RESEARCH ARTICLE

INTERNATIONAL MICROBIOLOGY (2004) 7:41–52

INTERNATIONAL MICROBIOLOGY

Emma Griffiths
Radhey S. Gupta*

Department of Biochemistry,
McMaster University, Hamilton,
Ontario, Canada

Signature sequences in diverse proteins provide evidence for the late divergence of the Order *Aquificales*

Summary. The *Aquificales* species are presently believed to be the earliest branching lineage within *Bacteria*. However, the branching order of this group in different phylogenetic trees is highly variable and not resolved. In the present work, the phylogenetic placement of *Aquificales* was examined by means of a cladistic approach based on the shared presence or absence of definite signature sequences (consisting of conserved inserts or deletions) in many highly conserved and important proteins, e.g. RNA polymerase β (RpoB), RNA polymerase β′ (RpoC), alanyl-tRNA synthetase (AlaRS), CTP synthase, inorganic pyrophosphatase (PPase), Hsp70 and Hsp60. For this purpose, fragments of the above genes that contained the signature regions were cloned from different *Aquificales* species (*Calderobacterium hydrogenophilum*, *Hydrogenobacter marinus*, and *Thermocrinis ruber*) and the sequence data were compared with those available from all other species. The presence in *Aquificales* species of distinctive inserts in Hsp70 and Hsp60 that are not found in any Firmicutes, Actinobacteria, or *Thermotoga*-*Clostridium* species excluded them from these groups of *Bacteria*. The shared presence of prominent indels in the RpoB (>100 amino acids), RpoC (>100 amino acids) and AlaRS (4 amino acids) proteins, which are only found in the various *Aquificales* species, the Chlamydiae, the CFBG (*Cytophaga*-Flavobacteria-*Bacteroides*-green sulfur bacteria) group, and Proteobacteria, strongly suggests their placement within these groups of *Bacteria*. A specific relationship between Proteobacteria and *Aquificales* is suggested by the presence in inorganic pyrophosphatase of a 2-amino-acid insert that is uniquely found in these phyla. However, the *Aquificales* species lacked a number of other protein signatures (e.g. indels in CTP synthase and Hsp70) that are characteristic of Proteobacteria, indicating that they constitute a distinct phylum related to Proteobacteria. These results provide strong and consistent evidence that the *Aquificales* diverged after the branching of Firmicutes, Actinobacteria, *Thermotoga*, *Deinococcus*-*Thermus*, green nonsulfur bacteria, Cyanobacteria, Spirochetes, Chlamydiae, and CFBG group, but before the emergence of the Proteobacteria. **[Int Microbiol** 2004; 7(1):41–52]

Key words: *Aquifex* · bacterial phylogeny · branching order

Received 14 September 2003
Accepted 30 October 2003

*Corresponding author:
Radhey S. Gupta
Department of Biochemistry
McMaster University
Hamilton, Ontario, L8N 3Z5 Canada
Tel. +1-905-5259140 ext. 22639
Fax +1-905-5229033
E-mail: gupta@mcmaster.ca

Introduction

The phylogenetic trees based on 16S rRNA provide the presently accepted framework for understanding the evolutionary relationships among *Bacteria* [29,33]. Based on these trees, a number of main groups or phyla within the *Bacteria* have been recognized. These include *Thermotoga*, green nonsulfur bacteria (GNS), *Deinococcus*-*Thermus*, Cyanobacteria, low G + C gram-positive (Firmicutes), high G + C gram-positive (Actinobacteria), Chlamydiae, CFBG (*Cytophaga*-Flavobacteria-*Bacteroides*-Green sulfur bacteria), *Planctomy-*

ces, and relatives, and the Proteobacteria [29,41]. In addition, a number of other phyla consisting of only a limited number of species are also recognized [29]. Although the branching orders of different phyla in rRNA or other phylogenetic trees have not been resolved, the *Aquificales* species (represented by the genus *Aquifex*) are thought to be the earliest branching lineage within *Bacteria* [1,4,6,8,14,29,30,33,35,36,40]. However, the deep branching of *Aquifex* is not supported by many protein phylogenies, in which it exhibits a closer relationship to the δ,ε-Proteobacteria and the Chlamydiae groups [5,7,10,18–20,27,31,36,39].

Despite the lack of a consistent picture concerning its phylogenetic placement, the deep branching of *Aquifex* has become a central aspect of the current view of bacterial phylogeny [8,11,29,33,36]. In this context, it is important to further investigate the branching position of this phylum relative to other groups of *Bacteria* using different approaches. In the present work, we used a cladistic approach based on shared conserved indels or signature sequences in various proteins to deduce the branching order of bacterial groups [19,23]. This approach has provided evidence that the major groups within *Bacteria* have branched off in the following order: low G + C gram-positives→high G + C gram-positives→ *Deinococcus/Thermus*→GNS→Cyanobacteria→Spirochetes →Chlamydiae, CFBG, *Aquifex*→δ,ε-Proteobacteria→α-Proteobacteria→β-Proteobacteria→γ-Proteobacteria [20,21,23]. By means of this approach, the genus *Aquifex* was found to branch at a position similar to that of the Chlamydiae and CFBG groups [20,21].

The present report describes work on a number of new as well as previously described protein signatures (i.e. Hsp70, Hsp60, CTP synthase, PPase, AlaRS, RpoB, and RpoC) that are helpful in understanding the phylogenetic placement of *Aquificale*s. The Order *Aquificales* comprises four genera: *Aquifex*, *Calderobacterium*, *Hydrogenobacter* and *Thermocrinis* [36]. Of these, sequence information is mainly available for *Aquifex* species. We have cloned and sequenced gene fragments of the above-mentioned proteins, containing the signature regions from species belonging to various *Aquificales* genera. Results of these studies provide consistent evidence that this group should be placed between the δ,ε-Proteobacteria and the Chlamydiae and CFBG groups and it constitutes a late-branching phylum within *Bacteria*.

Materials and methods

DNA. Purified *C. hydrogenophilum* (type strain; Z-829) DNA was generously provided by Dr. Karel Mikulik (Academy of Sciences, Czech Republic) [32]. The DNA for *H. marinus* (DSM 12046T) was kindly provided by Dr. Micheal Thomm (Institut für Allgemeine Mikrobiologie, Kiel, Germany) [38], and the DNA from *T. ruber* (DSM 12173) was a generous gift of Dr. K.O. Stetter (University of Regensburg, Berlin, Germany) [24]. The complete genome of *Aquifex aeolicus* have been sequenced [11] and sequence information for several *Aquifex pyrophilus* genes is available in the NCBI database.

PCR amplification and sequencing. Degenerate oligonucleotide primers, in opposite orientations, were designed for highly conserved regions that flanked the identified signatures in sequence alignments. The sequences of various PCR primers used in these studies are detailed in Table 1. Because these primers are based on highly conserved regions, they may also prove effective in amplifying these genes from other species. Ten-µl PCR reactions (approximately 0.2 mg DNA per reaction) were optimized for Mg^{2+} concentration (1.5–4 mM) for each set of primers. PCR was carried out using a Techne Progene thermocycler, over 30 cycles (15 s at 94°C, 15 s at 55°C, 1 min at 72°C) with an initial 1-min hot start at 94°C, and a final extension step (15 s at 94°C, 15 s at 55°C, 7 min at 72°C). The DNA fragments of the expected size were purified from 0.8 % (w/v) agarose gels and subcloned into the plasmid pDRIVE using a UA cloning kit (Qiagen). Due to DNA limitation, some gene fragments from *T. ruber* were not amplified. After transforming *E. coli* JM109 cells with the plasmids, inserts from a number of positive clones were sequenced. The sequence data for various *Aquificales* species have been deposited in the GenBank and the accession numbers for these sequences are included in the alignment figures.

Results

Determining the branching Order of *Aquificales* based on signatures sequences. Signature sequences provide a powerful means to deduce the relative branching order and interrelationships among different groups. By making use of conserved and defined indels that are commonly shared by different species, it is possible to group different species or taxa into distinct clades, which show specific relationships to each other [3,19,21,26,37]. The application of this approach for determining the phylogenetic placement of *Aquificales* is described below.

The Hsp70 (DnaK) family of proteins contain a 21–23-amino-acid (aa) insert in the N-terminal quadrant that distinguishes various gram-negative (or diderm) bacteria from gram-positive (or monoderm) bacteria (Fig. 1) [19]. The Hsp70 homologs are found in all *Bacteria* and the identified insert is present in all diderm (i.e., primarily gram-negative) bacteria, but not in *Thermotoga* or any Firmicutes and Actinobacteria. The insert in Hsp70 is not found in any archaeal homolog where this protein is found, which suggests that this indel constitutes an insert, and that the species lacking it are ancestral (Fig. 1) [19]. Both *Aquifex aeolicus* and *Aquifex pyrophilus*, whose sequences are available [11,17], contained this insert. The presence of this insert in *C. hydrogenophilum*, *H. marinus*, and *T. ruber* was examined by PCR amplifying the *dnaK* gene fragments covering the indel region. Sequences of the resulting fragments showed that this insert is present in all of these species (Fig. 1), indicat-

Table 1. PCR primers for amplifying different gene sequences

Gene	Primer	Primer sequence*	Fragment size
Hsp70	Forward	5'-GGNATHGAYYTNGGNACNAC-3'	1.1 kb
	Reverse	5'- GCNACNGCYTCRTCNGGRTT- 3'	
Hsp60	Forward	5'-AAGCTTTCNCCRAANCCNGGNGCY TTNACNGC-3'	0.5 kb
	Reverse	5'-GGYGAYGGYACYACHACWGC-3'	
Alanine tRNA synthetase	Forward	5'-TTYACNAAYGCNGGNATG-3'	170 bp
	Reverse	5'-CATYTCRAARAANGTRTG-3'	
CTP synthase (Proteobacterial insert)	Forward	5'-GGNATTHTGYYTNGGNATGCA-3'	420 bp
	Reverse	5'-AAYTCNGGRTGRAAYTG-3'	
Inorganic pyrophosphatase	Forward	5'-AARTAYGARMTIGAYAARGA-3'	333 bp
	Reverse	5'-TYYTTRTAIKKYTCRAARAA-3'	
RNA polymerase subunit	Forward		
	Calderobacterium	5'-TTYYTWGAGCACRAYGAYGCDAA YMGNGCNYTNATGGG-3'	830 bp
	Hydrogenothermus	5'-GGNTAYAAYTWYGARGAYGC-3'	1.2 kb
	Reverse	5'-ACCYTTRTTWCCRTGHCKTCCHGC CATYTTRTCDCC- 3'	
RNA polymerase β subunit	Forward	5'-ATHGGNGARCCNGGNACNCA-3'	642 bp
	Reverse	5'-GGNARNCCNCCNGTDATRTC-3'	
CTP synthase (*Aquifex*-Proteob. insert)	Forward	5'-GGNCAYTAYGARMGNTT-3'	357 bp
	Reverse	5'-TGYTGNGTNGGYTTNGTYTT-3'	

* Where N=A,T,C or G; H=A,C or T; Y=C or T; R=A or G; W=A or T; M=A or C; K=G or T; D=G,A or T; I=inositol.

ing that this group diverged after the branching of gram-positive bacteria (Actinobacteria and Firmicutes) and *Thermotoga* (Fig. 1).

Hsp60, which is present in all bacteria except a few mycoplasmas, contains a conserved 1-aa insert that is common to various Proteobacteria, Chlamydiae, CFBG group, Spirochetes, and Cyanobacteria, but absent from various Firmicutes, Actinobacteria, *Thermotoga*, *Deinococcus /Thermus* group, and GNS bacteria (Fig. 2) [19,20]. The sequence information for Hsp60 is available from >400 species and no exception to this pattern has been observed. A 0.5-kb fragment of *groEL* covering this region was amplified from the three *Aquificales* species, and all of them were found to contain this insert in the appropriate position (Fig. 2). This insert is also present in the published sequence of *A. aeolicus* [11]. The shared presence of this insert in different *Aquificales* as well as various Cyanobacteria, Spirochetes, Chlamydiae, CFBG and Proteobacteria provides evidence that it should be placed within these groups. The bacterial groups lacking this indel (Firmicutes, Actinobacteria, *Thermotoga, Deinococcus/ Thermus,* and GNS bacteria) are indicated to have branched off prior to the insertion of this indel.

A prominent signature was identified in the β'-subunit of RNA polymerase (RpoC) that consists of a large insert of *ca.* 200 aa in various Proteobacteria, Chlamydiae, CFBG group, and Spirochetes. In this position, the cyanobacteria contain a much larger insert (~600 aa), which could be of independent origin or may have changed subsequently. However, no insert is present in this position in various Archaea, Firmicutes, Actinobacteria, *Thermotoga, Deinococcus/Thermus* group, and GNS bacteria, indicating that these groups diverged prior to the introduction of this indel (Fig. 3). A 642-bp fragment of *rpoC* covering this region was amplified from *H. marinus* and *C. hydrogenophilum*, and both these species contained a 188-aa insert in this region. Similar inserts of 188 aa are also present in the published sequences of *A. aeolicus* and *A. pyrophilus* (Fig. 3) [11,27]. The shared presence of this large (~200 aa) insert in the Spirochetes, Chlamydiae, CFBG group, and Proteobacteria again suggests that *Aquificales* should be placed within these groups of *Bacteria*.

Alanyl-tRNA synthetase contains a highly conserved 4-aa insert that is common to various Proteobacteria, Chlamydiae, CFBG groups, and *A. aeolicus* and *A. pyrophilus*, but it is not found in any other group of prokaryotes, including Archaea (Fig. 4). A 170-bp fragment of AlaRS covering this region was amplified and sequenced from *C. hydrogenophilum, H. marinus,* and *T. ruber*. The results of these studies (Fig. 4) show that this insert is a common characteristic of all *Aquificales* species, thereby strongly supporting their placement within the Proteobacteria, Chlamydiae, and CFBG group and indicating that these phyla diverged after branching of groups lacking this indel (Fig. 4).

Another prominent signature showing a relationship similar to that of AlaRS was identified in the β-subunit of RNA polymerase (RpoB). This protein contains a large insert of ~120 aa that is present in Chlamydiae, the CFBG group,

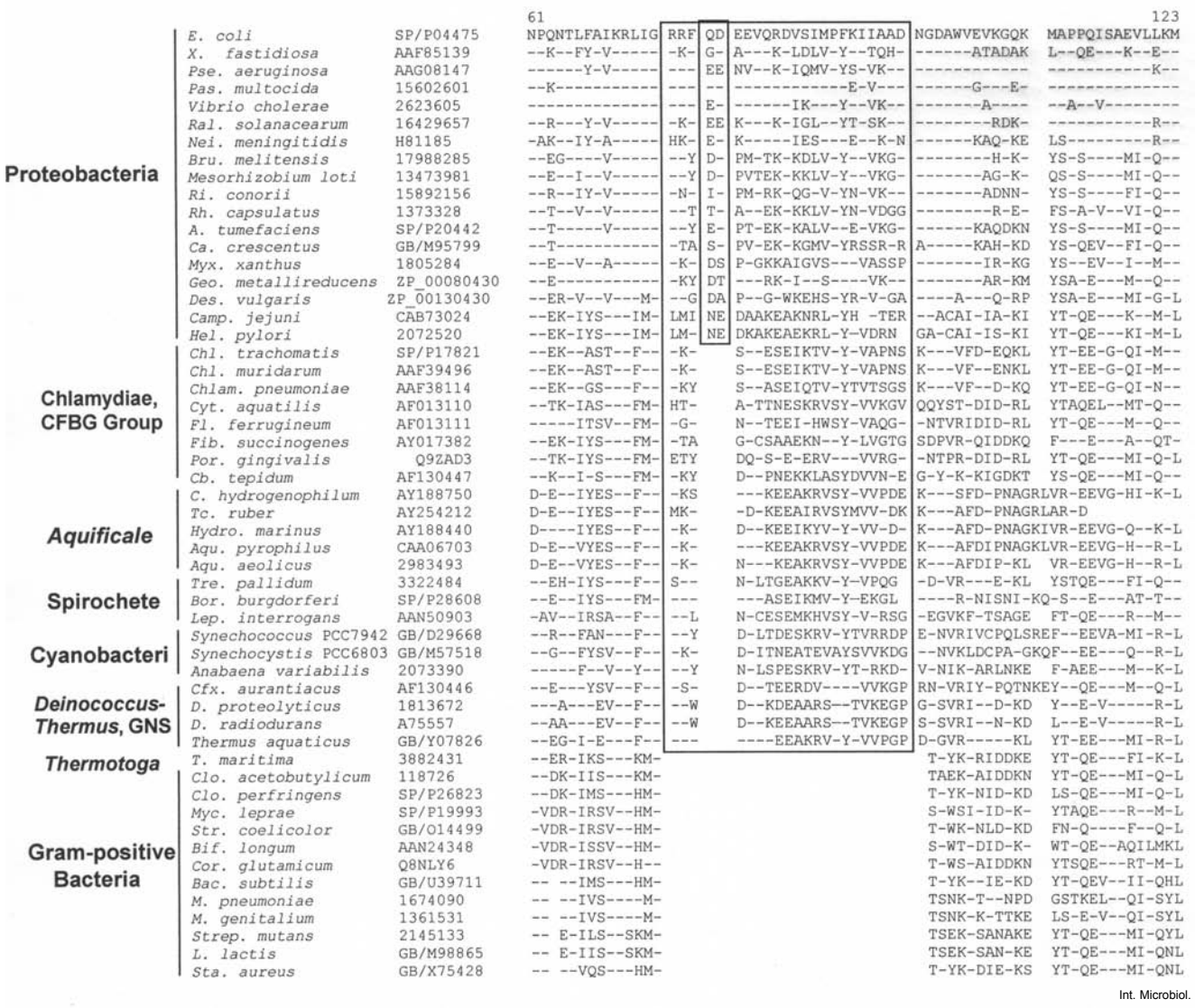

Fig. 1. Partial alignment of Hsp70 sequences showing a signature sequence consisting of an insert of 21–23-amino-acid (aa) (outer box), which is specific for diderm (characterized by the presence of an outer membrane) bacteria which comprise mainly gram-negative phyla. The smaller 2-aa insert (inner box) within the large insert is distinctive of the proteobacterial group. *Dashes* in the alignment indicate sequence identity with the amino acids found in *Escherichia coli* sequence shown on the top line. The accession numbers of various sequences are shown in the second column. Only representative sequences from different *Bacteria* are shown. The absence of the large insert in *Aquificales* species supports their placement within the diderm (gram-negative) bacteria. Abbreviations of species names are: *A.*, *Agrobacterium*; *Aqu.*, *Aquifex*; *Bac.*, *Bacillus*; *Bif.*, *Bifidobacterium*; *Bor.*, *Borrelia*; *Bru.*, *Brucella*; *C.*, *Calderobacterium*; *Ca.*, *Caulobacter*; *Camp.*, *Campylobacter*; *Cb.*, *Chlorobium*; *Chl.*, *Chlamydia*; *Chlam.*, *Chlamydophila*; *Cfx.*, *Chloroflexus*; *Clo.*, *Clostridium*; *Cor.*, *Corynebacterium*; *Cyt.*, *Cytophaga*; *D.*, *Deinococcus*; *Des.*, *Desulfovibrio*; *E.*, *Escherichia*; *Fib.*, *Fibrobacter*; *Fl.*, *Flavobacteria*; *Geo.*, *Geobacter*; *Hel.*, *Helicobacter*; *Hydro.*, *Hydrogenothermus*; *L.*, *Lactococcus*; *Lep.*, *Leptospira*; *M.*, *Mycoplasma*; *Myc.*, *Mycobacterium*; *Myx.*, *Myxococcus*; *Nei.*, *Neisseria*; *Pas.*, *Pasteurella*; *Por.*, *Porphyromonas*; *Pse.*, *Pseudomonas*; *Ral.*, *Ralstonia*; *Ri.*, *Rickettsia*; *Rh.*, *Rhodobacter*; *Sta.*, *Staphylococcus*; *Str.*, *Streptomyces*; *Strep.*, *Streptococcus*; *T.*, *Thermotoga*; *Tc.*, *Thermocrinis*; *Tre.*, *Treponema*; *X.*, *Xyllela. GNS* in bacterial groups name refers to green nonsulfur bacteria (*Cfr. aurantiacus*).

Proteobacteria, and *A. pyrophilus* and *A. aeolicus*, but is absent from all other *Bacteria* (Fig. 5). This insert is not found in the RpoB homologs from *Archaea*, providing evidence that the bacterial groups lacking this indel are ancestral. The shared presence of this insert in Proteobacteria and *A. pyrophilus* was first reported by Klenk et al. [27], but due to limited sequence information, its evolutionary significance was not clear. We have now amplified a 0.83-kb fragment of *rpoB* from *C. hydrogenophilum* and a 1.2-kb fragment of this gene from *H. marinus* containing the signature region. Sequencing of these fragments revealed that this insert is present in both species (Fig. 5). The shared presence of this

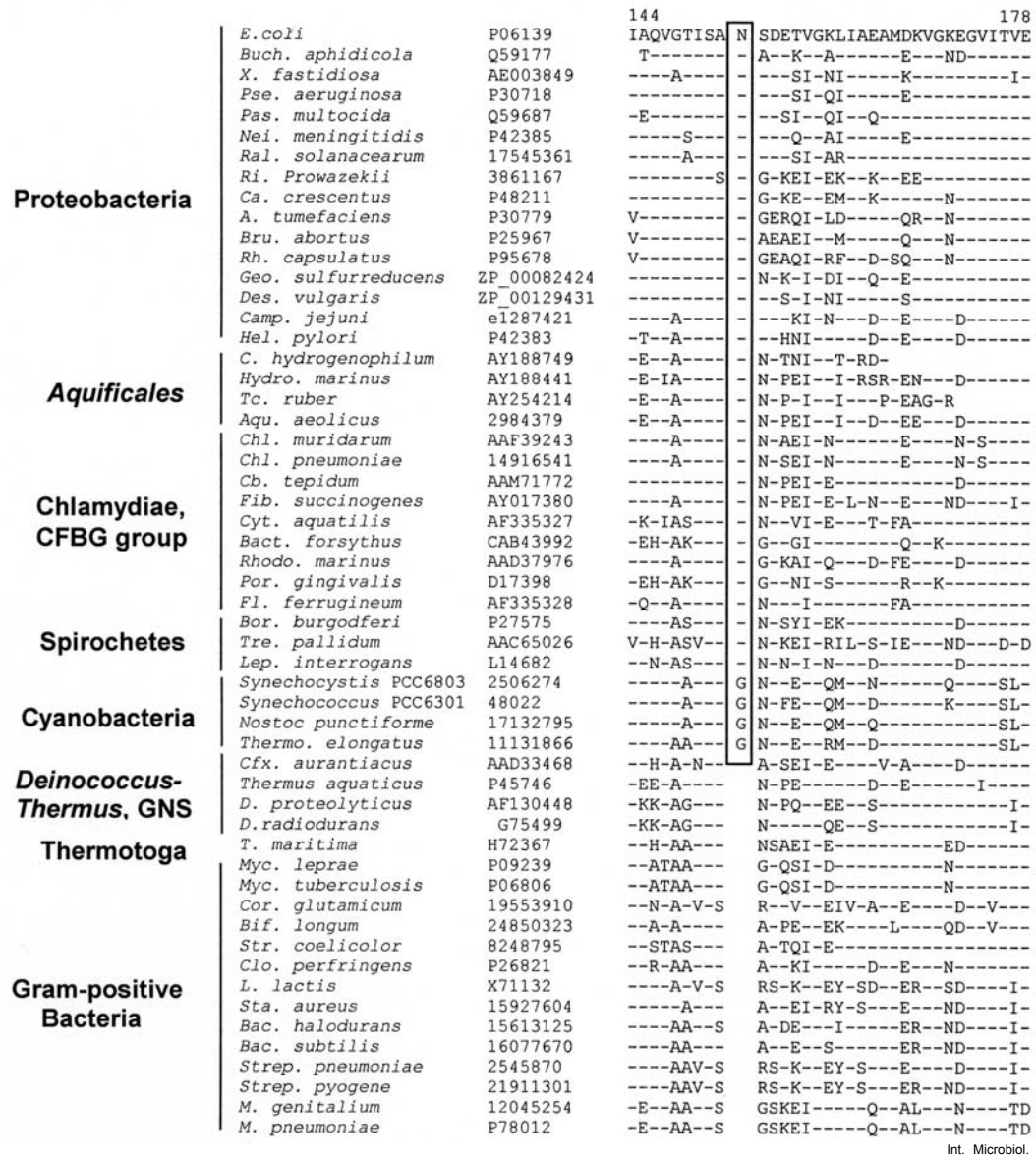

Fig. 2. Excerpts from Hsp60 sequence alignment showing a 1-aa insert that is commonly shared by Spirochetes, Cyanobacteria, Chlamydiae, CFBG group and Proteobacteria, but not present in *Deinococcus-Thermus*, green nonsulfur bacteria, *Thermotoga*, Actinobacteria and Firmicutes. Sequences from only representative species are shown. The presence of this insert in *Aquificales* species provides evidence for their placement within the former groups of *Bacteria*. For abbreviations of species names, see the legend to Fig. 1. Additional abbreviations: *Bact.*, *Bacteroides*; *Buch.*, *Buchnera*; *Rhodo.*, *Rhodothermus*; *Thermo.*, *Thermosynechococcus.*

prominent insert in the *Aquificales*, Proteobacteria, Chlamydiae and CFBG group of species provides evidence that these groups are related and that they diverged subsequent to other groups lacking the indel.

Another useful signature providing further clarification of the relationships among the Proteobacteria, Chlamydiae, CFBG group and *Aquificales* has now been identified in the enzyme inorganic pyrophosphatase (PPase), which catalyzes the hydrolysis of pyrophosphate. This protein contains a 2-aa insert in a conserved region that is shared by various Proteobacteria as well as *A. aeolicus*, but not present in Chlamydiae and CFBG group nor in any other *Bacteria* (Fig. 6). Approximately 300–335 bp fragments of *PPase* from *C. hydrogenophilum, T. ruber*, and *H. marinus* were PCR amplified and sequenced. The results (Fig. 6) show that PPases from *Aquificales* species contain this insert, which indicates that it is a common characteristic of the group. The shared presence of this indel in only the Proteobacteria and

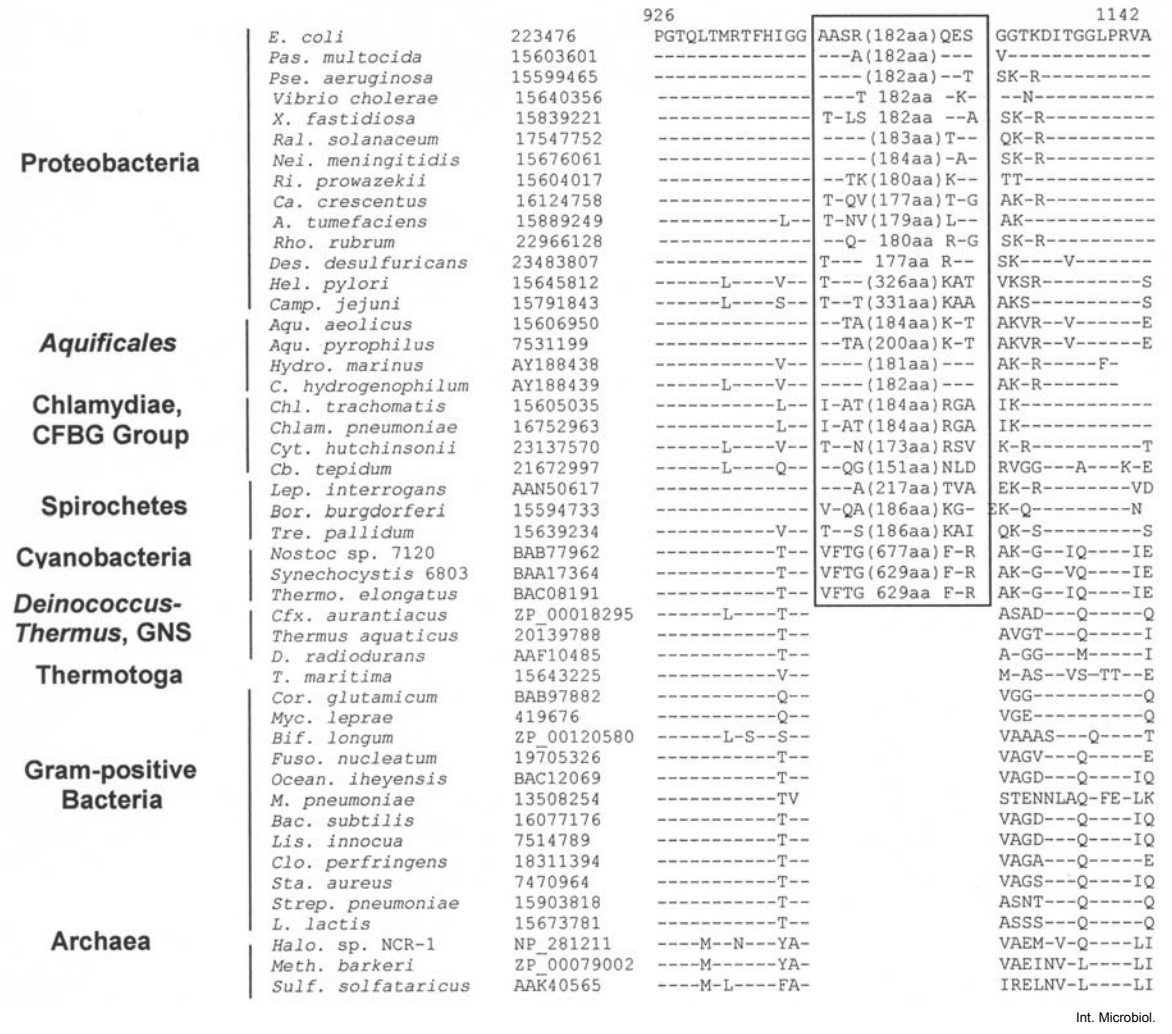

Fig. 3. The shared presence of a large insert in RNA polymerase β´ in Spirochetes, Chlamydiae, CFBG group, Proteobacteria, and various *Aquificales* species. The cyanobacteria contain a much larger insert (>600 aa) in this position. The observed indel is not present in *Archaea* or other bacterial groups. For abbreviations of species names, see previous figure legends. Additional abbreviations: *Fuso.*, *Fusobacterium*; *Halo.*, *Halobacterium*; *Lis.*, *Listeria*; *Meth.*, *Methanosarcina*; *Ocean.*, *Oceanobacillus*; *Rho.*, *Rhodospirillum*; *Sulf.*, *Sulfolobus.*

Aquificales species provides evidence that these two groups are specifically related and that they diverged after branching of the other groups.

We have previously described many signatures that are unique to Proteobacteria and which provide clear distinction among the species belonging to the α-, β-, γ- and the δ,ε-subdivisions [20,22]. Two of these signatures, one consisting of a 10-aa insert in CTP synthase (Fig. 7) and the other of a 2-aa insert in Hsp70 (see Fig. 1), are distinctive of the entire proteobacterial phylum and are not found in any other *Bacteria*. The proteobacterial signature in Hsp70 (2-aa insert) is present within the large insert in this protein, whose sequence for various *Aquificales* species is shown in Fig. 1. The 2-aa insert common to various Proteobacteria is not found in any of the *Aquificales* species, which indicates that the two are distinct from each other. We cloned and sequenced a 420-bp fragment of the CTP synthase gene from *C. hydrogenophilum*, *T. ruber*, and *H. marinus*. The CTP synthase from all these species, as well as from *A. aeolicus*, did not contain the 10-aa insert common to various Proteobacteria (Fig. 7). In addition to these signatures, the published sequences of *A. aeolicus* do not contain any of the other signatures distinctive of the α, β and γ-Proteobacteria [23]. These results provide strong evidence that the *Aquificales* phylum is distinct from Proteobacteria and that it branched off prior to the latter phylum.

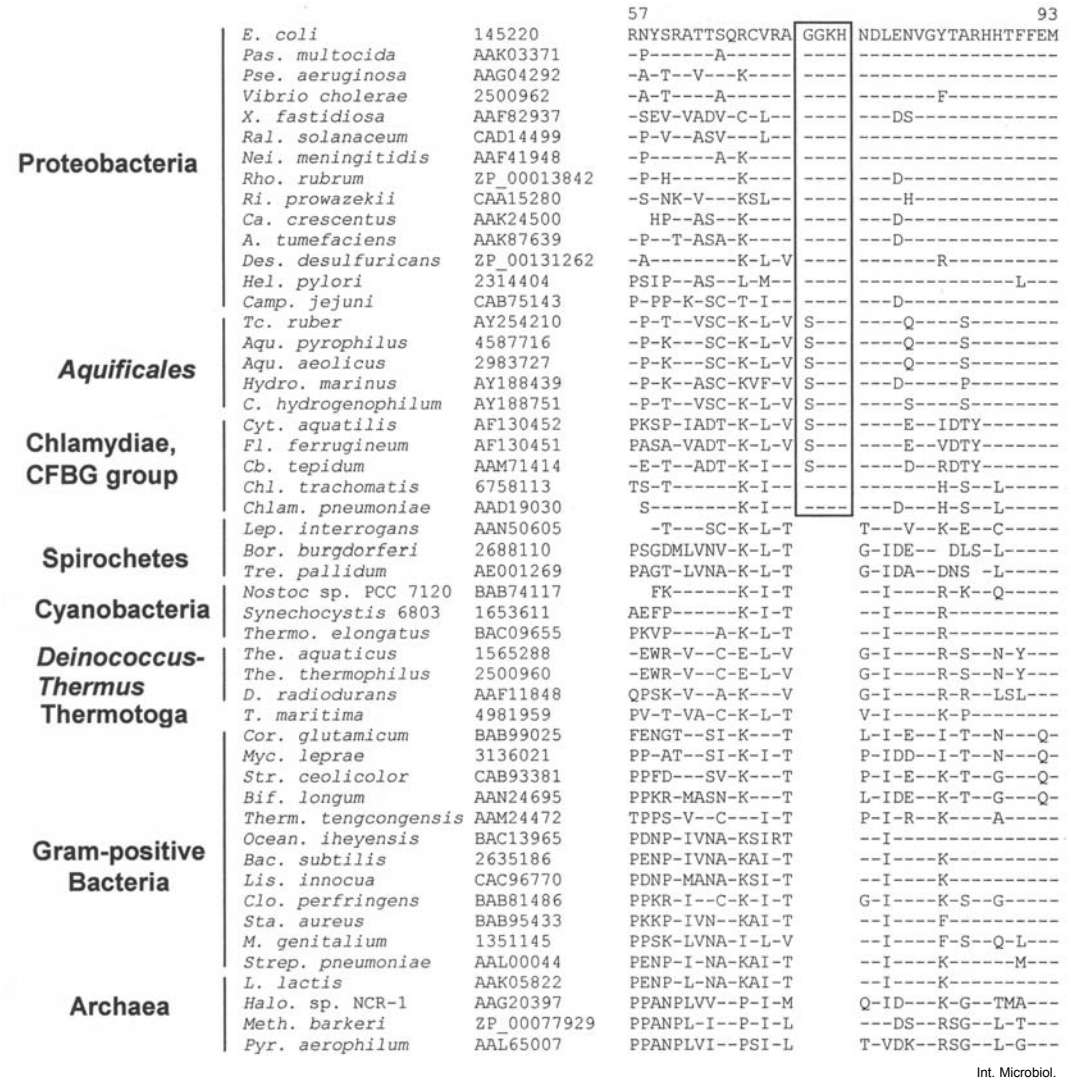

Fig. 4. A 4-aa insert in alanyl-tRNA synthetase that is a unique characteristic of Proteobacteria, Chlamydiae, CFBG group, and *Aquificales* homologs. This insert is not found in other groups of *Bacteria* or Archaea, indicating a specific relationship of *Aquificales* to the Proteobacteria, Chlamydiae and the CFBG group. For abbreviations of species names, see previous figure legends. Additional abbreviations: *Pyr.*, *Pyrobaculum*; *The.*, *Thermus*; *Therm.*, *Thermoanaerobacter.*

Discussion

In this work, we have used a cladistic approach involving the shared presence of conserved indels in widely distributed proteins to clarify the phylogenetic placement of *Aquificales* [29,33]. Unlike the phylogenetic trees in which the deduced relationships are dependent upon a large number of variables and often not resolved, the relationships inferred by this method are based on minimal assumptions and are unambiguous [3,19,23,37]. The phylum *Aquificales* is made up of four genera, *Aquifex*, *Hydrogenobacter*, *Calderobacterium*, and *Thermocrinis* [36]. The first two genera contain two and three species, respectively, whereas the last two genera are made up of only a single species [36]. Except for *A. aeolicus*, whose genome has been sequenced [11], very limited sequence information is available for other *Aquificales* species. This work describes a large number of signatures, many of them for the first time, that are relevant to understanding the phylogenetic placement of this phylum. The sequence information for various genes that were studied here was obtained from most of the *Aquificales* genera in order to ensure the general applicability of the derived inference to the entire phylum.

A summary diagram of the results obtained from different signatures is presented in Fig. 8. Based on their observed dis-

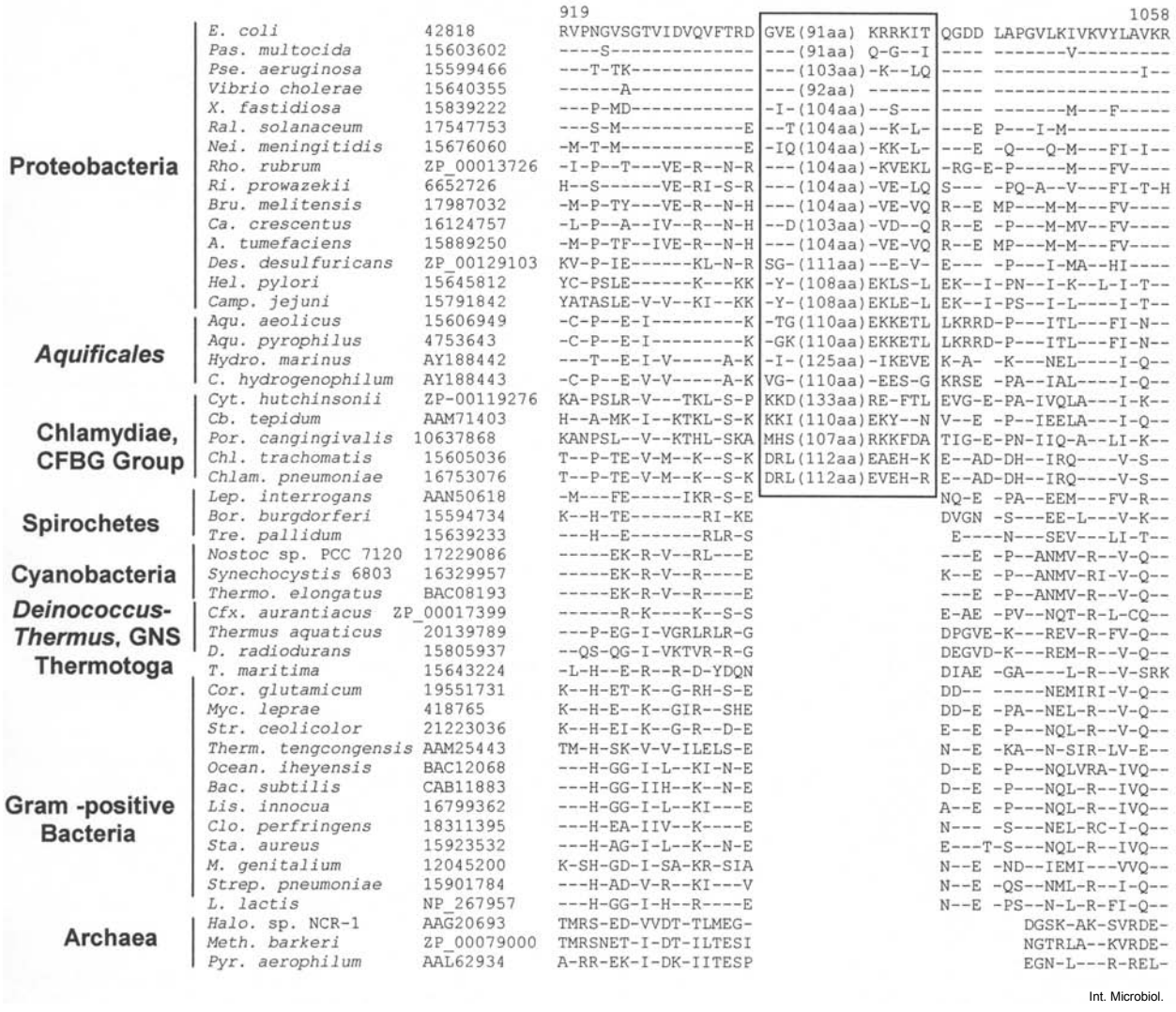

Fig. 5. Partial alignment of RNA polymerase β-subunit sequences showing a large insert (>100 aa) that is specific for the Proteobacteria, Chlamydiae, CFBG group, and *Aquificales* homologs. This insert is not found in other bacterial or archaeal homologs.

tribution in different bacterial phyla, the diagram depicts the inferred evolutionary stages where these signatures were introduced in these genes during the course of bacterial evolution. In this diagram, all of the marked signatures are present in the various bacterial groups above the indicated insertion points, but they are not found in any of the groups that lie below. The large inserts in Hsp70, RpoB, RpoC, and AlaRS, which are absent from various monoderm bacteria (Firmicutes, Actinobacteria and *Thermotoga*) are also not found in any archaeal homologs, providing evidence that within *Bacteria*, the gram-positive (monoderm) bacteria (Firmicutes, Actinobacteria) and *Thermotoga* constitute early branching lineages [28]. The presence in *Aquificales* of the large insert in Hsp70, which is a distinctive characteristic of diderm bacteria, supports their placement within this group. The signature in Hsp60 further excludes *Aquificales* from the *Deinococcus/Thermus* and GNS phyla. The large insert in RpoC, which in addition to its presence in *Aquificales* is found only in Cyanobacteria, Spirochetes, Chlamydiae, CFBG group, and Proteobacteria, places *Aquificales* within these groups. The shared conserved indels in AlaRS and RpoB further refine the placement of *Aquificales* to Chlamydiae, the CFBG group, and Proteobacterial groups. The signature in PPase, which is present only in Proteobacteria and *Aquificales* species, points to a specific relationship between these groups to the exclusion of all others. However, the absence of various signatures that are distinctive of the Proteobacteria in *Aquificales* homologs indicates that *Aquificales* constitutes a distinct phylum that diverged prior to the Proteobacteria. The only arrangement of the different bacte-

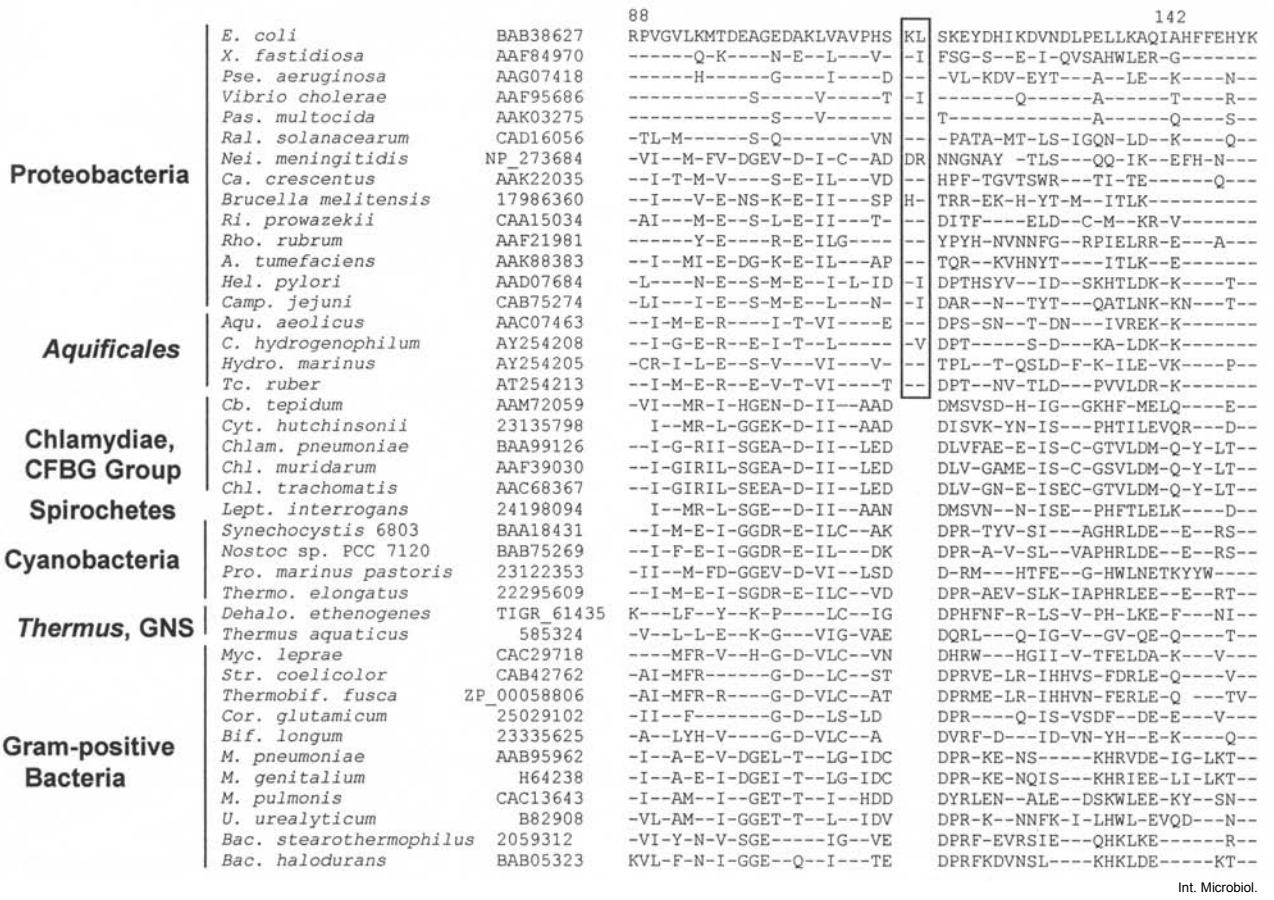

Fig. 6. Excerpts from sequence alignment of inorganic pyrophosphatase showing a 2-aa insert (boxed), which is found only in *Aquificales* and proteobacterial homologs, indicating a specific relationship between them. For abbreviations of species names, see previous figure legends. Additional abbreviations: *Dehalo.*, *Dehalococcoides*; *Pro.*, *Prochlorococcus*; *Thermobif.*, *Thermobifida*; *U.*, *Ureaplasma.*

rial groups that is compatible with the various signatures places *Aquificales* in a position between Chlamydiae and CFBG groups and Proteobacteria. All of the other main groups within *Bacteria* appear to have branched off at earlier stages.

The question can now be asked whether these results can be explained by any other reasonable means. In this context, note that most of the proteins in which the various signatures are found (e.g., RpoB, RpoC, AlaRS, Hsp60, Hsp70, CTP synthase) are single-copy essential genes present in virtually all *Bacteria*. For most of these proteins, >50 sequences are available from the bacterial groups containing the inserts (e.g., Proteobacteria, *Aquificales*, Chlamydiae, and CFBG group) and an equally large number from the bacterial groups lacking them (e.g., Spirochetes, Cyanobacteria, *Deinococcus -Thermus*, *Thermotoga*, gram-positives). If these genes (or indels) were subjects of frequent lateral gene transfers (LGTs) [9,13,25], then one would expect that some of the species from the former groups would be lacking the signatures and at the same time several species from the latter group would possess these indels, resulting in a more random arrangement. However, the fact that all species from the former groups contain these signatures, and none of these signatures are found in species from the latter groups, provides strong evidence that they were introduced only once in a common ancestor of the first group of species. (A second homolog of Hsp70 lacking the large indel has been found in *Borrelia burgdorferi*. This homolog is likely derived by means of LGT and it is readily distinguished from the normal Hsp70 homolog [21].) The χ^2 probability (assuming two degrees of freedom) that the observed distribution of these indels is due to random occurrence is virtually nil ($<10^{-10}$).

These results challenge and do not support the presently held view that the *Aquificales* group constitutes the deepest branching lineages within *Bacteria*. The deep branching of *Aquificales* in phylogenetic trees could result from a variety of factors, including the long branch-length effect and LGT [15,34]. The *Aquifex* genome appears to be rapidly evolving

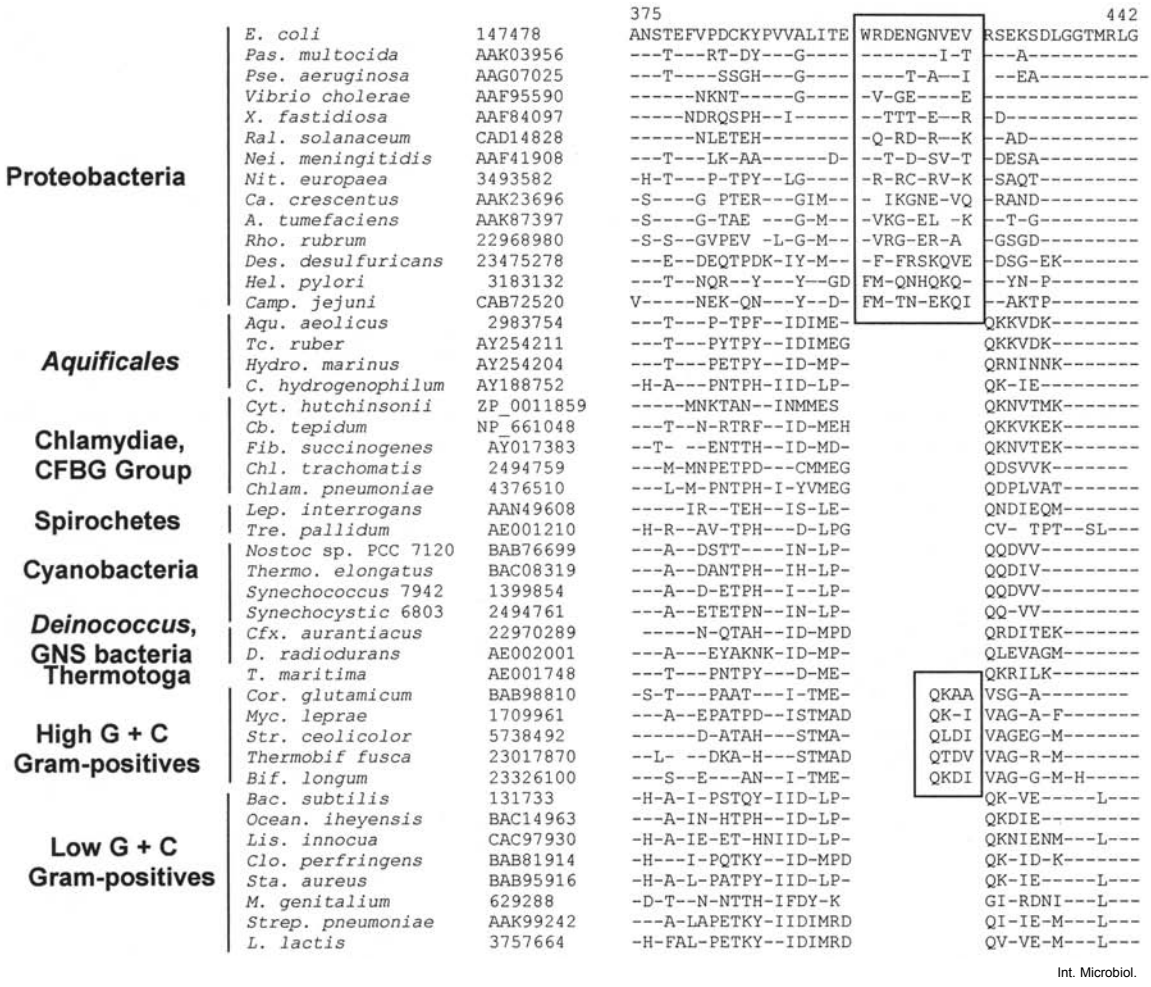

Fig. 7. Partial alignment of CTP synthase sequences showing a 10-aa insert that is distinctive of various Proteobacteria, but not found in *Aquifica les* species or other groups of *Bacteria*. A smaller 4-aa insert specific for *Actinobacteria* is also present in this region. For abbreviations of species names, see previous figure legends. *Nit.*, *Nitrosomas*.

[11,36], and thus many *Aquifex* genes are subjects of long branch-length effects leading to their abnormal branching in phylogenetic trees. LGT is another important factor that can lead to abnormal branching in phylogenetic trees. About 10% of the *A. aeolicus* genes exhibit extensive sequence identity to homologs from various Archaea, implicating massive LGT between these groups [2]. If this is the case, then in phylogenetic trees constructed from homologs of the transferred genes, *Aquificales* species will branch near the root of the tree, as their sequences would closely resemble those of the Archaea. Note in this regard that, in contrast to the *A. aeolicus* genome, which has a G + C content of 43.4%, the G + C content of 16S-23S-5S operons in this species is 65% [11], which suggests that either the rRNA genes in this species selectively evolved at a very rapid rate, or that they have been acquired from a high G + C species by means of LGT. However, the differences in evolutionary rates are not expected to have a significant effect on the placement of species into different clades based upon conserved indels in widely distributed proteins [19].

The late divergence of the *Aquificales*, as suggested by the present work, has important implications for bacterial /prokaryotic phylogeny. The clustering of the *Aquificales* and other hyperthermophiles at the base of the prokaryotic tree has provided the main argument for a hot origin of life [8,12]. However, a later divergence of *Aquificales* suggests that thermal adaptation within *Bacteria* probably occurred in many different lineages independently [16].

Acknowledgements. This work was supported by a research grant from the National Science and Engineering Research Council of Canada.

References

1. Acca M, Bocchetta M, Ceccarelli E, Creti R, Stetter KO, Cammarano P (1994) Updating mass and composition of archeal and bacterial ribosomes. Archeal-like features of ribosomes from the deep-branching bacterium *Aquifex pyrophilus*. System Appl Microbiol 16:629–637
2. Aravind L, Tatusov RL, Wolf YI, Walker DR, Koonin EV (1998) Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. Trends Genet 14:442–444
3. Baldauf SL, Palmer JD (1993) Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. Proc Natl Acad Sci USA 90:11558–11562
4. Bocchetta M, Gribaldo S, Sanangelantoni A, Cammarano P (2000) Phylogenetic depth of the bacterial genera *Aquifex* and *Thermotoga* inferred from analysis of ribosomal protein, elongation factor, and RNA polymerase subunit sequences. J Mol Evol 50:366–380
5. Brendel V, Brocchieri L, Sandler SJ, Clark AJ, Karlin S (1997) Evolutionary comparisons of RecA-like proteins across all major kingdoms of living organisms. J Mol Evol 44:528–541
6. Brown JR, Doolittle WF (1995) Root of the universal tree of life based on ancient aminoacyl- tRNA synthetase gene duplications. Proc Natl Acad Sci USA 92:2441–2445
7. Brown JR, Doolittle WF (1997) Archaea and the prokaryote-to-eukaryote transition. Microbiol Mol Biol Rev 61:456–502
8. Burggraf S, Olsen GJ, Stetter KO, Woese CR (1992) A phylogenetic analysis of *Aquifex pyrophilus*. System Appl Microbiol 15:353–356
9. Campbell AM (2000) Lateral gene transfer in prokaryotes. Theor Popul Biol 57:71–77
10. Daubin V, Gouy M, Perriere G (2002) A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. Genome Res 12:1080–1090
11. Deckert G, Warren PV, Gaasterland T et al. (1998) The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. Nature 392:353–358
12. Di Giulio M (2000) The universal ancestor lived in a thermophilic or hyperthermophilic environment. J Theor Biol 203:203–213
13. Doolittle WF (1999) Phylogenetic classification and the universal tree. Science 284:2124–2128
14. Eisen JA (1995) The RecA protein as a model molecule for molecular systematic studies of bacteria: comparison of trees of RecAs and 16S rRNAs from the same species. J Mol Evol 41:1105–1123
15. Felsenstein J (1978) Cases in which parsimony and compatibility methods will be positively misleading. System Zool 27:401–410
16. Galtier N, Tourasse N, Gouy M (1999) A nonhyperthermophilic common ancestor to extant life forms. Science 283:220–221
17. Gribaldo S, Lumia V, Creti R, De Macario EC, Sanangelantoni A, Cammarano P (1999) Discontinuous occurrence of the hsp70 (dnaK) gene among Archaea and sequence features of HSP70 suggest a novel outlook on phylogenies inferred from this protein. J Bacteriol 181:434–443
18. Gruber TM, Bryant DA (1998) Characterization of the group 1 and group 2 sigma factors of the green sulfur bacterium *Chlorobium tepidum* and the green non-sulfur bacterium *Chloroflexus auranticus*. Arch Microbiol 170:285–296
19. Gupta RS (1998) Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. Microbiol Mol Biol Rev 62:1435–1491
20. Gupta RS (2000) The phylogeny of proteobacteria: relationships to other eubacterial phyla and eukaryotes. FEMS Microbiol Rev 24:367–402
21. Gupta RS (2001) The branching order and phylogenetic placement of species from completed bacterial genomes, based on conserved indels found in various proteins. Int Microbiol 4:187–202

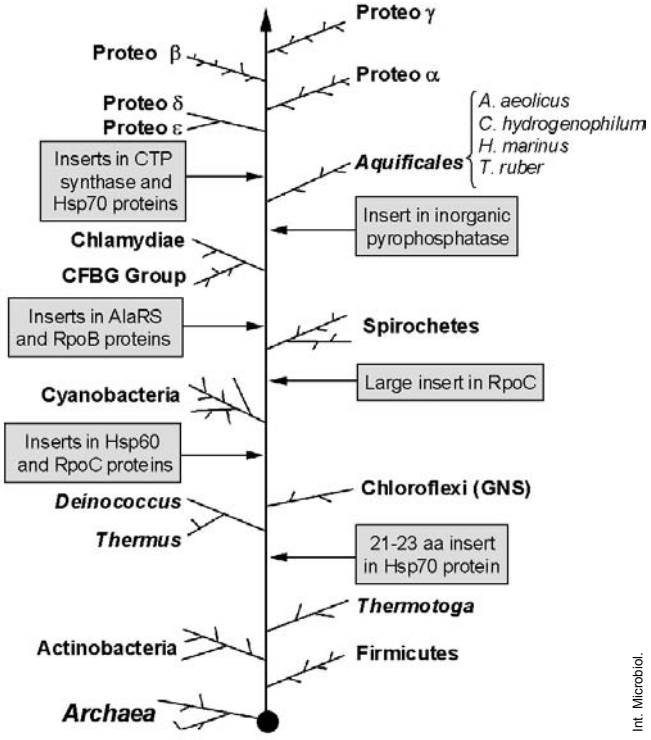

Fig. 8. A summary diagram indicating the evolutionary branch points where the signature sequences studied in the present work may have been introduced. All of the marked signatures are present in various groups above the indicated insertion points (i.e. those diverging later), but not in the groups lying below, which presumably diverged earlier. The insert in the RpoC protein, due to its different length, also distinguishes cyanobacteria from other *Bacteria*. The inserts in Hsp70, AlaRS, RpoB and RpoC are also not found in any archaeal homologs, supporting the inference that the groups indicated at the base of the tree are ancestral. The relative branching of other groups not studied here is based on signatures described in earlier work [21,23].

22. Gupta RS (2003) Evolutionary relationships among photosynthetic bacteria. Photosynth Res 76:173–183
23. Gupta RS, Griffiths E (2002) Critical issues in bacterial phylogenies. Theor Popul Biol 61:423–434
24. Huber R, Eder W, Heldwein S, et al. (1998) *Thermocrinis ruber* gen. nov., sp. nov., a pink-filament-forming hyperthermophilic bacterium isolated from Yellowstone National Park. Appl Environ Microbiol 64:3576–3583
25. Jain R, Rivera M, Lake JA (1999) Horizontal gene transfer among genomes: the complexity hypothesis. Proc Natl Acad Sci USA 96:3801–3806
26. Karlin S (1995) Statistical significance of sequence patterns in proteins. Curr Opin Struct Biol 5:360–371
27. Klenk HP, Meier TD, Durovic P, et al. (1999) RNA polymerase of *Aquifex pyrophilus*: implications for the evolution of the bacterial rpoBC operon and extremely thermophilic bacteria. J Mol Evol 48:528–541
28. Koch AL (2003) Were gram-positive rods the first bacteria? Trends Microbiol 11:166–170
29. Ludwig W, Klenk H-P Overview (2001) A phylogenetic backbone and taxonomic framework for prokaryotic systematics. In: Boone DR, Castenholz RW (eds) Bergey's Manual of Systematic Bacteriology. Springer-Verlag, Berlin, 2nd edn., pp. 49–65
30. Ludwig W, Schleifer KH (1999) Phylogeny of Bacteria beyond the 16S rRNA standard. ASM News 65:752–757

31. Macario AJL, Lange M, Ahring BK, De Macario EC (1999) Stress genes and proteins in the archaea. Microbiol Mol Biol Rev 63:923–967
32. Mikulik K, Qiao CL, Petrik T, Puscheva MA, Zavarzin GA (1988) Elongation factor Tu of the extreme thermophilic hydrogen oxidizing bacterium *Calderobacterium hydrogenophilum*. Biochem Biophys Res Commun 155:384–391
33. Olsen GJ, Woese CR, Overbeek R (1994) The winds of (evolutionary) change: breathing new life into microbiology. J Bacteriol 176:1–6
34. Philippe H, Laurent J (1998) How good are deep phylogenetic trees. Curr Opin Genet Dev 8:616–623
35. Pitulle C, Yang Y, Marchiani M et al. (1994) Phylogenetic position of the genus *Hydrogenobacter*. Int J Syst Bacteriol 44:620–626
36. Reysenbach A-L, Phylum BI (2001) Aquificae *phy. nov.* In: Boone DR, Castenholz RW (eds) Bergey's Manual of Systematic Bacteriology. Springer-Verlag, Berlin, 2nd edn., pp. 359–367
37. Rivera MC, Lake JA (1992) Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. Science 257:74–76
38. Stohr R, Waberski A, Volker H, Tindall BJ, Thomm M (2001) *Hydrogenothermus marinus* gen. nov., sp. nov., a novel thermophilic hydrogen-oxidizing bacterium, recognition of *Calderobacterium hydrogenophilum* as a member of the genus *Hydrogenobacter* and proposal of the reclassification of *Hydrogenobacter acidophilus* as *Hydrogenobaculum acidophilum* gen. nov., comb. nov., in the phylum 'Hydrogenobacter/Aquifex'. Int J Syst Evol Microbiol 51:1853–1862
39. Tekaia F, Lazcano A, Dujon B (1999) The genomic tree as revealed from whole proteome comparisons. Genome Res 9:550–557
40. Wetmur JG, Wong DM, Ortiz B, Tong J, Reichert F, Gelfand DH (1994) Cloning, sequencing, and expression of RecA proteins from three distantly related thermophilic eubacteria. J Biol Chem 269:25928–25935
41. Woese CR (1987) Bacterial evolution. Microbiol Rev 51:221–271

Las secuencias signatura de diversas proteínas demuestran la divergencia tardía del Orden *Aquificales*

Resumen. Actualmente se cree que las especies de *Aquificales* son las que primero se separaron dentro del dominio *Bacteria*. No obstante, el orden de ramificación de este grupo no está resuelto y en los diferentes árboles filogenéticos es altamente variable. En este trabajo hemos examinado la posición filogenética de *Aquificales* mediante un enfoque cladístico basado en la presencia o ausencia de secuencias signatura definidas (consistentes en adiciones o deleciones conservadas) en muchas proteínas importantes y muy conservadas, como son la RNA polimerasa β (RpoB), la RNA polimerasa β´ (RpoC), la alanil-tRNA sintetasa (AlaRS), la CTP sintasa, la pirofosfatasa inorgánica (PPasa), Hsp70 y Hsp60. Con este objeto, se clonaron fragmentos de los genes de las proteínas enumeradas que contenían las regiones signatura provenientes de diferentes especies de *Aquificales* (*Calderobacterium hydrogenophilum*, *Hydrogenobacter marinus* y *Thermocrinis ruber*) y se compararon las secuencias con las disponibles del resto de las especies. La presencia de insertos distintivos en las proteínas Hsp70 y Hsp60 de las especies de *Aquificales*, no presentes en ninguna especie de Firmicutes, Actinobacteria o *Thermotoga*-*Clostridium*, las excluyen de estos grupos del dominio *Bacteria*. La presencia compartida de importantes inserciones-deleciones en las proteínas RpoB (>100 aa), RpoC (>100 aa) y AlaRS (4 aa) que sólo se encuentran en varias especies de *Aquificales,* así como de Clamidias, el grupo CFBG (*Cytophaga*-Flavobacterias-*Bacteroides*-Bacterias verdes del azufre) y las Proteobacterias indica su pertenencia a estos grupos de *Bacteria*. Un inserto de 2 aa en la pirofosfatasa inorgánica, únicamente presente en los genes homólogos de *Aquificales* y Proteobacterias, parece indicar una relación específica entre estos dos filums. No obstante, las especies de *Aquificales* carecen de algunas otras signaturas de proteínas (por ejemplo, los indeles en CTP sintasa y Hsp70) características de las Proteobacterias, lo cual indica que constituyen un filum separado pero relacionado con las Proteobacterias. Estos resultados prueban intensa y consistentemente que las *Aquificales* se separaron después de la ramificación de los grupos Firmicutes, Actinobacterias, *Thermotoga*, *Deinococcus*-*Thermus*, Bacterias verdes del azufre, Cianobacterias, Espiroquetas y Clamidias-CFBG, pero antes de la emergencia de las Proteobacterias. [**Int Microbiol** 2004; 7(1):41–52]

Palabras clave: *Aquifex* · filogenia bacteriana · orden de ramificación

As seqüências assinaturas de diversas proteínas demonstram a divergência tardia da Ordem *Aquificales*

Resumo. Atualmente acredita-se que as espécies de *Aquificales* são as que primeiro se separaram dentro do dominio *Bacteria*. No entanto, a ordem de ramificação deste grupo não está resolvida e é altamente variável nas diferentes árvores filogenéticas. Neste trabalho foi examinada a posição filogenética de *Aquificales* mediante uma aproximação cladística baseada na presença ou ausência compartilhada de seqüências com assinaturas definidas (representando adições ou deleções conservadas), encontradas em muitas proteinas importantes e altamente conservadas como são a RNA polimerase β (RpoB), a RNA polimerase β' (RpoC), a alanil-tRNA sintetase (AlaRS), a CTP sintase, a pirosfosfatase inorgânica (PPase), Hsp 70 e Hsp60. Com esse obtetivo foram clonados fragmentos dos genes das proteínas enumeradas e que continham as regiões assinatura provenientes de diferentes espécies de Aquificales (*Calderobacterium hydrogenophilum*, *Hydrogenobacter marinus* e *Thermocrinis ruber*) e comparadas as seqüências com as disponíveis para as demais espécies. A presença de inserções distintas nas proteínas Hsp70 e Hsp60 das espécies de *Aquificales* as quais não foram encontrados em nenhuma espécie de Firmicutes, Actinobactérias ou *Thermotoga*-*Clostridium*, exclue estes grupos do domínio *Bacteria*. A presença compartilhada de importantes inserções-deleções nas proteínas RpoB (>100 aa), RpoC (>100 aa) e AlaRS (4 aa) em várias espécies de *Aquificales* assim como de Clamídeas, o grupo CFBG (*Cytophaga*-Flavobactérias-*Bacteroides*-Bactérias verdes do enxofre) e as Proteobactérias indica fortemente a inserção dessas espécies nos grupos de *Bacteria*. Uma adição de 2aa na pirofosfatase inorgânica, presente unicamente nos genes homólogos de *Aquificales* e Proteobactérias, parece indicar uma relação específica entre estes dois filos. Entretanto, as espécies de *Aquificales* não possuem assinatura em algumas proteínas (por exemplo, os indels em CTP sintase e Hsp70) que são características das Proteobactérias indicando que elas constituem um filo distinto, embora relacionado com as Proteobactérias. Estes resultados representam uma forte e consistente evidência de que os *Aquificales* se divergiram depois da ramificação dos Firmicutes, Actinobactérias, *Thermotoga*, *Deinococcus*-*Thermus*, Bactérias verdes do enxofre, Cianobactérias, Espioquetas e os grupos Clamídeas-CFBG, porém antes do surgimento das Proteobactérias. [**Int Microbiol** 2004; 7(1):41-52]

Palavras chave: *Aquifex* · filogenia bacteriana · ordem de ramificação