

# Molecular signatures in protein sequences that are characteristics of the phylum *Aquificae*

Emma Griffiths and Radhey S. Gupta

Department of Biochemistry, McMaster University, Hamilton, Ontario, Canada L8N 3Z5

Correspondence  
Radhey S. Gupta  
gupta@mcmaster.ca

Species of the phylum *Aquificae* are of great interest due to their strict extreme thermophilic growth characteristics. Presently, there is no known molecular characteristic which is unique to this group of bacteria. This work describes six conserved inserts and deletions (indels or signature sequences) in four widely distributed proteins that are distinctive features of species from the phylum *Aquificae*. These include three signatures consisting of a 2 aa insert, a 5–6 aa insert and a 6 aa deletion in DNA polymerase I (PolA), a 6–7 aa insert in glucose-inhibited protein A (GidA), a 52 aa insert in the RNA polymerase  $\beta'$ -subunit (RpoC) and a 4 aa insert in elongation factor Tu (EF-Tu). Fragments of these genes were amplified in most cases from *Hydrogenobacter hydrogenophilus*, *Hydrogenothermus marinus* and *Thermocrinis ruber* and combined with available sequence data from '*Aquifex aeolicus*' and *Sulfurihydrogenibium azorense*. The presence of the PolA, GidA and RpoC indels in all of the species sequenced provides evidence that they are probably distinctive characteristics of the entire phylum. The indel in EF-Tu, which is shared by *Aquifex* species and *Hydrogenobacter* but not *Hydrogenothermus* and *Sulfurihydrogenibium*, may provide a molecular marker for the family *Aquificaceae*. We have also identified a 51 aa insert in SecA preprotein translocase that is commonly shared by various species of the *Aquificae* as well as two *Thermotoga* species (*Thermotoga maritima* and *Thermotoga neapolitana*) which may be due to lateral gene transfer between these groups. In phylogenetic trees based on a concatenated dataset of fragments from eight different proteins as well as 16S rRNA, the observed branching pattern of these species was very similar and it was consistent with the relationships inferred from various indels. The identified indels provide a novel means for distinguishing species of the *Aquificae* from all other bacteria in molecular terms and may prove useful for functional studies aimed at understanding the unique biochemical and physiological characteristics of the *Aquificae*.

## INTRODUCTION

The species belonging to the order *Aquificales* constitute a distinct phylum, the *Aquificae*, within the *Bacteria* (Burggraf *et al.*, 1992; Reysenbach, 2001; Eder & Huber, 2002; Huber & Eder, 2002). Members of this group are non-spore-forming, Gram-negative rods which are strictly thermophilic with optimal growth usually occurring above 65 °C (Deckert *et al.*,

1998; Reysenbach, 2001; Huber & Eder, 2002). In terms of their metabolism, most species of the *Aquificae* are hydrogen-oxidizing bacteria which utilize hydrogen as the sole electron donor and oxygen as the electron acceptor (Reysenbach, 2001; Huber & Eder, 2002). Alternatively, thiosulfate or sulfur can also be used as energy sources. Due to their thermostability, many of the enzymes found in the genus *Aquifex* and other thermophilic bacteria are of interest for industrial and biotechnological applications (Huber & Stetter, 1998; van den Burg, 2003).

The phylum *Aquificae* presently comprises two families containing a total of eight genera. The genera *Hydrogenothermus*, *Sulfurihydrogenibium* and *Persephonella* are part of the family *Hydrogenothermaceae*, while *Hydrogenobacter*, *Thermocrinis*, *Aquifex* and the recently described genera *Hydrogenivirga* and *Hydrogenobaculum* form the family *Aquificaceae* (Eder & Huber, 2002; Götz *et al.*, 2002; Huber & Eder, 2002; Nakagawa *et al.*, 2003; Aguiar *et al.*, 2004; Nakagawa *et al.*, 2004). Although the physiological and

Published online ahead of print on 9 September 2005 as DOI 10.1099/ijs.0.63927-0.

Abbreviations: AlaRS, alanyl-tRNA synthetase; EF-Tu, elongation factor-Tu; GidA, glucose-inhibited division protein A; PolA, DNA polymerase I; RNAP, RNA polymerase; RpoC, RNA polymerase  $\beta'$ -subunit.

The GenBank/EMBL/DDBJ accession numbers for the gene fragments cloned and sequenced in this work are AY739864–AY739868, AY739870–AY739874, AY845414 and AY845946–AY845947.

Signature files for the proteins examined in this study are available as supplementary material in IJSEM Online.

metabolic characteristics described above are common among the *Aquificae*, they are not unique to this group of bacteria (Reysenbach, 2001; Huber & Eder, 2002). Hence, the assignment of novel species into the phylum *Aquificae* is at present based almost entirely on their branching patterns in trees based on 16S rRNA gene sequences (Ludwig & Klenk, 2001; Reysenbach, 2001; Eder & Huber, 2002; Huber & Eder, 2002; Takai *et al.*, 2003; Aguiar *et al.*, 2004; Nakagawa *et al.*, 2004). In this context, it is of much interest to identify other novel molecular characteristics that are unique to the *Aquificae* and which could prove useful in the identification and characterization of this group of bacteria. Rare genomic changes such as conserved inserts and deletions (indels; also referred to as signature sequences) in gene/protein sequences that are restricted to particular groups of bacteria provide a powerful and as yet largely untapped resource for developing definitive molecular markers that can clearly distinguish and circumscribe the major groups within the *Bacteria* (Gupta, 1998; Rokas & Holland, 2000; Coenye *et al.*, 2005). In our recent work, a large number of conserved indels in diverse proteins has been identified that are distinctive characteristics of various other bacterial groups such as the *Proteobacteria* (Gupta, 2000), *Chlamydiales* (Griffiths & Gupta, 2002), *Cyanobacteria* (Gupta *et al.*, 2003), the *Fibrobacter–Chlorobi–Bacteroidetes* (Gupta, 2004), *Alphaproteobacteria* (Gupta, 2005), *Actinobacteria* and the *Deinococcus–Thermus* groups (Griffiths & Gupta, 2004a). These taxon- or phylum-specific genetic changes provide valuable molecular markers for taxonomic as well as genetic and biochemical characterization of these groups of bacteria.

In this work, we describe seven rare genomic changes consisting of conserved indels in widely distributed proteins that were originally found to be unique to ‘*Aquifex aeolicus*’ VF5. Among the *Aquificae*, ‘*A. aeolicus*’ is the only species whose genome has been completely sequenced so far (Deckert *et al.*, 1998). Although sequencing of the genome of *Sulfurihydrogenibium azorense* Az-Fu1<sup>T</sup> is currently under way (<http://www.tigr.org/>), very little sequence information is available for other species of the *Aquificae*. To determine whether these indels are distinctive characteristics of the *Aquificae*, sequence information for these gene fragments was obtained in most cases from *Thermocrinis ruber* DSM 12173<sup>T</sup> (Huber *et al.*, 1998), *Hydrogenothermus marinus* (Stöhr *et al.*, 2001) and *Hydrogenobacter hydrogenophilus* INMI Z-829<sup>T</sup> (Kawasumi *et al.*, 1984; Mikulik *et al.*, 1988; Stöhr *et al.*, 2001), covering a number of genera within the phylum *Aquificae* (Reysenbach, 2001; Huber & Eder, 2002). A phylogenetic tree for the *Aquificae* was also constructed based on a concatenated dataset of fragments from eight different proteins and compared with the tree constructed on the basis of 16S rRNA gene sequences.

## METHODS

**PCR amplification and sequencing.** Purified *Hydrogenobacter hydrogenophilus* INMI Z-829<sup>T</sup> DNA was generously made available

to us by Dr Karel Mikulik, *Hydrogenothermus marinus* DSM 12046<sup>T</sup> was kindly donated by Dr Michael Thomm. DNA from *Thermocrinis ruber* DSM 12173<sup>T</sup> was purchased from Dr H. Huber, University of Regensburg, Berlin, Germany.

Degenerate oligonucleotide primers, in opposite orientations, were designed for highly conserved regions that flanked sequences identified as potential indels in the *GidA*, *SecA*, *RpoC*, *PolA* and *EF-Tu* genes. Degeneracy was incorporated into these primers to account for codon variability. Sequences of various PCR primers that were used in these studies are listed in Table 1. All primers were synthesized at MOBIX (Central Facility of the Institute for Molecular Biology and Biotechnology, McMaster University).

**PCR conditions.** PCR of 10 µl (approx. 0.2 µg DNA per reaction; DMSO added to 2% final volume) were optimized for Mg<sup>2+</sup> concentration (in the range 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 45 °C, 1 min at 72 °C, unless otherwise specified) with an initial 1 min hot start at 94 °C and a final extension step (15 s at 94 °C, 15 s at 45 °C, 7 min at 72 °C). *Taq* polymerase (MBI Fermentas) was used for all amplifications. DNA fragments of the expected size were purified from 0.8% (w/v) agarose gels using a GENECLAN kit (MP Biomedicals) and subcloned into the plasmid pDRIVE using a UA cloning kit (Qiagen Inc.). After transformation of *Escherichia coli* JM109 cells with the plasmids, the inserts from a number of positive clones were sequenced. Sequences of all cloned fragments were run through a BLAST search to ensure that each positive clone was from a novel source. All sequenced clones for a given gene and species were found to be identical in the present work.

**Phylogenetic analysis.** Multiple sequence alignments for each of the proteins for which sequence information was available from different species of the *Aquificae* (*A. aeolicus*<sup>2</sup>, *S. azorense*, *Thermocrinis ruber*, *Hydrogenothermus marinus* and *Hydrogenobacter hydrogenophilus*), as well as *Thermus aquaticus*, were created using the Align Plus 4 program package (Scientific and Educational Software). As most of the amplified fragments were small (400 bp–1.0 kb), sequenced fragments of different proteins (*GidA*, *SecA*, *PolA*, inorganic pyrophosphatase, *Hsp60*, *Hsp70*, *CTP synthetase* and *AlaRS*) were concatenated to create an alignment consisting of 886 aa positions. The sequences of all species were trimmed to the same length as the fragments of the species of the *Aquificae* and hypervariable regions were removed in order to maximize conservation for analysis. Genetic distances between bootstrapped datasets (100 replications) of aligned sequences were calculated using Kimura’s method to correct for multiple amino acid substitutions. A neighbour-joining consensus tree was constructed from these data using the TREECON for Windows software package (Van de Peer & De Wachter, 1994). Phylogenetic analyses in the presence or absence of the signature regions yielded similar results. A neighbour-joining tree based on 16S rRNA gene sequences, which included representative species from each genus of the *Aquificae*, was also constructed in the same manner.

**Sequence accession numbers and alignment.** The sequence information for various conserved indels and their flanking conserved regions from different species of the *Aquificae* as well as representative species from other groups of bacteria were compiled into signature files as shown in Fig. 1. Due to space considerations, signature files for other proteins are not included here but they are available as supplementary material in IJSEM Online. These sequence alignments are also available at <http://www.bacterialphylogeny.com>

**Table 1.** List of PCR primers

Degenerate bases in the primer sequences are indicated as: N=A, C, G or T; 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; S=C or G; B=C, G or T; V=A, C or G.

Target	Forward	Reverse	Fragment size
Glucose-inhibited division protein A	5'-AARACNGGNACNCCNGC-3'	5'-TCRTAYTCDATNGCRTCNGC-3'	~468 bp
DNA polymerase I	5'-ASAYCCWAATWTVCARAAAYATHC-3'	5'-CCCCTNCCYTGWATNGGRTARTT-3'	500 bp
RNA polymerase $\beta'$ subunit	<i>Thermocrinis</i> and <i>Hydrogenobacter</i> ; 5'-GAYGTSCTGGCAGGATACC-3'	5'-ATYTCNAARTGYTTRTCRTT-3'	573 bp
SecA protein	<i>Hydrogenothermus</i> ; 5'-GAAATGAAAACY-GGBGAAGG-3'	5'-GGNGTNCNGCYTCMTC-3'	519 bp
	<i>Hydrogenobacter</i> and <i>Thermocrinis</i> ; 5'-GT-NACNGTNAAYGAYTA-3'	5'-TCRTCACDATNGCRTA-3'	399 bp
Elongation factor-Tu	<i>Hydrogenothermus</i> and <i>Hydrogenobacter</i> ; 5'-ACNATHGGNCAYGTNGAYCAYGG-3'	<i>Hydrogenothermus</i> ; 5'-GTYTGNGGCATNG-GNCCRTC-3'	300 bp
		<i>Hydrogenobacter</i> ; 5'-ACCATRTCAGYTTT-TCATRAA-3'	387 bp

## RESULTS

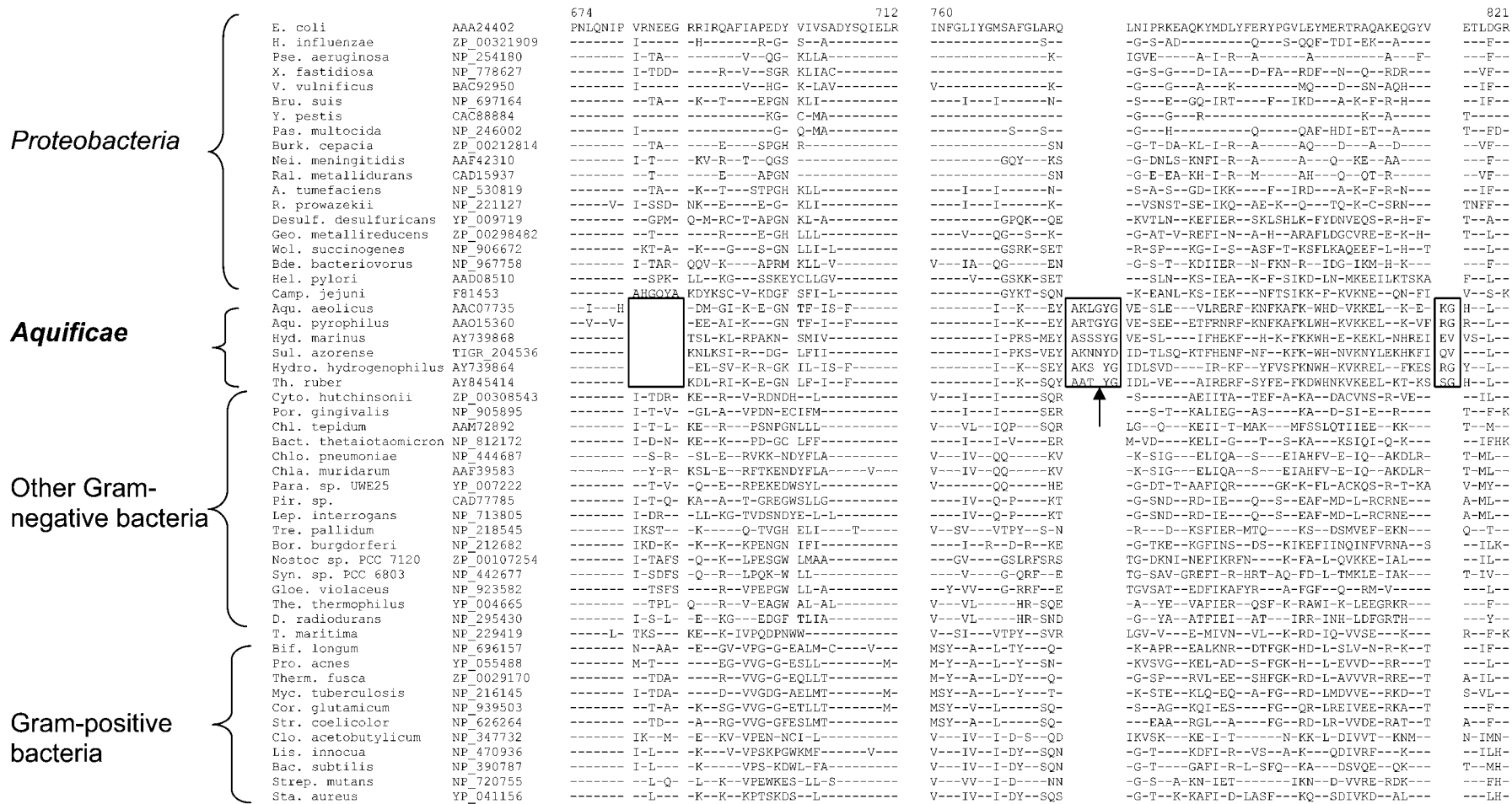
### Identification and characterization of conserved indels specific for the *Aquificae*

The aim of these studies was to identify novel molecular markers that are uniquely shared by species from the phylum *Aquificae* and which would enable distinction of species from this phylum from all other known bacteria in molecular terms. Until recently, sequence information for the phylum *Aquificae* was only available from '*A. aeolicus*' VF5 (Deckert *et al.*, 1998). Sequence alignments of various proteins constructed in our earlier work were examined to identify any indels that were uniquely present in '*A. aeolicus*' (and *Aquifex pyrophilus*, if available) and which were flanked on both sides by conserved regions. When an indel is present in only a single species, it is difficult to determine whether it is unique to that particular species or whether it is a shared characteristic of related bacteria. However, based upon our experience with other conserved indels, it was reasoned that if the indel was present in a conserved region there was a good chance that it might also be present in other related bacteria (Gupta, 2000, 2004; Griffiths & Gupta, 2002, 2004a). Based on this rationale, the sequence alignments of different proteins constructed in our earlier work were examined to identify any indel in a conserved region that was present uniquely in '*A. aeolicus*' VF5. These studies led to identification of a number of indels specific to '*A. aeolicus*' VF5 which could potentially be present in other members of the *Aquificae*. The presence of these indels in other *Aquificae* species was determined by means of PCR amplification of the corresponding fragments from a number of other members of the *Aquificae* (*Thermocrinis ruber*, *Hydrogenothermus marinus* and *Hydrogenobacter hydrogenophilus*) covering some of the other genera within this phylum (Reysenbach, 2001; Huber & Eder, 2002). A brief description of the proteins where these indels are found and the results of our

studies examining the presence of these indels in other members of the *Aquificae* are given below.

There are three highly conserved families of DNA polymerases in bacteria corresponding to homologues of *E. coli* DNA polymerase I (PolA), II (PolB) and III (PolC) (Savic *et al.*, 1990). In the sequence alignment of PolA, three conserved indels, consisting of a 6 aa deletion, a 5–6 aa insertion and a 2 aa insertion, are present in '*A. aeolicus*', *A. pyrophilus* and *S. azorensis* but are not found in any other available sequence. The specificity of these indels was determined by amplification and sequencing of ~0.5 kb fragments of the *polA* genes from *Hydrogenobacter hydrogenophilus*, *Thermocrinis ruber* and *Hydrogenothermus marinus* using the PCR primers listed in Table 1. The amplified fragments were unique in all cases and BLAST searches indicated '*A. aeolicus*' as their closest match. All three cloned fragments were found to contain all of these indels, indicating that they are distinctive for this group of species. The sequence information for all of these species of the *Aquificae* is included in the sequence alignment shown in Fig. 1. It is of interest that, when compared with *Hydrogenothermus*, *Sulfurihydrogenibium* and the two *Aquifex* species which contain a 6 aa insert (second box in Fig. 1), the corresponding inserts in both *Hydrogenobacter* and *Thermocrinis* are only 5 aa long and apparently lack an amino acid in the middle of the 6 aa insert. This suggests a closer relationship between *Hydrogenobacter* and *Thermocrinis* in comparison with the other members of the group.

Homologues of the glucose-inhibited division protein A (GidA homologues) have been classified into two groups on the basis of size. The larger protein, GidA<sub>L</sub>, which is present in most bacteria (except actinobacteria) and eukaryotes, consists of 611–679 aa residues (White *et al.*, 2001). The smaller protein, GidA<sub>S</sub>, which is only present in some bacteria, contains between 435 and 482 aa residues (White



**Fig. 1.** Partial sequence alignment of DNA polymerase I showing three distinct indels that are found uniquely in species of the *Aquificae* (boxed). The sequences for *Hydrogenothermus marinus*, *Hydrogenobacter hydrogenophilus* and *Thermococcus ruber* were obtained in the present work. A 1 aa deletion is present in *Thermococcus* and *Hydrogenobacter* within the 6 aa insert (denoted by arrow), indicating a closer relationship between these two species. Dashes in all sequence alignments indicate identity to the amino acid sequence on the top line. The numbers at the top indicate the position of the sequence in the *E. coli* homologue. Abbreviations of genus names are as follows: A., *Agrobacterium*; Aqu., *Aquifex*; Bac., *Bacillus*; Bact., *Bacteroides*; Bde., *Bdellovibrio*; Bif., *Bifidobacterium*; Bor., *Borrelia*; Bru., *Brucella*; Burk., *Burkholderia*; Camp., *Campylobacter*; Chl., *Chlorobium*; Chla., *Chlamydia*; Chlo., *Chlamydomonas*; Clo., *Clostridium*; Cor., *Corynebacterium*; Cyto., *Cytophaga*; D., *Deinococcus*; Desulf., *Desulfobacterium*; E., *Escherichia*; Gloe., *Gloeobacter*; Geo., *Geobacter*; H., *Haemophilus*; Hel., *Helicobacter*; Hyd., *Hydrogenothermus*; Hydro., *Hydrogenobacter*; Lep., *Leptospira*; Lis., *Listeria*; Myc., *Mycobacterium*; Nei., *Neisseria*; Para., *Parachlamydia*; Pas., *Pasteurella*; Pir., *Pirellula*; Por., *Porphyromonas*; Pro., *Propionibacterium*; Pse., *Pseudomonas*; R., *Rickettsia*; Ral., *Ralstonia*; Sta., *Staphylococcus*; Str., *Streptomyces*; Strep., *Streptococcus*; Sul., *Sulfurihydrogenibium*; Syn., *Synechococcus*; T., *Thermotoga*; Th., *Thermococcus*; The., *Thermus*; Therm., *Thermobifida*; Tre., *Treponema*; V., *Vibrio*; Wol., *Wolinella*; X., *Xylella*; Y., *Yersinia*.

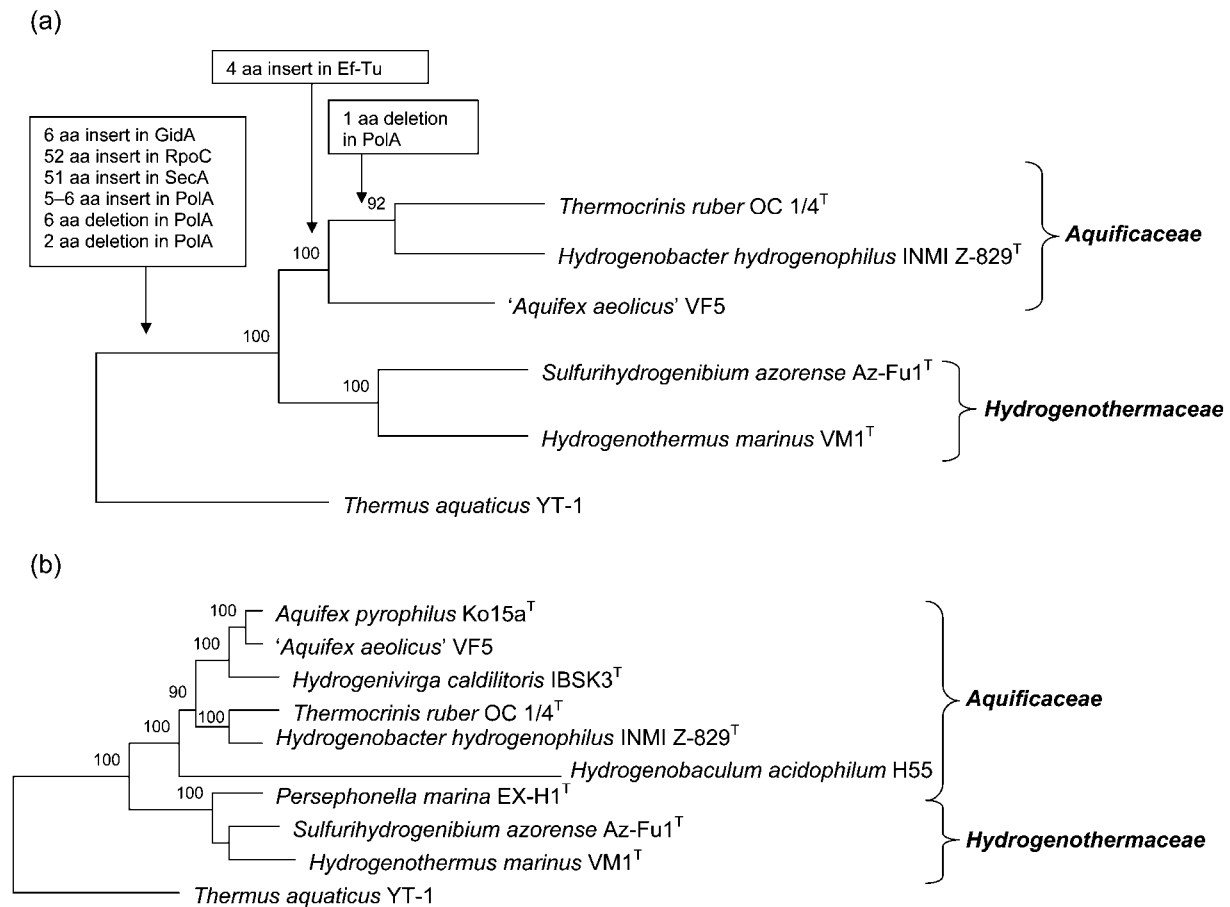
*et al.*, 2001). These homologues are quite divergent in terms of their sequences and are readily distinguished. Although the exact cellular role of the GidA proteins is presently unknown, they have been implicated in a variety of functions including initiation of chromosome replication and as regulatory molecules controlling virulence-associated factors of certain enteric pathogens (Sha *et al.*, 2004). The GidA<sub>L</sub> homologue of '*A. aeolicus*' was found to contain a 7 aa insert in a conserved region that was unique to this species. We amplified ~0.5 kb fragments of the GidA<sub>L</sub> gene covering this region from *Hydrogenobacter hydrogenophilus*, *Thermocrinis ruber* and *Hydrogenothermus marinus*. The amplified fragments were unique in all cases and BLAST searches indicated '*A. aeolicus*' as the closest hit. The sequence information for all of the species of the *Aquificae* as well as representative species from other groups of bacteria is available as Supplementary Fig. S1 in IJSEM Online (see Methods). In contrast with other members of the *Aquificae*, which contained a 7 aa insert, the insert in *S. azorensis* was only 6 aa long, but its sequence and position were highly similar to those from the other species, indicating a common origin.

DNA-dependent RNA polymerase (RNAP) is a central enzyme necessary for gene expression in all organisms (Olsen & Woese, 1997; Klenk *et al.*, 1999). The prototypical eubacterial RNAP contains four core polypeptides, the  $\beta'$ -subunit,  $\beta$ -subunit and a dimer of the  $\alpha$ -subunit, plus one of several possible  $\sigma$  factors, which confer promoter specificity. The core subunits of RNA polymerase are evolutionarily highly conserved in species from all three domains (Olsen & Woese, 1997). In the  $\beta'$ -subunit (RpoC) of the *Aquifex* and *Sulfurihydrogenibium* RNAPs, a large 52 aa insert was identified in the C-terminal region which was unique to these species (see Supplementary Fig. S2 in IJSEM Online). The sequence to the left of this insert shows a much higher degree of conservation compared with that at the C terminus. However, there is enough similarity on both sides to indicate that the identified insert is unique for this group. The specificity of this insert for the phylum *Aquificae* was determined by PCR amplification of portions of the *rpoC* gene from *Hydrogenobacter hydrogenophilus* and *Thermocrinis ruber*. PCR amplification was not successful with *Hydrogenothermus marinus*, possibly due a greater sequence divergence in the primer-binding regions of this species. However, for both *Hydrogenobacter hydrogenophilus* and *Thermocrinis ruber*, the signature insert was present and its sequence was highly conserved (>65% amino acid similarity) (see Supplementary Fig. S2 in IJSEM Online). As the RpoC insert is found in *S. azorensis*, which branches closely with *Hydrogenothermus* in 16S rRNA trees (Huber & Eder, 2002; Aguiar *et al.*, 2004), it is likely that this insert will also be present in *Hydrogenothermus marinus*. Due to its large size and high degree of sequence conservation, this indel provides a good molecular marker for species of the *Aquificae* and it probably plays an important role in the function of RNAP from this group of thermophiles (Iyer *et al.*, 2004).

In the protein synthesis translation elongation factor EF-Tu, which is found ubiquitously in species from all three domains (Hashimoto & Hasegawa, 1996), a conserved 4 aa insert was present in '*A. aeolicus*' and *A. pyrophilus* near the N terminus (see Supplementary Fig. S3 in IJSEM Online). We were successful in amplifying the EF-Tu gene fragment covering this region from *Hydrogenobacter hydrogenophilus* and *Hydrogenothermus marinus* but not from *Thermocrinis ruber* DNA. These studies revealed that, while the insert in EF-Tu was present in *Hydrogenobacter hydrogenophilus*, '*A. aeolicus*' and *A. pyrophilus*, it was lacking in *Hydrogenothermus marinus* and *S. azorensis* (see Supplementary Fig. S3 in IJSEM Online). Since the former three species form part of the family *Aquificaceae* (Reysenbach, 2001; Huber & Eder, 2002), it is possible that this signature is only a distinctive characteristic for this family.

We also found a 51 aa insert in the SecA preprotein translocase which was present in '*A. aeolicus*' and *S. azorensis* and also *Thermotoga maritima* and *Thermotoga neapolitana* (see Supplementary Fig. S4 in IJSEM Online). *secA* gene fragments were amplified from *Thermocrinis ruber*, *Hydrogenothermus marinus* and *Hydrogenobacter hydrogenophilus* and were all found to contain this large insert (Supplementary Fig. S4) indicating that it is a shared characteristic of the phylum *Aquificae* (indel amino acid similarity >49%). It is possible that the presence of this large insert in both these groups of thermophilic organisms is a consequence of lateral gene transfer (Aravind *et al.*, 1998; Nelson *et al.*, 1999).

Apart from 16S rRNA gene sequences, no sequence information has been available for many members of the *Aquificae*. However, in the present study and another recent study (Griffiths & Gupta, 2004b), we have cloned and sequenced fragments of several genes from species covering several genera from the two families of the phylum *Aquificae*. Hence, it was of interest to compare the interrelationships within the phylum *Aquificae* based on these protein sequences (Fig. 2a) with that of 16S rRNA gene sequences (Fig. 2b) and to determine whether the relationships indicated by these trees were in accordance with those suggested by the indels (Fig. 2a). For this purpose, phylogenetic trees were constructed based on a concatenated dataset of fragments from eight proteins [GidA, SecA and PolA from the present work and inorganic pyrophosphatase, AlaRS, CTP synthase, Hsp70 and Hsp60 proteins from Griffiths & Gupta (2004b)] as well as 16S rRNA sequences. The two trees showed very similar topologies with good resolution for most of the internal branches. Two main clades were identified in both trees, corresponding to the two recognized families within this order (Reysenbach, 2001; Eder & Huber, 2002; Huber & Eder, 2002; Aguiar *et al.*, 2004; Nakagawa *et al.*, 2004). *Hydrogenothermus*, *Sulfurihydrogenibium* and *Persephonella* constituted one clade (family *Hydrogenothermaceae*) while *Hydrogenobacter*, *Thermocrinis*, *Aquifex*, *Hydrogenivirga* and *Hydrogenobaculum* made up the family *Aquificaceae* (Eder & Huber, 2002;



**Fig. 2.** Summary diagram illustrating the evolutionary stages where different signatures identified in the present work were probably introduced. The sequence information for other signatures shown in this diagram is available as supplementary material in IJSEM Online and at <http://www.bacterialphylogeny.com> (a) Bootstrapped phylogenetic tree based on a concatenated sequence alignment of fragments from eight proteins (GidA, SecA, PolA, inorganic pyrophosphatase, Hsp70, Hsp60, AlaRS and CTP synthetase). The bootstrap values (out of 100) are marked on various internodes. (b) Neighbour-joining bootstrapped tree for 16S rRNA gene sequences. The accession numbers of 16S RNA gene sequences used in construction of this tree are as follows: M83548, *Aquifex pyrophilus* Ko15a<sup>T</sup>; NC\_000918, '*Aquifex aeolicus*' VF5; AB120294, *Hydrogenivirga caldillitoris* IBSK3<sup>T</sup>; AJ005640, *Thermocrinis ruber* DSM 12173<sup>T</sup>; Z30242, *Hydrogenobacter hydrogenophilus* INMI Z-829<sup>T</sup>; AY268103, *Hydrogenobaculum acidophilum* H55; AF188332, *Persephonella marina* EX-H1<sup>T</sup>; AF528192, *Sulfurihydrogenibium azorense* Az-Fu1<sup>T</sup>; AJ292525, *Hydrogenothermus marinus* DSM 12046<sup>T</sup>; X58343, *Thermus aquaticus* YT-1.

Huber & Eder, 2002; Aguiar *et al.*, 2004; Nakagawa *et al.*, 2004). Although sequence information for all of the members of the *Aquificae* is not available for all of the proteins examined, the relationships seen in these trees are identical for various common species.

## DISCUSSION

In this study, we have identified a number of rare genomic changes consisting of conserved indels in protein sequences that are distinctive characteristics of various species of the *Aquificae*. For the signatures described here, sequence information was obtained, or was already available, for

species covering at least two genera from each of the two families. The five conserved indels in PolA, GidA and RpoC proteins that are unique to various species of the *Aquificae* probably provide phylum-specific molecular signatures. The presence of these signatures in various sequences from species of both families of the *Aquificae* indicates that they may also be found in other species of this group for which sequence information is not yet available. In addition to these probable phylum-specific signatures, a conserved insert in the EF-Tu protein is specific to the family *Aquificaceae* and provides a means to distinguish members of this family from the *Hydrogenothermaceae*. Further, a 1 aa deletion in the PolA protein that is unique to the genera

*Hydrogenobacter* and *Thermocrinis* suggests a closer relationship between these species, which is supported by phylogenetic analysis (Fig. 2).

The taxonomic groups within the phylum *Aquificae* are presently identified solely on the basis of their branching pattern in 16S rRNA gene sequence trees (Reysenbach, 2001; Huber & Eder, 2002). The molecular signatures described here provide molecular markers for identifying and circumscribing some of these taxonomic clades in more definitive terms (Coenye *et al.*, 2005). Species of the *Aquificae* have been previously reported to contain a high proportion of the unusual fatty acids *n*-C<sub>20</sub> and *cy*-C<sub>21</sub> in their membranes (Jahnke *et al.*, 2001). However, the presence of such fatty acids is not a unique characteristic of members of the *Aquificae* and the amounts of these compounds in different species are variable (Jahnke *et al.*, 2001). Another putative signature for the *Aquificae* consists of a single nucleotide bulge at positions 198–219 (*E. coli* numbering) in the secondary structure of the 16S rRNA (L'Haridon *et al.*, 1998). However, this motif is also present in the *Deltaproteobacteria*, indicating that it is not a distinctive characteristic of the *Aquificae*. In contrast to these markers, the molecular markers described here appear to be specific for this phylum and their presence or absence can also be readily ascertained. It has been proposed that the *Aquificae* should be grouped with the *Epsilonproteobacteria* (Cavalier-Smith, 2002). However, none of the *Aquificae* signatures identified in our work were found in *Epsilonproteobacteria*. Further, we have also identified a number of protein signatures that are distinctive characteristics of the *Epsilonproteobacteria*. None of these signatures are present in the *Aquificae* (Gupta, 2000; R. S. Gupta, unpublished results). These results indicate that these two groups are distinct from each other (Coenye & Vandamme, 2004; Griffiths & Gupta, 2004b; Shatalkin, 2004). Signature sequences in a number of different proteins provide evidence that the *Aquificae* is a late branching group within the *Bacteria* (Gupta & Griffiths, 2002; Griffiths & Gupta, 2004b). However, the late branching of the *Aquificae* is not supported by the 16S rRNA gene sequence trees and certain other lines of evidence (Reysenbach, 2001; Coenye & Vandamme, 2004). Some possible reasons for this discrepancy have been discussed in our earlier work (Griffiths & Gupta, 2004b).

We have also come across a large insert in the SecA protein that is shared uniquely by various species of the *Aquificae* as well as two *Thermotoga* species (*Thermotoga maritima* and *Thermotoga neapolitana*). The presence of this insert in various *Aquificae* species indicates that it is a distinctive characteristic of this entire phylum and that it was introduced into a common ancestor of this phylum in a similar manner to a number of other signatures described here (e.g. GidA, RpoC and PolA). It is unclear at present whether this insert is also present in other species of the *Thermotogae* or if it is limited to either some or all species from the genus *Thermotoga*. The shared presence of this

insert in two thermophilic groups could be interpreted as either indicative of a specific relationship between these two groups (Bocchetta *et al.*, 2000; Coenye & Vandamme, 2004) or as an example of lateral gene transfer from one group to the other. The signature sequences in a number of other highly conserved proteins, Hsp60, Hsp70, RNA polymerase  $\beta$ , RNA polymerase  $\beta'$ , alanyl-tRNA synthetase, CTP synthetase and inorganic pyrophosphatase, indicate that these two phyla are distinct from each other (Griffiths & Gupta, 2004b). In view of these results, we favour the possibility that the shared presence of this insert in the SecA protein in these two thermophilic bacteria phyla is due to lateral gene transfer. Analyses of genome sequence also strongly indicate that many of the genes in these hyperthermophilic bacteria have undergone lateral gene transfer from *Archaea* as well as other hyperthermophiles (Aravind *et al.*, 1998; Nelson *et al.*, 1999; Nesbo *et al.*, 2001).

In addition to their usefulness for taxonomic studies, the *Aquificae*-specific signatures identified in the present work are also of much interest for genetic and biochemical investigations. The identified signatures represent major 'site-specific' genetic changes in highly conserved and in some cases ubiquitous proteins that are limited to species from the phylum *Aquificae*. Because these inserts or deletions, once introduced in a common ancestor of the *Aquificae*, have not been lost from any of the sequenced species, it is highly likely that they perform essential biological roles in the functioning of these organisms. In view of the evolutionarily conserved nature of these genetic changes, it is reasonable to assume that they are modifying the cellular functions of these proteins (either their biochemical activity or their regulation) in some unique manner which is distinctive for this group of bacteria. It is of interest to note that the indel in PolA falls in a region which affects the accuracy of replication (Minnick *et al.*, 1999), whereas that in the EF-Tu protein lies close to a region that acts as a sensor for the presence or absence of GTP and transmits this signal to other domains of the protein (Vogele *et al.*, 2001). The large insert in the SecA protein is also present in a region that is proximal to a nucleotide-binding fold essential for its function (Osborne *et al.*, 2004). Hence, further studies aimed at understanding the functional significance of these genetic changes should be of much interest and could provide useful insights concerning the unique biochemical and physiological characteristics of species from this phylum.

## ACKNOWLEDGEMENTS

We are grateful to Dr Karel Mikulik of the Academy of Sciences of the Czech Republic, Czech Republic and Dr Michael Thomm of the Institut für Allgemeine Mikrobiologie, Kiel, Germany, for the generous gifts of DNA from *Hydrogenobacter* (*Calderobacterium*) and *Hydrogenothermus*, respectively. This work was supported by a research grant from the National Science and Engineering Research Council of Canada.

## REFERENCES

- Aguiar, P., Beveridge, T. J. & Reysenbach, A. L. (2004). *Sulfurihydrogenibium azorense* sp. nov., a thermophilic hydrogen-oxidizing microaerophile from terrestrial hot springs in the Azores. *Int J Syst Evol Microbiol* **54**, 33–39.
- Aravind, L., Tatusov, R. L., Wolf, Y. I., Walker, D. R. & Koonin, E. V. (1998). Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. *Trends Genet* **14**, 442–444.
- 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.
- Burggraf, S., Olsen, G. J., Stetter, K. O. & Woese, C. R. (1992). A phylogenetic analysis of *Aquifex pyrophilus*. *Syst Appl Microbiol* **15**, 353–356.
- Cavaliere-Smith, T. (2002). The neomuran origin of archaeobacteria, the negibacterial root of the universal tree and bacterial mega-classification. *Int J Syst Evol Microbiol* **52**, 7–76.
- Coenye, T. & Vandamme, P. (2004). A genomic perspective on the relationship between the *Aquificales* and the epsilon-Proteobacteria. *Syst Appl Microbiol* **27**, 313–322.
- Coenye, T., Gevers, D., Van de Peer, Y., Vandamme, P. & Swings, J. (2005). Towards a prokaryotic genomic taxonomy. *FEMS Microbiol Rev* **29**, 147–167.
- Deckert, G., Warren, P. V., Gaasterland, T. & 12 other authors (1998). The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* **392**, 353–358.
- Eder, W. & Huber, R. (2002). New isolates and physiological properties of the *Aquificales* and description of *Thermocrinis albus* sp. nov. *Extremophiles* **6**, 309–318.
- Götz, D., Banta, A., Beveridge, T. J., Rushdi, A. I., Simoneit, B. R. & Reysenbach, A. L. (2002). *Persephonella marina* gen. nov., sp. nov. and *Persephonella guaymasensis* sp. nov., two novel, thermophilic, hydrogen-oxidizing microaerophiles from deep-sea hydrothermal vents. *Int J Syst Evol Microbiol* **52**, 1349–1359.
- Griffiths, E. & Gupta, R. S. (2002). Protein signatures distinctive of chlamydial species: horizontal transfer of cell wall biosynthesis genes *glmU* from archaea to chlamydiae, and *murA* between chlamydiae and *Streptomyces*. *Microbiology* **148**, 2541–2549.
- Griffiths, E. & Gupta, R. S. (2004a). Distinctive protein signatures provide molecular markers and evidence for the monophyletic nature of the *Deinococcus-Thermus* phylum. *J Bacteriol* **186**, 3097–3107.
- Griffiths, E. & Gupta, R. S. (2004b). Signature sequences in diverse proteins provide evidence for the late divergence of the Order *Aquificales*. *Int Microbiol* **7**, 41–52.
- Gupta, R. S. (1998). Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among archaeobacteria, eubacteria, and eukaryotes. *Microbiol Mol Biol Rev* **62**, 1435–1491.
- Gupta, R. S. (2000). The phylogeny of proteobacteria: relationships to other eubacterial phyla and eukaryotes. *FEMS Microbiol Rev* **24**, 367–402.
- Gupta, R. S. (2004). The phylogeny and signature sequences characteristics of *Fibrobacteres*, *Chlorobi* and *Bacteroidetes*. *Crit Rev Microbiol* **30**, 123–143.
- Gupta, R. S. (2005). Protein signatures distinctive of alpha proteobacteria and its subgroups and a model for alpha proteobacterial evolution. *Crit Rev Microbiol* **31**, 101–135.
- Gupta, R. S. & Griffiths, E. (2002). Critical issues in bacterial phylogeny. *Theor Popul Biol* **61**, 423–434.
- Gupta, R. S., Pereira, M., Chandrasekera, C. & Johari, V. (2003). Molecular signatures in protein sequences that are characteristic of cyanobacteria and plastid homologues. *Int J Syst Evol Microbiol* **53**, 1833–1842.
- Hashimoto, T. & Hasegawa, M. (1996). Origin and early evolution of eukaryotes inferred from the amino acid sequences of translation elongation factors 1alpha/Tu and 2/G. *Adv Biophys* **32**, 73–120.
- Huber, R. & Eder, W. (2002). *Aquificales*. In *The Prokaryotes: an Evolving Electronic Resource for the Microbiological Community*. Release 3.8. Edited by M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer & E. Stackebrandt. New York: Springer. <http://link.springer-ny.com/link/service/books/10125/>
- Huber, H. & Stetter, K. O. (1998). Hyperthermophiles and their possible potential in biotechnology. *J Biotechnol* **64**, 39–52.
- Huber, R., Eder, W., Heldwein, S., Wanner, G., Huber, H., Rachel, R. & Stetter, K. O. (1998). *Thermocrinis ruber* gen. nov., sp. nov., a pink-filament-forming hyperthermophilic bacterium isolated from Yellowstone national park. *Appl Environ Microbiol* **64**, 3576–3583.
- Iyer, L. M., Koonin, E. V. & Aravind, L. (2004). Evolution of bacterial RNA polymerase: implications for large-scale bacterial phylogeny, domain accretion, and horizontal gene transfer. *Gene* **335**, 73–88.
- Jahnke, L. L., Eder, W., Huber, R., Hope, J. M., Hinrichs, K. U., Hayes, J. M., Des Marais, D. J., Cady, S. L. & Summons, R. E. (2001). Signature lipids and stable carbon isotope analyses of Octopus Spring hyperthermophilic communities compared with those of *Aquificales* representatives. *Appl Environ Microbiol* **67**, 5179–5189.
- Kawasumi, T., Igarshi, Y., Kodama, T. & Minoda, Y. (1984). *Hydrogenobacter thermophilus* gen. nov., sp. nov., an extremely thermophilic, aerobic, hydrogen-oxidizing bacterium. *Int J Syst Bacteriol* **34**, 5–10.
- Klenk, H. P., Meier, T. D., Durovic, P., Schwass, V., Lottspeich, F., Dennis, P. P. & Zillig, W. (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.
- L'Haridon, S., Cilia, V., Messner, P., Ragueneas, G., Gambacorta, A., Sleytr, U. B., Prieur, D. & Jeanthon, C. (1998). *Desulfurobacterium thermolithotrophum* gen. nov., sp. nov., a novel autotrophic, sulphur-reducing bacterium isolated from a deep-sea hydrothermal vent. *Int J Syst Bacteriol* **48**, 701–711.
- Ludwig, W. & Klenk, H.-P. (2001). Overview: a phylogenetic backbone and taxonomic framework for prokaryotic systematics. In *Bergey's Manual of Systematic Bacteriology*, 2nd edn, vol. 1, pp. 49–65. Edited by D. R. Boone & R. W. Castenholz. Berlin: Springer.
- Mikulik, K., Qiao, C. L., Petrik, T., Puscheva, M. A. & Zavarzin, G. A. (1988). Elongation factor Tu of the extreme thermophilic hydrogen oxidizing bacterium *Calderobacterium hydrogenophilum*. *Biochem Biophys Res Commun* **155**, 384–391.
- Minnick, D. T., Bebenek, K., Osheroff, W. P., Turner, R. M., Jr, Astatke, M., Liu, L., Kunkel, T. A. & Joyce, C. M. (1999). Side chains that influence fidelity at the polymerase active site of *Escherichia coli* DNA polymerase I (Klenow fragment). *J Biol Chem* **274**, 3067–3075.
- Nakagawa, S., Takai, K., Horikoshi, K. & Sako, Y. (2003). *Persephonella hydrogeniphila* sp. nov., a novel thermophilic, hydrogen-oxidizing bacterium from a deep-sea hydrothermal vent chimney. *Int J Syst Evol Microbiol* **53**, 863–869.
- Nakagawa, S., Nakamura, S., Inagaki, F., Takai, K., Shirai, N. & Sako, Y. (2004). *Hydrogenivirga caldilitoris* gen. nov., sp. nov., a novel extremely thermophilic, hydrogen- and sulfur-oxidizing bacterium from a coastal hydrothermal field. *Int J Syst Evol Microbiol* **54**, 2079–2084.

- Nelson, K. E., Clayton, R. A., Gill, S. R. & 26 other authors (1999).** Evidence for lateral gene transfer between *Archaea* and *Bacteria* from genome sequence of *Thermotoga maritima*. *Nature* **399**, 323–329.
- Nesbo, C. L., L'Haridon, S., Stetter, K. O. & Doolittle, W. F. (2001).** Phylogenetic analyses of two "Archaeal" genes in *Thermotoga maritima* reveal multiple transfers between *Archaea* and *Bacteria*. *Mol Biol Evol* **18**, 362–375.
- Olsen, G. J. & Woese, C. R. (1997).** Archaeal genomics: an overview. *Cell* **89**, 991–994.
- Osborne, A. R., Clemons, W. M., Jr & Rapoport, T. A. (2004).** A large conformational change of the translocation ATPase SecA. *Proc Natl Acad Sci U S A* **101**, 10937–10942.
- Reysenbach, A.-L. (2001).** Phylum BI. *Aquificae* phy. nov. In *Bergey's Manual of Systematic Bacteriology*, 2nd edn, vol. 1, pp. 359–367. Edited by D. R. Boone & R. W. Castenholz. Berlin: Springer.
- Rokas, A. & Holland, P. W. (2000).** Rare genomic changes as a tool for phylogenetics. *Trends Ecol Evol* **15**, 454–459.
- Savic, D. J., Jankovic, M. & Kostic, T. (1990).** Cellular role of DNA polymerase I. *J Basic Microbiol* **30**, 769–784.
- Sha, J., Kozlova, E. V., Fadl, A. A., Olano, J. P., Houston, C. W., Peterson, J. W. & Chopra, A. K. (2004).** Molecular characterization of a glucose-inhibited division gene, *gidA*, that regulates cytotoxic enterotoxin of *Aeromonas hydrophila*. *Infect Immun* **72**, 1084–1095.
- Shatalkin, A. I. (2004).** Highest level of division in classification of organisms. 3. Monodermata and Didermata. *Zh Obshch Biol* **65**, 195–210 (in Russian).
- Stöhr, R., Waberski, A., Volker, H., Tindall, B. J. & 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.
- Takai, K., Kobayashi, H., Neelson, K. H. & Horikoshi, K. (2003).** *Sulfurihydrogenibium subterraneum* gen. nov., sp. nov., from a subsurface hot aquifer. *Int J Syst Evol Microbiol* **53**, 823–827.
- Van de Peer, Y. & De Wachter, R. (1994).** TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Comput Appl Biosci* **10**, 569–570.
- van den Burg, B. (2003).** Extremophiles as a source for novel enzymes. *Curr Opin Microbiol* **6**, 213–218.
- Vogele, L., Palm, G. J., Mesters, J. R. & Hilgenfeld, R. (2001).** Conformational change of elongation factor Tu (EF-Tu) induced by antibiotic binding. Crystal structure of the complex between EF-Tu.GDP and aurodox. *J Biol Chem* **276**, 17149–17155.
- White, D. J., Merod, R., Thomasson, B. & Hartzell, P. L. (2001).** *GidA* is an FAD-binding protein involved in development of *Myxococcus xanthus*. *Mol Microbiol* **42**, 503–517.