

EXTENDED GENOME REPORT

Open Access



# Complete genome sequence of the haloalkaliphilic, obligately chemolithoautotrophic thiosulfate and sulfide-oxidizing $\gamma$ -proteobacterium *Thioalkalimicrobium cyclicum* type strain ALM 1 (DSM 14477<sup>T</sup>)

Ulrike Kappler<sup>1\*</sup>, Karen Davenport<sup>2</sup>, Scott Beatson<sup>1</sup>, Alla Lapidus<sup>3</sup>, Chongle Pan<sup>4,5</sup>, Cliff Han<sup>2,4</sup>, Maria del Carmen Montero-Calasanz<sup>6</sup>, Miriam Land<sup>4,5</sup>, Loren Hauser<sup>4,5</sup>, Manfred Rohde<sup>7</sup>, Markus Göker<sup>8</sup>, Natalia Ivanova<sup>4</sup>, Tanja Woyke<sup>4</sup>, Hans-Peter Klenk<sup>6</sup> and Nikos C. Kyrpides<sup>4,9</sup>

## Abstract

*Thioalkalimicrobium cyclicum* Sorokin et al. 2002 is a member of the family *Piscirickettsiaceae* in the order *Thiotrichales*. The  $\gamma$ -proteobacterium belongs to the colourless sulfur-oxidizing bacteria isolated from saline soda lakes with stable alkaline pH, such as Lake Mono (California) and Soap Lake (Washington State). Strain ALM 1<sup>T</sup> is characterized by its adaptation to life in the oxic/anoxic interface towards the less saline aerobic waters (mixolimnion) of the stable stratified alkaline salt lakes. Strain ALM 1<sup>T</sup> is the first representative of the genus *Thioalkalimicrobium* whose genome sequence has been deciphered and the fourth genome sequence of a type strain of the *Piscirickettsiaceae* to be published. The 1,932,455 bp long chromosome with its 1,684 protein-coding and 50 RNA genes was sequenced as part of the DOE Joint Genome Institute Community Sequencing Program (CSP) 2008.

**Keywords:** Aerobic, Obligate chemolithoautotroph, Sulfur oxidizer, Gram-negative, Mono Lake, *Piscirickettsiaceae*, CSP 2008

**Abbreviations:** CSP: Community Sequencing Program; DSM: Deutsche Sammlung von Mikroorganismen; DSMZ: Deutsche Sammlung von Mikroorganismen und Zellkulturen (German Collection of Microorganisms and Cell Cultures); DSR: dissimilatory sulfite reductase; HSP: high-scoring segment pair; IMG-ER: integrated microbial genomes - expert review; IUMS: international union of microbiological societies; JCM: Japan collection of microorganisms; ML: maximum likelihood; PEP: phosphoenolpyruvate; SOX: sulfur oxidation; SQR: Sulfide: quinone reductase.

\* Correspondence: u.kappler@uq.edu.au

<sup>1</sup>School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Australia

Full list of author information is available at the end of the article

## Introduction

Strain ALM 1<sup>T</sup> (= DSM 14477 = JCM 11371) is the type strain of the species *Thioalkalimicrobium cyclicum* [1], one of four species in the genus *Thioalkalimicrobium* [2]. The most prominent feature of *T. cyclicum* is its ability to live chemolithoautotrophically in the aerobic surface waters of a mixolimnion lake. Cultures of strain ALM 1<sup>T</sup> were first isolated from Mono Lake water samples taken from the sulfide-oxygen interface layer at a depth of 19–25 m [1]. The species epithet for the organism was derived from the Latin adjective *cyc.li'cum*, *cyc*lus, pertaining to the circle-like shape of the cells. For a short time after the initial description of the organism it was known as “*Thialkalimicrobium cyclicum*” until the Judicial Commission of the International Committee on Systematics of Prokaryotes restored the correct genus name at the X<sup>th</sup> International IUMS Congress of Bacteriology and Applied Microbiology in Paris (France) [3]. Here we present a summary classification and a set of features for *T. cyclicum* ALM 1<sup>T</sup> (DSM 14477<sup>T</sup>), together with the description of the genomic sequencing and annotation of the genome. Sequencing was done within the DOE JGI CSP 2008 for analysis of three type strains of alkaliphilic sulfur oxidizers.

## Organism information

### Classification and features

A representative genomic 16S rDNA sequence of *T. cyclicum* ALM 1<sup>T</sup> was compared using NCBI BLAST [4] under default settings (e.g., considering only the HSPs from the best 250 hits) with the most recent release of the Greengenes database [5] and the relative frequencies of taxa and keywords (reduced to their stem [6]) were determined, weighted by BLAST scores. The most frequently occurring genera were *Thiomicrospira* (74.7), *Thioalkalimicrobium* (11.2), ‘*Thialkalimicrobium*’ (8.4), *Hydrogenovibrio* (3.8) and ‘*Thiovibrio*’ (1.9 %) (49 hits in total). Regarding the single hit to sequences from members of the species, the average identity within HSPs was 98.7 %, whereas the average coverage by HSPs was 96.4 %. Regarding the single hit to sequences from other members of the genus, the average identity within HSPs was 98.5 %, whereas the average coverage by HSPs was 92.6 %. Among all other species, the one yielding the highest score was ‘*Thialkalimicrobium sibiricum*’ (AF126549), which corresponded to an identity of 98.6 and an HSP coverage of 96.3 %. (Note that the Greengenes database uses the INSDC (= EMBL/NCBI/DDBJ) annotation, which is not an authoritative source for nomenclature or classification, inverted commas indicate species names that are not approved.) The highest-scoring environmental sequence was DQ900619 (Greengenes short name ‘Sulfur-oxidizing Soap Lake (Washington State) meromictic haloalkaline unprecedented sulfide content lake water

isolate ASL1ASL1 str. ASL1’ where ‘meromictic’ denotes a lake with separate, oxic and anoxic waterzones that do not intermix), which showed an identity of 99.7 % and an HSP coverage of 89.3 %. Environmental samples which yielded hits of a higher score than the highest scoring species were not found.

Figure 1 shows the phylogenetic neighborhood of *T. cyclicum* in a 16S rRNA based tree. The sequences of the two identical 16S rRNA gene copies in the genome differ by one nucleotide and a nine bp long gap from the previously published 16S rRNA sequence (AF329082), which contained nine ambiguous base calls.

The paraphyletic structure of the genus *Thiomicrospira* in Fig. 1 and the location of *Hydrogenovibrio marinus* and *Galenea microaerophila* within *Thiomicrospira* might indicate the need for genome sequence-based reclassifications once enough reference sequences become available.

Cells of *T. cyclicum* ALM 1<sup>T</sup> are non-motile, Gram-negative staining, irregular spheres often in the form of open rings with a diameter of 0.5–0.8 μm and a cell width of 0.3–0.4 μm (Table 1 and Fig. 2) [1]. Carboxysome-like structures were frequently observed (see [1]). Colonies of strain ALM 1<sup>T</sup> are reddish, transparent with a diameter up to 3 mm [1]. Cells oxidize thiosulfate and sulfide but grow less actively on polysulfide and tetrathionate [1]. The pH range for growth is 6.5 to 11 (optimum 9.5) with a moderate salt concentration (about 0.6 M NaCl) [1].

### Chemotaxonomic data

The original description of strain ALM 1<sup>T</sup> [1] did not provide any chemotaxonomic information. No new chemotaxonomical data were generated for this report.

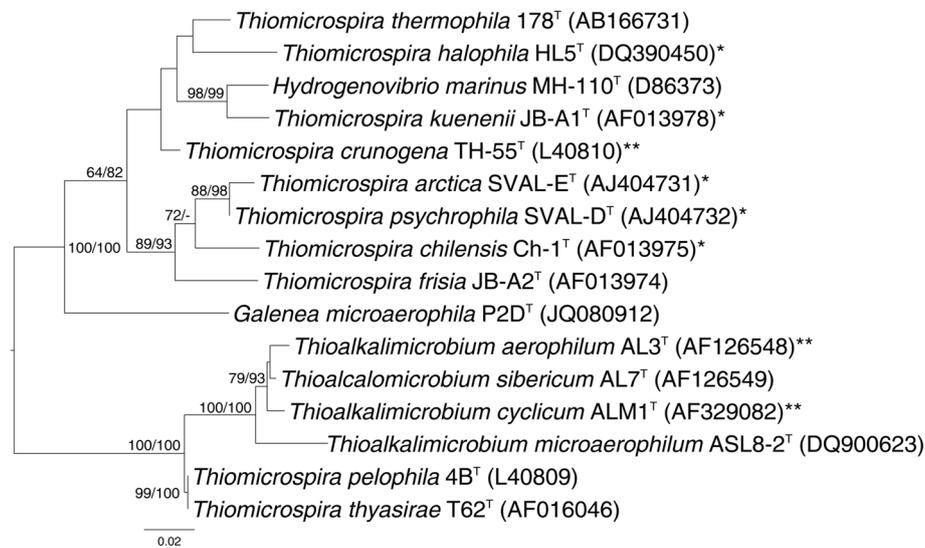
## Genome sequencing information

### Genome project history

This organism was selected for sequencing as part of the DOE JGI CSP 2008. The genome project is deposited in the Genomes On Line Database [7] and the complete genome sequence is deposited in GenBank. Sequencing, finishing and annotation were performed by the DOE JGI using state of the art sequencing technology [8]. A summary of the project information is shown in Table 2.

### Growth conditions and genomic DNA preparation

Strain ALM 1<sup>T</sup> was grown from a culture of DSM 14477<sup>T</sup> in DSMZ medium 925 at 28 °C. gDNA was purified using the Genomic-tip 100 System (Qiagen) following the directions provided by the supplier. The purity, quality and size of the bulk gDNA preparation were assessed by JGI according to DOE-JGI guidelines which included electrophoretic separation of samples and comparison against standards of known molecular masses, analysis of UV absorption spectra and sequencing of the 16S rDNA.



**Fig. 1** Phylogenetic tree highlighting the position of *T. cyclicum* within the family *Piscirickettsiaceae*. *T. cyclicum* is shown relative to the type strains of *Thioalkalimicrobium* and all species from the three closest related genera (*Thiomicrospira*, *Hydrogenovibrio* and *Galenea*) within the family. The tree was inferred from 1,530 aligned characters [58, 59] of the 16S rRNA gene sequence under the ML criterion [60]. The rooting shown was inferred by the midpoint-rooting method [36]. The branches are scaled in terms of the expected number of substitutions per site (bar). Numbers adjacent to the branches are support values from 1,000 ML bootstrap replicates [61] (left) and from 1,000 maximum-parsimony bootstrap replicates [62] (right) if larger than 60 %. Lineages with type strain genome sequencing projects registered in GOLD [7] are labeled with one asterisk, those also listed as 'Complete and Published' with two asterisks (see [37] and AGFA00000000 for *T. aerophilum*)

### Genome sequencing and assembly

The genome was sequenced using a combination of Illumina and 454 sequencing platforms. All general aspects of library construction and sequencing can be found at the JGI website [64]. Pyrosequencing reads were assembled using the Newbler assembler [9]. The initial Newbler assembly consisted of 15 contigs in one scaffold and the consensus contigs were computationally shredded to form 2 kb overlapping reads. Illumina GAii sequencing data (3,091 Mb) were assembled with Velvet [10] and the consensus sequences were computationally shredded into 1.5 kb overlapping reads. The computational shreds from both assemblies were assembled together with the 454 long-insert paired end reads using phrap [11, 12]. The 454 draft assembly was based on 171.4 Mb 454 draft data and all of the 454 paired end data. The Phred/Phrap/Consed software package [11–13] was used for sequence assembly and quality assessment in the subsequent finishing process. After the shotgun stage, reads were assembled with parallel phrap (High Performance Software, LLC). Possible mis-assemblies were corrected with gapResolution [14, 65], Dupfinisher [15], or sequencing cloned bridging PCR fragments with subcloning. Gaps between contigs were closed by editing in Consed, by PCR and by bubble PCR primer walks [16] (J.-F. Chang, unpublished). A total of 74 additional reactions and one shatter library were necessary to close gaps and to raise the quality of the final sequence. Illumina reads were also used to correct potential base errors and increase

consensus quality using a software Polisher developed at JGI [17]. The error rate of the final genome sequence is less than 1 in 100,000. Together, the combination of the Illumina and 454 sequencing platforms provided 1559.4 × coverage of the genome. The final assembly contained 216,642 pyrosequence and 38,029,488 Illumina reads.

### Genome annotation

Genes were identified using Prodigal [18] as part of the DOE-JGI [8] genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline [19]. The predicted CDSs were translated and used to search the NCBI non-redundant database, UniProt [20], TIGRFam [21], Pfam [22], PRIAM [23], KEGG [24], COG [25], and InterPro [26] databases. These data sources were combined to assert a product description for each predicted protein. Additional gene prediction analysis and functional annotation was performed within the IMG-ER platform [27].

### Genome properties

The genome consists of a circular 1,932,455 bp chromosome with 47 % G + C content (Table 3 and Fig. 3). Of the 1734 genes predicted, 1684 were protein-coding genes, and 50 RNAs; 19 pseudogenes were also identified. The majority of the protein-coding genes (78.5 %) were assigned a putative function while the remaining ones were annotated as hypothetical proteins. The distribution

**Table 1** Classification and general features of *T. cyclicum* ALM1<sup>T</sup> in accordance with the MIGS recommendations [48] (published by the Genome Standards Consortium [49]) and the NamesforLife database [50]

MIGS ID	Property	Term	Evidence code <sup>a</sup>
	Classification	Domain <i>Bacteria</i>	TAS [51]
		Phylum ' <i>Proteobacteria</i> '	TAS [52]
		Class <i>Gammaproteobacteria</i>	TAS [53, 54]
		Order <i>Thiotrichales</i>	TAS [54, 55]
		Family <i>Piscirickettsiaceae</i>	TAS [54, 56]
		Genus <i>Thioalkalimicrobium</i>	TAS [2]
		Species <i>Thioalkalimicrobium cyclicum</i> (Type) strain: ATM <sup>T</sup> (AF329082)	TAS [1]
	Gram stain	negative	TAS [1, 2]
	Cell shape	open ring-shaped	TAS [1]
	Motility	non-motile	TAS [1]
	Sporulation	not reported	
	Temperature range	mesophile, about 28 °C	NAS
	Optimum temperature	not reported	
	pH range; Optimum	not reported	
	Carbon source	CO <sub>2</sub>	NAS
MIGS-6	Habitat	water	TAS [1]
MIGS-6.3	Salinity	Moderate, 0.6 to 1.5 M NaCl	TAS [1]
MIGS-22	Oxygen requirement	aerobic	TAS [1]
MIGS-15	Biotic relationship	free-living	TAS [1]
MIGS-14	Pathogenicity	none	NAS
MIGS-4	Geographic location	Lake Mono (California)	TAS [1]
MIGS-5	Sample collection	1999	TAS [1]
MIGS-4.1	Latitude	38.012	TAS [1]
MIGS-4.2	Longitude	-118.976	TAS [1]
MIGS-4.4	Altitude	1926 m	TAS [1]

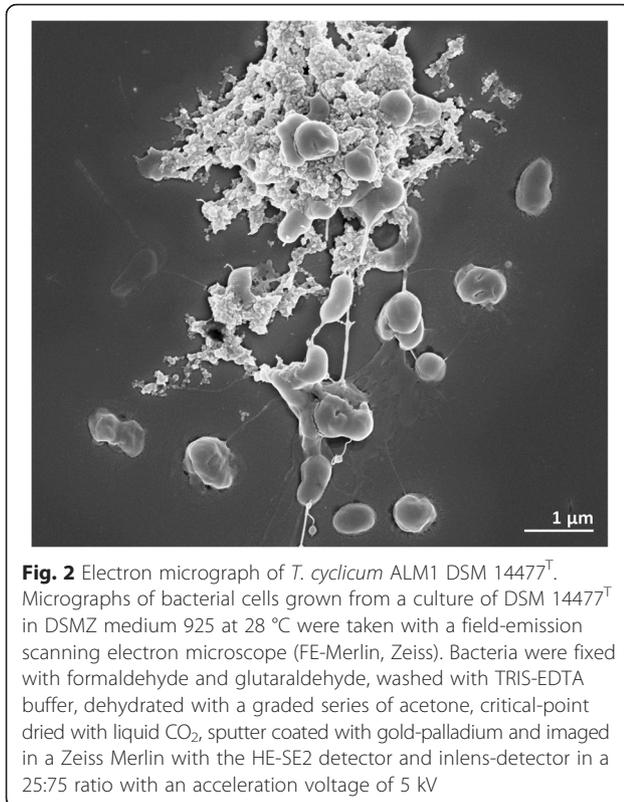
<sup>a</sup>Evidence codes - *IDA* Inferred from Direct Assay, *TAS* Traceable Author Statement (i.e., a direct report exists in the literature), *NAS* Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [57]

of genes into COGs functional categories is presented in Table 4.

### Insights from the genome sequence

*T. cyclicum* has been described as an obligate chemolithoautotroph, and its genome contains only 1684 protein encoding genes indicating a reduction in gene contents possibly in adaptation to this lifestyle. Reductions in genome size are a common feature in bacteria from many specialized ecological niches where relatively stable growth conditions are encountered. Examples include other free-living bacteria in the family *Piscirickettsiaceae*, *Thioalkalimicrobium* species, *Thiomicrospira crunogena* and related species, although in most cases the genome reduction is not as extreme as in this case and ~2000 proteins are present.

During chemolithoautotrophic growth, strain ALM 1<sup>T</sup> oxidizes reduced sulfur compounds such as thiosulfate or sulfide without the formation of sulfur as an intermediate, which suggests that it uses a Sox-type sulfur oxidation pathway [28] rather than a combination of DSR and Sox proteins which leads to the formation of elemental sulfur as an intermediate and seems to be common in other *Gammaproteobacteria*, such as the phototrophic purple sulfur bacteria [29]. The Sox sulfur oxidation pathway relies on four essential enzyme complexes, SoxAX, SoxYZ, SoxB, and SoxCD, to oxidize reduced sulfur compounds to sulfate without the formation of free intermediates [28, 30, 31], and all of these proteins are encoded in the *T. cyclicum* genome. In the reaction cycle of the Sox multienzyme complex the SoxAX cytochrome (encoded by Thicy\_216 & Thicy\_219) that catalyzes covalent attachment of reduced sulfur compounds such as thiosulfate to



the SoxYZ carrier protein (Thicy\_217, Thicy\_218), the manganese-containing SoxB (Thicy\_0833) protein then removes the fully oxidized sulfur residues from SoxYZ through hydrolysis, while the SoxCD sulfane dehydrogenase (Thicy\_50, Thicy\_51) a heterotetrameric complex of the molybdenum protein and a cytochrome, catalyzes a six electron oxidation of reduced sulfur residues bound to SoxYZ [32–36] (Fig. 4). All *T. cyclicum* Sox proteins

are most closely related to homologues found in *Thioalkalimicrobium aerophilum* (94–95 % amino acid identity) as well as *Thiomicrospira* sp. and *Hydrogenovibrio marinus*, which is in keeping with the phylogenetic position of this bacterium.

In *T. cyclicum* the *sox* genes encoding the essential components of the Sox multienzyme complex are distributed in three separate genomic gene loci (Fig. 4). This is similar to what has been seen in the related *Thiomicrospira* species [37], but differs from the situation in other *Proteobacteria*, e.g. the  $\alpha$ -*Proteobacterium* *Paracoccus pantotrophus*, where the genes encoding the core Sox enzymes as well as genes encoding accessory Sox proteins such as SoxV, W, H, S or the SoxF sulfide dehydrogenase are located in the same gene region and often even within only one or two major operons [28, 38].

In *T. cyclicum* ALM 1, most of genes encoding accessory Sox proteins appear to be absent, only a homologue of the SoxF flavocytochrome (Thicy\_0003, Thicy\_0004) (Fig. 4) and a gene encoding a homologue of the SoxH protein (Thicy\_0412), the exact function of which is unknown, were detected during our analyses of the genome. This leads to the question of how essential these accessory proteins are for the function of the Sox complex. *T. cyclicum*, unlike many of the sulfur oxidizing  $\alpha$ -*Proteobacteria*, is an obligate chemolithoautotroph and thus relies on the optimal function of this pathway for energy generation, and yet does not appear to rely on the accessory proteins to keep the core Sox enzymes functional.

Studies of the SoxAX cytochromes in various *Proteobacteria* have led to the realization that these proteins are extremely diverse. There are currently three recognized types of proteins that vary significantly in terms of the redox cofactors present as well as their subunit structure and specifically the sequences of the SoxX proteins

**Table 2** Project information

MIGS ID	Property	Term
MIGS 31	Finishing quality	Finished
MIGS-28	Libraries used	One 454 pyrosequence standard library, one 454 PE library (9 kb insert size), one Illumina library
MIGS 29	Sequencing platforms	Illumina GAii, 454 GS FLX Titanium
MIGS 31.2	Fold coverage	1,521.2 × Illumina, 38.2 × pyrosequence
MIGS 30	Assemblers	Newbler version 2.3, Velvet version 0.7.63, phrap version SPS – 4.24
MIGS 32	Gene calling method	Prodigal
	Locus Tag	Thicy
	Genbank ID	CP002776
	GenBank Date of Release	July 6, 2012
	GOLD ID	Gc01777
	BIOPROJECT	PRJNA52629
MIGS 13	Source Material Identifier	DSM 14477
	Project relevance	Biotechnological, Bioremediation

**Table 3** Genome statistics

Attribute	Value	% of Total
Genome size (bp)	1,932,455	100.0
DNA coding (bp)	1,818,441	94.1
DNA G + C (bp)	907,872	47.0
DNA scaffolds	1	
Total genes	1,734	100.0
Protein coding genes	1,684	97.1
RNA genes	50	2.9
Pseudo genes	19	1.1
Genes in internal clusters	543	31.3
Genes with function prediction	1,361	78.5
Genes assigned to COGs	1,459	84.1
Genes with Pfam domains	1,502	86.6
Genes with signal peptides	124	7.2
Genes with transmembrane helices	357	20.6
CRISPR repeats	0	

involved in the SoxAX complexes [30, 31]. In comparison to biochemically characterized SoxA proteins, the protein encoded by the *T. cyclicum* *soxA* gene is most closely related to the Type III SoxA protein from *Allochromatium vinosum* (33.9 % amino acid sequence identity, as opposed to 21.8 and 17.4 % identity with the Type I and Type II SoxA proteins from *P. pantotrophus* and *S. novella*) (Fig. 5). Type III SoxAX proteins normally contain three subunits, SoxA, SoxX and SoxK [39]. The low molecular weight SoxK protein is required to stabilize the complex of SoxA and SoxX. In *T. cyclicum*, however, no gene encoding a protein homologous to SoxK appears to be present, which indicates that there is even more diversity of SoxAX proteins than previously assumed. A similar situation was already described by Ogawa et al. [39] for the SoxAX protein from the related bacterium, *Thiomicrospira crunogena*, and has also been discussed in depth, including a phylogenetic analysis across all groups of SoxA related proteins in two recent reviews [30, 31] (Fig. 5).

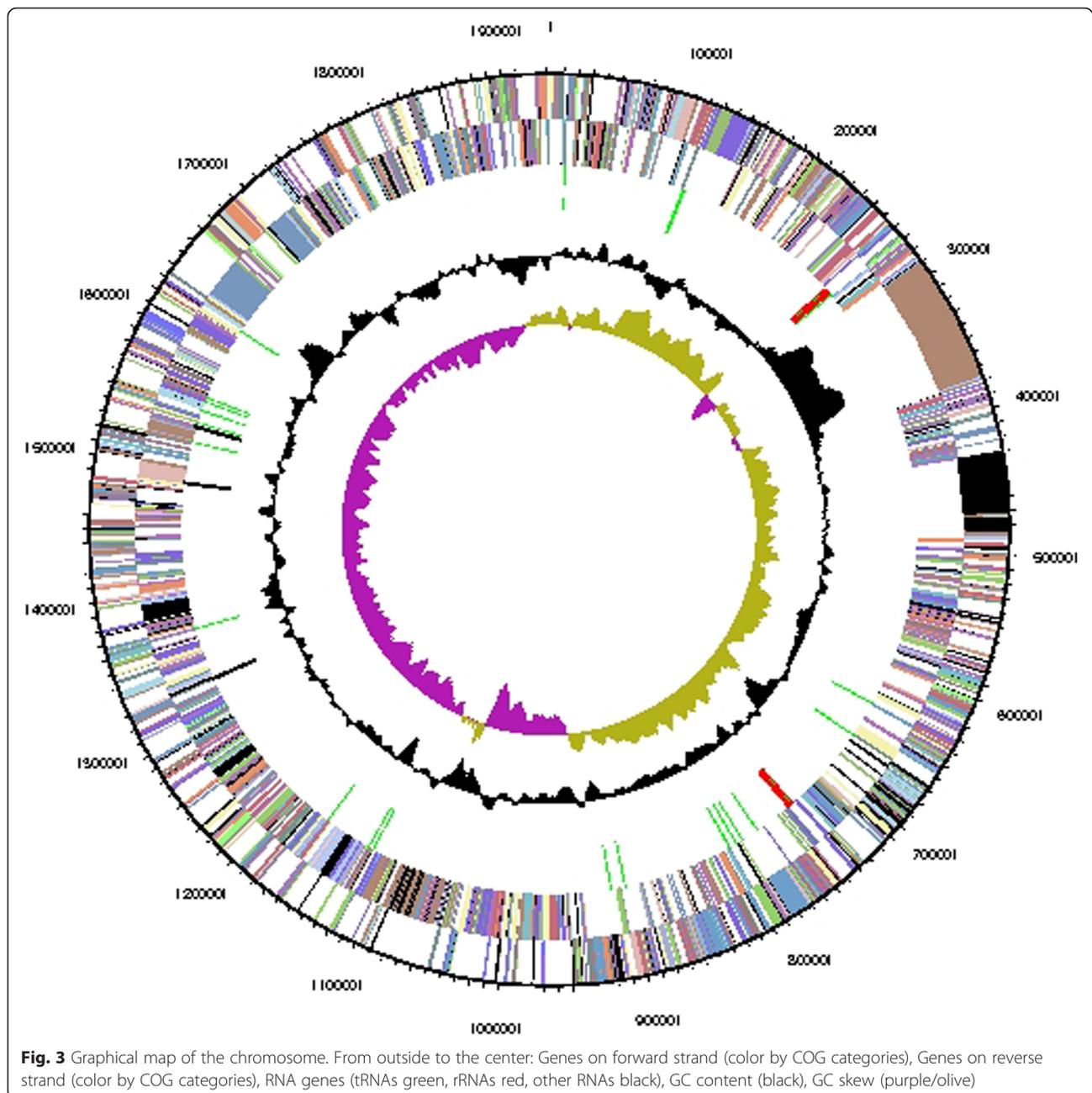
Although the Sox sulfur oxidation pathway has been recognized as a key pathway in microbial sulfur chemolithotrophy, issues still exist with the annotation of the various genes in automated annotation pipelines. For example, *soxB* genes are often annotated as encoding a '5' nucleotidase', which is correct as SoxB does belong to this larger group of enzymes, but at the same time creates confusion as to the actual nature of the encoded protein. Dedicated SoxB protein domains exist (e.g. cd07411, abbreviated as "MPP\_soxB\_N" or SoxB proteins with an N-terminal metallophosphatase domain), and recently a dedicated full length domain, TIGR04486 (thiosulf\_soxB) has been defined.

Another curiosity is the annotation of the SoxCD sulfane dehydrogenase (Thicy\_0050/0051) as a 'SO (sulfite oxidase) family protein' (which is correct), and then as a 'nitrate reductase (NADH)'. In as far as is currently known, the sulfite oxidase enzyme family only contains nitrate reductases from plants, and no prokaryotic nitrate reductases have ever been found in this enzyme family. Clearly, there is scope for improving the specificity of current COGs/cd patterns to avoid such obvious errors in the future, although the conserved domain cd\_02113 is diagnostic for SoxC proteins, regardless of their annotation.

We also analyzed the *T. cyclicum* genome for other proteins known to be involved in sulfur oxidation that are not part of the Sox multienzyme complex. Such proteins are frequently found in sulfur oxidizing bacteria and enhance their ability to use different sulfur compounds, including those that are not generally recognized as substrates of the Sox pathways (e.g. tetrathionate) or to cope with toxic sulfur compounds that can be byproducts of abiotic sulfur conversions (e.g. sulfite and sulfide converting enzymes). No genes encoding homologues of DSR, APS reductase (*aprABM*), tetrathionate hydrolase (*tth* gene) or sulfite dehydrogenases (*sorAB*) were identified. However, we did identify two genes (Thicy\_0064 & Thicy\_1132) that encode proteins with strong similarities to proteins annotated as SQRs in *T. crunogena* (Tcr\_1170, Tcr\_1381) [37].

The protein encoded by the putative SQR gene Thicy\_1132 is actually related to Ndh NADH dehydrogenase-type proteins, while Thicy\_0064 shows homology to 'HcaD uncharacterized FAD dependent dehydrogenases, COG0446'. Using the SQR classification system of [40], the two *T. cyclicum* SQRs could be classified as a periplasmic (34 aa Tat- leader peptide; 52 % conserved aa) SqrB type protein (Thicy\_0064), and a soluble, likely cytoplasmic SqrF-like protein (Thicy\_1132, 49 % conserved aa). Interestingly, the Thicy\_1132 encoded protein only has homology to the SqrF like proteins, while the Thicy\_0064 encoded protein exhibited significant homologies to all SQR types except SqrE and SqrF. It would thus appear that *T. cyclicum* contains two SQRs of different types, as representatives from both SqrB and SqrF groups have been enzymatically characterized.

Overall despite the fact that sulfur oxidation is a key element of *T. cyclicum* metabolism, the actual number of genes supporting this process is very small and shows very little redundancy or diversity. All genes encoding essential proteins of the Sox pathway are present as single copies, and genes encoding other enzymes known to support chemolithotrophic growth on sulfur compounds are absent. This is in contrast to other sulfur oxidizing bacteria such as the haloalkaliphilic *Thioalkalivibrio sulfidophilus* [41] which contains several copies of genes encoding SoxAX proteins and *Starkeya novella* where



two copies of SoxAX and SoxYZ encoding genes are present [42], as well as additional genes encoding sulfur converting enzymes that are not part of the Sox complex.

With chemolithoautotrophy being the major growth mode for *T. cyclicum*, we also investigated the carbon dioxide fixation pathways present in this bacterium. Of the known microbial pathways for carbon dioxide fixation only the Calvin Benson Bassham cycle was present, and carbon dioxide incorporation into phosphoenolpyruvate to form oxaloacetate, a required intermediate of the TCA cycle, was also identified using the KEGG pathway database [24].

Central carbon metabolism in *T. cyclicum* includes a complete set of genes encoding glycolysis and the pentose phosphate pathway as well as a pyruvate dehydrogenase enzyme complex and several routes by which pyruvate can be converted into oxaloacetate (PEP synthase, Thicy\_1283, EC 2.7.9.2; PEP carboxylase, Thicy\_1240, EC 4.1.1.31) or lactate (D-lactate dehydrogenase, Thicy\_1457, 1.1.1.28)

The TCA cycle of *T. cyclicum* is incomplete, with genes encoding the 2-oxoglutarate dehydrogenase or homologous enzymes (e.g. 2-oxoglutarate:ferredoxin oxidoreductase, KorAB) not having been identified in the genome.

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

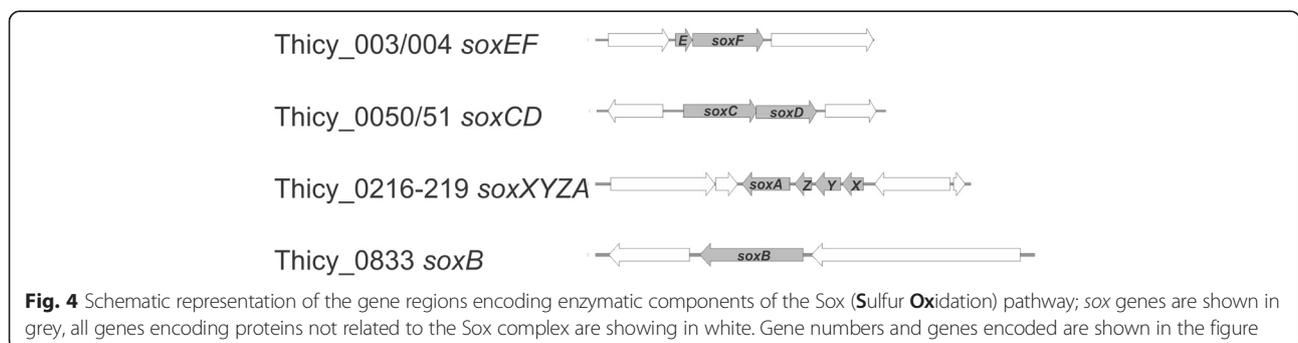
Code	Value	%age	Description
J	153	9.7	Translation, ribosomal structure and biogenesis
A	1	0.1	RNA processing and modification
K	71	4.5	Transcription
L	86	5.4	Replication, recombination and repair
B	0	0.0	Chromatin structure and dynamics
D	27	1.7	Cell cycle control, Cell division, chromosome partitioning
V	28	1.8	Defense mechanisms
T	48	3.0	Signal transduction mechanisms
M	117	7.4	Cell wall/membrane biogenesis
N	12	0.8	Cell motility
U	56	3.5	Intracellular trafficking and secretion
O	88	5.6	Posttranslational modification, protein turnover, chaperones
C	95	6.0	Energy production and conversion
G	59	3.7	Carbohydrate transport and metabolism
E	123	7.8	Amino acid transport and metabolism
F	49	3.1	Nucleotide transport and metabolism
H	109	6.9	Coenzyme transport and metabolism
I	40	2.5	Lipid transport and metabolism
P	89	5.6	Inorganic ion transport and metabolism
Q	26	1.6	Secondary metabolites biosynthesis, transport and catabolism
R	170	10.8	General function prediction only
S	135	8.5	Function unknown
-	275	15.9	Not in COGs

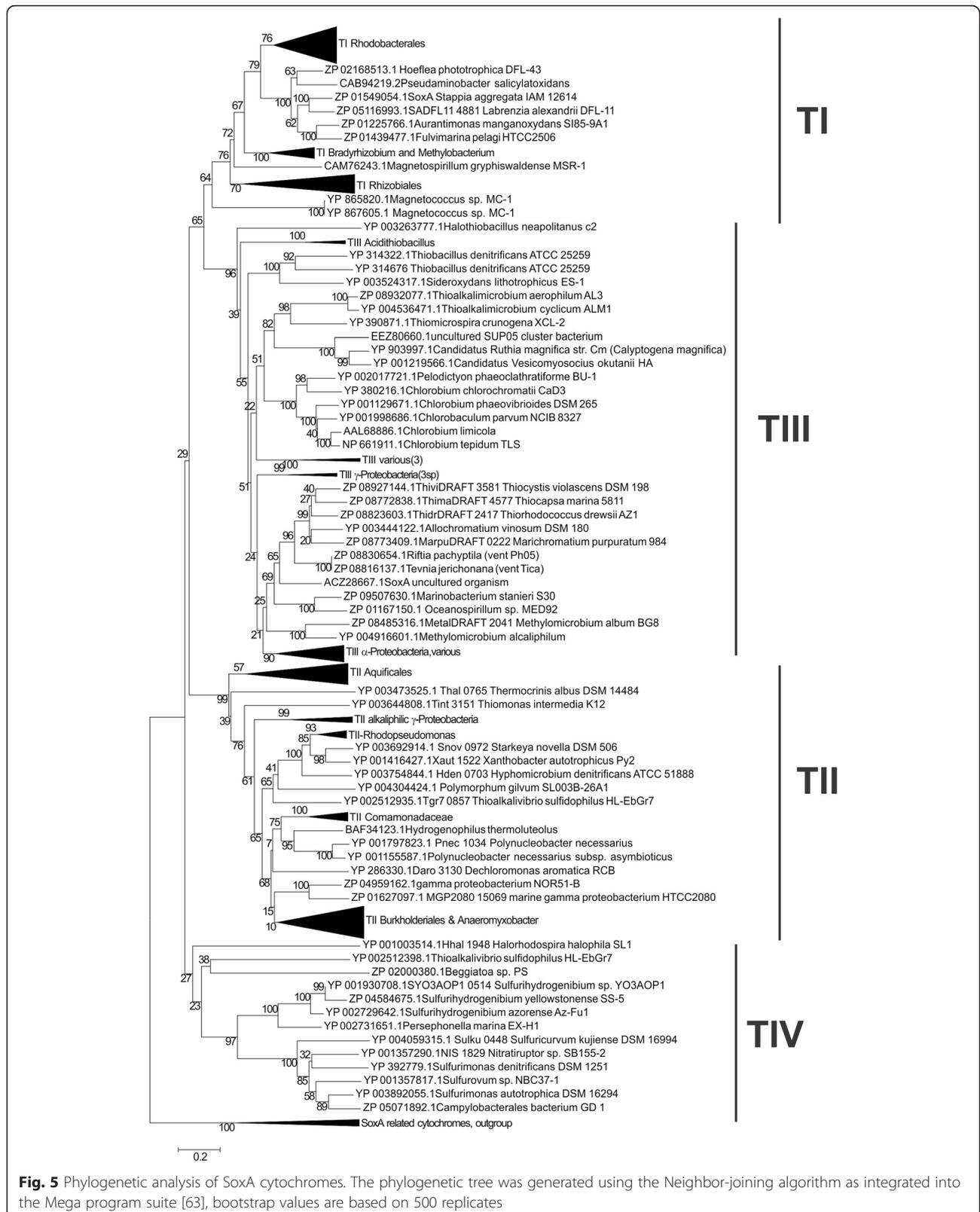
The total is based on the total number of protein coding genes in the genome

This indicates that in *T. cyclicum* the TCA cycle mainly serves biosynthetic purposes rather than being part of general energy generation, which is in keeping with the chemolithoautotrophic lifestyle of this bacterium, as sulfur oxidation by the Sox pathway or via SQRs will feed electron directly into the respiratory chain for energy generation.

The respiratory chain of *T. cyclicum* is of a very linear architecture, with only complex I being represented by three different types of NADH dehydrogenases. A multi-subunit (‘mitochondrial type’) NADH dehydrogenase

(EC 1.6.5.3) is encoded by the *nuo* gene cluster (Thicy\_0637–0650), while the other two are encoded by two genes (EC 1.6.99.1, Thicy\_1224–1225) and a single gene (EC 1.6.99.3, Thicy\_0083), respectively. Complex II/succinate dehydrogenase is encoded by genes Thicy\_0875–0878, while Complex III/cytochrome bc1-complex is encoded by Thicy\_0482–484. Only a single gene cluster encoding a cytochrome *c* oxidase appears to be present (Thicy\_1535–1529), which encodes a *cbb*<sub>3</sub>-type cytochrome oxidase. This type of cytochrome oxidase is known to have a high affinity for oxygen and





**Fig. 5** Phylogenetic analysis of SoxA cytochromes. The phylogenetic tree was generated using the Neighbor-joining algorithm as integrated into the Mega program suite [63], bootstrap values are based on 500 replicates

thus has been associated with microaerophilic growth conditions [43, 44], suggesting that in its natural environment *T. cyclicum* encounters medium to low oxygen tensions. In addition to this function, *cbb<sub>3</sub>*-type cytochrome oxidases have been implicated in affecting various regulatory processes in bacterial cells [45–47], including redox regulation and responses to environmental conditions, and it is possible that the enzyme from *T. cyclicum* also fulfills additional, regulatory functions. An F-type ATPase (Thicy\_1606–1612) completes the respiratory chain.

## Conclusions

With only about ~1700 encoded genes the genome of *T. cyclicum* ALM1<sup>T</sup> is relatively small compared to genomes from related sulfur oxidizing bacteria such as *Thiomicrospira* sp. or *Thioalkalivibrio* sp. which generally contain ~2000 or more protein encoding genes. The reduction in genome size becomes even more obvious in comparison to other sulfur chemolithoautotrophic bacteria (e.g. *Starkeya novella*, *Paracoccus* sp., *Rhizobiales* sp.) that often have more than 4000 encoded genes and also tend to encode redundant pathways. This again is likely to reflect the limited availability of substrates for energy generation in the organism's natural habitat, which is an extreme environment with high alkalinity and salinity.

Despite the reliance of *T. cyclicum* on autotrophy for acquiring cell carbon, only a single pathway for carbon dioxide fixation was found, and only the Sox pathway for sulfur oxidation and a few additional proteins that enable efficient use of sulfide as an energy source (SQRs and flavocytochromes) were identified. This is in keeping with a direct oxidation of sulfur substrate such as thiosulfate and sulfide to sulfate without intermediate formation of elemental sulfur which is a trait of the other major sulfur oxidation pathway that uses the DsrAB dissimilatory sulfite reductase. It is also supported by our observation on aerobic cultures of *T. cyclicum* supplement with thiosulfate as an energy source, which showed no sign of sulfur formation, which would have led to increased, optically apparent, turbidity of the culture during growth.

However, with about 20 % of genes having either unknown functions or not being assigned a COG category, there are clearly many things that can still be discovered regarding this organism. The apparent absence of accessory genes aiding in the maturation of the essential Sox sulfur oxidation enzymes is unusual, and should be further investigated, as should the effect of a high pH environment on the physical and catalytic properties of the periplasmic Sox proteins. *T. cyclicum* also prefers moderate salt concentrations, and it would be interesting to carry out comparative studies on compatible solutes and other adaptations between species of haloalkaliphilic sulfur oxidizers.

## Taxonomic and nomenclatural proposals

The difference in the reported G + C content of *T. cyclicum* (49.6 %) [1] to the one calculated from the genome sequence (47.0 %) calls for an emendation of the species description. The genome sequence-derived G + C content is also outside of the 48 to 51.2 % G + C range reported for the genus *Thioalkalimicrobium* [2].

## Emended description of the species

### *Thioalkalimicrobium cyclicum* Sorokin et al. 2002

The description of the species *Thioalkalimicrobium cyclicum* is the one given by Sorokin et al. 2002 [1], with the following modification. The G + C content, rounded to zero decimal places, is 47 %.

### Acknowledgements

The work conducted by the U.S. Department of Energy Joint Genome Institute was supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231, and a Fellowship and grant to UK (DP 0878525).

### Authors' contributions

UK, SB, M<sup>4</sup>CM-C, MG, HPK, and NCK drafted the manuscript. KD, AL, CP, CH, ML, LH, NI and TW sequenced, assembled and annotated the genome. MR provided the electron micrograph. All authors read and approved the final manuscript.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Australia. <sup>2</sup>Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA. <sup>3</sup>Centre for Algorithmic Biotechnology, St. Petersburg State University, St. Petersburg, Russia. <sup>4</sup>DOE Joint Genome Institute, Walnut Creek, California, USA. <sup>5</sup>Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA. <sup>6</sup>School of Biology, Newcastle University, Newcastle upon Tyne, UK. <sup>7</sup>Central Facility for Microscopy, HZI – Helmholtz Centre for Infection Research, Braunschweig, Germany. <sup>8</sup>Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany. <sup>9</sup>Department of Biological Sciences, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia.

Received: 23 October 2015 Accepted: 24 May 2016

Published online: 03 June 2016

### References

- Sorokin DY, Gorlenko VM, Tourova TP, Tsapin A, Nealon KH, Kuenen GJ. *Thioalkalimicrobium cyclicum* sp. nov. and *Thioalkalivibrio jannaschii* sp. nov., novel species of haloalkaliphilic, obligately chemolithoautotrophic sulfur-oxidizing bacteria from hypersaline alkaline Mono Lake (California). *Int J Syst Evol Microbiol.* 2002;52:913–20.
- Sorokin DY, Lysenko AM, Mityushina LL, Tourova TP, Jones BE, Rainey FA, Robertson LA, Kuenen GJ. *Thioalkalimicrobium aerophilum* gen. nov., sp. nov. and *Thioalkalimicrobium sibiricum* sp. nov., and *Thioalkalivibrio versutus* gen. nov., sp. nov., *Thioalkalivibrio nitratis* sp. nov. and *Thioalkalivibrio denitrificans* sp. nov., novel obligately alkaliphilic and obligately chemolithoautotrophic sulfur-oxidizing bacteria from soda lakes. *Int J Syst Evol Microbiol.* 2001;51:565–80.
- De Vos P, Truper HG, Tindall BJ. Judicial commission of the international committee on systematics of prokaryotes Xth international (IUMS) congress of bacteriology and applied microbiology. Minutes of the meetings, 28, 29 and 31 July and 1 August 2002, Paris, France. *Int J Syst Evol Micro.* 2005;55:525–35.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25:3389–402.

5. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Daley D, Hu P, Andersen GL. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol.* 2006;72(7):5069–72.
6. Porter MF. An algorithm for suffix stripping. *Program: electronic library and information systems.* 1980;14:130–7.
7. Pagani I, Liolios K, Jansson J, Chen I-MA, Smirnova T, Nosrat B, Markowitz VM, Kyrpides NC. The Genomes OnLine Database (GOLD) v. 4: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res.* 2012;40(D1):D571–9.
8. Mavromatis K, Land ML, Brettin TS, Quest DJ, Copeland A, Clum A, Goodwin L, Woyske T, Lapidus A, Klenk HP and others. The fast changing landscape of sequencing technologies and their impact on microbial genome assemblies and annotation. *PLoS One.* 2012;7(12):e48837.
9. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z and others. Genome sequencing in open microfabricated high density picoliter reactors. *Nature.* 2005;437(7057):376–80.
10. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 2008;18(5):821–9.
11. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces UsingPhred. I. Accuracy assessment. *Genome Res.* 1998;8(3):175–85.
12. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 1998;8(3):186–94.
13. Gordon D, Abajian C, Green P. Conseq: a graphical tool for sequence finishing. *Genome Res.* 1998;8(3):195–202.
14. LaButti K, Foster B, Han CG, Brettin TS, Lapidus A. Gap Resolution: A software package for improving Newbler genome assemblies. Report Number: LBNL-1899E Abs. 2009.
15. Han C, Chain P. Finishing repeat regions automatically with Dupfinisher. In: Arabnia HR, Valafar H, editors. *Proceeding of the 2006 international conference on bioinformatics & computational biology: CSREA PRes.* 2006. p. 141–6.
16. Smith DR. Ligation-mediated PCR, of restriction fragments from large DNA molecules. *Genome Res.* 1992;2:21–7.
17. Lapidus A, LaButti K, Foster BSL, Trong SEG. POLISHER: an effective tool for using ultra short reads in microbial genome assembly and finishing. *AGTB.* 2008. Marco Island FT.
18. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics.* 2010;11:119–9.
19. Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. *Nat Methods.* 2010;7(6):455–7.
20. Consortium TU. UniProt: a hub for protein information. *Nucleic Acids Res.* 2015;43(D1):D204–12.
21. Haft DH, Selengut JD, White O. The TIGRFAMS database of protein families. *Nucleic Acids Res.* 2003;31(1):371–3.
22. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J and others. Pfam: the protein families database. *Nucleic Acids Res.* 2014;42(D1):D222–30.
23. Claudel-Renard C, Chevalet C, Faraut T, Kahn D. Enzyme-specific profiles for genome annotation: PRIAM. *Nucleic Acids Res.* 2003;31(22):6633–9.
24. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 2012;40(D1):D109–14.
25. 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(1):33–6.
26. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L and others. InterPro: the integrative protein signature database. *Nucleic Acids Res.* 2009;37(Database issue): D211–5.
27. 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(17):2271–8.
28. Friedrich CG, Bardischewsky F, Rother D, Quentmeier A, Fischer J. Prokaryotic sulfur oxidation. *Current Opinion in Microbiol.* 2005;8(3):253–9.
29. Frigaard NU, Dahl C. Sulfur metabolism in phototrophic sulfur bacteria. In: Poole RK, editor. *Advances in Microbial Physiology.* Volume Volume 54: Academic Press; 2008. p 103–200.
30. Kappler U, Maher M. The bacterial SoxAX cytochromes. *Cell Mol Life Sci.* 2013;70(6):977–92.
31. Kappler U, Maher MJ. SoxAX cytochromes. In: Scott RA, editor. *Encyclopedia of Inorganic and Bioinorganic Chemistry.* Chichester: John Wiley; 2013. doi:10.1002/9781119951438.eibc2169.
32. Sauve V, Roversi P, Leath KJ, Garman EF, Antrobus R, Lea SM, Berks BC. Mechanism for the hydrolysis of a sulfur-sulfur bond based on the crystal structure of the thiosulfohydrolase SoxB. *J Biol Chem.* 2009;284(32):21707–18.
33. Sauve V, Bruno S, Berks BC, Hemmings AM. The SoxYZ complex carries sulfur cycle intermediates on a peptide swinging arm. *J Biol Chem.* 2007;282(32):23194–204.
34. Bamford VA, Bruno S, Rasmussen T, Appia-Ayme C, Cheesman MR, Berks BC, Hemmings AM. Structural basis for the oxidation of thiosulfate by a sulfur cycle enzyme. *EMBO J.* 2002;21(21):5599–610.
35. Zander U, Faust A, Klink BU, de Sanctis D, Panjikar S, Quentmeier A, Bardischewsky F, Friedrich CG, Scheidig AJ. Structural basis for the oxidation of protein-bound sulfur by the sulfur cycle molybdohemo-enzyme sulfane dehydrogenase SoxCD. *J Biol Chem.* 2011;286(10):8349–60.
36. Kilmartin JR, Maher MJ, Krusong K, Noble CJ, Hanson GR, Bernhardt PV, Riley MJ, Kappler U. Insights into structure and function of the active site of SoxAX cytochromes. *J Biol Chem.* 2011;286(28):24872–81.
37. Scott KM, Sievert SM, Abril FN, Ball LA, Barrett CJ, Blake RA, Boller AJ, Chain PSG, Clark JA, Davis CR and others. The genome of deep-sea vent chemolithoautotroph *Thiomicrospira crunogena* XCL-2. *PLoS Biol.* 2006;4(12):2196–212.
38. Kappler U, Aguey-Zinsou KF, Hanson GR, Bernhardt PV, McEwan AG. Cytochrome *c*<sub>551</sub> from *Starkeya novella*: characterization, spectroscopic properties, and phylogeny of a diheme protein of the SoxAX family. *J Biol Chem.* 2004;279(8):6252–60.
39. Ogawa T, Furusawa T, Nomura R, Seo D, Hosoya-Matsuda N, Sakurai H, Inoue K. SoxAX binding protein, a novel component of the thiosulfate-oxidizing multienzyme system in the green sulfur bacterium *Chlorobium tepidum*. *J Bact.* 2008;190(18):6097–110.
40. Gregersen LH, Bryant DA, Frigaard N-U. Mechanisms and evolution of oxidative sulfur metabolism in green sulfur bacteria. *Front Microbiol.* 2011;2. doi: 10.3389/fmicb.2011.00116.
41. Muyzer G, Sorokin DY, Mavromatis K, Lapidus A, Clum A, Ivanova N, Pati A, d'Haeseleer P, Woyske T, Kyrpides NC. Complete genome sequence of “*Thioalkalivibrio sulfidophilus*” HL-EbGr7. *Stand Genomic Sci.* 2011;4(1):23–35.
42. Kappler U, Davenport K, Beatson S, Lucas S, Lapidus A, Copeland A, Berry KW, Glavina Del Rio T, Hammon N, Dalin E and others. Complete genome sequence of the facultatively chemolithoautotrophic and methylotrophic alpha-Proteobacterium *Starkeya novella* type strain (ATCC 8093 T). *Stand Genomic Sci.* 2012;7(1):44–58.
43. Cosseau C, Batut J. Genomics of the *ccoNOQP*-encoded *ccb3* oxidase complex in bacteria. *Arch Microbiol.* 2004;181(2):89–96.
44. Marchal K, Sun J, Keijers V, Haaker H, Vanderleyden J. A Cytochrome *ccb3* (Cytochrome *c*) terminal oxidase in *Azospirillum brasilense* Sp7 supports microaerobic growth. *J Bact.* 1998;180(21):5689–96.
45. Hamada M, Toyofuku M, Miyano T, Nomura N. *ccb3*-type cytochrome *c* oxidases, aerobic respiratory enzymes, impact the anaerobic life of *Pseudomonas aeruginosa* PAO1. *J Bact.* 2014;196(22):3881–9.
46. Colburn-Clifford J, Allen C. A *ccb3*-Type Cytochrome C oxidase contributes to *Ralstonia solanacearum* R3bv2 growth in microaerobic environments and to bacterial wilt disease development in tomato. *Mol Plant-Microbe Interactions.* 2010;23(8):1042–52.
47. Kim YJ, Ko U, Lee JM, Kang HY, Kim YM, Kaplan S, Oh JI. Dominant role of the *ccb3* oxidase in regulation of photosynthesis gene expression through the PrfBA system in *Rhodospirillum rubrum* 2.4.1. *J Bact.* 2007;189(15):5617–25.
48. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV and others. The minimum information about a genome sequence (MIGS) specification. *Nat Biotech.* 2008;26(5):541–7.
49. Deck J, Barker K, Beaman R, Buttigieg PL, Dröge G, Guralnick R, Miller C, Tuama ÉÓ, Murrell Z, Parr C and others. Clarifying concepts and terms in biodiversity informatics. *Stand Genomic Sci.* 2013;8(2):352–9.
50. Garrity G. NamesforLife. BrowserTool takes expertise out of the database and puts it right in the browser. *Microbiology Today.* 2010;37:9.
51. 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(12):4576–9.
52. Garrity GM, Bell JA, Lilburn T. Phylum XIV. Proteobacteria phyl. nov. In: Brenner DJ, Krieg NR JTS, Garrity GM, editors. *Bergey's Manual of Systematic Bacteriology.* New York: Springer; 2005. p. 1. Volume 2 pt B.

53. Garrity GM, Bell JA, Lilburn T. Class III. *Gammaproteobacteria* class. nov. In: Brenner DJ, Krieg NR JTS, Garrity GM, editors. *Bergey's Manual of Systematic Bacteriology*. secondth ed. New York: Springer; 2005. p. 1–59. Volume 2 pt B.
54. Validation List No. 106. List of new names and new combinations previously effectively, but not validly, published. *Int J Syst Evol Micro*. 2005;55:2235–8.
55. Garrity GM, Bell JA, Lilburn T, Order V. *Thiotrichales* ord. nov. In: Brenner DJ, Krieg NR JTS, Garrity GM, editors. *Bergey's Manual of Systematic Bacteriology*. secondth ed. New York: Springer; 2005. p. 131–210. Volume 2 pt B.
56. Fryer JL. CN. L. Family II. *Piscirickettsiaceae* fam. nov. In: Brenner DJ, Krieg NR JTS, Garrity GM, editors. *Bergey's Manual of Systematic Bacteriology*. secondth ed. New York: Springer; 2005. p. 180. Volume 2 pt B.
57. The Gene Ontology C, Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS and others. Gene Ontology: tool for the unification of biology. *Nat Genet*. 2000;25(1):25–9.
58. Lee C, Grasso C, Sharlow MF. Multiple sequence alignment using partial order graphs. *Bioinformatics*. 2002;18(3):452–64.
59. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*. 2000;17(4):540–52.
60. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML web servers. *Syst Biol*. 2008;57(5):758–71.
61. Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. How many bootstrap replicates are necessary? *J Comp Biol*. 2010;17(3):337–54.
62. Swofford DL. PAUP\*: Phylogenetic Analysis Using Parsimony (\*and Other Methods), version 4.0 b 10. Sunderland: Sinauer Associates; 2002.
63. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011;28:2731–9.
64. JGI website: [www.jgi.doe.gov](http://www.jgi.doe.gov).
65. Website for gapResolution: [jgi.doe.gov/data-and-tools/gap-resolution/](http://jgi.doe.gov/data-and-tools/gap-resolution/)

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](http://www.biomedcentral.com/submit)

