatis    GGGTFDISILEI---------GDGVFEVLSTNGDTHLGGDDFDGVIINWM 32
1PCC7942         GGSTLDVS--IL-------QLGDSVFEVKATAGNNHLGGDDFDAVIVDWL 32
2PCC7942         GGGTFDVS--VL-------EVGDGVFEVLATSGDTHLGGDDFDKKIVDFL 31
3PCC7942         GGGTFDVS--VL-------KVGNGVFEVKATSGDTQLGGNDFDRRIVDWL 37
M.tuberculosis   GGGTFDVSLLEI---------GEGVVEVRATSGDNHLGGDDWDQRVVDWL 34
H.marismortui    GGGTFDVSILDL---------GGGVYEVVATNGDNDLGGDDWDHAIIDYL 36
B.subtilis       GGGTFDVSILEL---------GDGVFEVRSTAGDNRLGGDDFDQVIIDHL 34
Me.mazei         GGGTFDVSILEL---------GGGVFEVKSTSGDTHLGGDDFDQRVIDYL 32
Z.mays(C)        GGGTFDVSLLTI---------EEGIFEVKATAGDTHLGGEDFDNRMVNHF 37
```

FIG. 1. Alignment of partial HSP70 sequences from planctomycete species and *V. spinosum* with other HSP70 sequences from members of the *Bacteria*, *Archaea*, and *Eucarya*. Amino acid positions 148 to 364 (*E. coli* sequence) are shown. The species names are as follows: E.coli, *Escherichia coli*; P.limnophilus, *Planctomyces limnophilus*; 1P.maris, *Planctomyces maris* clone 1; 4V.spinosum, *Verrucomicrobium spinosum* clone 4; 1V.spinosum, *Verrucomicrobium spinosum* clone 1; Strain140, strain 140; Pi.staleyi, *Pirellula staleyi*; Pi.marina, *Pirellula marina*; 8P.maris, *Planctomyces maris* clone 8; S.cerevisiae(m), *Saccharomyces cerevisiae* (mitochondrial); C.trachomatis, *Chlamydia trachomatis*; 1PCC7942, *Synechococcus* sp. strain PCC7942 *dnaK1*; 2PCC7942, *Synechococcus* sp. strain PCC7942 *dnaK2*; 3PCC7942, *Synechococcus* sp. strain PCC7942 *dnaK3*; M.tuberculosis, *Mycobacterium tuberculosis*; H.marismortui, *Haloarcula marismortui*; B.subtilis, *Bacillus subtilis*; Me.mazei, *Methanosarcina mazei*; Z.mays(C), *Zea mays* (cytoplasmic).

within the domains *Archaea*, *Bacteria*, and *Eucarya*, including the three *Synechococcus dnaK* sequences described by Nimura et al. (25, 26) (Fig. 1). The reference sequences in data set 1 were truncated to include only the positions present in the planctomycete sequences, i.e., amino acid residues 148 through 364 (*E. coli* nomenclature). Data set 2 contained the truncated reference sequences as in data set 1 but none of the sequences determined in this study.

Comparison of sequences in data set 1. Phylogenetic analysis with data set 1 (truncated reference sequences plus planctomycete partial sequences) suggested that the *dnaK* sequences of the planctomycetes and *V. spinosum* fall in two different places in the tree (Fig. 2). The consensus neighbor-joining tree presented in Fig. 2 shows that *Pirellula marina*, *Pirellula staleyi* 140, and *P. maris* clone 8 form a cluster within the radiation of the group containing the members of the class *Proteobacteria*, the *Chlamydia* species, and *Borrelia burgdorferi* as representatives of the gram-negative bacteria, and the mitochondrial HSP70s. *V. spinosum* clone 1 is also evident as a lineage within this group, branching adjacent to the mitochondrial sequences. The planctomycetes appear to form a coherent phylogenetic group, as evidenced by this group being recovered in 98% of trees generated (Fig. 2). The phylogenetic positions of these strains agree with the tree topologies produced from 16S rDNA sequence data (20, 34, 35, 44); i.e., the planctomycetes and *V. spinosum* appear not to be closely related to other members of the domain *Bacteria*. These planctomycete sequences and the sequence of *V. spinosum* clone 1 will be referred to from here on as the group I sequence type. The other planctomycete sequences (*P. limnophilus* and *P. maris* clone 1), together with *V. spinosum* clone 4 and *Mycoplasma genitalium* (from now on referred to as the group II sequence type), appear to occupy a position adjacent to the members of the domain *Eucarya*. The grouping of these sequences (group II) with the *Eucarya* was recovered in only 33% of the trees generated from the bootstrapped data set, indicating that this is not a stable topology. The phylogenetic distance between the two different *dnaK* sequences of *P. maris* and between those of *V. spinosum* is at least as great as that previously reported for the three *dnaK* homologs of *Synechococcus* sp. strain PCC7942 (25, 26).

Comparison of partial reference sequences only (data set 2). To determine whether the inclusion in the analysis of partial *dnaK* sequences from the planctomycetes and *V. spinosum* had affected the phylogenetic relationships between the reference sequences, a consensus tree was constructed from neighbor-joining trees based on 100 resamplings of data set 2 (truncated reference sequences). The tree obtained (data not shown) recovered the same groupings as were found when the sequences determined in this study were included (Fig. 2), with the exception that *Mycoplasma genitalium* was found to group with the other low-GC gram-positive bacteria, as reported recently

```
E.coli           VEEFKKDQGIDLRNDPLAMQRLKEAAEKAKIELSSAQQTDVNLPYI-TA- 56
P.limnophilus    LNQAGMKFEHTEMQEPLRVSRLRRECEQAKRRLTNEASTEVRLPN---S- 53
1P.maris         LTAQNLQVEIAEVKHPLLVSRLKQECEAAKCILARETEAKIRIPNL---- 52
4V.spinosum      WDNCGLDPAMATKTSVLQRMRLIAAAEEAKKHLSSVSATEVLLPFF--L- 52
1V.spinosum      ISDFKADSR-HPQRSARRPATHQGRAEKAKIALSSSQSYDLNLPFI-TA- 51
Strain140        ADEFRK-TRDRLRNDAMALQRLQEACEKAKKELSTLPETDINLPFI-TM- 49
Pi.staleyi       AGEFQKENSVDLRKDPWHCSACK-SLRTAKKELSTLPQTEINLPFI-TM- 49
Pi.marina        ADEFKKEHGVVLRNDTMALQRLQEACEKAKKELSSQAQTDINLPFI-TA- 50
8P.maris         ADSFKKEQGIDLRSDAMALQRLREAAEKAKKELSSSQTTDINLPFI-TA- 44
S.cerevisiae(m)  VSRFKTETGIDLENDRMAIQRIREAAEKAKIELSSTVSTEINLPFI-TA- 48
C.trachomatis    LDEFKKQEGIDLSKDNMALQRLKDAAEKAKIELSGVSSTEINQPFI-TI- 43
1PCC7942         ADNFLKAESIDLRQDKMAIQRLREASEQAKIDLSTLPTTTINLPFIATAT 50
2PCC7942         AGEFQKNEGIDLRKDKQALQRLTEAAEKAKIELSSATQTEINLPF-IT-A 48
3PCC7942         AEQFLEAEGIDLRRDRQALQRLIEAAEKAKIELSGVSVTDINLPFI-TAT 54
M.tuberculosis   VDKFKGTSGIDLTKDKMAMQRLREAAEKAKIELSSSQSTSINLPYI-TV- 48
H.marismortui    ADEFEAEHGIDLRDDRQALQRLTEAAEEAKIELSSRKETRINLPFI-AT- 53
B.subtilis       VSEFKKENGIDLSKDKMALQRLKDAAEKAKKDLSGVSSTQISLPFI-TA- 45
Me.mazei         LAEFKKSEGIDLSKDKAVLQRLKDAAEKAKIELSGVANTNINLPFL-TV- 46
Z.mays(C)        VQEFKRKNKKDISGNPRALRRLRTACERAKRTLSSTAQTTIEI-----D- 57

E.coli           --D-ATGPKHMNIKVTRAKLESLVEDLVNRSIEPLKVALQDAGLSVSDID 68
P.limnophilus    -QG-EIEPDAPRYAITREMFDLWTKPTLDRILSPIRRALGDAGLKRQEID 70
1P.maris         -QG-KMEADAATHTMTRDEFLELADPLVKRLARPIAKAVRDSRIAPQDFD 69
4V.spinosum      ------EQQSVQVTVSRSDLEGLTQPLMERTRKHCLRALADARCEVTDLH 67
1V.spinosum      --D-ATGPKHIMKSLSRAKMEQLTDDLFERTVKPVRECLNAAKLDASKID 65
Strain140        --D-ASGPKHLTMKITRSKFEELIDKLVDRCRGPVLQALKDAGMSPSDID 61
Pi.staleyi       --A-GGSPKHLQMTITRARFEELVDGLIERCRKPVLQALKDAKMSPSDID 63
Pi.marina        --D-ASGPKHLQMSISRSKFEELTDSLIQRCRVPVEKALADANLKPSDID 61
8P.maris         --D-SSGAKHLQMAITRSEFEKLIDPLVERCRKPVEQAMKDAGLSASEID 56
S.cerevisiae(m)  --D-ASGPKHINMKFSRAQFETLTAPLVKRTVDPVKKALKDAGLSTSDIS 62
C.trachomatis    --D-ANGPKHLALTLTRAQFEHLASSLIERTKQPCAQALKDAKLSASDID 59
1PCC7942         VDG-APEPKHIEVELQREQFEVLASNLVQATIEPIQQALKDSNLTIDQID 60
2PCC7942         TQD-G--PKHLDLTLTRAKFEELASDLIDRCRIPVEQAIKDAKLALSEID 61
3PCC7942         -ED---EPKHLETRLTRSEFEALCEDLLERMMRPLRRALKDARLQPQDID 67
M.tuberculosis   --DADKNPLFLDEQLTRAEFQRITQDLLDRTRKPFQSVIADTGISVSEID 61
H.marismortui    --T-DDGPLDLEQKITRAKFESLTEDLIERTLGPTEQALADADYTKSDID 68
B.subtilis       --G-EAGPLHLELTLTRAKFEELSSHLVERTMGPVRQALQDAGLSASEID 60
Me.mazei         --GTDGEPKHMDIDLTRAQFQKMTEDLLEKTLVSMRRALSDAKLTPNDLD 60
Z.mays(C)        --S-LFEGIDFTPRSSRARFEELNMDLFRKCMEPVEKCLRDAKMDKSSVH 69

E.coli           DVILVGGQTRMP-MVQKKVAEFF-GKEP                       74
P.limnophilus    EVILAGGASRMP-SLIKRIEELF--ERP                       77
1P.maris         SVILVGGATRME-AVRGFIREFF-GVEP                       79
4V.spinosum      DVILVGGSTRMP-LVRETVRSIF-QREP                       75
1V.spinosum      ELVLVGGMTRMP-KVVETARKLA-GKEP                       74
Strain140        EIVLVGGSTRVP-KVRQAVKEIF-GKEP                       68
Pi.staleyi       EVVLVGGSTRVP-KVQKLVKDIF-GKEP                       68
Pi.marina        EVVLVGGSTRIP-KVAEMVKKIF-GKDP                       67
8P.maris         EVVLVGGSTRVP-KVQEFVKKIF-GKEP                       61
S.cerevisiae(m)  EVLLVGGMSRMP-KVVETVKSLF-GKDP                       67
C.trachomatis    DVLLVGGMSRMP-AVQAVVKEIF-GKEP                       65
1PCC7942         RILLVGGSSRIPAIQQAVQKFFG-GKTP                       69
2PCC7942         EIVLVGGSTRIP-AVQAIVKQMT-GKEP                       69
3PCC7942         EVVLVGGSTRMPMVQQLVRSLI--GREP                       74
M.tuberculosis   HVVLVGGSTRMP-AVTDLVKELTGGKEP                       70
H.marismortui    EVILVGGSTRMP-QVQDQVEEMT-GQEP                       74
B.subtilis       KVILVGGSTRIP-AVQEAIKKET-GKEA                       69
Me.mazei         KVILVGGATRMP-AVVELVENFT-GKKP                       69
Z.mays(C)        DVVLVGGSTRIP-KVQQL-QDFFNGKEL                       75
```

FIG. 1—*Continued.*

by Falah and Gupta (7), rather than adjacent to the *Eucarya* as indicated, albeit with low statistical significance, in Fig. 2. The topology of Fig. 2 suggests that the inclusion of partial sequences from the planctomycetes and *V. spinosum* has distorted the reconstruction of phylogenetic relationships between *Mycoplasma genitalium* and other members of the *Bacteria.*

DISCUSSION

Comparison with previously published HSP-based phylogenies. The tree presented here reproduced approximately the same clusters of higher relationship seen in previous reports of HSP70-based phylogenies. The internal branching order of the clusters containing representatives of the actinomycete line of descent, the class *Proteobacteria*, and the domain *Eucarya* were almost exactly as previously depicted (17). Relationships observed in other analyses, such as between the chloroplasts and the cyanobacteria, between the *Chlamydia* species and *B. burgdorferi*, and between the mitochondria and the *Proteobacteria* (7), between the actinomycetes and the halophilic *Archaea* (14, 17), and between the *Clostridium* species and the methanogenic archaeon *Methanosarcina mazei* (16, 23), were also found in the analyses performed in this study. However, at the level of lower relationship, there were profound differences between the tree presented by Gupta and Singh (17) and that obtained in this study, as described above. It is not possible to determine with certainty the cause of discrepancies between the tree topologies produced in this study and those of previous investigations, but the use of partial sequences may be responsible.

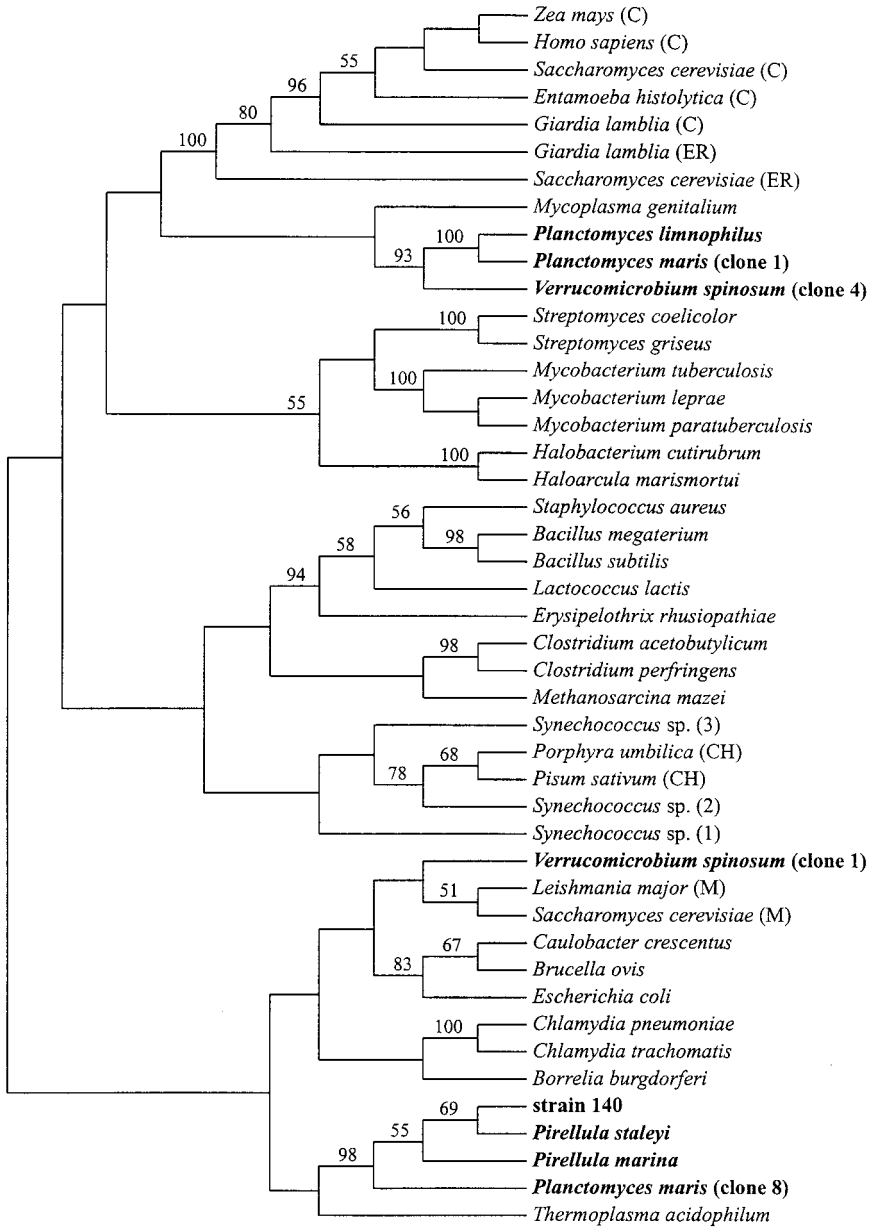

FIG. 2. HSP70-based consensus neighbor-joining tree, generated from 100 resamplings of data set 1 (partial sequence data), indicating the phylogenetic relationships among members of the order *Planctomycetales*, *V. spinosum*, and representatives of the domains *Bacteria*, *Archaea*, and *Eucarya*. The lengths of the branches do not indicate phylogenetic distance. Numbers at nodes refer to bootstrap values. Bootstrap values lower than 50% are not shown. C, cytoplasmic; CH, chloroplast; M, mitochondrial; ER, endoplasmic reticulum.

Phylogenetic conclusions. The first conclusion of this study (from examination of the group I type sequences) is that analysis of HSP70 sequences suggests that the planctomycetes and *V. spinosum* form two phylogenetic lineages within the domain *Bacteria* that are not closely related to members of other lineages. The second observation (from examination of the group II sequences) is that the planctomycetes and *V. spinosum* possess a HSP70 protein that appears to have some phylogenetic relationship to those of members of the domain *Eucarya.*

The latter finding suggests two possible explanations. The ancestral gene of the group II sequences could be viewed as a candidate for the gram-negative partner in the fusion event that gave rise to the eukaryotic cell, as proposed by Gupta and coworkers (12, 13, 15). A common ancestry of the planctomycetes and *V. spinosum* is suggested, albeit not with high statistical significance, by analyses of 16S rRNA and rDNA sequences (1, 44).

An alternative explanation for the apparent relationship between the group II sequences and the *Eucarya* is a higher than normal rate of evolution in the *dnaK* genes of the planctomycetes and *Verrucomicrobium*, which caused the loss of positions that are conserved in bacterial and archaeal *dnaK* sequences and, as a consequence, a false similarity to the eukaryotes and to each other. Rapid evolution in planctomycetes has been previously suggested to be the origin of the large phylogenetic distances that separate planctomycetes from other bacteria on

the basis of 16S rRNA and rDNA sequence analysis (35). This explanation for the phylogenetic position of the group II sequences is supported by the presence in these sequences of a 4-amino-acid insertion at position 320 (in the *E. coli* sequence), which is reported to be a signature for members of the domains *Bacteria* and *Archaea* (13), and the absence of any of the signatures identified as unique to the domain *Eucarya* (13). Further evidence that the phylogenetic position of the group II sequences does not indicate common ancestry with the *Eucarya* is provided by the overall sequence similarity values. The group II sequences do not have a higher similarity to the *Eucarya* sequences than to the *Bacteria* and *Archaea* sequences (data not shown), which would be expected if group II and the *Eucarya* had a common ancestor. Although the possibility that the planctomycetes and *V. spinosum* acquired the group II sequence type by lateral transfer cannot be excluded, some authors (15) have argued against this route on the grounds that phylogenies reconstructed from other genes, e.g., glutamine synthetase, concur with those produced from *dnaK*/HSP70.

HSP70 multigene family. Three of the planctomycete strains investigated (*Pirellula marina*, *Pirellula staleyi*, and strain 140) were found to have only the group I type HSP70, one (*P. limnophilus*) had only the group II type, and both types were found in *P. maris* and *V. spinosum*. Two chromosomal loci for *dnaK* have been demonstrated for *P. limnophilus* (45) and strain 140 (unpublished data), suggesting that a group I type HSP70 sequence for *P. limnophilus* and a group II type HSP70 sequence for strain 140 might be present, and escaped detection in this study. The presence of only one *dnaK* locus in *Pirellula marina* (unpublished data) indicates that only the group I type sequence may be present in this strain. The number of *dnaK* loci in *P. maris* is not known. To determine whether the apparent similarity between the group II sequences and the domain *Eucarya* is genuine or artifactual, *dnaK* sequences should be obtained from more planctomycete strains and the presence of group I or group II sequence types should be determined. The same analyses should be performed for members of other phylogenetic lineages within the domain *Bacteria*, including those for which (group I type) *dnaK* sequences have already been found. The presence of undetected group II *dnaK* sequences is possible.

As a result of this investigation, nine additional partial *dnaK*/HSP70 sequences have been added to the database of available HSP70 sequences. This will allow HSP70 sequence comparison to be performed with a more representative selection of phylogenetic diversity within the domain *Bacteria*—an important consideration given that HSP70-based phylogenies have been cited as evidence for a model of evolution of the eukaryotic cell.

Although the use of partial sequences to reconstruct phylogenetic relationships, as described in this study, is not ideal and may contribute to branching-point instability, the results of our analyses suggest that *dnaK*-based phylogenetic analysis in planctomycetes and *V. spinosum* is worthy of further study. Complete *dnaK* sequence data from these strains may allow a more stable phylogeny to be reconstructed in the future.

ACKNOWLEDGMENTS

This work was supported by funds from the Deutsche Forschungsgemeinschaft (German Research Council) (grants Sta 12-1 and 12-2).

REFERENCES

1. **Albrecht, W., A. Fischer, J. Smida, and E. Stackebrandt.** 1987. *Verrucomicrobium spinosum*, a eubacterium representing an ancient line of descent. Syst. Appl. Microbiol. **10:**57–62.
2. **Bardwell, J. C. A., and E. A. Craig.** 1984. Major heat shock gene of *Drosophila* and the *Escherichia coli* heat inducible *dnaK* gene are homologous. Proc. Natl. Acad. Sci. USA **81:**848–852.
3. **Benachenhou-Lahfa, N., P. Forterre, and B. Labedan.** 1993. Evolution of glutamate dehydrogenase genes: evidence for two paralogous protein families and unusual branching patterns of the archaebacteria in the universal tree of life. J. Mol. Evol. **36:**335–346.
4. **Boorstein, W. R., T. Zeigelhoffer, and E. A. Craig.** 1994. Molecular evolution of the HSP70 multigene family. J. Mol. Evol. **38:**1–17.
5. **Brown, J. R., Y. Masuchi, F. T. Robb, and W. F. Doolittle.** 1994. Evolutionary relationships of bacterial and archaeal glutamine synthetase genes. J. Mol. Evol. **38:**566–576.
6. **Craig, E. A., P. J. Kang, and W. Boorstein.** 1990. A review of the role of 70kDa heat shock proteins in protein translocation across membranes. Antonie Leeuwenhoek **58:**137–146.
7. **Falah, M., and R. S. Gupta.** 1997. Phylogenetic analysis of mycoplasmas based on Hsp70 sequences: cloning of the *dnaK* (*hsp70*) gene region of *Mycoplasma capricolum*. Int. J. Syst. Bacteriol. **47:**33–37.
8. **Felsenstein, J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:**783–793.
9. **Felsenstein, J.** 1993. PHYLIP (phylogenetic inference package) version 3.5.1. Department of Genetics, University of Washington, Seattle.
10. **Gething, M. J., K. McCammon, and J. Sambrook.** 1992. Expression of wild-type and mutant forms of influenza hemagglutinin: the role of folding in intracellular transport. Cell **46:**939–950.
11. **Gogarten, J. P., H. Kibak, P. Dittrich, L. Taiz, E. J. Bowman, B. J. Bowman, M. F. Manolson, R. J. Poole, T. Date, T. Oshima, J. Konishi, K. Denda, and M. Yoshida.** 1989. Evolution of the vacuolar H^+-ATPase: implications for the origin of eukaryotes. Proc. Natl. Acad. Sci. USA **86:**6661–6665.
12. **Golding, G. B., and R. S. Gupta.** 1995. Protein-based phylogenies support a chimeric origin for the eukaryotic genome. Mol. Biol. Evol. **12:**1–6.
13. **Gupta, R. S., K. Aitken, M. Falah, and B. Singh.** 1994. Cloning of *Giardia lamblia* heat shock protein HSP70 homologs: implications regarding origin of eukaryotic cells and of endoplasmic reticulum. Proc. Natl. Acad. Sci. USA **91:**2895–2899.
14. **Gupta, R. S., and G. B. Golding.** 1993. Evolution of HSP70 gene and its implications regarding relationships between Archaebacteria, Eubacteria and Eukaryotes. J. Mol. Evol. **37:**573–582.
15. **Gupta, R. S., and G. B. Golding.** 1996. The origin of the eukaryotic cell. Trends Biochem. Sci. **21:**166–171.
16. **Gupta, R. S., and B. Singh.** 1992. Cloning of the HSP70 gene from *Halobacterium marismortui*: relatedness of archaebacterial HSP70 to its eubacterial homologs and a model for the evolution of the HSP70 gene. J. Bacteriol. **174:**4594–4605.
17. **Gupta, R. S., and B. Singh.** 1994. Phylogenetic analysis of 70 kD heat shock protein sequences suggests a chimeric origin for the eukaryotic cell nucleus. Curr. Biol. **4:**1104–1114.
18. **Ingolia, T. D., and E. A. Craig.** 1982. Drosophila gene related to the major heat-shock-induced gene is transcribed at normal temperatures and not induced by heat shock. Proc. Natl. Acad. Sci. USA **79:**525–529.
19. **Iwabe, N., K. Kuma, M. Hasegawa, S. Osawa, and T. Miyata.** 1989. Evolutionary relationship of archaebacteria, eubacteria and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc. Natl. Acad. Sci. USA **86:**9355–9359.
20. **Liesack, W., R. Söller, T. Stewart, H. Haas, S. Giovannoni, and E. Stackebrandt.** 1992. The influence of tachytelically (rapidly) evolving sequences on the topology of phylogenetic trees—intrafamily relationships and the phylogenetic position of the *Planctomycetaceae* as revealed by comparative analysis of 16S ribosomal RNA sequences. Syst. Appl. Microbiol. **15:**357–362.
21. **Lindquist, S., and E. A. Craig.** 1988. The heat shock proteins. Annu. Rev. Genet. **22:**631–677.
22. **Ludwig, W., J. Neumaier, N. Klugbauer, E. Brockmann, C. Roller, S. Jilg, K. Reetz, I. Schachtner, A. Ludvigsen, M. Bachleitner, U. Fischer, and K. H. Schleifer.** 1993. Phylogenetic relationships of *Bacteria* based on comparative sequence analysis of elongation factor Tu and ATP-synthase β-subunit genes. Antonie Leeuwenhoek **64:**285–305.
23. **Macario, A. J. L., C. B. Dugan, and E. C. D. Macario.** 1991. A *dnaK* homolog in the archaebacterium *Methanosarcina mazei* S6. Gene **108:**133–137.
24. **Maidak, B. L., N. Larsen, M. J. McCaughey, R. Overbeck, G. J. Olsen, K. Fogel, J. Blandy, and C. R. Woese.** 1994. The ribosomal database project. Nucleic Acids Res. **22:**3483–3487.
25. **Nimura, K., H. Yoshikawa, and H. Takahashi.** 1994. Identification of *dnaK* multigene family in *Synechococcus* sp. strain PCC 7942. Biochem. Biophys. Res. Commun. **201:**466–471.
26. **Nimura, K., H. Yoshikawa, and H. Takahashi.** 1994. Sequence analysis of the third *dnaK* homolog in *Synechococcus* sp. strain PCC 7942. Biochem. Biophys. Res. Commun. **201:**848–854.
27. **Olsen, G. J., and C. R. Woese.** 1993. Ribosomal RNA: a key to phylogeny. FASEB J. **7:**113–123.
28. **Rainey, F. A., N. Ward-Rainey, R. M. Kroppenstedt, and E. Stackebrandt.** 1996. The genus *Nocardiopsis* represents a phylogenetically coherent taxon

and a distinct actinomycete lineage: proposal of *Nocardiopsaceae* fam. nov. Int. J. Syst. Bacteriol. **46:**1088–1092.
29. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.
30. **Schlesner, H.** 1986. *Pirella marina* sp. nov., a budding, peptidoglycan-less bacterium from brackish water. Syst. Appl. Microbiol. **8:**177–180.
31. **Schlesner, H., and E. Stackebrandt.** 1986. Assignment of the genera *Planctomyces* and *Pirella* to a new family *Planctomycetaceae* fam. nov. and description of the order *Planctomycetales* ord. nov. Syst. Appl. Microbiol. **8:**174–176.
32. **Seaton, B. L., and L. E. Vickery.** 1994. A gene encoding a new Dnak/hsp70 homolog in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **91:**2066–2070.
33. **Smith, M. W., D. F. Feng, and R. F. Doolittle.** 1992. Evolution by acquisition: the case for horizontal gene transfer. Trends Biochem. Sci. **17:**489–493.
34. **Stackebrandt, E., A. Fischer, P. Hirsch, T. Roggentin, and H. Schlesner.** 1986. The phylogeny of an ancient group of budding peptidoglycan-less eubacteria: the genera *Planctomyces* and *Pirella*. Endocyt. C. Res. **3:**29–40.
35. **Stackebrandt, E., W. Ludwig, W. Schubert, F. Klink, H. Schlesner, T. Roggentin, and P. Hirsch.** 1984. Molecular genetic evidence for early evolutionary origin of budding peptidoglycan-less eubacteria. Nature **307:**735–737.
36. **Stackebrandt, E., and F. A. Rainey.** 1995. Partial and complete 16S rDNA sequences, their use in generation of 16S rDNA phylogenetic trees and their implications in molecular ecological studies 3.1.1, p. 1–17, *In* A. D. L. Akkermans, J. D. van Elsas, and F. J. De Bruijn (ed.), Molecular microbial ecology manual. Kluwer Academic Publishers, Dordrecht, The Netherlands.
37. **Staley, J. T.** 1968. *Prosthecomicrobium* and *Ancalomicrobium*: new prosthecate freshwater bacteria. J. Bacteriol. **95:**1921–1942.
38. **Staley, J. T.** 1973. Budding bacteria of the *Pasteuria-Blastobacter* group. Can. J. Microbiol. **19:**609–614.
39. **Stratagene.** 1996. pCR-Script™ SK(+) cloning kit—instruction manual. Stratagene, La Jolla, Calif.
40. **Tiboni, O., P. Cammarano, and A. M. Sanangelantoni.** 1993. Cloning and sequencing of the gene encoding glutamine synthetase I from the Archaeum *Pyrococcus woesei*: anomalous phylogenies inferred from analysis of archaeal and bacterial glutamine synthetase I sequences. J. Bacteriol. **175:**2961–2969.
41. **Tilly, K., R. Hauser, J. Campbell, and G. J. Ostheimer.** 1993. Isolation of *dnaJ*, *dnaK*, and *grpE* homologues from *Borrelia burgdorferi* and complementation of *Escherichia coli* mutants. Mol. Microbiol. **7:**359–369.
42. **Wadsworth, S. C.** 1982. A family of related proteins is encoded by the major *Drosophila* heat shock gene family. Mol. Cell. Biol. **2:**286–292.
43. **Ward, N., F. A. Rainey, E. Stackebrandt, and H. Schlesner.** 1995. Unraveling the extent of diversity within the order *Planctomycetales*. Appl. Environ. Microbiol. **61:**2270–2275.
44. **Ward-Rainey, N., F. A. Rainey, H. Schlesner, and E. Stackebrandt.** 1995. Assignment of hitherto unidentified 16S rDNA species to a main line of descent within the domain Bacteria. Microbiology **141:**3247–3250.
45. **Ward-Rainey, N., F. A. Rainey, and E. Stackebrandt.** 1996. Physical map of the genome of *Planctomyces limnophilus*, a representative of the phylogenetically distinct planctomycete lineage. J. Bacteriol. **178:**1908–1913.
46. **Werner-Washburne, M., D. E. Stone, and E. A. Craig.** 1987. Complex interactions among members of an essential family of hsp70 genes in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **7:**2568–2577.
47. **Woese, C. R.** 1987. Bacterial evolution. Microbiol. Rev. **51:**221–271.
48. **Woese, C. R., O. Kandler, and M. L. Wheelis.** 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc. Natl. Acad. Sci. USA **87:**4576–4579.
49. **Yamamoto, S., and S. Harayama.** 1995. PCR amplification and direct sequencing of *gyrB* genes with universal primers and their application to the detection and taxonomic analysis of *Pseudomonas putida* strains. Appl. Environ. Microbiol. **61:**1104–1109.
50. **Yamamoto, S., and S. Harayama.** 1996. Phylogenetic analysis of *Acinetobacter* strains based on the nucleotide sequences of *gyrB* genes and on the amino acid sequences of their products. Int. J. Syst. Bacteriol. **46:**506–511.