

SHORT GENOME REPORT

Open Access



Complete genome sequence and whole-genome phylogeny of *Kosmotoga pacifica* type strain SLHLJ1^T from an East Pacific hydrothermal sediment

Lijing Jiang^{1,2,3}, Stéphane L'Haridon^{4,5,6}, Mohamed Jebbar^{4,5,6}, Hongxiu Xu^{1,2,3}, Karine Alain^{5,4,6*} and Zongze Shao^{1,2,3*}

Abstract

Kosmotoga pacifica strain SLHLJ1^T is a thermophilic chemoorganoheterotrophic bacterium isolated from a deep-sea hydrothermal sediment. It belongs to the physiologically homogeneous *Thermotogaceae* family. Here, we describe the phenotypic features of *K. pacifica* together with its genome sequence and annotation. The chromosome has 2,169,170 bp, organized in one contig. A total of 1897 candidate protein-encoding genes and 177 RNA genes were identified. The 16S rRNA gene sequence of this strain is distantly related to sequences of some relatives classified in the same genus (*K. olearia* 7.02% and *K. shengliensis* 7.83%), with dissimilarity percentages close to the threshold generally described for genus delineation. Nevertheless, the percentage of conserved proteins (POCP), which is much higher than 50% (around 70%), together with phenotypic features of the isolates, confirm the affiliation all *Kosmotoga* species described so far to the same genus.

Keywords: Marine, Hydrothermal vent, *Thermotogales*, Chemoorganoheterotroph, Thermophile

Introduction

The phylum *Thermotogae* is currently composed of 50 species spread across 13 genera, distinguishable mainly by their characteristic outer membrane known as the 'toga'. These genera are named *Athalassotoga*, *Defluviitoga*, *Fervidobacterium*, *Geotoga*, *Kosmotoga*, *Marinitoga*, *Mesoaciditoga*, *Mesotoga*, *Oceanotoga*, *Petrotoga*, *Pseudothermotoga*, *Thermosiphon* and *Thermotoga* [1–12]. They are grouped into 5 families [1, 10]: (i) *Thermotogaceae*, comprising the genera *Thermotoga* and *Pseudothermotoga*; (ii) *Fervidobacteraceae*, comprising the genera *Fervidobacterium* and *Thermosiphon*; (iii) *Petrotogaceae*, comprising the genera *Petrotoga*, *Defluviitoga*, *Geotoga*, *Marinitoga* and *Oceanotoga*; (iv) *Kosmotogaceae*, comprising the genera *Kosmotoga* and *Mesotoga*; and (v)

Mesoaciditogaceae, comprising the genera *Mesoaciditoga* and *Athalassotoga*. The first representatives of this phylum described from the mid-1990s were all neutrophilic, thermophilic or hyperthermophilic fermentative bacteria from a range of hot anaerobic microbial environments such as deep-sea and terrestrial vents, anaerobic digesters or oil reservoirs. They are relatively homogeneous in terms of physiology. In the last few years, the description of the genera *Mesotoga*, *Mesoaciditoga* and *Athalassotoga*, corresponding to three divergent lineages among the *Thermotogae*, showed that there are also representatives of this order that grow under mesophilic or slightly acidic conditions [1, 7, 8]. The different genera of *Thermotogae* display different tolerances to oxygen and salts, and can produce L-alanine or reduce different sulfur species to prevent the toxic effect of H₂ produced during fermentation. Phylogenetic analyses of the 16S rRNA gene and of concatenated ribosomal proteins place *Thermotogae* as a sister group to *Aquificales*, representing a deeply-branching lineage of the

* Correspondence: Karine.Alain@univ-brest.fr; shaozz@163.com
⁵CNRS, UMR 6197-Laboratoire de Microbiologie des Environnements Extrêmes (LM2E), Institut Universitaire Européen de la Mer (IUEM), rue Dumont d'Urville, 29 280 Plouzané, France
¹Key Laboratory of Marine Genetic Resources, Third Institute of Oceanography, SOA, Xiamen 361005, China
Full list of author information is available at the end of the article

bacterial tree that emerges close to the first delineation between bacterial and archaeal branches [13]. However, the evolutionary history of these bacteria is also characterized by numerous lateral gene transfer events with *Firmicutes* and with *Thermococcales* [13, 14].

The genus *Kosmotoga* was proposed by DiPippo et al. in 2009 [5] and belongs to the family *Kosmotogaceae*, one of the five families of the phylum *Thermotogae*. The genus is currently composed of four type species, *K. olearia* [5], *K. arenicorallina* [15], *K. shengliensis* [15] and *K. pacifica* [16]. *Kosmotoga* species have been isolated from oil reservoirs as well as shallow and deep-sea hydrothermal vents. Strain SLHLJ1^T (=DSM 26965^T = JCM 19180^T = UBOCC 3254^T =MCCC 1A00641^T) is the type strain of the species *K. pacifica*, which was isolated from sediments of an active hydrothermal vent on the East Pacific Rise (102°55'W, 3°58'S) [16]. Here, we present a summary of the physiological features of *K. pacifica* SLHLJ1^T, together with a description of the complete genomic sequence and annotation. A brief genomic comparison was made between *K. pacifica* SLHLJ1^T and *K. olearia* TBF 19.5.1^T and we also calculated (i) ANI and (ii) POCP values among pairs of genomes of *Thermotogae* for which complete genomic sequences were available.

Organism information

Classification and features

Strain SLHLJ1^T was isolated by repeated streaking on plates as described elsewhere [16]. In this study, a whole-genome phylogeny of the *Thermotogae* lineage was constructed based on the core genome (499 core genes) from 20 complete genomes. The core genes were chosen based on identified orthologous genes, which were also single-copy genes from 20 genomes (Additional file 1: Table S1). The result indicated that *K. pacifica* SLHLJ1^T was affiliated to the genus *Kosmotoga*, which formed a deep branch in the phylogenetic tree constructed with the neighbor-joining algorithm (Fig. 1). *K. pacifica* SLHLJ1^T was closely related to *K. arenicorallina*, sharing 97.93% 16S rRNA gene sequence similarity, and was distantly related (<93%) to the other species of the genus *Kosmotoga*. Phylogenetic comparison of 16S rRNA gene sequences of *K. pacifica* SLHLJ1^T and other *Thermotogae* also supported the result that *K. pacifica* SLHLJ1^T clusters with other *Kosmotoga* species (Additional file 2: Figure S1) [16].

K. pacifica SLHLJ1^T cells are Gram-negative non-motile short rods or ovoid cocci (~1 µm long by ~0.6 µm wide) surrounded by a typical toga. They appear singly or occasionally in chains of 3–4 cells within the sheath (Fig. 2). Spores were never observed. Strain SLHLJ1^T grows between 33 and 78 °C, but the optimal

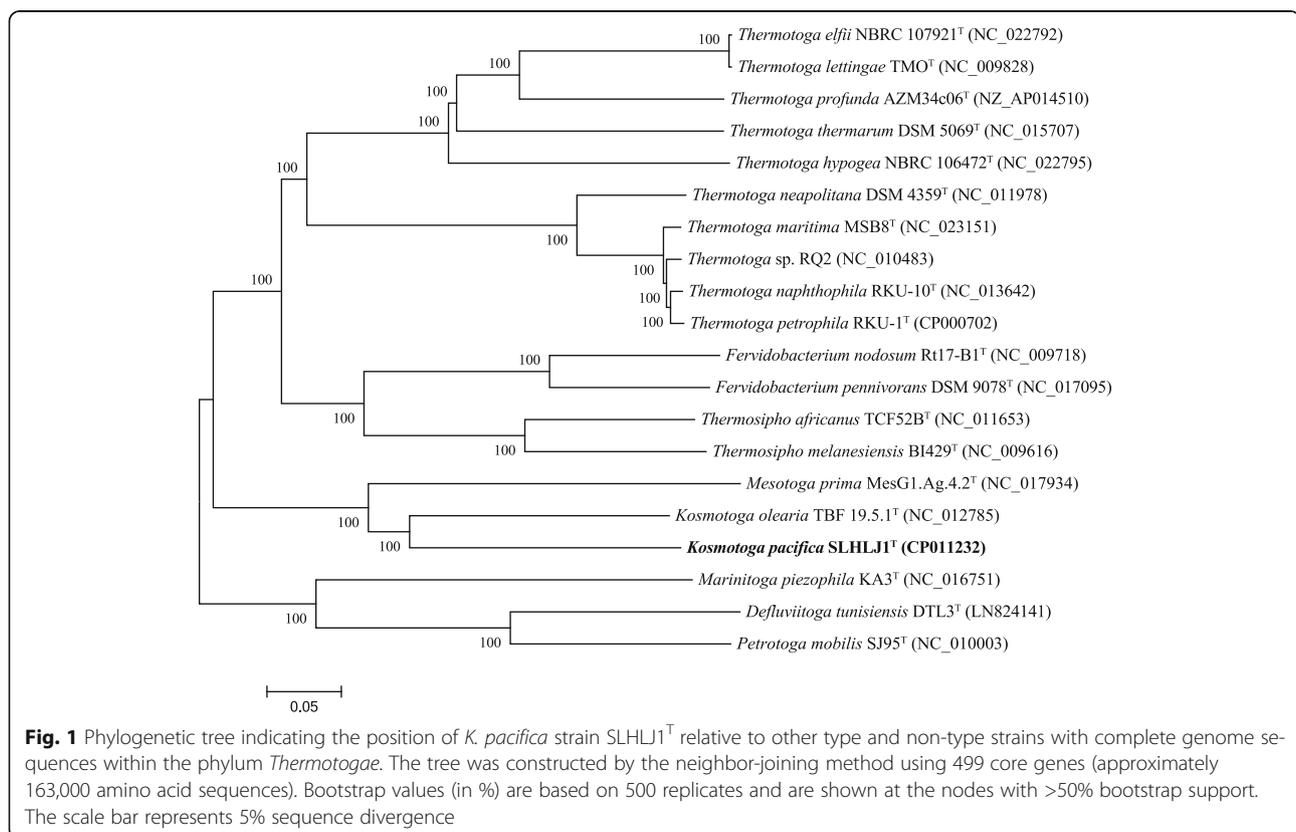


Fig. 1 Phylogenetic tree indicating the position of *K. pacifica* strain SLHLJ1^T relative to other type and non-type strains with complete genome sequences within the phylum *Thermotogae*. The tree was constructed by the neighbor-joining method using 499 core genes (approximately 163,000 amino acid sequences). Bootstrap values (in %) are based on 500 replicates and are shown at the nodes with >50% bootstrap support. The scale bar represents 5% sequence divergence

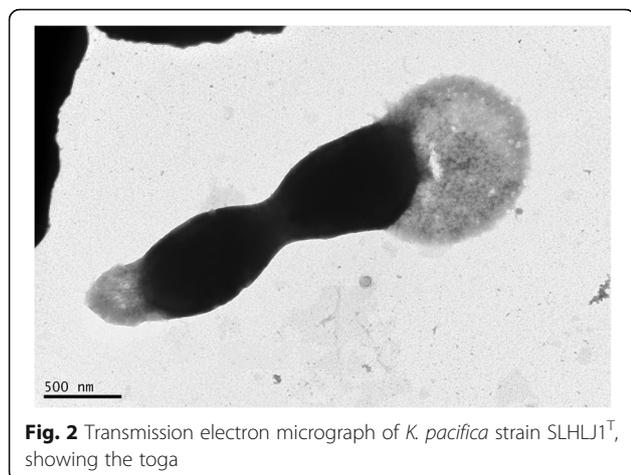


Fig. 2 Transmission electron micrograph of *K. pacifica* strain SLHLJ1^T, showing the toga

growth temperature is 70 °C. Growth occurs under strictly anaerobic and obligate chemoorganoheterotrophic conditions. A small amount of yeast extract is required for growth. The following substrates support growth in the presence of 0.02% yeast extract: peptone, brain–heart infusion, tryptone, glycerol, maltose, xylose, glucose, fructose, cellobiose, trehalose, lactate, propionate and glutamate. The strain can reduce L-cystine and elemental sulfur [16]. A summary of the classification and general features of *K. pacifica* SLHLJ1^T is presented in Table 1.

Genome sequencing information

Genome project history

This organism was selected for sequencing based on its phylogenetic position. The complete genome sequence was deposited in GenBank under the accession number CP011232. Sequencing, finishing and annotation of the *K. pacifica* SLHLJ1^T genome were performed by the Shanghai Majorbio Bio-pharm Technology Co., Ltd (Shanghai, China). Table 2 presents the main project information and its association with MIGS version 2.0 compliance [25].

Growth conditions and DNA isolation

Strain SLHLJ1^T was grown anaerobically for 24 h at 70 °C in 50 mL DSMZ medium 282 (with yeast extract as a carbon and energy source), supplemented with 12 g/L L-cystine. DNA was isolated from the liquid phase without L-cystine, using a standard phenol/chloroform/isoamyl alcohol extraction protocol [26]. The quality and quantity of the extracted DNA were analyzed using agarose gel electrophoresis and NanoDrop. A total of around 20 µg DNA was obtained.

Genome sequencing and assembly

The genome was sequenced using a combination of an Illumina MiSeq (2 × 300 bp) and 454 sequencing

platforms. Libraries were prepared in accordance with manufacturer's instructions. The Newbler V2.8 software package was used for sequence assembly and quality assessment [27]. The draft genome sequence was generated using 454 data. The 454 draft assembly was based on 243,758,031 bp 454 draft data. Newbler parameters were -consed, -a 50, -l 350, -g, -m, and -ml 20. The Phred/Phrap/Consed software package [28] was used for sequence assembly and quality assessment in the subsequent finishing process. Illumina reads were used for gap-filling, correcting potential base errors and increasing consensus quality. Gaps were then filled in by sequencing the PCR products using an ABI 3730xl capillary sequencer. A total of four additional reactions were necessary to close gaps and to improve the quality of the finished sequence. Together, the combination of the Illumina and 454 sequencing platforms provided 676 × coverage of the genome. The final assembly contained 637,426 pyrosequences and 4,870,336 Illumina reads.

Genome annotation

The protein-coding genes, structural RNAs (5S, 16S, and 23S), tRNAs and small non-coding RNAs were predicted using the NCBI PGAP server online [29]. The functional annotation of predicted ORFs was performed using RPS-BLAST [30] against the COG database [31] and Pfam database [32]. The TMHMM program was used for gene prediction with transmembrane helices [33] and the signalP program for gene prediction from peptide signals [34]. ANI values were calculated using JSpecies software [35] and the ANI tool of the Integrated Microbial Genome (IMG) system [36]. POCP indexes were calculated as described elsewhere [37].

Genome properties

The properties and statistics about the genome are summarized in Table 3. The genome is organized in one circular chromosome of 2,169,170 bp (42.52% GC content). In total, 2074 genes were predicted, 1897 of which were protein-coding genes, and 177 of which were RNA genes; 124 pseudogenes were also identified. Most protein-coding genes (83.75%) were assigned putative functions and the remaining ones were annotated as hypothetical proteins. The distribution of genes between COG functional categories is presented in Table 4 and Fig. 3.

Insights from the genome sequence

In the genome sequence of *K. pacifica* SLHLJ1^T, a relatively large number of genes were observed to be assigned to the COG functional categories for transport and metabolism of carbohydrates (6.75%), amino acids (5.54%), translation, ribosomal structure and biogenesis (6.8%), and energy production and conversion (5.75%).

Table 1 Classification and general features of *K. pacifica* SLHLJ1^T [12]

| MIGS ID | Property | Term | Evidence code ^a |
|-----------|------------------------|--|----------------------------|
| | | Domain <i>Bacteria</i> | TAS [17] |
| | | Phylum <i>Thermotogae</i> | TAS [18, 19] |
| | | Class <i>Thermotogae</i> | TAS [18, 20] |
| | Current classification | Order <i>Kosmotogales</i> | TAS [18, 21] |
| | | Family <i>Kosmotogaceae</i> | TAS [18, 22] |
| | | Genus <i>Kosmotoga</i> | TAS [12, 23] |
| | | Species <i>Kosmotoga pacifica</i> | TAS [16] |
| | | Type strain SLHLJ1 ^T | TAS [16] |
| | Gram stain | Negative | TAS [16] |
| | Cell shape | Coccobacilli with a 'toga' (a sheath-like structure) | TAS [16] |
| | Motility | non-motile | TAS [16] |
| | Sporulation | Non-sporulating | TAS [16] |
| | Temperature range | 33-78 °C | TAS [16] |
| | Optimum temperature | 70 °C | TAS [16] |
| | pH range; Optimum | 5.5-8.5;7 | |
| | Carbon source | Yeast extract, peptone, brain–heart infusion, tryptone, glycerol, maltose, xylose, glucose, fructose, cellobiose, trehalose, lactate, propionate and glutamate | TAS [16] |
| | Energy metabolism | Chemoorganoheterotrophic | TAS [16] |
| MIGS-6 | Habitat | Hydrothermal vent environment | TAS [16] |
| MIGS-6.3 | Salinity | 0.5-6% NaCl (w/v) | TAS [16] |
| MIGS-22 | Oxygen requirement | Anaerobic | TAS [16] |
| MIGS-15 | Biotic relationship | Free living | TAS [16] |
| MIGS-14 | Pathogenicity | None | NAS |
| MIGS-23.1 | Isolation | Sediment | TAS [16] |
| MIGS-4 | Geographic location | East Pacific Rise | TAS [16] |
| MIGS-5 | Sample collection time | July 2011 | TAS [16] |
| MIGS-4.1 | Latitude | 3°58'S | TAS [16] |
| MIGS-4.2 | Longitude | 102°55'W | TAS [16] |
| MIGS-4.3 | Altitude | -2891 m | TAS [16] |

^a Evidence codes - TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or on anecdotal evidence). These evidence codes are from the Gene Ontology project [24]. * The rank of phylum is not covered by the Rules of the International Code of Nomenclature of Prokaryotes

Further genome analysis of *K. pacifica* SLHLJ1^T revealed it contained genes for the Embden-Meyerhof-Parnas pathway to convert glucose into pyruvate, but not for the complete pentose phosphate pathway and Entner-Doudoroff pathway due to the lack of several key genes (such as glucose 6-phosphate dehydrogenase and 2-keto-3-deoxy-6-phosphogluconate aldolase). In addition, the tricarboxylic acid cycle was also found to be incomplete in *K. pacifica* SLHLJ1^T. The strain is capable of breaking down substrates such as xylose, cellobiose or trehalose, which is not surprising since an abundance of genes coding for carbohydrate breakdown has been predicted in its genome.

Prior to this study, the only available genome for the genus *Kosmotoga* was *K. olearia* TBF 19.5.1^T. Here, we

compared the genome of *K. pacifica* SLHLJ1^T with *K. olearia* TBF 19.5.1^T (Table 5). *K. olearia* and *K. pacifica* share 92.98% 16S rRNA gene sequence similarity based on full 16S rRNA sequences. The genome size of strain SLHLJ1^T is slightly smaller than that of strain TBF 19.5.1^T. These two strains have nearly identical G + C contents: 42.52% for strain SLHLJ1^T against 41.5% for strain TBF 19.5.1^T. Strain SLHLJ1^T has a slightly smaller gene content than strain TBF 19.5.1^T (2074 vs 2194). *K. pacifica* SLHLJ1^T shares 1524 orthologous genes with *K. olearia* TBF 19.5.1^T.

Furthermore, we wanted to confirm the affiliation of *K. pacifica* SLHLJ1^T to the genus *Kosmotoga* with genomic data. Indeed, there are two lineages within the

Table 2 Project information

| MIGS ID | Property | Term |
|-----------|----------------------------|---|
| MIGS-31 | Finishing quality | Finished |
| MIGS-28 | Libraries used | Two genomic libraries: one 454 PE library (3 kb insert size) and one Illumina PE library (500 bp insert size) |
| MIGS-29 | Sequencing platforms | Illumina Miseq, 454 GS FLX Titanium |
| MIGS-31.2 | Fold coverage | 564 × Illumina; 112 × 454 3 K-PE |
| MIGS-30 | Assemblers | Newbler version 2.8 |
| MIGS-32 | Gene calling method | NCBI PGAP pipeline |
| | Locus Tag | IX53 |
| | Genbank ID | CP011232.1 |
| | GenBank Date of Release | June 3, 2015 |
| | GOLD Project ID | Gp0119521 |
| | BIOPROJECT | PRJNA256122 |
| MIGS-13 | Source material identifier | UBOCC 3254 and MCCC 1A00641 |
| | Project relevance | Thermophile, GEBA |

Kosmotoga genus (*K. pacifica* SLHLJ1^T and *K. arenicorallina* S304^T on the one hand, and *K. shengliensis* 2SM-2^T and *K. olearia* TBF 19.5.1 T on the other) and these are distantly related based on 16S rRNA gene sequence comparisons (they share between 91.7 and 92.4% 16S rRNA gene sequence similarity) [38]. ANI is a useful index for species circumscription [35], and it was recently proposed that a prokaryotic genus could be defined as a group of species with all pairwise POCP values higher than 50% [37]. We therefore performed these two types of analyses to address the issue of the limits of the genus *Kosmotoga*. The POCP index and

Table 3 Genome statistics

| Attribute | Value | % of Total |
|----------------------------------|-----------|------------|
| Genome size (bp) | 2,169,170 | 100.0 |
| DNA coding (bp) | 1,814,445 | 83.65 |
| DNA G + C (bp) | 922,242 | 42.52 |
| DNA scaffolds | 1 | |
| Total genes | 2074 | 100.00 |
| Protein coding genes | 1897 | 91.47 |
| RNA genes | 177 | 8.53 |
| Pseudo genes | 124 | 5.98 |
| Genes in internal clusters | ND | ND |
| Genes with function prediction | 1737 | 83.75 |
| Genes assigned to COGs | 1124 | 54.19 |
| Genes assigned Pfam domains | 1770 | 85.34 |
| Genes with signal peptides | 37 | 1.78 |
| Genes with transmembrane helices | 538 | 25.94 |
| CRISPR repeats | 0 | 0 |

ND not determined

Table 4 Number of genes associated with the general COG functional categories

| Code | Value | % ^a | Description |
|------|-------|----------------|--|
| J | 129 | 6.80 | Translation, ribosomal structure and biogenesis |
| A | 0 | 0.00 | RNA processing and modification |
| K | 53 | 2.79 | Transcription |
| L | 53 | 2.79 | Replication, recombination and repair |
| B | 2 | 0.11 | Chromatin structure and dynamics |
| D | 13 | 0.69 | Cell cycle control, cell division, chromosome partitioning |
| Y | 0 | 0.00 | Nuclear structure |
| V | 20 | 1.05 | Defense mechanisms |
| T | 24 | 1.27 | Signal transduction mechanisms |
| M | 45 | 2.37 | Cell wall/membrane biogenesis |
| N | 2 | 0.11 | Cell motility |
| Z | 0 | 0.00 | Cytoskeleton |
| W | 0 | 0.00 | Extracellular structures |
| U | 17 | 0.90 | Intracellular trafficking and secretion, and vesicular transport |
| O | 49 | 2.58 | Posttranslational modification, protein turnover, chaperones |
| C | 109 | 5.75 | Energy production and conversion |
| G | 128 | 6.75 | Carbohydrate transport and metabolism |
| E | 105 | 5.54 | Amino acid transport and metabolism |
| F | 40 | 2.11 | Nucleotide transport and metabolism |
| H | 34 | 1.79 | Coenzyme transport and metabolism |
| I | 26 | 1.37 | Lipid transport and metabolism |
| P | 80 | 4.22 | Inorganic ion transport and metabolism |
| Q | 14 | 0.74 | Secondary metabolites biosynthesis, transport and catabolism |
| R | 96 | 5.06 | General function prediction only |
| S | 85 | 4.48 | Function unknown |
| - | 773 | 40.75 | Not in COGs |

^a The total is based on the total number of protein coding genes in the genome

ANI value between *K. pacifica* SLHLJ1^T and *K. olearia* TBF 19.5.1^T were respectively 70.2% and 68.5% (with JSpecies) (Fig. 4), or 72.5% (with the IMG system), supporting the assignment of these two isolates to two different species of the same genus.

A total of 20 complete genomic sequences belonging to the phylum *Thermotogae* are publicly available in the NCBI database, including representatives of the genera *Defluviitoga*, *Fervidobacterium*, *Kosmotoga*, *Marinitoga*, *Mesotoga*, *Petrotoga*, *Thermosiphon* and *Thermotoga*. To gain a thorough understanding of the evolutionary relationships and phenotypic distances among the different groups in the *Thermotogae*, a phylogenomic analysis was conducted based on core gene sequences from these 19 genomic sequences and the one of *K. pacifica*. In

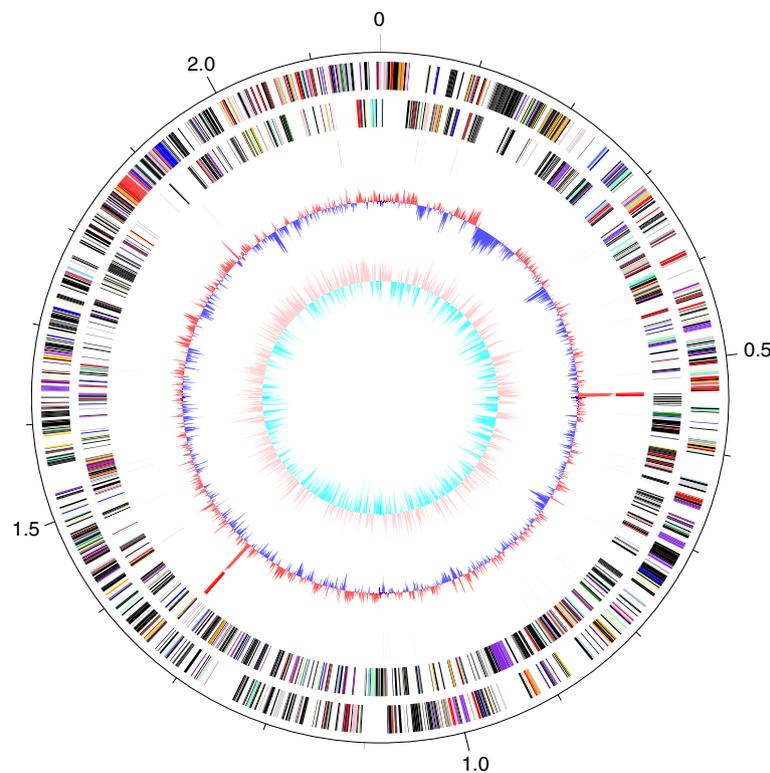


Fig. 3 Graphical map of the chromosome of *K. pacifica* strain SLHLJ1^T. From the edge to the center: Genes on forward strand (colored by COG categories), Genes on reverse strand (colored by COG categories), RNA genes (tRNAs purple and rRNAs red), GC content and GC skew

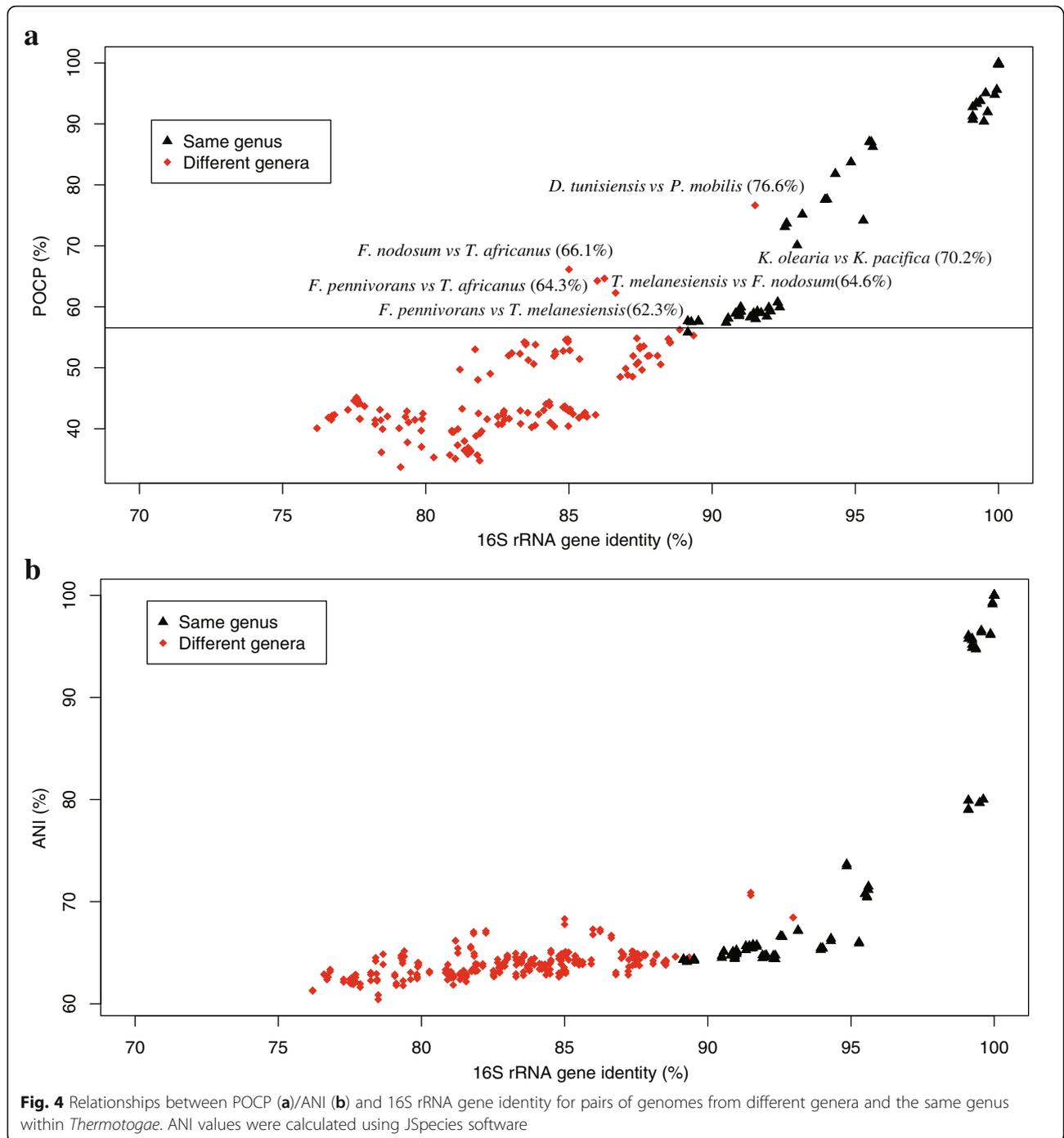
addition, POCP and ANI values between pairs of strains were also calculated. Results are shown in Figs. 1 and 4. The interspecies ANI values calculated using JSpecies and IMG system software ranged from 64 to 99% and from 6 to 99%, respectively, while the intergenera ANI values were in the ranges of 60–70% and 65–86%. Thirty six percent of the intergenera ANI values overlapped with the interspecies ANI values; a result showing, in agreement with [37], that ANI cannot be used as a boundary for genus delineation. Interspecies POCP values were between 55.8 and 95.6%, with a large majority above 57%. Intergenera POCP values ranged from 33.7 to 76.6%, with a majority below 57% (Fig. 4, Additional file 3: Table S3). POCP analyses revealed that there were several high percentages of conserved proteins between representatives of different genera, such as *Defluviitoga tunisiensis* vs *Petrotoga mobilis* (76.6%), *Fervidobacterium nodosum* vs *Thermosipho africanus* (66.1%), *Fervidobacterium pennivorans* vs *Thermosipho africanus* (64.3%) or *Thermosipho melanesiensis* vs

Fervidobacterium nodosum (64.6%). This result was surprising for us, knowing that 16S rRNA gene sequence dissimilarities among *Thermotogae* genera (>11%) are much higher than in the vast majority of bacterial orders, but that physiology is homogeneous among the *Thermotogae*, with only a few or minor differences between genera (Additional file 4: Table S2).

Representatives of the four following groups: *Defluviitoga*/*Petrotoga*, *Fervidobacterium*/*Thermotoga*/*Thermosipho*, *Marinitoga*/*Petrotoga*, and *Kosmotoga*/*Mesotoga* (two genera characterized by their distinctly different temperature ranges for growth), shared all pairwise POCP values higher than 50%, which is the pairwise POCP value suggested as a threshold for genus delineation [37]. These clusters of genera are in agreement with the well-resolved clades identified in a previous comparative genomic analysis and supported by multiple conserved signature indels [10]. The compilation of physiological and genotypic features of the different genera (Additional file 4: Table S2), together with the POCP

Table 5 Comparative genomic characteristics of *K.pacifica* SLHLJ1^T and *K. olearia* TBF 19.5.1^T

| Genome Name | Genome size (bp) | %GC | Gene count | Protein coding | Plasmid number | rRNA | tRNAs | Orthologous genes |
|---|------------------|------|------------|----------------|----------------|------|-------|-------------------|
| <i>K. pacifica</i> SLHLJ1 ^T | 2,169,170 | 42.5 | 2074 | 1897 | 0 | 6 | 46 | 1534 |
| <i>K. olearia</i> TBF 19.5.1 ^T | 2,302,126 | 41.5 | 2194 | 2116 | 0 | 6 | 46 | 1579 |



index (Fig. 4 and Additional file 3: Table S3) and 16S rRNA phylogenetic distance (Additional file 2: Figure S1) tend to indicate that the pairs of *Defluviitoga*-*Petrotoza* and *Fervidobacterium*-*Thermosiphon* representatives are less genotypically distant and also have less differentiating characteristics than the other pairs of genera. The results of POCP values together with the physiology of these taxa call into question the classification of the *Thermotogae* at the genus phylogenetic level and suggest that either (i)

there might be fewer genera of *Thermotogae* than currently described, and that *Thermotogae* could be reclassified at the genus level by taking into account genomic information, evolutionary history and discriminative physiological characteristics; or (ii) the POCP might not be a sufficiently resolved genomic index for the delineation of genera within a homogeneous phenotype. In the light of these observations, it could be interesting to perform deep phylogenetic analyses of the *Thermotogae* (with

a maximum of genomes) to study the evolutionary history and parallel evolution of genotypes and phenotypes within this *family*.

Conclusions

Strain SLHLJ1^T is the first strain of the genus *Kosmotoga* to be isolated from the deep-sea hydrothermal vent environment. Its physiology and genetic content were compared to those of other *Thermotogae*. This comprehensive analysis showed that genomic information is necessary to understand the evolutionary relationships of the different groups in this well-defined lineage characterized by homogeneous physiology.

Additional files

Additional file 1: Table S1. List of the core genes chosen for the whole genome phylogenetic analysis. This list is composed of 499 orthologous genes from 20 genomes within the phylum *Thermotogae*. (XLS 210 kb)

Additional file 2: Figure S1. Phylogenetic tree based on 16S rRNA gene sequences showing the position of *K. pacifica* strain SLHLJ1^T within the phylum *Thermotogae*. The alignment was performed with 16S rDNA sequences of related species and environmental sequences. The topology shown was obtained with the neighbor-joining algorithm. Bootstrap values (from 1000 replicates) are indicated at the branch nodes. The scale bar represents 2% sequence divergence. (PDF 462 kb)

Additional file 3: Table S3. Comparison of POCP value and 16S rRNA gene identity for pairs of genomes from different genera of *Thermotogae*. (DOCX 21 kb)

Additional file 4: Table S2. Differential characteristics of eight genera of *Thermotogae*, with genome sequences. Data were taken from *Defluviitoga* [2], *Fervidobacterium* [3, 39], *Kosmotoga* [5, 15, 16], *Marinitoga* [6, 40], *Mesotoga* [8], *Petrotoga* [4, 41], *Thermosipho* [10, 42], and *Thermotoga* [12, 43]. ND, No data available; +, positive; -, negative; ±, positive for some species, but not all. (DOCX 18 kb)

Abbreviations

ANI: Average nucleotide identity; PGAP: Prokaryotic genome annotation pipeline; POCP: Percentage of conserved proteins

Acknowledgements

We gratefully acknowledge Weiming Xiong for his help with the analysis of the genomic data. This work was financially supported by the National Program on Key Basic Research Project (973 Program) (No.2012CB417300), COMRA project (No.DY125-15-R-01), National Natural Science Foundation of China (No. 41672333), the EU FP7 program MaCuMBA (Grant agreement no 311975), the PICS-INEE Phypress, the PHC Cai Yuanpei Pandore (N° 30412WG), the PHC Cai Yuanpei Provirvent (No. 34634WE), the "Laboratoire d'Excellence" LabexMER (ANR-10-LABX-19) and the program "Investissements d'Avenir". We are indebted to Helen McCombie [Bureau de Traduction de l'Université (BTU), Université de Bretagne Occidentale, Brest, e-mail: btu@univ-brest.fr] for improving the English.

Authors' contributions

LJJ carried out the genome sequence analysis and drafted the manuscript. SLH and XHX performed the DNA extraction. KA, MJ and ZZS participated in the study design and finalized the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Key Laboratory of Marine Genetic Resources, Third Institute of Oceanography, SOA, Xiamen 361005, China. ²State Key Laboratory Breeding Base of Marine Genetic Resources, Xiamen 361005, China. ³Key Laboratory of Marine Genetic Resources of Fujian Province, Xiamen 361005, China. ⁴Université de Bretagne Occidentale, UMR 6197-Laboratoire de Microbiologie des Environnements Extrêmes (LM2E), Institut Universitaire Européen de la Mer (IUEM), rue Dumont d'Urville, 29 280 Plouzané, France. ⁵CNRS, UMR 6197-Laboratoire de Microbiologie des Environnements Extrêmes (LM2E), Institut Universitaire Européen de la Mer (IUEM), rue Dumont d'Urville, 29 280 Plouzané, France. ⁶Ifremer, UMR 6197-Laboratoire de Microbiologie des Environnements Extrêmes (LM2E), Technopôle Brest-Iroise, BP70, 29 280 Plouzané, France.

Received: 17 September 2015 Accepted: 10 December 2016

Published online: 05 January 2017

References

- Itoh T, Onishi M, Kato S, Iino T, Sakamoto M, Kudo T, Takashina T, Ohkuma M. *Athalassotoga saccharophila* gen. nov. sp. nov. isolated from an acidic terrestrial hot spring of Japan, and proposal of *Mesoaciditogales* ord. nov., *Mesoaciditogaceae* fam. nov. in the phylum *Thermotogae*. *Int J Syst Evol Microbiol*. 2016;66:1045–51. PubMed <http://dx.doi.org/10.1099/ijsem.0.000833>.
- Ben Hania W, Godbane R, Postec A, Hamdi M, Ollivier B, Bardeau ML. *Defluviitoga tunisiensis* gen. nov., sp. nov., a thermophilic bacterium isolated from a mesothermic and anaerobic whey digester. *Int J Syst Evol Microbiol*. 2012;62:1377–82. PubMed <http://dx.doi.org/10.1099/ijms.0.033720-0>.
- Patel BKC, Daniel RM. *Fervidobacterium nodosum* gen. nov. and spec. nov., a new chemoorganotrophic, aldoactive, anaerobic bacterium. *Arch Microbiol*. 1985;141:63–9. <http://dx.doi.org/10.1007/BF00446741>.
- Davey ME, Key R, Nakamura K, Stahl D. Isolation of three species of *Geotoga* and *Petrotoga*: two new genera, representing a new lineage in the bacterial line of descent distantly related to the *Thermotogales*. *Syst Appl Microbiol*. 1993;16:191–200. [http://dx.doi.org/10.1016/S0723-2020\(11\)80467-4](http://dx.doi.org/10.1016/S0723-2020(11)80467-4).
- Dipippo JL, Nesbo CL, Dahle H, Doolittle WF, Birkland NK, Noll KM. *Kosmotoga olearia* gen. nov., sp. nov., a thermophilic, anaerobic heterotroph isolated from an oil production fluid. *Int J Syst Evol Microbiol*. 2009;59:2991–3000. PubMed <http://dx.doi.org/10.1099/ijms.0.0008045-0>.
- Wery N, Lesongeur F, Pignet P, Derennes V, Cambon-Bonavita MA, Godfroy A, Barbier G. *Marinitoga camini* gen. nov., sp. nov., a rod-shaped bacterium belonging to the order *Thermotogales*, isolated from a deep-sea hydrothermal vent. *Int J Syst Evol Microbiol*. 2001;51:495–504. PubMed <http://dx.doi.org/10.1099/00207713-51-2-495>.
- Reysenbach AL, Liu Y, Lindgren AR, Wagner ID, Sislak CD, Mets A, Schouten S. *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*. 2013;63:4724–9. PubMed <http://dx.doi.org/10.1099/ijms.0.050518-0>.
- Nesbo CL, Bradnan DM, Adebusi A, Dlutek M, Petrus AK, Foght J, Doolittle WF, Noll KM. *Mesotoga prima* gen. nov., sp. nov., the first described mesophilic species of the *Thermotogales*. *Extremophiles*. 2012;16:387–93. PubMed <http://dx.doi.org/10.1007/s00792-012-0437-0>.
- Jayasinghearachchi HS, Lal B. *Oceanotoga teriensis* gen. nov., sp. nov., a thermophilic bacterium isolated from offshore oil-producing wells. *Int J Syst Evol Microbiol*. 2011;61:554–60. PubMed <http://dx.doi.org/10.1099/ijms.0.018036-0>.
- Bhandari V, Gupta RS. 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*. 2014;105:143–68. PubMed <http://dx.doi.org/10.1007/s10482-013-0062-7>.
- Huber RWC, Langworthy TA, Fricke H, Stetter KO. *Thermosipho africanus* gen. nov., represents a new genus of thermophilic eubacteria within the *Thermotogales*. *Syst Appl Microbiol*. 1989;12:32–7. [http://dx.doi.org/10.1016/S0723-2020\(89\)80037-2](http://dx.doi.org/10.1016/S0723-2020(89)80037-2).
- Huber RLT, König H, Thomm M, Woese CR, Sleytr UB, Stetter KO. *Thermotoga maritima* sp. nov. represents a new genus of unique extremely

- thermophilic eubacteria growing up to 90. *Arch Microbiol.* 1986;144:324–33. <http://dx.doi.org/10.1007/BF00409880>.
13. Zhaxybayeva O, Swithers KS, Lapiere P, Fournier GP, Bickhart DM, DeBoy RT, et al. On the chimeric nature, thermophilic origin, and phylogenetic placement of the *Thermotogales*. *Proc Natl Acad Sci USA.* 2009;106:5865–70. PubMed <http://dx.doi.org/10.1073/pnas.0901260106>.
 14. Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, et al. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature.* 1999;399:323–9. PubMed <http://dx.doi.org/10.1038/20601>.
 15. Nunoura T, Hirai M, Imachi H, Miyazaki M, Makita H, Hirayama H, Furushima Y, Yamamoto H, Takai K. *Kosmotoga arenicorallina* sp. nov. a thermophilic and obligately anaerobic heterotroph isolated from a shallow hydrothermal system occurring within a coral reef, southern part of the Yaeyama Archipelago, Japan, reclassification of *Thermococcoides shengliensis* as *Kosmotoga shengliensis* comb. nov., and emended description of the genus *Kosmotoga*. *Arch Microbiol.* 2010;192:811–9. PubMed <http://dx.doi.org/10.1007/s00203-010-0611-7>.
 16. L'Haridon S, Jiang LJ, Alain K, Chalopin M, Rouxel O, Beauverger M, Xu HX, Shao ZZ, Jebbar M. *Kosmotoga pacifica* sp. nov., a thermophilic chemorganoheterotrophic bacterium isolated from an East Pacific hydrothermal sediment. *Extremophiles.* 2014;18:81–8. PubMed <http://dx.doi.org/10.1007/s00792-013-0596-7>.
 17. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA.* 1990;87:4576–9. PubMed <http://dx.doi.org/10.1073/pnas.87.12.4576>.
 18. Validation list no. 85. Validation of publication of new names and new combinations previously effectively published outside the IJSEM. *Int J Syst Evol Microbiol.* 2002; 52:3–4. PubMed <http://dx.doi.org/10.1099/ijs.0.02358-0>
 19. Reysenbach AL, Phylum II B. *Thermotogae* phyl. nov. In: Boone DR, Castenholz RW, Garrity GM, editors. *Bergey's Manual of Systematic Bacteriology*, second edition, vol. 1 (The *Archaea* and the deeply branching and phototrophic *Bacteria*). New York: Springer; 2001. p. 369–87.
 20. Reysenbach AL, Class I. *Thermotogae* class nov. In: Boone DR, Castenholz RW, Garrity GM, editors. *Bergey's Manual of Systematic Bacteriology*, second edition, vol. 1 (The *Archaea* and the deeply branching and phototrophic *Bacteria*). New York: Springer; 2001. p. 369–70.
 21. Reysenbach AL, Order I. *Thermotogales* ord. nov. In: Boone DR, Castenholz RW, Garrity GM, editors. *Bergey's Manual of Systematic Bacteriology*, second edition, vol. 1 (The *Archaea* and the deeply branching and phototrophic *Bacteria*). New York: Springer; 2001. p. 369–70.
 22. Reysenbach AL, Family I. *Thermotogaceae* fam. nov. In: Boone DR, Castenholz RW, Garrity GM, editors. *Bergey's Manual of Systematic Bacteriology*, second edition, vol. 1 (The *Archaea* and the deeply branching and phototrophic *Bacteria*). New York: Springer; 2001. p. 370.
 23. List Editor. Validation List no. 22. Validation of the publication of new names and new combinations previously effectively published outside the IJSB. *Int J Syst Bacteriol.* 1986; 36:573–576. <http://dx.doi.org/10.1099/00207713-36-4-573>
 24. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000;25:25–9. PubMed <http://dx.doi.org/10.1038/75556>.
 25. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, et al. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol.* 2008;26:541–7. PubMed <http://dx.doi.org/10.1038/nbt1360>.
 26. Charbonnier F, Erauso G, Barbeyron T, Prieur D, Forterre P. Purification of plasmids from thermophilic and hyperthermophilic archaeobacteria. In F. T. Robb, K. R. Sowers, S. DasSarma, A. R. Place, H. J. Schreier, and E. M. Fleischmann (ed), *Archaea: a laboratory manual*, in press. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
 27. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen ZT, Dewell SB, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature.* 2005;437:376–80. PubMed <http://dx.doi.org/10.1038/nature03959>.
 28. Phred, Phrap and Consed software packages. <http://www.genome.washington.edu>
 29. Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, Garrity G, Kodira CD, Kyrpides N, Madupu R, Markowitz V, et al. Toward an online repository of Standard Operating Procedures (SOPs) for (meta)genomic annotation. *OMICS.* 2008;12:137–41. PubMed <http://dx.doi.org/10.1089/omi.2008.0017>.
 30. Marchler-Bauer A, Anderson JB, Derbyshire MK, DeWeese-Scott C, Gonzales NR, Gwatz M, Hao L, He S, Hurwitz DI, Jackson JD, et al. CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res.* 2007;35:D237–40. PubMed <http://dx.doi.org/10.1093/nar/gkl951>.
 31. Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 2000;28:33–6. PubMed <http://dx.doi.org/10.1093/nar/28.1.33>.
 32. Sonnhammer EL, Eddy SR, Durbin R. Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins* 1997; 28:405–420. PubMed [http://dx.doi.org/10.1002/\(SICI\)1097-0134\(199707\)28:3 <405::AID-PROT10 > 3.0.CO;2-L](http://dx.doi.org/10.1002/(SICI)1097-0134(199707)28:3 <405::AID-PROT10 > 3.0.CO;2-L)
 33. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 2001;305:567–80. PubMed <http://dx.doi.org/10.1006/jmbi.2000.4315>.
 34. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol.* 2004;340:783–95. PubMed <http://dx.doi.org/10.1016/j.jmb.2004.05.028>.
 35. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A.* 2009;106:19126–31. PubMed <http://dx.doi.org/10.1073/pnas.0906412106>.
 36. Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, Pati A. Microbial species delineation using whole genome sequences. *Nucleic Acids Res.* 2015;43:6761–71. doi:10.1093/nar/gkv657. PubMed <http://dx.doi.org/>
 37. Qin QL, Xie BB, Zhang XY, Chen XL, Zhou BC, Zhou J, Oren A, Zhang YZ. A proposed genus boundary for the prokaryotes based on genomic insights. *J Bacteriol.* 2014;196:2210–5. PubMed <http://dx.doi.org/10.1128/JB.01688-14>.
 38. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer KH, Whitman WB, Euzéby J, Amann R, Rosselló-Móra R. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Rev Microbiol.* 2014;12:635–45. PubMed <http://dx.doi.org/10.1038/nrmicro3330>.
 39. Podosokorskaya OA, Merkel AY, Kolganova TV, Chernyh NA, Miroshnichenko ML, Bonch-Osmolovskaya EA, Kublanov IV. *Ferriobacterium riparium* sp. nov., a thermophilic anaerobic cellulolytic bacterium isolated from a hot spring. *Int J Syst Evol Microbiol.* 2011;61:2697–701. PubMed <http://dx.doi.org/10.1099/ijs.0.026070-0>.
 40. Postec A, Ciobanu M, Birrien JL, Bienvu N, Prieur D, Romancer ML. *Marinitoga litoralis* sp. nov., a thermophilic, heterotrophic bacterium isolated from a coastal thermal spring on Ile Saint-Paul, Southern Indian Ocean. *Int J Syst Evol Microbiol.* 2010;60:1778–82. PubMed <http://dx.doi.org/10.1099/ijs.0.017574-0>.
 41. Miranda-Tello E, Fardeau ML, Joulian C, Magot M, Thomas P, Tholozan JL, Ollivier B. *Petrotoga halophila* sp. nov., a thermophilic, moderately halophilic, fermentative bacterium isolated from an offshore oil well in Congo. *Int J Syst Evol Microbiol.* 2007;57:40–4. PubMed <http://dx.doi.org/10.1099/ijs.0.064516-0>.
 42. Podosokorskaya OA, Bonch-Osmolovskaya EA, Godfroy A, Gavrillov SN, Beskorovaynaya DA, Sokolova TG, Kolganova TV, Toshchakov SV, Kublanov IV. *Thermosiphon activus* sp. nov., a thermophilic, anaerobic, hydrolytic bacterium isolated from a deep-sea sample. *Int J Syst Evol Microbiol.* 2014; 64:3307–13. PubMed <http://dx.doi.org/10.1099/ijs.0.063156-0>.
 43. Mori K, Yamazoe A, Hosoyama A, Ohji S, Fujita N, Ishibashi J, Kimura H, Suzuki K. *Thermotoga profunda* sp. nov. and *Thermotoga caldifontis* sp. nov., anaerobic thermophilic bacteria isolated from terrestrial hot springs. *Int J Syst Evol Microbiol.* 2014;64:2128–36. PubMed <http://dx.doi.org/10.1099/ijs.0.060137-0>.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

