Permanent draft genome of Thermithiobacillus tepidarius DSM 3134T, a moderately thermophilic, obligately chemolithoautotrophic member of the Acidithiobacillia

Boden, Rich

http://hdl.handle.net/10026.1/6677

10.1186/s40793-016-0188-0
Standards in Genomic Sciences
Springer Science and Business Media LLC

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.
Permanent draft genome of Thermithiobacillus tepidarius DSM 3134\textsuperscript{T}, a moderately thermophilic, obligately chemolithoautotrophic member of the Acidithiobacillia

Rich Boden\textsuperscript{1,2*}, Lee P. Hutt\textsuperscript{1,2}, Marcel Huntemann\textsuperscript{3}, Alicia Clum\textsuperscript{3}, Manoj Pillay\textsuperscript{3}, Krishnaveni Palaniappan\textsuperscript{3}, Neha Varghese\textsuperscript{3}, Natalia Mikhailova\textsuperscript{3}, Dimitrios Stamatis\textsuperscript{3}, Tatiparthi Reddy\textsuperscript{3}, Chew Yee Ngan\textsuperscript{3}, Chris Daum\textsuperscript{3}, Nicole Shapiro\textsuperscript{3}, Victor Markowitz\textsuperscript{3}, Natalia Ivanova\textsuperscript{3}, Tanja Woyke\textsuperscript{3} and Nikos Kyrpides\textsuperscript{3}

Abstract

Thermithiobacillus tepidarius DSM 3134\textsuperscript{T} was originally isolated (1983) from the waters of a sulfidic spring entering the Roman Baths (Temple of Sulis-Minerva) at Bath, United Kingdom and is an obligate chemolithoautotroph growing at the expense of reduced sulfur species. This strain has a genome size of 2,958,498 bp. Here we report the genome sequence, annotation and characteristics. The genome comprises 2,902 protein coding and 66 RNA coding genes. Genes responsible for the transaldolase variant of the Calvin-Benson-Bassham cycle were identified along with a biosynthetic horseshoe in lieu of Krebs’ cycle sensu stricto. Terminal oxidases were identified, viz. cytochrome c oxidase (cbb\textsubscript{3}, EC 1.9.3.1) and ubiquinol oxidase (bd, EC 1.10.3.10). Metalloresistance genes involved in pathways of arsenic and cadmium resistance were found. Evidence of horizontal gene transfer accounting for 5.9 % of the protein-coding genes was found, including transfer from Thiobacillus spp. and Methylococcus capsulatus Bath, isolated from the same spring. A sox gene cluster was found, similar in structure to those from other Acidithiobacillia – by comparison with Thiobacillus thioparus and Paracoccus denitrificans, an additional gene between soxA and soxB was found, annotated as a DUF302-family protein of unknown function. As the Kelly-Friedrich pathway of thiosulfate oxidation (encoded by sox) is not used in Thermithiobacillus spp., the role of the operon (if any) in this species remains unknown. We speculate that DUF302 and sox genes may have a role in periplasmic trithionate oxidation.

Keywords: Thermithiobacillus tepidarius, Acidithiobacillia, Sulfur oxidation, Chemolithoautotroph, Thiosulfate, DUF302

Introduction

Thermithiobacillus tepidarius DSM 3134\textsuperscript{T} [1, 2] is a moderately thermophilic sulfur-oxidising obligately chemolithoautotrophic member of the Acidithiobacillia originally published as ‘Thiobacillus tepidarius’ and assigned to the Betaproteobacteria – this has since been resolved by proteogenomic studies and the species moved firstly to a new genus – Thermithiobacillus – [3] in the Gammaproteobacteria and later to a separate Class, along the the genus Acidithiobacillia [4]. To date it is the only species of the genus with a validly published name and one of only two strains in cultivation [5]. The obligately aerobic chemolithoautotroph was demonstrated [1, 2] to use sulfur oxyanions as sole energy sources. It has a temperature optimum of 44 °C, reflecting its environment of isolation. Chemostat-based studies have demonstrated unusually high specific growth yields compared to...
other chemolithoautotrophs and biochemical studies have demonstrated the presence of a range of sulfur oxidation enzymes including rhodanese (EC 2.8.1.1), trithionate hydrolase (EC 3.12.1.1), thiosulfate dehydrogenase (EC 1.8.2.2), a tetrahionate-proton symport system [6] and at least 3 of each cytochromes c and b [7]. Proton translocations per mole of energy source were significantly higher than in other sulfur-oxidising autotrophs, potentially explaining the high yields. *T. tepidarius* DSM 3134^T^ was selected for genome sequencing as part of the Department of the Environment DOE-CSP 2012 initiative – as a type species of a genus.

**Organism information**

**Classification and features**

This strain was isolated from sulfidic groundwater flowing into a Roman bathhouse (Temple of Sulis-Minerva, now The Roman Baths, Bath, UK) – the only other strain of this genus held in a culture collection (*Thermithiobacillus* sp. NCIMB 8349) came from decomposing concrete in the Melbourne sewers in the 1940s [5]. The authors have detected at least 6 OTUs representing probably other *Thermithiobacillus* spp. in 16S rRNA gene libraries from the Roman Baths and have isolated a number of strains to date, indicating that *Thermithiobacillus* spp. are no more difficult to isolate than other sulfur-oxidising autotrophs and may thus simply be rare or confined to rare ecosystems. It forms white colonies of 2–5 mm diameter in 48 h that smell faintly of elemental sulfur if grown on thiosulfate-containing basal salts agar. In batch cultures, thiosulfate is oxidized stoichiometrically to tetrahionate early in the exponential phase, resulting in an increase in culture pH from pH 6.8 to pH 7.5–8.0 – a hallmark of the genus – before being fully oxidized to sulfate, with concomitant fall in culture pH, usually ending at pH 5.2. In continuous cultures, no intermediates accumulate in the medium. In the authors’ hands, trithionate has also been observed very early in the growth phase in batch culture, prior to tetrahionate production. Substrate-level phosphorylation appears not to participate in the energy conservation of this strain and all ATP is thus formed through oxidative phosphorylation [2]. The type – and only – strain was isolated from an enrichment culture comprising water obtained from the inflow of the Great Bath (Roman Baths, Bath, UK) in 1983 (Ann P. Wood, personal communication) added to a basal salts medium supplemented with thiosulfate and monomethylamine hydrochloride, before plating onto basal salts agar containing 5 mM thiosulfate as sole energy source and incubated under air enriched with 5 % (v/v) carbon dioxide as sole carbon source. Key features of this organism are summarized in Table 1. A phylogenetic tree based on the 16S rRNA gene sequence, showing the position of the organism with regard to the *Acidithiobacillia*, rooted with *Thiobacillus thioparus*, is given in Fig. 1.

Cells are 0.6 – 1.0 by 0.2 to 0.4 μm and stain Gram negative. They are rapidly motile by means of a single polar flagellum up to 4 μm in length, as shown in Fig. 2. Ubiquinone-8 is the dominant respiratory quinone and cells fix carbon dioxide via the Calvin-Benson-Bassham cycle at the expense of inorganic sulfur oxidation. Cells accumulate polyphosphate (‘volutin’) granules when grown in batch culture but are typically free from storage granules when grown in energy-source-limited chemostats. Anaerobic growth is not observed with tetrahionate as the electron donor and nitrate, nitrite, nitrous oxide, elemental sulfur, sulfate, tetrahionate or pyruvate as terminal electron acceptors, but cultures can reduce nitrate to nitrite. Experimental estimations of G + C content of genomic DNA are 66.6 ± 0.5 mol% by buoyant density [1] or 65.9 ± 0.8 mol% by acid denaturation [9] in our hands.

Dry biomass is 47 % (w/w) C regardless of the energy source used. *T. tepidarius* DSM 3134^T^ does not grow on any organic carbon compound tested, including sugars (glucose, ribose, fructose, sucrose), intermediates of Krebs cycle (citrate, succinate, fumarate, malate, oxaloacetate), carboxylates (glycolate, formate, acetate, propionate, pyruvate), C^1^ compounds (monomethylamine, dimethylamine, trimethylamine, methanol, methane), structural amino acids (all 20), substituted thiophenes (thiophene-2-carboxylate, thiophene-3-carboxylate) or complex media (yeast extract, nutrient broth, brain-heart infusion, Columbia sheep blood agar, chocolate agar). Energy sources that support autotrophic growth are elemental sulfur, sulfide, trithionate, tetrahionate, hexathionate, heptathionate and thiosulfate. Fe(II), Mn(II), Cu(I), U(IV), pentathionate, dithionate, thiocyanate, sulfite, carbon disulfide, carbonyl sulfide, dimethylsulfide, dimethylsulfoxide, dimethylsulfone and formate do not support autotrophic growth as energy sources. The high growth yields and tetrahionate-accumulation in the early phases of growth make this strain a very interesting model organism for elucidation of sulfur oxidation pathways and their evolution.

**Genome sequencing information**

**Genome project history**

This organism was selected for sequencing on the basis of its role in sulfur cycling, physiological, biochemical, evolutionary and biogeochemical importance, and is part of the Genomic Encyclopedia of *Bacteria* and * Archaea*, 1,000 Microbial Genomes project at the U.S. Department of Energy, Joint Genome Institute (JGI). The genome project is deposited in the Genomes OnLine Database [10] and a high-quality permanent draft genome sequence in IMG [11]. Sequencing, finishing and annotation were performed by the JGI using state of the art sequencing
technology [12]. A summary of the project information is shown in Table 2.

**Growth conditions and genomic DNA preparation**

*T. tepidarius* DSM 3134\textsuperscript{T} DNA was obtained from Dr Hans-Peter Klenk at the Deutsche Sammlung von Mikroorganismen und Zellkulturen Gmbh (DSMZ) having been grown on basal salts medium pH 6.9, supplemented with 10 mM tetrathionate as the sole energy source (DSM Medium 333). DNA was extracted using the JETFLEX Genomic DNA Purification Kit from Genomed (Löhne, Germany) into TE Buffer.

**Genome sequencing and assembly**

The draft genome of *Thermithiobacillus tepidarius* DSM 3134\textsuperscript{T} was generated at the DOE Joint Genome Institute (JGI) using the Illumina technology [13]. An Illumina standard shotgun library was constructed and sequenced using the Illumina HiSeq 2000 platform which generated 13,370,056 reads totaling 2,005.5 Mbp. Library construction and sequencing were performed at the JGI - details are on their website [14]. All raw Illumina sequence data was passed through JGI’s DUK filtering program, which removes known Illumina sequencing and library preparation artifacts (Mingkun L, Copeland A, Han J, Unpublished). Filtered Illumina reads were assembled using Velvet (version 1.1.04) [15]; 1–3 Kbp simulated paired end reads were created from Velvet contigs using wgsim [16] and Illumina reads were assembled with simulated read pairs using Allpaths-LG (version r42328) [17]. Parameters for assembly steps were: Velvet (velveth: 63 –shortPaired and velvetg: -very clean yes – exportFiltered yes –min contig lgth 500 –scaffolding no – cov cutoff 10); wgsim (–e 0 –1 100 –2 100 –r 0 –R 0 –X 0);

\begin{table}[h]
\centering
\begin{tabular}{llll}
MIGS ID & Property & Term & Evidence code\textsuperscript{a} \\
\hline
Classification & Domain & Bacteria & TAS [34] \\
Phylum & Proteobacteria & TAS [4, 35] \\
Class & Acidithiobacillia & TAS [4] \\
Order & Acidithiobacillales & TAS [4] \\
Family & Thermithiobactaceae & TAS [4] \\
Genus & Thermithiobacillus & TAS [3] \\
Species & *Thermithiobacillus tepidarius* & TAS [1–5] \\
\textit{(Type) strain: DSM 3134} & & TAS [1–5] \\
Gram stain & Negative & TAS [1, 2] \\
Cell shape & Rod & TAS [1, 2] \\
Motility & Motile & TAS [1, 2] \\
Sporulation & None & TAS [1, 2] \\
Temperature range & 20–52 °C & TAS [1, 2, 5] \\
Optimum temperature & 44 °C & TAS [1, 2] \\
P\text{H} range; Optimum & 5.2–8.0; 6.8 & TAS [1, 2] \\
Carbon source & Carbon dioxide & TAS [1, 2] \\
MIGS-6 & Habitat & Thermal sulfidic springwater & TAS [1] \\
MIGS-6.3 & Salinity & N.D. & NAS [1–5] \\
MIGS-22 & Oxygen requirement & Aerobic & TAS [1, 2] \\
MIGS-15 & Biotic relationship & Free-living & TAS [1, 2] \\
MIGS-14 & Pathogenicity & Non-pathogen & NAS \\
MIGS-4 & Geographic location & United Kingdom/England & TAS [1, 2] \\
MIGS-5 & Sample collection & 1983 & NAS \\
MIGS-4.1 & Latitude & 51.381072 & TAS [1, 2] \\
MIGS-4.2 & Longitude & -2.359619 & TAS [1, 2] \\
MIGS-4.4 & Altitude & 31 m & TAS [1, 2] \\
\end{tabular}
\caption{Classification and general features of *Thermithiobacillus tepidarius* DSM 3134\textsuperscript{T} according to MIGS recommendations [8].}
\end{table}

\textsuperscript{a}Evidence codes - IDA Inferred from direct assay, 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 anecdotal evidence). These evidence codes are from the Gene Ontology project [28, 29].
Allpaths–LG (PrepareAllpathsInputs: PHRED_64 = 1 PLOIDY = 1 FRAG_COVERAGE = 125 JUMP_COVERAGE = 25 LONG_JUMP_COV = 50, RunAllpathsLG: THREADS = 8 RUN = std_shredpairs TARGETS = standard VAPI_WARN_ONLY = True OVERWRITE = True). The final draft assembly contained 44 contigs in 43 scaffolds. The total size of the genome is 2.96 Mbp and the final assembly is based on 3,44.8 Mbp of Illumina data, which provides an average 116.4× coverage of the genome.

Figure 1: Maximum-likelihood phylogenetic tree based on CLUSTALW alignment of 16S rRNA gene sequences of the Acidithiobacillia. Type strains of each species of Acidithiobacillus are used, along with that of Thermithiobacillus (emboldened). Thermithiobacillus sp. NCIMB 8349 (the only other Thermithiobacillus sp. in culture) is given for the sake of completeness. Sequences pertaining to organisms for which a publically available genome sequence exists are underlined. Accession numbers for the GenBank database are in parentheses. Alignment and tree were constructed in MEGA 6 [30] using 1,509 positions and pairwise deletion. Tree was drawn using the Tamura-Nei model for maximum-likelihood trees [31]. Values at nodes are based on 5,000 bootstrap replicates. Scale-bar indicates 2 substitutions per 100.

Table 2: Project information

| MIGS ID | Property                     | Term                                      |
|---------|------------------------------|-------------------------------------------|
| MIGS 31 | Finishing quality           | Improved High-Quality Draft              |
| MIGS-28 | Libraries used              | Illumina Standard PE                     |
| MIGS 29 | Sequencing platforms        | Illumina HiSeq 2000/2500                  |
| MIGS 32 | Fold coverage               | 116.4                                    |
| MIGS 30 | Assemblers                  | Allpaths/Velvet                          |
| MIGS 31.2 | Gene calling method       | NCBI Prokaryotic Genome Annotation Pipeline |
| Locus Tag |                             | G579DRAFT                                |
| Genbank ID |                             | AUIS0100000000                           |
| GenBank Date of Release |                             | August 15, 2015                          |
| GOLD ID |                             | GA002306                                  |
| BIOPROJECT |                             | PRJNA185671                              |
| MIGS 13 | Source Material Identifier | DSM 3134†                                |
| Project relevance |                     | GEBAS-KMG                                |

Figure 2: Transmission electron micrograph of T. tepidarius from a thiosulfate-limited chemostat (20 mM, 0.15 h⁻¹). Cells were obtained from a chemostat-culture at steady-state by centrifugation and were washed and suspended in sterile 150 mM sodium chloride solution and applied to Formvar® and carbon coated copper grid before washing with further saline and staining in 50 mM uranyl acetate for 5 mins and washing again. Stained grids were visualized in a JEOL JEM-1400Plus transmission electron microscope, operating at 120 kV.
Genome annotation
Genes were identified using Prodigal [18] as part of the DOE-JGI genome annotation pipeline [19]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information non-redundant database, UniProt, TIGR-Fam, Pfam, KEGG, COG, and InterPro database. These data sources were combined to assert a product description for each predicted protein. tRNA ScanSE was used to find tRNA genes and rRNA genes were found using searches against models of the ribosomal RNA genes built from SIVLA [20, 21]. Additional gene prediction analysis and functional annotation was performed within the IMG-ER platform [22, 23]. For each gene discussed in this publication, the annotation was manually checked against the GenBank database using the BLASTn and BLASTp algorithms - both of the gene from T. tepidarius and using the equivalent gene from members of the Acidithiobacillia or Escherichia coli.

Genome properties
The genome of T. tepidarius DSM 3134T is 2,958,498 bp-long with a 66.8 mol% G + C content (Table 3). Of the 2,968 predicted genes, 2,902 were protein-coding genes and 66 were RNA genes, including 2 rRNA operons. A total of 2,348 genes (79.1 %) were assigned a putative function. A total of 3.4 % were identified as pseudogenes – the remainder annotated as hypothetical proteins. The properties and the statistics of the genome are given in Table 3. The distribution of genes into COGs functional categories is presented in Table 4. The genome is one of the smaller genomes of those sequenced thus far from chemolithoautotrophic Proteobacteria (Table 5).

Insights from the genome sequence
As an obligate autotroph, it would be anticipated that genes encoding a complete Calvin-Benson-Bassham cycle and, in lieu of Krebs cycle, a biosynthetic horseshoe [24] would be present. A complete CBB cycle is present, and owing to the presence of a transaldolase (EC 2.2.1.2) and absence of a sedoheptulose-1,7-bisphosphatase (EC 3.1.3.37) gene, we can conclude that it is a transaldolase-variant CBB cycle [25]. Of Krebs’ cycle genes, citrate synthase (EC 2.3.3.16), aconitase (EC 4.2.1.3), isocitrate dehydrogenase (NADP⁺, EC 1.1.1.42), succinyl coenzyme A synthase (ADP-forming, EC 2.6.1.5) and malate dehydrogenase (oxaloacetate decarboxylating, NADP⁺, EC 1.1.1.40) were present. No fumarase or succinate dehydrogenase genes could be identified. The E1 subunit of α-ketoglutarate dehydrogenase was missing and the closest BLASTp match to the E2 subunit is annotated as a pyruvate dehydrogenase. These lesions are consistent with other obligate autotrophs and confirm the presence of a biosynthetic horseshoe in T. tepidarius [24].
In terms of respiration, 2 cytochrome c oxidases (cbb3 EC 1.9.3.1) and 2 ubiquinol oxidases (bd, EC 1.10.3.10) could be identified, which is consistent with previous physiological studies [7]. Three cytochromes b563 and three cytochromes c553 were identified, along with other cytochromes c, again constant with previous studies [7].

Extended insights

Two pairs of genes encoding ribulose-1,5-bisphosphate carboxylase (RuBisCO) could be identified, each comprising a large and small subunit gene. One pair is found close to cbbO and cbbQ genes, with no other cbb genes close by — this is consistent with Acidithiobacillus spp. and other obligate chemolithoautotrophs and indicates a Form IAq RuBisCO. The other pair is found close to cbb genes and in that sense is perhaps more similar to Form II RuBisCO [26]. Metalloresistance genes including those for arsenite efflux and arsenate reductase (arsB and arsC, respectively) were identified along with those implicated in tellurite, cadmium, cobalt, zinc, copper and silver resistance. Sulfur-oxidation genes are obviously of paramount interest in an obligate chemolithoautotroph, however, a number of proposed enzymes of sulfur metabolism have no genes identified thus far. It is known that the Acidithio-

bacillia [1, 2, 4–6] do not use the Kelly-Friedrich or “Sox” pathway of thiosulfate oxidation, and instead oxidise thiosulfate to tetrahionate via a poorly understood dehydrogenase — more than one form of which may exist. Some Kelly-Friedrich pathway genes are present in the genome and these are given in Fig. 3, showing comparison with those from other organisms that do not use the Kelly-Friedrich pathway versus one (Paracoccus denitrificans) that does. It can be seen from Fig. 3 that the non-Kelly-Friedrich organisms lack the soxC and soxD genes that are involved in a 6-electron capture during thiosulfate oxidation and all contain a gene encoding DUF302-family protein of unknown function 191 amino acids in length (G579DRAFT_01426 in T. tepidarius). Assuming these proteins are found in the periplasm of T. tepidarius as they are in Paracoccus spp., they could play a role in trithionate or higher polythionate oxidation (tetrathionate being oxidized solely in the cytoplasm [6]. The DUF302 protein of T. tepidarius would have a mass of 20.6 kDa based on the amino acyl sequence but contains a potential dimerization domain, so could be 41.2 kDa. It is worth noting that the periplasmic trithionate hydrolase (EC 3.12.1.1, gene unknown) of Acidiphilium acidophilum was 35 kDa [27].

One hundred seventy eight genes (5.9 % of genome) were flagged as potentially horizontally transferred from the species Thiobacillus thioparus, Thiobacillus denitrificans and Sulfuricella denitrificans in the Hydrogenophi-

laceae. This is particularly interesting since Thiobacillus aquaesulis DSM 4255T (= ATCC 43788T, no genome available) is closely related to these 3 species and was isolated originally from the Roman Baths and thus inhabits the exact same location [28]. A further 55 genes (1.9 %) were potentially transferred from Methyllococcus capsulatus, a strain of which (Bath = NCIMB 11132) was also isolated from the Roman Baths [25]. There is no clear pattern in the proteins encoded by the genes marked as potentially transferred.

Conclusions

The genome of T. tepidarius DSM 3134T is the first for this genus and one of very few available for the Class Acidithiobacillia. The genome gives evidence and insight into the carbon dioxide fixation pathway, biosynthesis and sulfur oxidation as well as metal resistance and potential gene transfer from other species also isolated from the Roman Baths from which this organism was obtained. These data confirm that a transaldolase variant of the Calvin-Benson-Bassham cycle is used for carbon dioxide fixation. Sulfur oxidation genes of the sox operon are present but soxC and soxD are missing, though a DUF302-family protein was present — and also found across obligate chemolithoautotrophs in the Proteobacter-

ia that use the Kelly-Trudinger (aka S4I) pathway of sulfur oxidation, rather than the Kelly-Friedrich (aka Sox) pathway. This genome sequence has already been utilized to propose the Class Acidithiobacillia [4] for Thermithio-
bacillus and Acidithiobacillus and to determine their
evolutionary relationship with the *Gammaproteobacteria*. Thus far, the type species of each genus of the *Acidithiobacillia* is now sequenced, along with several other *Acidithiobacillus* spp. and other obligate chemolithoautotrophic Bacteria such as *Thiobacillus* spp. and *Halothiobacillus* spp. (Table 5), of these, *T. tepidarius* DSM 3134^T^ has one of the smaller genomes, presumably because it lacks the salt-tolerance systems of *Halothiobacillus* spp. or the iron-oxidation or acid-tolerance of *Acidithiobacillus* spp. This genome sequence will enable further evolutionary studies into the nature of the *Acidithiobacillia* and chemolithoautotrophs in general, along with ecological studies including organism-organism interactions in the environment owing to the evidence for horizontal gene transfer evident in this genome.

**Abbreviations**

KMG: 1,000 microbial genomes; S4I: Tetrathionate intermediate pathway (aka Kelly-Trudinger pathway); Sox: Sulfur oxidation pathway (aka Kelly-Friedrich pathway)

**Acknowledgements**

We acknowledge Dr. Hans-Peter Klenk at the DSMZ for the provision of genomic DNA for sequencing and thank Dr Ann P. Wood (formerly of King’s College London, UK) for access to unpublished data pertaining to *T. tepidarius* DSM 3134^T^ and its isolation.

**Funding**

The sequencing and annotation was performed under the auspices of the United States Department of Energy Joint Genome Institute (JGI), a DOE Office of Science User Facility and is supported by the Office of Science of the United States Department of Energy under Contract Number DE-AC02-05CH11231. The authors wish to acknowledge the School of Biological Sciences, University of Plymouth, United Kingdom for Ph.D. studentship funding to LH and the Royal Society for a research grant (RG120444) to RB that supported the analysis of the genome.

**Authors’ contributions**

RB and LH analyzed and mined the genome data in public databases for genes of interest and performed BLASTn/BLASTp searches to verify and validate the annotation etc and made comparisons of the sulfur oxidation operons with those in other organisms. RB constructed the phylogenetic tree. LH grew the organism and performed analyses thereof, and performed electron microscopy at the Electron Microscopy Centre, University of Plymouth. All other authors contributed to the sequencing, assembly and
annotation of the genome sequence. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Author details
1 School of Biological Sciences, University of Plymouth, Drake Circus, Plymouth PL4 8AA, UK. 2 Sustainable Earth Institute, University of Plymouth, Drake Circus, Plymouth PL4 8AA, UK. 3 DOE Joint Genome Institute, Walnut Creek CA 94598, USA.

Received: 4 June 2016 Accepted: 31 August 2016

References

1. Wood AP, Kelly DP. Physiological characteristics of a new thermophilic obligately chemolithotrophic Thiobacillus species, *Thiobacillus tepidarius*. Int J Syst Bact. 1985;35:434–7.

2. Wood AP, Kelly DP. Chemolithotrophic metabolism of the newly-isolated moderately thermophilic, obligately autotrophic *Thiobacillus tepidarius*. Arch Microbiol. 1986;147:1–7.

3. Kelly DP, Wood AP. Reclassification of some species of *Thiobacillus* to the newly designated genera *Acidithiobacillus* gen. nov., *Halothiobacillus* gen. nov. and *Thermothiobacillus* gen. nov. Int J Syst Evol Microbiol. 2000;50:511–6.

4. Williams KP, Kelly DP. Proposal for a new class within the phylum *Proteobacteria*, *Acidithiobacillus* classis nov., with the type order *Acidithiobacillales*, and emended description of the class *Gammaproteobacteria*. Int J Syst Evol Microbiol. 2013;63:2901–6.

5. Boden R, Cleland D, Green PN, Katayama Y, Uchino Y, Murrell JC, Kelly DP. Phylogenetic assessment of culture collection strains of *Thiobacillus thioparus*, and definitive 16S rRNA gene sequences for *T. thioparus*, *T. denitrificans*, and *Haloferax mediterraneus*. Arch Microbiol. 2012;194:187–95.

6. Lu WP, Kelly DP. Cellular location and partial purification of the 'thiosulphate-oxidizing enzyme' and 'trithionate hydrolase' from *Thiobacillus tepidarius*. J Gen Microbiol. 1988;134:877–85.

7. Kelly DP, Lu WP, Poole RK. Cytochromes of *Thiobacillus tepidarius* and the respiratory chain involved in the oxidation of thiosulfate and tetrathionate. Arch Microbiol. 1993;160:87–95.

8. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, et al. The minimum information about a genome sequence (MIGS) specification. Nucleic Acids Res. 2008;36:541–7.

9. Fredericq E, Oth A, Fontaine F. The ultraviolet spectrum of deoxyribonucleic acid and its constituents. J Mol Biol. 1963;11:1–7.

10. Pagani L, Liolios K, Jansson J, Chen IM, Smirnova T, Nosrat B, et al. The Genomes OnLine Database (GOLD) v. 4: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res. 2012;40:D571–3.

11. Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Pillay M, et al. IMG 4 version of the integrated microbial genomes comparative analysis system. Nucleic Acids Res. 2014;42:D560–7.

12. Mavromatis K, Land ML, Brettin TS, Quest DJ, Copeland A, Mavromatis K, et al. The fast changing landscape of sequencing technologies and their impact on microbial genome assemblies and annotation. PLoS One. 2012;7, e48837.

13. Bennett S. Solexa Ltd. Pharmacogenomics. 2004;5:433–8.

14. JGI: Joint genome institute [http://www.jgi.doe.gov]

15. Zerbino D, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18:821–9.

16. Reads simulator wgsim [https://github.com/lh3/wgsim]

17. Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BL, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. P Natl A Sci. 2011;108:1513–8.

18. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119.

19. Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Palaniappan K, Szeto E, Pillay M, Chen I-MA, Pati A, Nielsen T, Markowitz VM, Kyrpides NC. The Standard Operating Procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v4). Stand Genom Sci. 2015;10:86.

20. Lowe TM, Eddy SR. RNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25:955–64.

21. Puersse E, Quast C, Knittell FB, Ludwig W, Peplies J, Glickner FO. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nuc Acids Res. 2007;35:2188–7196.

22. Markowitz VM, Mavromatis K, Ivanova NN, Chen I-MA, Chu K, Kyrpides NC. IMG ER: a system for microbial genome annotation expert review and curation. Bioinformatics. 2009;25:2271–8.

23. Chen I-MA, Markowitz VM, Chu K, Anderson I, Mavromatis K, Kyrpides NC, Ivanova NN. Improving microbial genome annotations in an integrated database context. PLoS One. 2013;8, e54859.

24. Wood AP, Aurikko JP, Kelly DP. A challenge for 21st century molecular biology and biochemistry: what are the causes of obligate autotrophy and methanotrophy. EEMS Microbiol Rev. 2004;28:335–52.

25. Anthony C. The biochemistry of methylotrophs. 1st ed. London: Academic Press; 1982.

26. Badger MR, Bek EJ. Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO2 acquisition by the CBB cycle. J Exp Bot. 2008;59:1525–41.

27. Meulenberg R, Pronk JT, Frank J, Hazeu W, Bos P, Kuenen JG. Purification and partial characterization of a thermostable trithionate hydrolase from the acidophilic sulphur oxidizer *Thiobacillus acidophilus*. Eur J Biochem. 1992;209:367–74.

28. Wood AP, Kelly DP. Isolation and characterisation of *Thiobacillus aquaesulis* sp. nov., a novel facultatively autotrophic moderate thermophile. Arch Microbiol. 1988;149:339–43.

29. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium Nat Genet. 2000;25:25–9.

30. Guide to GO evidence codes [http://www.geneontology.org/GO.evidence.shtml]

31. Tamura K, Stecher G, Peterson D, Filipski A, N银川 S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013;30:2725–19.

32. Tamura K, N银川 M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 1993;10:512–6.

33. Kelly DP. Oxidation of sulphur compounds. In: Cole JA, Ferguson SJ, editors. The nitrogen and sulphur cycles. Cambridge: Cambridge University Press; 1988. p. 65–98.

34. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A. 1990;87:4576 [PubMed].

35. Garrity GM, Bell JA, Lilburn T, Phylum XIV. *Proteobacteria* phylog. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Bergey's Manual of Systematic Bacteriology, vol. 2. Second ed. New York: Springer; 2005. p. 1. Part B.