Towards a congruent reclassification and nomenclature of the thermophilic species of the genus Pseudothermotoga within the order Thermotogales

Hassiba Belahbib, Zarath Summers, Marie-Laure Fardeau, Manon Joseph, Christian Tamburini, Alain Dolla, Bernard Ollivier, Fabrice Armougom

Towards a congruent reclassification and nomenclature of the thermophilic species of the genus *Pseudothermotoga* within the order *Thermotogales*

Hassiba Belahbib[a], Zarath M. Summers[b], Marie-Laure Fardeau[a], Manon Joseph[a], Christian Tamburini[a], Alain Dolla[c], Bernard Ollivier[a], Fabrice Armougom[a,*]

[a] *Aix Marseille Univ, Université de Toulon, CNRS, IRD, MIO, Marseille, France*
[b] *ExxonMobil Research and Engineering Company, 1545 Route 22 East, Annandale, NJ 08801, United States*
[c] *Aix Marseille Univ, CNRS, LCB, Marseille, France*

ARTICLE INFO

Article history:
Received 15 December 2017
Received in revised form 20 April 2018
Accepted 27 April 2018

Keywords:
Thermotogae
Pseudothermotoga
Taxonomy
Classification

ABSTRACT

The phylum *Thermotogae* gathers thermophilic, hyperthermophic, mesophilic, and thermo-acidophilic anaerobic bacteria that are mostly originated from geothermally heated environments. The metabolic and phenotypic properties harbored by the *Thermotogae* species questions the evolutionary events driving the emergence of this early branch of the universal tree of life. Recent reshaping of the *Thermotogae* taxonomy has led to the description of a new genus, *Pseudothermotoga,* a sister group of the genus *Thermotoga* within the order *Thermotogales*. Comparative genomics of both *Pseudothermotoga* and *Thermotoga* spp., including 16S-rRNA-based phylogenetic, pan-genomic analysis as well as signature indel conservation, provided evidence that *Thermotoga caldifontis* and *Thermotoga profunda* species should be reclassified within the genus *Pseudothermotoga* and renamed as *Pseudothermotoga caldifontis* comb. nov. (type strain = AZM44c09^T) and *Pseudothermotoga profunda* comb. nov. (type strain = AZM34c06^T), respectively. In addition, based upon whole-genome relatedness indices and DNA–DNA Hybridization results, the reclassification of *Pseudothermotoga lettingae* and *Pseudothermotoga subterranea* as latter heterotypic synonyms of *Pseudothermotoga elfii* is proposed. Finally, potential genetic elements resulting from the distinct evolutionary story of the *Thermotoga* and *Pseudothermotoga* clades are discussed.

Introduction

Thermotoga maritima [19] embodies the original member of the phylum *Thermotogae,* which currently harbors at least 65 cultivated species distributed within 13 genera [23]. Although their true phylogenetic position remains a matter of debate [6,13], the *Thermotogales,* which are closely related to *Aquificales* [36], are commonly depicted as slowly evolving organisms defining an early branching lineage within the *Bacteria* [1]. The phylogenetic relatedness among the *Thermotogales* species do not reflect however the genetic variability of their genomes which undergo pervasive horizontal gene transfer (HGT), including gene influx from *Firmicutes* and *Archaea* [32,34,58].

The phylum *Thermotogae* was originally described from singular characteristics carried by its representatives which thrive most often at high-temperature and exhibit a single-unit lipid membrane as well as a unique loose-fitting sheath structure (called "toga") surrounding the cell [40]. Until recently, the phylum *Thermotogae* was believed to mainly comprise anaerobic, thermophilic ($T_{opt} \geq 70\,^{\circ}C$) and hyperthermophilic ($T_{opt} \geq 80\,^{\circ}C$), fermentative rod-shaped bacteria with the highest growth temperature (65–90 °C) under near-neutral pH conditions [14]. This view was dispelled with the discovery within the order *Thermotogales* of mesophilic [18,33] and moderately thermophilic acidophilic lineages comprising the genus *Mesotoga* and genera *Mesoaciditoga* and *Athalassotoga* [21,45], respectively.

The phylum *Thermotogae* initially consisted within a single class *Thermotogae*; a single order, *Thermotogales*; and a single family, *Thermotogaceae* [44]. Recently however, Bhandari and Gupta have revisited the taxonomy of the *Thermotogae* by identifying molecular signatures from whole genome-based comparisons [5]. The new taxonomic proposal resulted from the identification of Con-

* Corresponding author at: Aix Marseille Univ, Université de Toulon, CNRS, IRD, Mediterranean Institute of Oceanography (MIO), 163 Avenue Luminy, 13288, Marseille, France.
E-mail address: fabrice.armougom@mio.osuphytheas.fr (F. Armougom).

served Signature Indels (CSIs) at several taxonomic levels that was fairly congruent with the inference of 16S rRNA-based phylogenetic relatedness of the *Thermotogae* species [5]. According to their findings, the phylum *Thermotogae* was redefined and divided into four separate orders (*Thermotogales*, *Kosmotogales*, *Petrotogales* and *Mesoaciditogales*) and five families (*Thermotogaceae*, *Fervidobacteriaceae*, *Kosmotogaceae*, *Petrotogaceae* and *Mesoaciditogaceae*). At the genus level, the original genus *Thermotoga* was proposed to be split into two genera *Thermotoga* and *Pseudothermotoga*. This conclusion was supported by 16S rRNA-based phylogenetic reconstruction and identification of clade-specific CSIs that indicated two distinct evolutionary histories of cultivated members of the original genus *Thermotoga* [5]. To date, the genus *Pseudothermotoga* contains *P. elfii*, *P. lettingae*, *P. subterranea, P. hypogea* and *P. thermarum* [35]. Bhandari and Gupta's original work [5] echoes the modern taxonomy practice that refers to the use of the complete genome DNA sequence as the reference standard to determine both phylogeny and taxonomy of species [48]. With the rapid advances in affordable full genome sequencing, the current view of taxonomy thereby integrates whole genome-based comparisons [42,53] with an inflow of Overall Genome Relatedness indices (OGRI) [8] including digital DNA–DNA Hybridization (DDH) [30], Average Nucleotide Identity (ANI) [15,26], or tetra nucleotide regression [52]. The ANI measure is currently the most accepted OGRI by the scientist community [25].

Recently, two novel thermophilic anaerobic species were isolated from terrestrial hot springs in Japan and were classified as belonging to the genus *Thermotoga* as *T. caldifontis* and *T. profunda* [31]. Using multiphasic molecular approaches based on 16S rRNA, core-genome phylogeny, and identification of the *Pseudothermotoga*-specific CSIs, we provide evidence for reclassifying *T. profunda* and *T. caldifontis* within the genus *Pseudothermotoga* as *Pseudothermotoga profunda* comb. nov., and *Pseudothermotoga caldifontis* comb. nov., respectively. We also demonstrate by experimental DDH and OGRI measures that *Pseudothermotoga lettingae* and *Pseudothermotoga subterranea* should be reclassified as latter heterotypic synonyms of *Pseudothermotoga elfii*.

Methods

Genome feature analysis

Thermotoga and *Pseudothermotoga* genomes were retrieved from the NCBI ftp site (ftp://ftp.ncbi.nlm.nih.gov/) under the accession numbers listed in Table 1. The coding sequences (CDS) information was extracted from the gbk and GFF files. The GC content was estimated by the geecee function of the EMBOSS package 6.6.0.0 [46]. The hypothetical protein fractions were deduced from the protein annotation files (faa). The *T. caldifontis* and *T. profunda* genomes were included in the genus *Pseudothermotoga* for the genome feature analysis. Mean comparison of *Thermotogaceae* clades was performed by the non-parametric Mann–Whitney–Wilcoxon test. The Mann–Whitney–Wilcoxon test can be applied on small sample size and does not require any assumption about the variance equality and population distributions. The Mann–Whitney–Wilcoxon tests, genome size, gene and GC contents were performed by GraphPad prism Software version 5.0. M, SD indicate mean and standard deviation, respectively.

SSU rRNA-based phylogeny

The near-full length 16S rDNA sequences (>1300 pb) of the *Thermotogae* species were retrieved from the NCBI database (https://www.ncbi.nlm.nih.gov/) and aligned with the MAFFT software v7.123b [24]. A maximum likelihood (ML) phylogenetic tree was constructed from 1624 nucleotide positions using the PhyML 3.1 algorithm [17] implemented in the Seaview package v4.6 [16] with GTR + I + G substitution model and 1000 bootstraps replicates. The phylogenetic tree was rooted using the *Coprothermobacter proteolyticus* IT3 (GU363592) and *Coprothermobacter platensis* (Y08935.1) 16S rDNA sequences since the genus *Coprothermobacter* was the closest branching out group in update SSU rRNA phylogenetic tree [57].

Core-genome-based phylogeny

The genome sequences of cultivated *Thermotogaceae* were retrieved from the NCBI accession numbers summarized in Table 1. A core/pan-genome analysis was performed using the Roary pangenome pipeline v3.6.1 [37]. The Roary pipeline generated a core gene alignment from PRANK v.140110 [29] with a 50% identity cut-off level, as well as a gene presence/absence matrix. A maximum-likelyhood (ML) core-genome tree was inferred using PhyML 3.1 algorithm implemented in Seaview v4.6 with generalized time-reversible (GTR) and gamma distribution. PhyML approximated likelihood ratio (aLRT) SH-like values were indicated in a consensus tree. The roary plots.py scripts were used to combine the core-genome-based tree and the gene presence/absence matrix.

Molecular signatures within the genus Pseudothermotoga

The CSIs identified as unique to the genus *Pseudothermotoga* were originated from reference proteins including the Glycerol-3-phosphate dehydrogenase, the hypothetical protein Theth_0845, the hypothetical protein TM1140 and the hypothetical protein

Table 1
Genomic and growth features within the *Thermotogaceae*.

Strains	Genome size (bp)	GC content (%)	Predicted CDS	CDS without predicted function	Optimal growth temperature	Accession number
P. elfii DSM 9442^T	2,169,860	39	2063	497 (24%)	66	AP014507
P. lettingae TMO^T DSM 14385	2,135,342	39	2047	414 (20%)	65	CP000812
T. profunda AZM34c06^T	2,187,612	39	2073	600 (29%)	60	AP014510
P. hypogea DSM 11164^T	2,165,416	49	2073	663 (32%)	70	CP007141
T. caldifontis AZM44c09^T	2,014,912	51	1940	564 (29%)	70	AP014509
P. thermarum DSM 5069^T	2,039,943	40	1965	428 (21%)	70	CP002351
T. naphthophila RKU-10^T	1,809,823	46	1805	333 (19%)	80	CP001839
T. petrophila KU-1^T	1,823,511	46	1772	335 (19%)	80	CP000702
Thermotoga sp. Cell2	1,749,904	46	1632	321 (18%)	–	CP003409
Thermotoga sp. RQ2	1,877,693	46	1832	337 (18%)	80	CP000969
Thermotoga sp. RQ7	1,851,618	47	1812	347 (19%)	80	CP007633
T. neapolitana DSM4359^T	1,884,562	47	1829	354 (19%)	80	CP000916
Thermotoga sp. 2812B	1,843,731	46.2	1808	361 (20%)	–	CP003408
T. maritima MSB8^T	1,869,612	46	1858	369 (20%)	80	AE000512

Tlet_0907 [5]. The reference protein sequences were retrieved from their corresponding genomes, *P. lettingae* TMO^T (Tlet_907), *T. maritima* MSB8^T (TM1140), *P. thermarum* LA3^T (Therth_0845) and *P. lettingae* TMO^T (Glycerol-3-phosphate dehydrogenase). The identification of the homologous proteins in the other *Thermotogaceae* genomes was performed by BLASTp similarity search [2] using each of the four reference proteins as seed. A protein set was built with each reference protein and their homologs within *Thermotogaceae* spp. The multiple sequence alignments of the protein sets were performed by MAFFT v.7123b [24]. The multiple sequence alignment was visualized by Jalview 2.9.0b2 [54] using ClustalX amino acid colors. Only the alignment part containing the CSI region was displayed. The CSI positions were reported according to the reference sequence in each CSI alignment.

DNA–DNA Hybridization dissimilarity measures

DNA–DNA Hybridization (DDH) studies were performed using *P. lettingae* TMO^T (DSM 14385) [4] and *P. elfii* SEBR 6459^T (DSM 9442) [43]. The DDH studies were carried out by the DSMZ as described by De Ley et al. [27] with some modifications [11,20] and using a model 2600 spectrophotometer equipped with a 2527-R thermoprogrammer and plotter (Gilford Instruments). Renaturation rates were computed with the TRANSFER.BAS program [22].

Whole-genome relatedness indices

The Average Nucleotide Identity (ANI) and Average Amino Acid Identity (AAI) indices were deduced from pairwise genome [26] and conserved coding proteins [47] comparisons, respectively, using the available full *Thermotogaceae* genomes. The ANIb measure was preferred to ANIm since it was more appropriate for the comparison of distant related genomes and for *Thermotoga* spp. [28]. The ANIb relied on the sequence alignment tool MUMer [9] and was implemented in the JSpecies 1.2.1 software package (www.imedea.uib.es/jspecies). The ANIb range values (95–96%) correspond to the threshold for species delineation (32). The AAI calculation was performed by CompareM v0.0.21 software (https://github.com/dparks1134/CompareM), which uses DIAMOND 0.8.14 [7] for best reciprocal hits. Homology was defined with sequence similarity above 50%, sequence coverage above 75% and a minimum E-value of 1e-05. The resulting ANIb and AAI values were transformed in distance matrix and the hierarchical clustering was performed by hclust (R package) using a complete linkage method. The heatmap.2 function (R ggplots package) was used to show the ANI/AAI heatmap with a color scheme relied on ANI values.

Results and discussion

Genomic characteristics within the Thermotogaceae

Previous work revealed that the former genus *Thermotoga* consisted of two distinct groups of species according to their optimal growth temperature (T_{opt}) higher than 77 °C and lower than 70 °C for *Thermotoga* spp. and *Pseudothermotoga* spp., respectively [14]. With a T_{opt} comprised between 60 °C and 70 °C, *Thermotoga caldifontis* and *Thermotoga profunda* shared growth temperature characteristics of *Pseudothermotoga* spp. In addition to this phenotypic feature, *T. caldifontis* and *T. profunda* carried genomic attributes more similar to the *Pseudothermotoga* than to the *Thermotoga* representatives (Table 1, Fig. 1). The mean genome size reached 2.12 Mb (SD = 0.073) and 1.84 Mb (SD = 0.044) within the genera *Pseudothermotoga* and *Thermotoga*, respectively (Fig. 1A). The genome size of *Pseudothermotoga* spp. was therefore significantly larger (Mann–Whitney–Wilcoxon test, p = 0.0007) and more variable than the *Thermotoga* genomes. Moreover, the genome size and growth temperature were negatively associated (r = −0.95) within *Thermotogaceae* (Table 1). Recently, the "streamlining" concept has been proposed, i.e. that thermal adaptation involves reduction in the genome size [50]. As expected, the mean gene content of *Pseudothermotoga* species compared to *Thermotoga* species consisted of more coding sequences (Mann–Whitney–Wilcoxon

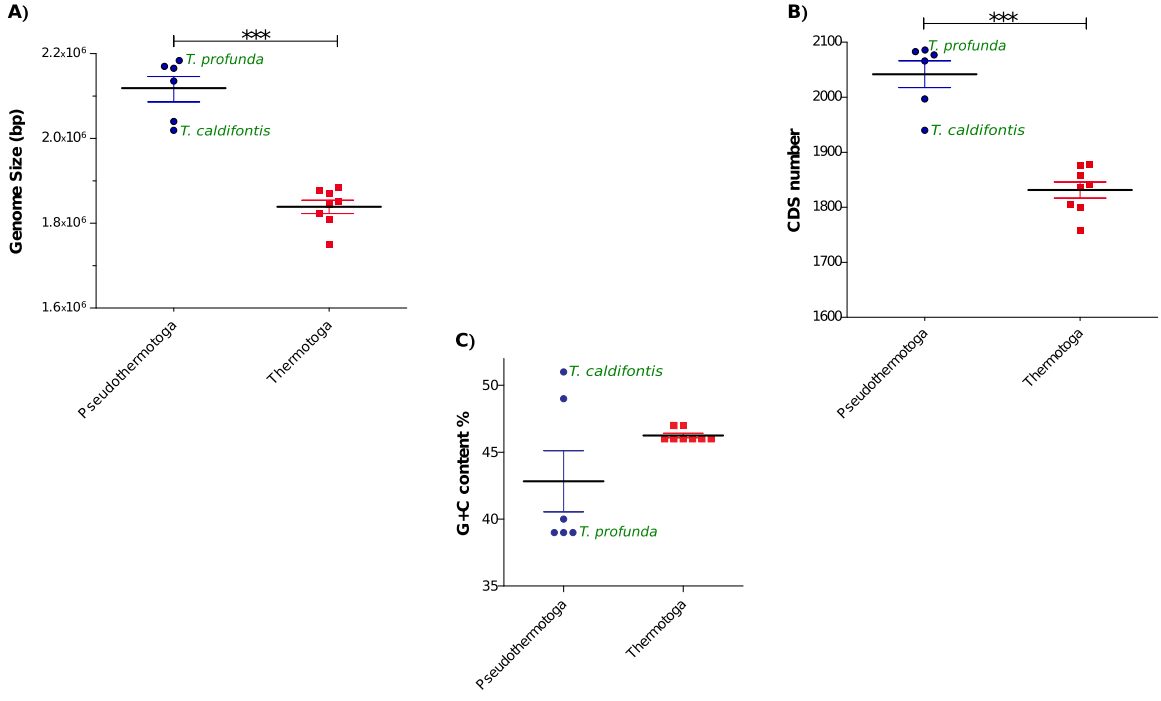

Fig. 1. Genome features within *Thermotogaceae* family.
(A) Genome size variability. (B) CDS number distribution. (C) GC content distribution. Black line indicates mean ± SD. *T. caldifontis* and *T. profunda* species are included within the genus *Pseudothermotoga*.

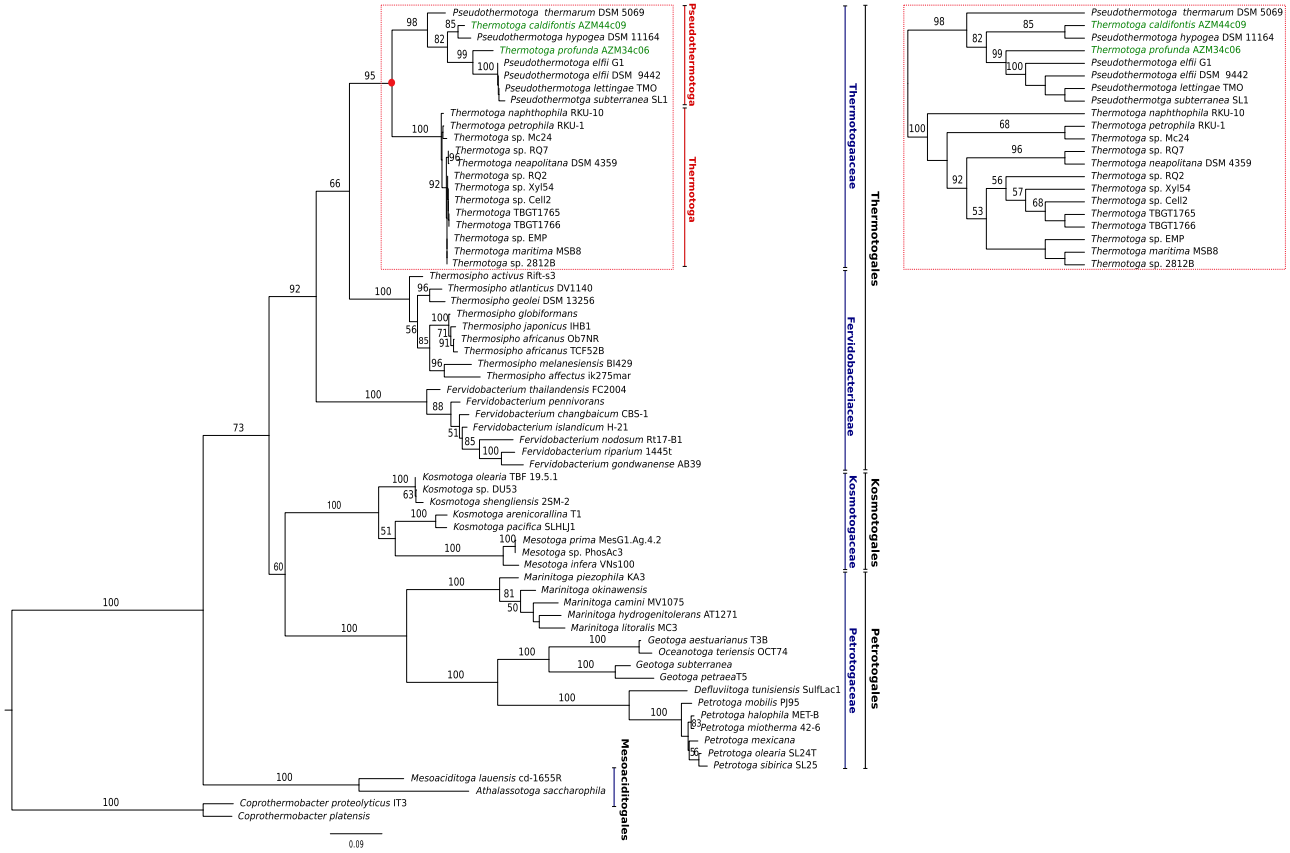

Fig. 2. 16S rRNA-based phylogenetic tree of the *Thermotogae*.
ML tree built from 65 16S rRNA sequences and 1624 positions. Bootstrap values from 1000 replicates are indicated only when >50%. The tree is rooted with 16S of *Coprothermobacter proteolyticus* IT3 and *Coprothermobacter platensis*. Bar scale, 9% nucleotide substitutions. Order and family taxonomic levels are indicated in black and blue, respectively. The *Pseudothermotoga* and *Thermotoga* genera are boxed in red; the red circle indicates their exclusive common ancestor. Details of the branch support values of phylogenetic relatedness within *Pseudothermotoga* and *Thermotoga* is shown by a cladogram on the top right.

test, p = 0.0007) (Fig. 1B), without affecting genome compactness (similar gene density, data not shown). The hypothetical protein fraction deduced from the CDS functional annotation was larger within the *Pseudothermotoga* spp. than within the *Thermotoga* spp. (Table 1). Compared to the GC content of the *Pseudothermotoga* spp. (M = 42.83, SD = 5.601), those of *Thermotoga* spp. were slightly higher (not significant, p = 0.3308), and more similar to each other (M = 46.25, SD = 0.463). *T. caldifontis* and *P. hypogea* species had the highest GC content (Table 1, Fig. 1C), were closely related but slightly differed from the rest of the *Pseudothermotoga* spp. because of their lower genome size and a higher GC content.

The phylogenetic relatedness of Thermotogae

The evolutionary relationships of 65 species affiliated to the phylum *Thermotogae* were established by 16S rRNA-based phylogenetic inference (Fig. 2). The recent description of thermoacidophilic *Mesoaciditogales* comprising only *Mesoaciditoga lauensis* and *Athalassotoga saccharophila* formed the deepest diverging lineage within this phylum [21,45]. The next well-supported branch separated both *Petrotogales* and *Kosmotogales* orders from the *Thermotogales*. The *Petrotogales* clade consists of *Marinitoga* spp. as well as the *Geotoga* and *Petrotoga* sister groups. Within the order *Kosmotogales,* the thermophilic *Kosmotoga* lineage notably harbors the *K. olearia* and the recent *K. pacifica* species, which demonstrate an outstandingly wide growth temperature range, 20–79 °C and 33–78 °C, respectively [10,23]. The closest relative of *Kosmotoga* was the mesophilic *Mesotoga* lineage, with lower growth temperature range (30–40 °C). However, the low bootstrap support (60, Fig. 2) of the *Kosmotogales* and *Petrotogales* dichotomy brought into question their true phylogenetic placement [41,45] despite a recent core-gene-based phylogeny with 17 *Thermotogae* species [23]. Finally, *Thermotogales* appeared to be the most diverse lineage, encompassing the genera *Thermotoga*, *Pseudothermotoga*, *Thermosipho, Oceanotoga* and *Fervidobacterium*. Within the order *Thermotogales*, the *Thermotoga* and *Pseudothermotoga* clades shared an exclusive common ancestor (red circle, Fig. 2); they were sister groups and were therefore more closely related to one another than to *Thermosipho* spp. From their most recent ancestor, more evolutionary change was observed within the *Pseudothermotoga* than within the *Thermotoga* lineage. However, the weakness of the phylogenetic signal from the 16S rRNA-based phylogeny did not offer an acceptable means to determine the relationships between *Thermotoga* species (Fig. 2). The poorly resolved phylogeny of the *Thermotoga* representatives indicates a need to investigate additional phylogenetic markers (RpoB), core-genes phylogeny [23] or concatenated housekeeping proteins [5]. Additionally, a full comparative genomics study of the gene flow events undergone by the members of the two genera would shed light on their evolutionary history. This first phylogenetic approach suggested that *Thermotoga profunda* and *Thermotoga caldifontis* belong to the *Pseudothermotoga* clade rather than to *Thermotoga* (Fig. 2).

The core-genome-based phylogeny of Thermotogaceae

To confirm the taxonomic affiliation of *Thermotoga caldifontis* and *Thermotoga profunda* within the genus *Pseudothermotoga*, we investigated large-scale features of their genomes. The inference of

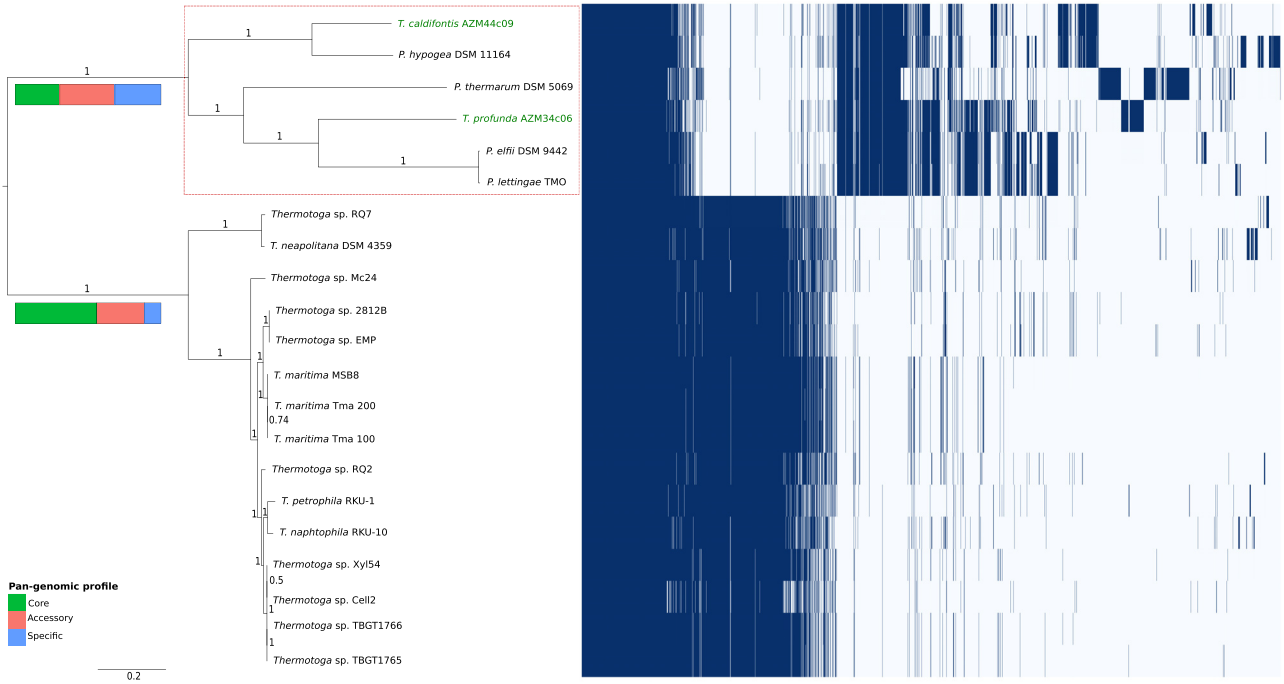

Fig. 3. Core-genome-based phylogeny of the *Thermotogaceae* family.
The core-genes consensus midpoint tree (left) is deduced from 603,752 positions of 614 shared genes. PhyML branch support values are indicated in black respectively. The gene presence/absence matrix is deduced from 5052 total genes (right). Blue tile shows the presence of gene in a genome. Aligned blue tiles are genes shared by a least two genomes. *Pseudothermotoga* genus is boxed in red. The pan-genomic profile for the *Pseudothermotoga* and *Thermotoga* clades are shown as bar charts.

core-genome-based phylogeny of the family *Thermotogaceae* relied on the analysis of the gene content of eight *Thermotoga* and six *Pseudothermotoga* species. The identification of the genes shared by all the available *Thermotogaceae* genomes (the core-genome) was deduced by gene orthology and resulted in 614 genes used for the *Thermotogaceae* phylogeny reconstruction. The core-genome-based phylogeny preserved *P. hypogea* as the closest relative of *T. caldifontis*, and the sister taxa *P. elfii* as the closest relatives of *T. profunda* (Fig. 3), as was observed in the 16S rRNA-based phylogeny (Fig. 2). The lack of consistency in the tree topology of *Thermotoga* spp. between 16S-rRNA-based and core-genome-based phylogeny can be explained by the low bootstrap branch supports for this clade (Fig. 2). In Fig. 3, primary genetic indicators of the dichotomy of genera *Pseudothermotoga* and *Thermotoga* were denoted with a gene presence (blue tile)/absence matrix, where the genes shared by two or more species are shown by aligned blue tiles or blocks. Similar gene content profiles of *Thermotoga* species resulted in a considerable core-genome size (number of genes shared by all the members) to the detriment of the accessory genome (shared by at least two members but not by all) and species-specific genes (Fig. 3). In contrast, the *Pseudothermotoga* species showed more variability in their gene content profiles, which was exhibited by a lower core genome size but a greater accessory genome and more species-specific genes (Fig. 3). The gene content showed that *T. profunda* and *T. caldifontis* were genetically more closely related to the *Pseudothermotoga* than to *Thermotoga* species. As discussed above, deciphering the distinct evolutionary history of the *Thermotoga* and *Pseudothermotoga* lineages will require more complex compara-

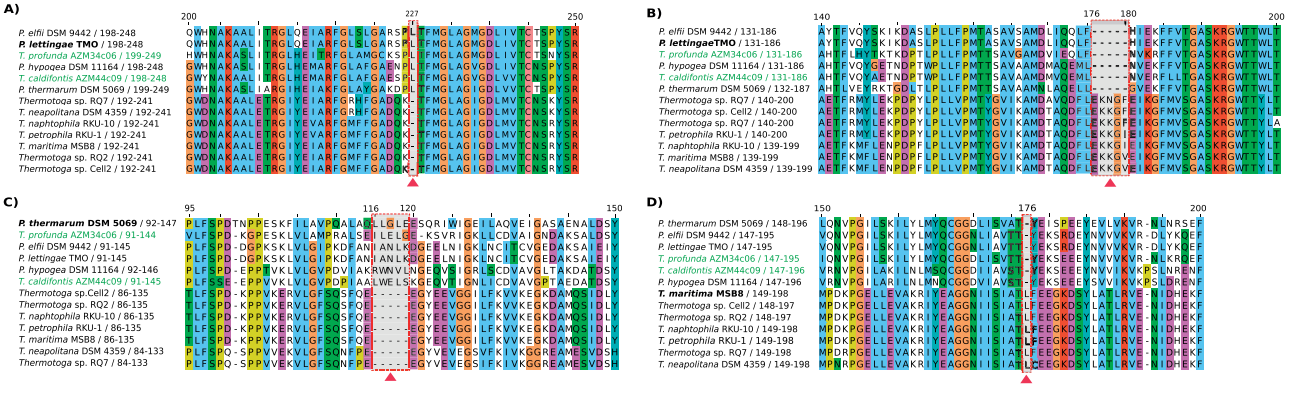

Fig. 4. Partial sequence alignments of the *Pseudothermotoga*-specific Conserved Signature Indels.
Partial alignments extracted from full protein sequence alignments. The positions added to sequence labels mean start/end positions according to the full aligned sequences. Positions on the top of the CSIs boxed in dashed red lines target the reference sequence (labeled in bold). *T. caldifontis* and *T. profunda* hold the four *Pseudothermotoga*-specific CSIs. (**A**) Glycerol-3-phosphate dehydrogenase. A Leucine insertion conserved in all *Pseudothermotoga* members. (**B**) Hypothetical protein Tlet_0907. *Pseudothermotoga* members show a specific five-amino acid deletion. (**C**) Hypothetical protein Theth_0845. A five-amino-acid insertion is conserved within the *Pseudothermotoga* clade. (**D**) Hypothetical protein TM1140. A Leucine deletion conserved within the *Pseudothermotoga* clade.

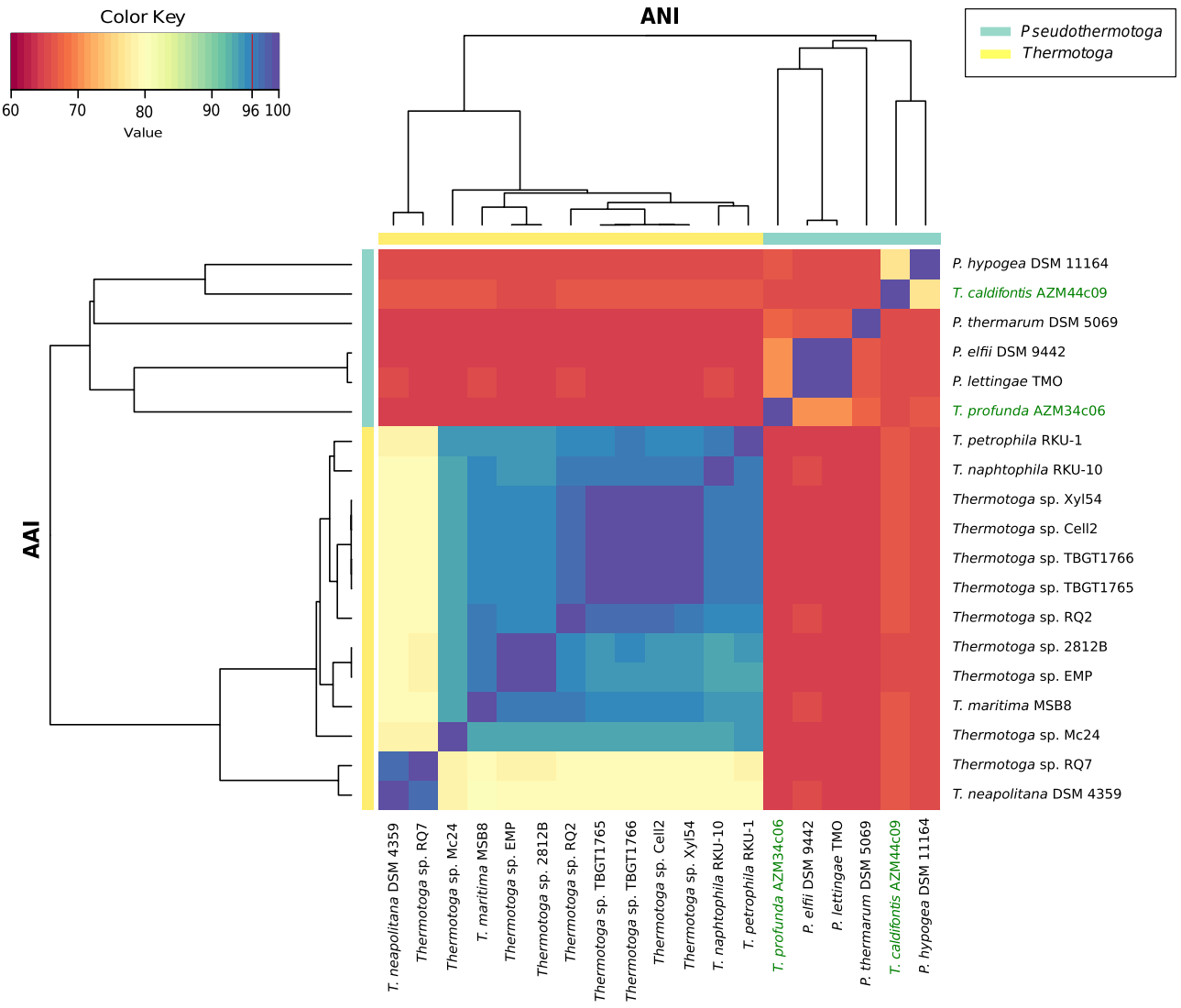

Fig. 5. Whole-genome relatedness within the *Thermotogaceae* family.
Hierarchical clustering from Average Nucleotide Identity (vertically) and Average Amino Acid Identity (horizontally) measures using 19 *Thermotogaceae* genomes. Value scale is depicted in red, yellow and green colors for ANI genome pairwise comparison. The value scale ranges from 0 to 100% according to sequence identity level. The vertical line at the value 96 indicates the ANI species demarcation threshold.

tive genomics including gene family reconstruction, identification of species-specific genes, functional systems, and horizontal gene transfer, a dominant force of *Thermotogales* evolution [34].

Molecular signatures of the Pseudothermotoga *clade*

Previous genomic insights on the *Thermotogae* have brought out the discovery of four Conserved Signature Indels (CSIs) specific to the members of the *Pseudothermotoga* clade [5]. The *Pseudothermotoga*-specific CSIs have been identified from *Thermotogales* protein family alignments of the Glycerol-3-phosphate dehydrogenase (one amino-acid insertion), the hypothetical Theth_0845 (5/6 amino-acid insertion), the hypothetical TM1140 (one amino-acid deletion) and the hypothetical Tlet_0907 (five amino-acid deletion) [5]. Presence of the four *Pseudothermotoga*-specific CSIs within *Thermotoga profunda* and *Thermotoga caldifontis* offers additional evidence of their reclassification proposal as *Pseudothermotoga* species. The multiple sequence alignment (MSA) of the homologs to the Glycerol-3-phosphate dehydrogenase (Fig. 4A) and to the hypothetical protein TM1140 (Fig. 4D) showed that the two *Pseudothermotoga*-specific CSIs, a Leucine insertion and a Leucine deletion, respectively, were also conserved in *T. caldifontis* and *T. profunda* (Fig. 4A, D). The next *Pseudothermotoga*-specific CSI, a five-amino-acid deletion derived from the MSA of homologs to the Tlet_0907 protein, was also shared by both *T. caldifontis* and *T. profunda* (Fig. 4B). However, Bhandari and Gupta reported a six-amino-acid deletion in their MSA of homologs to Tlet_907 [5]. In the same way, the MSA of homologs to the Theth_0845 protein exhibited also a slight modified pattern from the original *Pseudothermotoga*-specific CSI described in supplementary Fig. S39 of Bhandari and Gupta work [5] (Fig. 4C).

Our analysis led us to identify two slight modified *Pseudothermotoga*-specific CSIs (Fig. 4B,C) compared to the original work [5], that could be explained by the MSA methods used and by the divergence of the selected sequences in the MSA. Using the T-COFFEE MSA method [3] (data not shown), the CSIs identified were congruent with our MAFFT-based analysis. The accuracy evaluation of the MSA methods (benchmark) commonly shows that consistency-based programs such as T-COFFEE and MAFFT exceed the other methods [38]. It is also important to emphasize that the previous identification of *Pseudothermotoga*-specific CSIs [5] involved only two *Pseudothermotoga* spp. (*P. lettingae* and *P.*

thermarum) whereas all *Pseudothermotoga* spp. were included in our analysis (Fig. 4). Finally, the pairwise identity matrix of the Theth_0845 protein family alignment (Fig. 4C) provided low sequence identity values (<35%) close to the twilight zone [49] that could stimulate debate about the true evolutionary relatedness of these proteins and especially the biological relevance of this CSI. The occurrence of the four *Pseudothermotoga*-specific CSIs in *T. caldifontis* and *T. profunda* is thereby an additional argument to reclassify these two organisms within the genus *Pseudothermotoga* as *P. caldifontis* comb. nov. and *P. profunda* comb. nov., respectively.

Species delineation within the genus Pseudothermotoga

Fardeau et al. [12], previously reported that DNA–DNA homologies between *P. elfii* strain G1 and *P. subterranea* DSM 9912^T, *P. lettingae* DSM 14385^T and *P. elfii* DSM 9442^T were 83%, 71.4% and 86%, respectively. This study also indicated a DNA–DNA homology of 83% between *P. subterranea* DSM 9912^T and *P. elfii* DSM 9442^T, and of 78% between *P. subterranea* DSM 9912^T and *P. lettingae* DSM 14385^T [12]. In our study, complementary DDH experiments were carried out in duplicate with *P. lettingae* DSM 14385^T and *P. elfii* DSM 9442^T, they indicated DNA–DNA homology of 94% (91.5% and 96.7%) between these two species. All DDH values of the pairwise species comparison were thereby above 70%. Consequently, all the *Pseudothermotoga* strains could not be differentiated at the species level by DDH when taking into account the recommendation of the *ad hoc* committee [55], which advised a threshold value of 70% DDH similarity for the definition of bacterial species.

The close relationship linking DDH and ANI values [15,25] led us to calculate ANI indices for estimating overall genome similarity and clarifying the genetic relationships within the genus *Pseudothermotoga*. In addition to ANI indices, the Amino Acid Identity (AAI) measure was also considered for its robustness for divergent species [47]. Within the *Thermotogaceae* genomes, the estimated ANI and AAI values ranged from 64.19 to 99.96% and from 63.32 to 99.98%, respectively. The double ANI–AAI clustering distinguished the *Pseudothermotoga* and *Thermotoga* genera (Fig. 5). The *Thermotoga* clustering showed intra-species relationships for *Thermotoga* sp. Xyl 54, *Thermotoga*. sp Cell2, *Thermotoga*. sp. TBGT1765 and *Thermotoga*. sp TBGT1766 (AAI and ANI values >98). Strikingly, both ANI and AAI values revealed intra-species relationships between *P. elfii* and *P. lettingae* with 99.12% and 99.5% of identity level, respectively. Based upon DDH and ANI–AAI results, *P. elfii*, *P. lettingae* and *P. subterranea* species should be considered as heterotypic synonyms. Herein we propose to unite these three species. According to rules 38, 42 and 24b(2) of the Bacteriological Code [39], the name *Pseudothermotoga elfii* has priority and hence should be used for the unified taxon.

Conclusions

With the significant advances in full genome sequencing, modern taxonomy practice promotes the use of gene marker-based/core-genome-based phylogenetic inferences, comparative genomics, and whole genome relatedness indices as reliable methods for taxonomic classification [8,42,51,53]. As a result, investigation of the *Thermotoga* and *Pseudothermotoga* genera by modern taxonomy practice and classical DDH measures puts forward congruent classification and nomenclature of these genera within the *Thermotogaceae* family. Our data show that *Thermotoga caldifontis* and *Thermotoga profunda* species share similar features (genome size, CDS number) with the *Pseudotermotoga* species. They are also phylogenetically related to *Pseudothermotoga* species, and carry the four *Pseudothermotoga*-specific CSIs. Moreover, a preliminary pan-genomic analysis shows that *T. profunda* and *T. caldifontis* are genetically more closely related to the *Pseudothermotoga* spp. We thereby propose to reclassify *Thermotoga caldifontis* and *Thermotoga profunda* as *Pseudothermotoga caldifontis* comb. nov. and *Pseudothermotoga profunda* comb. nov., respectively. We also demonstrate that *Pseudothermotoga elfii, P. subterranea* and *P. lettingae* are heterotypic synonyms and should therefore be united with the priority name of *Pseudothermotoga elfii*. Despite the considerable effort needed to understand the *Thermotogae* phylogenetic relationships [13,58], the evolutionary events driving *Thermotogae* sub-groups remain blurred. Further complex comparative genomics may assist the deciphering of the evolutionary mechanisms within *Thermotogae*.

Description of Pseudothermotoga profunda *comb. nov.*
Pseudothermotoga profunda *(pro.fun'da L. fem. adj.* profunda *deep, pertaining to the source of isolation of the type strain)*

Basonym: *Thermotoga profunda* Mori et al. 2014.

The description of this taxon is the same as that given previously [31].

The type strain is AZM34c06^T = NBRC 106115^T = DSM 23275^T.

The sequence accession number (16S rRNA gene) for the type strain is NR_133904.1.

The accession number of the genome sequence is AP014510.

Description of Pseudothermotoga caldifontis *comb. nov.*
Pseudothermotoga caldifontis *(cal.di.fon'tis. L. adj.* caldus *hot; L. masc. n.* fons, fontis *a spring, N.L. gen. n.* caldifontis *of a hot spring, pertaining to the source of isolation of the type strain)*

Basonym: *Thermotoga caldifontis* Mori et al. 2014.

The description of this taxon is the same as that given previously [31].

The type strain is AZM44c09^T = NBRC 106116^T = DSM 23272^T.

The sequence accession number (16S rRNA gene) for the type strain is NR_133903.1.

The accession number of the genome sequence is AP014509.

Emended description of the genus Pseudothermotoga *Bhandari and Gupta 2014*

The morphological and physiological description of this genus, whose type species is *Pseudothermotoga thermarum* [5], remains the same as described by Windberger et al. [56]. Other species belonging this genus are *P. profunda* (Basonym: *T. profunda*) and *P. caldifontis* (Basonym: *Thermotoga caldifontis*) (Mori et al. [31]).

Emended description of the species Pseudothermotoga elfii *Bhandari and Gupta 2014*

The characteristics of this species are as described by Ravot et al. [43] with the following amendments. The following substrates may be used: methanol, lactate, pyruvate, galactose, cellobiose, glycerol, starch, xylan, pectin, methylamine, dimethylamine, trimethylamine, 2-oxoglutarate, serine, yeast extract, peptone, gelatin and casamino acids. Acetate, betaine, leucine, isoleucine, valine, formate and H_2/CO_2 (80/20, v/v) may be oxidized either in the presence of thiosulfate or a hydrogenotrophic methanogen as electron acceptor. Elemental sulfur, Fe(III) and anthraquinone-2,6-disulfonate may serve as electron acceptors. The type strain is DSM 9442^T = ATCC 51869T.

Acknowledgement

We thank the Professor William B. Whitman, University of Georgia, for the revision of the manuscript.

Funding

This project has been funded with support from ExxonMobil Research and Engineering.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.syapm.2018.04.007.

References

[1] Achenbach-Richter, L., Gupta, R., Stetter, K.O., Woese, C.R. (1987) Were the original eubacteria thermophiles? Syst. Appl. Microbiol. 9, 34–39.
[2] Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic local alignment search tool. J. Mol. Biol. 215, 403–410.
[3] Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., Keduas, V., Notredame, C. (2006) Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. Nucleic Acids Res. 34, 600–603.
[4] Balk, M., Weijma, J., Stams, A.J.M. (2002) *Thermotoga lettingae* sp. nov., a novel thermophilic, methanol-degrading bacterium isolated from a thermophilic anaerobic reactor. Int. J. Syst. Evol. Microbiol. 52, 1361–1368.
[5] Bhandari, V., Gupta, R.S. (2014) Molecular signatures for the phylum (class) *Thermotogae* and a proposal for its division into three orders (*Thermotogales*, *Kosmotogales* ord. nov. and *Petrotogales* ord. nov.) containing four families (*Thermotogaceae*, *Fervidobacteriaceae* fam. nov., *Kosmotogaceae* fam. nov. and *Petrotogaceae* fam. nov.) and a new genus *Pseudothermotoga* gen. nov. with five new combinations. Antonie van Leeuwenhoek Int. J. Gen. Mol. Microbiol. 105, 143–168.
[6] Brochier, C., Philippe, H. (2002) Phylogeny: a non-hyperthermophilic ancestor for *Bacteria*. Nature 417, 244.
[7] Buchfink, B., Xie, C., Huson, D.H. (2015) Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60.
[8] Chun, J., Rainey, F.A. (2014) Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. Int. J. Syst. Evol. Microbiol. 64, 316–324.
[9] Delcher, A.L. (2002) Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Res. 30, 2478–2483.
[10] DiPippo, J.L., Nesbø, C.L., Dahle, H., Doolittle, W.F., Birkland, N.K., Noll, K.M. (2009) *Kosmotoga olearia* gen. nov., sp. nov., a thermophilic, anaerobic heterotroph isolated from an oil production fluid. Int. J. Syst. Evol. Microbiol. 59, 2991–3000.
[11] Escara, J.F., Hutton, J.R. (1980) Thermal stability and renaturation of DNA in dimethyl sulfoxide solutions: acceleration of the renaturation rate. Biopolymers 19, 1315–1327.
[12] Fardeau, M.-L., Goulhen, F., Bruschi, M., Khelifi, N., Cayol, J.-L., Ignatiadis, I., Guyot, F., Ollivier, B. (2009) *Archaeoglobus fulgidus* and *Thermotoga elfii*, Thermophilic isolates from deep geothermal water of the Paris Basin. Geomicrobiol. J. 26, 119–130.
[13] Forterre, P. (2015) The universal tree of life: an update. Front. Microbiol. 6, 1–18.
[14] Frock, A.D., Notey, J.S., Kelly, R.M. (2010) The genus *Thermotoga*: recent developments. Environ. Technol. 31, 1169–1181.
[15] Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., Tiedje, J.M. (2007) DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int. J. Syst. Evol. Microbiol. 57, 81–91.
[16] Gouy, M., Guindon, S., Gascuel, O. (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. 27, 221–224.
[17] Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.
[18] Ben Hania, W., Ghodbane, R., Postec, A., Brochier-Armanet, C., Hamdi, M., Fardeau, M.L., Ollivier, B. (2011) Cultivation of the first mesophilic representative ("mesotoga") within the order *Thermotogales*. Syst. Appl. Microbiol. 34, 581–585.
[19] Huber, R., Langworthy, T.A., König, H., Thomm, M., Woese, C.R., Sleytr, U.B., Stetter, K.O. (1986) *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90 °C. Arch. Microbiol. 144, 324–333.
[20] Huss, V.A.R., Festl, H., Schleifer, K.H. (1983) Studies on the spectrophotometric determination of DNA hybridization from renaturation rates. Syst. Appl. Microbiol. 4, 184–192.
[21] Itoh, T., Onishi, M., Kato, S., Iino, T., Sakamoto, M., Kudo, T., Takashina, T., Ohkuma, M. (2016) *Athalassotoga saccharophila* gen. nov., sp. nov., isolated from an acidic terrestrial hot spring, and proposal of *Mesoaciditogales* ord. nov. and *Mesoaciditogaceae* fam. nov. in the phylum *Thermotogae*. Int. J. Syst. Evol. Microbiol. 66, 1045–1051.
[22] Jahnke, D. (1992) BASIC computer program for evaluation of GILFORD SYSTEM 2600 spectrophotometer on a PC/XT/AT type personal computer. System 15, 61–73.
[23] Jiang, L., L'Haridon, S., Jebbar, M., Xu, H., Alain, K., Shao, Z. (2017) Complete genome sequence and whole-genome phylogeny of *Kosmotoga pacifica* type strain SLHLJ1T from an East Pacific hydrothermal sediment. Stand. Genom. Sci. 12, 3.
[24] Katoh, K., Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.
[25] Kim, M., Oh, H.S., Park, S.C., Chun, J. (2014) Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int. J. Syst. Evol. Microbiol. 64, 346–351.
[26] Konstantinidis, K.T., Tiedje, J.M. (2005) Genomic insights that advance the species definition for prokaryotes. Proc. Natl. Acad. Sci. U. S. A. 102, 2567–2572.
[27] De Ley, J., Cattoir, H., Reynaerts, A. (1970) The quantitative measurement of DNA hybridization from renaturation rates. Eur. J. Biochem. 12, 133–142.
[28] Li, X., Huang, Y., Whitman, W.B. (2015) The relationship of the whole genome sequence identity to DNA hybridization varies between genera of prokaryotes. Antonie van Leeuwenhoek Int. J. Gen. Mol. Microbiol. 107, 241–249.
[29] Löytynoja, A. (2014) Phylogeny-aware alignment with PRANK. Methods Mol. Biol. (Clifton N.J.) 1079, 155–170.
[30] Meier-Kolthoff, J.P., Klenk, H.P., Göker, M. (2014) Taxonomic use of DNA G + C content and DNA-DNA hybridization in the genomic age. Int. J. Syst. Evol. Microbiol. 64, 352–356.
[31] Mori, K., Yamazoe, A., Hosoyama, A., Ohji, S., Fujita, N., Ishibashi, J.I., Kimura, H., Suzuki, K.I. (2014) *Thermotoga profunda* sp. nov. and *Thermotoga caldifontis* sp. nov., anaerobic thermophilic bacteria isolated from terrestrial hot springs. Int. J. Syst. Evol. Microbiol. 64, 2128–2136.
[32] Nesbø, C.L., Bapteste, E., Curtis, B., Dahle, H., Lopez, P., Macleod, D., Dlutek, M., Bowman, S., Zhaxybayeva, O., Birkeland, N.-K., Doolittle, W.F. (2009) The genome of *Thermosipho africanus* TCF52B: lateral genetic connections to the *Firmicutes* and *Archaea*. J. Bacteriol. 191, 1974–1978.
[33] Nesbø, C.L., Bradnan, D.M., Adebusuyi, A., Dlutek, M., Petrus, A.K., Foght, J., Doolittle, W.F., Noll, K.M. (2012) *Mesotoga prima* gen. nov., sp. nov., the first described mesophilic species of the *Thermotogales*. Extremophiles 16, 387–393.
[34] Nesbø, C.L., Swithers, K.S., Dahle, H., Haverkamp, T.H.A., Birkeland, N.-K., Sokolova, T., Kublanov, I., Zhaxybayeva, O. (2015) Evidence for extensive gene flow and *Thermotoga* subpopulations in subsurface and marine environments. ISME J. 9, 1532–1542.
[35] Oren, A., Garrity, G.M. (2014) List of new names and new combinations previously effectively, but not validly, published. List no. 158. Int. J. Syst. Evol. Microbiol. 64, 2184–2187.
[36] Oshima, K., Chiba, Y., Igarashi, Y., Arai, H., Ishii, M. (2012) Phylogenetic position of *Aquificales* based on the whole genome sequences of six *Aquificales* species. Int. J. Evol. Biol. 2012, 1–9.
[37] Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., Fookes, M., Falush, D., Keane, J.A., Parkhill, J. (2015) Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31, 3691–3693.
[38] Pais, F.S.-M., Ruy, P. de C., Oliveira, G., Coimbra, R.S. (2014) Assessing the efficiency of multiple sequence alignment programs. Algorithms Mol. Biol. 9, 4.
[39] Parker, C.T., Tindall, B.J., Garrity, G.M. (2015) International code of nomenclature of prokaryotes. Int. J. Syst. Evol. Microbiol.
[40] Petrus, A.K., Swithers, K.S., Ranjit, C., Wu, S., Brewer, H.M., Gogarten, J.P., Pasa-Tolic, L., Noll, K.M. (2012) Genes for the major structural components of *Thermotogales* species' togas revealed by proteomic and evolutionary analyses of OmpA and OmpB homologs. PLoS One 7, 1–9.
[41] Pollo, S.M.J., Zhaxybayeva, O., Nesbø, C.L. (2015) Insights into thermoadaptation and the evolution of mesophily from the bacterial phylum *Thermotogae*. Can. J. Microbiol. 61, 655–670.
[42] Ramasamy, D., Mishra, A.K., Lagier, J.C., Padhmanabhan, R., Rossi, M., Sentausa, E., Raoult, D., Fournier, P.E. (2014) A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. Int. J. Syst. Evol. Microbiol. 64, 384–391.
[43] Ravot, G., Magot, M., Fardeau, M.L., Patel, B.K., Prensier, G., Egan, A., Garcia, J.L., Ollivier, B. (1995) *Thermotoga elfii* sp. nov., a novel thermophilic bacterium from an African oil-producing well. Int. J. Syst. Bacteriol. 45, 308–314.
[44] Reysenbach, A.L. (2001) Phylum BII. *Thermotogae* phy. nov. Bergey's Manual of Systematic Bacteriology. In: Boone, D.R., Castenholz, R.W. (Eds.), Springer-Verlag, pp. 369–387.
[45] Reysenbach, A.L., Liu, Y., Lindgren, A.R., Wagner, I.D., Sislak, C.D., Mets, A., Schouten, S. (2013) *Mesoaciditoga lauensis* gen. nov., sp. nov., a moderately thermoacidophilic member of the order *Thermotogales* from a deep-sea hydrothermal vent. Int. J. Syst. Evol. Microbiol. 63, 4724–4729.
[46] Rice, P., Longden, I., Bleasby, A. (2000) EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. 16, 276–277.
[47] Rosselló-Móra, R. (2005) Updating prokaryotic taxonomy. J. Bacteriol. 187, 6255–6257.
[48] Rosselló-Móra, R., Amann, R. (2015) Past and future species definitions for *Bacteria* and *Archaea*. Syst. Appl. Microbiol. 38, 209–216.
[49] Rost, B. (1999) Twilight zone of protein sequence alignments. Protein Eng. 12, 85–94.

[50] Sabath, N., Ferrada, E., Barve, A., Wagner, A. (2013) Growth temperature and genome size in bacteria are negatively correlated, suggesting genomic streamlining during thermal adaptation. Genome Biol. Evol. 5, 966–977.
[51] Schleifer, K.H., Amann, R., Rosselló-Móra, R. (2015) Taxonomy in the age of genomics. Introduction. Syst. Appl. Microbiol. 38, 207–208.
[52] Teeling, H., Meyerdierks, A., Bauer, M., Amann, R., Glöckner, F.O. (2004) Application of tetranucleotide frequencies for the assignment of genomic fragments. Environ. Microbiol. 6, 938–947.
[53] Vandamme, P., Peeters, C. (2014) Time to revisit polyphasic taxonomy. Antonie van Leeuwenhoek Int. J. Gen. Mol. Microbiol. 106, 57–65.
[54] Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J. (2009) Jalview Version 2-A multiple sequence alignment editor and analysis workbench. Bioinformatics 25, 1189–1191.
[55] Wayne, L.G., Brenner, D.J., Colwell, R.R., Grimont, P.A.D., Kandler, O., Krichevsky, M.I., Moore, L.H., Moore, W.E.C., Murray, R.G.E., Stackebrandt, E., Starr, M.P., Truper, H.G. (1987) Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int. J. Syst. Bacteriol. 37, 463–464.
[56] Windberger, E., Huber, R., Trincone, A., Fricke, H., Stetter, K.O. (1989) *Thermotoga thermarum* sp. nov. and *Thermotoga neapolitana* occurring in African continental solfataric springs. Arch. Microbiol. 151, 506–512.
[57] Yarza, P., Ludwig, W., Euzéby, J., Amann, R., Schleifer, K.H., Glöckner, F.O., Rosselló-Móra, R. (2010) Update of the all-species living tree project based on 16S and 23S rRNA sequence analyses. Syst. Appl. Microbiol. 33, 291–299.
[58] Zhaxybayeva, O., Swithers, K.S., Lapierre, P., Fournier, G.P., Bickhart, D.M., Deboy, R.T., Nelson, K.E., Nesbø, C.L., Doolittle, W.F., Gogarten, J.P., Noll, K.M. (2009) On the chimeric nature, thermophilic origin, and phylogenetic placement of the *Thermotogales*. Proc. Natl. Acad. Sci. U. S. A. 106, 5865–5870.