Comparative Genome Analysis Reveals Metabolic Versatility and Environmental Adaptations of Sulfobacillus thermosulfidooxidans Strain ST

Xue Guo1,2, Huaqun Yin1,2, Yili Liang1,2, Qi Hu1,2, Xishu Zhou1,2, Yunhua Xiao1, Liyuan Ma1, Xian Zhang1, Guanzhou Qiu1,2, Xueduan Liu1,2*

1 School of Minerals Processing and Bioengineering, Central South University, Changsha, China, 2 Key Laboratory of Biometallurgy of Ministry of Education, Central South University, Changsha, China

Abstract

The genus Sulfobacillus is a cohort of mildly thermophilic or thermostolerant acidophiles within the phylum Firmicutes and requires extremely acidic environments and hypersalinity for optimal growth. However, our understanding of them is still preliminary partly because few genome sequences are available. Here, the draft genome of Sulfobacillus thermosulfidooxidans strain ST was deciphered to obtain a comprehensive insight into the genetic content and to understand the cellular mechanisms necessary for its survival. Furthermore, the expressions of key genes related with iron and sulfur oxidation were verified by semi-quantitative RT-PCR analysis. The draft genome sequence of Sulfobacillus thermosulfidooxidans strain ST, which encodes 3225 predicted coding genes on a total length of 3,333,554 bp and a 48.35% G+C, revealed the high degree of heterogeneity with other Sulfobacillus species. The presence of numerous transposases, genomic islands and complete CRISPR/Cas defence systems testifies to its dynamic evolution consistent with the genome heterogeneity. As expected, S. thermosulfidooxidans encodes a suit of conserved enzymes required for the oxidation of inorganic sulfur compounds (ISCs). The model of sulfur oxidation in S. thermosulfidooxidans was proposed, which showed some different characteristics from the sulfur oxidation of Gram-negative A. ferrooxidans. Sulfur oxygenase reductase and heterodisulfide reductase were suggested to play important roles in the sulfur oxidation. Although the iron oxidation ability was observed, some key proteins cannot be identified in S. thermosulfidooxidans. Unexpectedly, a predicted sulfocyanin is proposed to transfer electrons in the iron oxidation. Furthermore, its carbon metabolism is rather flexible, can perform the transformation of pentose through the oxidative and non-oxidative pentose phosphate pathways and has the ability to take up small organic compounds. It encodes a multitude of heavy metal resistance systems to adapt the heavy metal-containing environments.

Some of these species that have been tested also have the ability of iron oxidation [3,7].

Over the past few years, an increasing number of acidophile genomes have been sequenced. There are now at least 56 draft or completely sequenced genomes of acidophiles including 30 bacteria and 26 archaea, providing a first glimpse of the genomics of acidophilic life over a range of environmental conditions [8]. However, most of these genomes belong to Gram-negative S. acidophilus in the genus Sulfobacillus has been investigated genomically, revealing that Sulfobacillus exploits a surprisingly different enzymatic repertoire for energy and carbon metabolism compared with the Acidithiobacillus counterparts. For example, S. acidophilus lacks homologues for rusticyanin typically found in iron-oxidizing acidophiles [9], suggesting different components for electron transport in the iron oxidation. Furthermore, members of the genus Sulfobacillus also exhibit remarkably different abilities of iron (II) oxidation and sulfide oxidation [10]. However, a comprehensive understanding of the metabolic versatility and environmental adaptations of
Sulfobacillus will require comparative genomic analysis with more members of the genus Sulfobacillus.

In this study, *S. thermosulfidooxidans* strain ST isolated from an acid hot spring in Tengchong, Yunnan (Southwestern China) presents interesting physiological and metabolic capacities (Fig. S1 in File S1 and Table 1). A deep coverage draft genome of *S. thermosulfidooxidans* strain ST was sequenced and analyzed. It was next compared to the genomes of other thermophiles to explore the physiology of *S. thermosulfidooxidans* at the whole genome level. Furthermore, semi-quantitative RT-PCR analysis was performed to verify the expressions of key genes related with iron and sulfur oxidation. A detailed analysis of energy metabolism and central carbon metabolism in *S. thermosulfidooxidans* was described. This genome exploration revealed a very dynamically evolving genome contributing to an unexpected physiological versatility.

**Materials and Methods**

**Growth of *S. thermosulfidooxidans* Strain ST and DNA Extraction**

*Sulfobacillus thermosulfidooxidans* strain ST was grown at 45°C aerobically in 9 K basal salt medium [11] with 4.5% (w/v) ferrous sulfate. Microbial cells were harvested by centrifugation (12,900 rpm) for 10 min at 4°C. Genomic DNA was extracted from the pelleted cells using TIANamp Bacteria DNA kit (TIANGEN, China) according to the manufacturer’s instruction and finally suspended in MilliQ water. The genomic DNA was stored in –80°C until used for genome sequencing.

**RNA Extraction and Semi-quantitative RT-PCR Analysis**

45 g/L ferrous sulfate and 10 g/L elemental sulfur (S0) were separately used as substrate in the cultivation. Microbial cells were harvested in mid exponential growth phase. Total RNA was extracted using TRIzol reagent (Invitrogen, Carlsbad, USA), harvested in mid exponential growth phase. Total RNA was extracted using TRIzol reagent (Invitrogen, Carlsbad, USA), treated with RNase-free DNAse I (Qiagen, Valencia, USA) and purified with a RNeasy kit (Qiagen, Valencia, USA). Then, single-stranded cDNA was synthesized with ReverTra Ace qPCR RT Kit (Toyobo, Japan), according to the manufacturer’s protocol.

The cDNA was stored at –80°C until used for semi-quantitative RT-PCR analysis.

Primers targeting selected genes putatively involved in ISC metabolism and iron oxidation were designed for semi-quantitative RT-PCR (Table S1 in File S1). The semi-quantitative RT-PCR was performed in 25 μl reaction mixture containing 12.5 μl universal TqPCR Master Mix (Tiangen Biotech, China), 0.5 μl single-stranded cDNA, and 1 μl each of 10 μM forward and reverse primers, and 10 μl deionized water. The specific amplification protocol was as follows: 95°C for 5 min, then 40 cycles of 95°C for 20 s, 55°C for 15 s, and 72°C for 15 s and a final incubation of 72°C for 10 min. PCR products were visualized on 2% agarose gels and sequenced bidirectionally.

**Genome Sequencing, Assembly, Annotation**

Genomic library construction, sequencing, and assembly were performed at the Beijing Genomics Institute (BGI; Shenzhen, China) using Illumina Hiseq 2000 sequencing platform and yielded approximately 350 Mb sequence information. Finally, the raw reads were assembled into 33 supercontigs using SOAPdenovo package [12]. Coding sequences (CDSs) were predicted with the ORF finders Glimmer [13]. All CDSs were manually curated and verified by comparison with the publicly available databases NCBI non-redundant[14], KEGG[15], COG[16] using the annotation software BLAST[17]. The unassigned CDSs were further annotated using the hmmpfam program of the HMMER package 26.0 [18]. The hidden Markov models for the protein domains were obtained from the Pfam database 26.0. And the identifications of tRNA and rRNA were performed using the tRNAscan-SE [19] and RNAmer programs [20], respectively. CRISPR loci were identified using CRISPRFinder [21]. Transporter gene annotations were performed by additionally taking into account the information of transporter classification database [22]. Detailed information on the genes ordered by functional category was summarized in Tables S2, S3, S4, and S5 in File S1. The whole genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number PRJNA203261.

**Table 1. General features of the *S. thermosulfidooxidans* genome in comparison with other Clostridiales family XVII incertae sedis genomes.**

| Organism                  | *Sulfobacillus thermosulfidooxidans* ST | *Sulfobacillus acidophilus* DSM 10332 | *Sulfobacillus acidophilus* TPY | *Thermoanaerobacter arianoiensis* DSM 12885 |
|---------------------------|----------------------------------------|--------------------------------------|---------------------------------|--------------------------------------------|
| Habitat                   | Acid hot spring                        | Acidic sulfidic and sulfurous sites  | Hydrothermal vent in the Pacific Ocean | Mud from the Mariana Trench in the Pacific Ocean |
| Temperature range         | Moderate thermophilic (48°C optimum)    | Moderate thermophilic (50°C optimum)  | Moderate thermophilic (Approximately 50°C) | Thermophilic (50–80°C) |
| pH range                  | 1.2–2.4                                | 1.6–2.3                              | 1.6–2.3                         | 5.4–9.5                                    |
| Motility                  | Non-motile                             | Non-motile                           | Motile                          | Non-motile                                 |
| Nutrition type            | Mixotrophic                            | Mixotrophic                          | Mixotrophic                     | Chemo-heterotroph                          |
| Genome size in Mb         | 3.33                                   | 3.56                                 | 3.55                            | 2.84                                       |
| G+C content               | 48.35%                                 | 56.8%                                | 56.8%                           | 72.5%                                      |
| Protein-coding genes      | 3225                                   | 3585                                 | 3837                            | 2435                                       |
| 16S-23S rRNA genes        | 1                                      | 5                                    | 5                               | 2                                          |
| Number of tRNAs           | 48                                     | 69                                   | 52                              | 60                                         |

doi:10.1371/journal.pone.0099417.t001
Genome Synteny Comparisons

Pairwise alignments for dot plot representations were performed on six-frame amino acid translation of the genome sequences of *Sulfobacillus thermosulfidooxidans*, *Sulfobacillus acidophilus* and *Thermaerobacter marianensis* using the Promer program in the MUMmer 3.23 package [23]. The default parameters in all analyses were applied, so exact matches longer that six amino acids were identified and adjacent exact matches were joined if a gap no longer than 30 amino acids occurred. And the resulting clusters were further processed if their matches were longer than 20 amino acids and then aligned using a BLOSUM62 amino acid substitution matrix. Furthermore, the GGDC-Genome-to-Genome Distance Calculator was used to estimate the overall similarity among the three genomes. The inferred digital DNA-DNA hybridization values were calculated [24].

Calculation of COGs for Venn Diagrams

The predicted proteome sequences of selected genomes (*S. acidophilus* and *T. marianensis*) except for *S. thermosulfidooxidans* were retrieved from the NCBI database. The best sequence similarities were obtained by BLAST against COG database using maximal *E*-value = 1e^{-3}. Proteins that were not group into COGs were represented as specific proteins for each organism. The calculation of COGs was performed and visualized with the R package [25].

Phylogenetic Analyses

Predicted amino acid sequences of selected genes were aligned with reference sequences using multiple sequence alignment tool ClustalW 2.0 [26]. If not mentioned otherwise, phylogenetic trees were constructed using Molecular Evolutionary Genetics Analysis 4.0 software (MEGA, version 4.0).

Results and Discussion

Sequencing and Automatic Annotation of the Strain ST Genome

The draft genome sequence of *S. thermosulfidooxidans* strain ST contained 53 supercontigs, ranging from 507 bp to 547,747 bp (the N50 and N90 contig sizes are 376,394 bp and 36,184 bp, respectively), with a total length of 3,333,554 bp. Since most contigs end with repeated sequences, further assembly was not possible with current data. However, given that the present draft has 100x sequence coverage, it is reasonable to assume that the majority of genes in genome of the strain ST are identified from the current draft. It differs from those members of the *Clostridiales* family XVII incertae sedis by its lower overall G+C content (48.35%) (Table 1). Furthermore, within 3225 predicted open reading frames (ORFs), 704 ORFs (22%) are annotated as hypothetical proteins and more than 1190 predicted proteins of *S. thermosulfidooxidans* do not have the best hits within the members of *Clostridiales* family XVII incertae sedis. Likewise, few larger syntenic regions were observed between the *S. thermosulfidooxidans* genome and those of other *Clostridiales* family XVII incertae sedis (Fig. 1). The inferred digital DNA-DNA hybridization value by the GGDC-Genome-to-Genome Distance Calculator for *S. thermosulfidooxidans* and *S. acidophilus* is <30%. Altogether, these findings suggest that the gene complement of *S. thermosulfidooxidans* is significantly different from those of other *Clostridiales* family XVII incertae sedis. Furthermore, we also performed a comparative COG analysis among *S. thermosulfidooxidans*, *T. marianensis* and *S. acidophilus* (Fig. S2 in File S1). In general, *S. thermosulfidooxidans* shared more orthologous genes with *S. acidophilus* than *T. marianensis*. But more than one third of genes in *S. thermosulfidooxidans* genome are unique, which suggests that some physiological features of *S. thermosulfidooxidans* differ a lot from the other two bacteria. Besides, one rRNA operon containing 16S, 23S and 5S rRNA genes was found in the draft genome of *S. thermosulfidooxidans* strain ST (Table S2 in File S1). The 16S rRNA sequence of *S. thermosulfidooxidans* has 93% and 86% nucleotide identity to the 16S rRNAs of *S. acidophilus* and *T. marianensis*. In addition, *S. thermosulfidooxidans* also encodes a similar collection of information processing genes like the other *Sulfobacillus* members (Table S2 in File S1).

Integrative Elements, CRISPR Defence System

More than 47 transposase genes with transposase signatures are identified in the *S. thermosulfidooxidans* genome and can be assigned to at least nine different types of transposase (Table S3 in File S1) that are associated with insertion sequence families IS3, IS4, IS21, IS66, IS110, IS200, IS605, IS1477 and ISChy4 [27]. Interestingly,
the IS element of S. thermosulfidooxidans belonging to group IS605 occurs 14 copies in the genome, but they have best BLAST hits with 11 different genera excluding the genus Sulfobacillus. Furthermore, most of these IS elements occur in poorly conserved genomic regions, and several elements appear to have inserted into protein-encoding genes indicating that they significantly contribute to genome evolution. Although twenty-seven transposase genes have best hits within the genus Sulfobacillus, the organizations surrounding the IS elements are significantly different from those of S. acidophilus indicating that these IS elements markedly changed genomic structure of S. thermosulfidooxidans. Except for IS elements, some other genes relative with DNA transposition were also identified in the S. thermosulfidooxidans genome (Table S3 in File S1). For example, a site-specific recombinase that can catalyze sequential DNA strand exchange reactions [28] was found in the genome. Besides, a cassette chromosome recombinase gene was also identified in the genome (Table S3 in File S1). This cassette chromosome recombinase typically controls resistance gene transmissions [29]. These site-specific recombinases may also greatly contribute genomic variability and physiological versatility.

Interestingly, CRISPR/Cas (clustered regularly interspaced short palindromic repeats/CRISPR-associated genes) viral defense systems [30,31,32] were identified in the S. thermosulfidoxidans genome (Table S4 in File S1). Eight cas genes orienting in the same direction formed a typical CRISPR/Cas system with 44 spacers in the downstream genomic sequence. According to a recent classification of the CRISPR/Cas systems into three major types (I–III), the CRISPR/Cas systems in S. thermosulfidooxidans can be assigned into type I CRISPR/Cas systems that is proposed to function in virus defense by directly targeting DNA [33]. Moreover, six CRISPR-associated cmr genes orienting in the opposite direction with cas genes, all belonging to the repair-associated mysterious protein (RAMP) superfamily [34], are identified in the immediate upstream of cas genes. These cmr genes may compose a CRISPR viral defense system with 39 spacers in the further upstream, though the functions of these cmr genes are largely unknown [33]. Furthermore, the similarities of these cmr genes supported the view that they may be obtained by gene horizontal transfer. Besides, one more CRISPR locus is identified in the distant genomic region. But only 5 spacers were found in the region and no cas or cmr genes were found in the region. Furthermore, putative CRISPR loci in the genomes of S. acidophilus and T. marianensis were also detected by CRISPRfinder [21]. Although we could also identify associated cas genes, which are required for viral defense in the two genomes, associated cmr genes were not identified in the two genomes. In contrast to S. thermosulfidooxidans, the predicted CRISPR sequences of S. acidophilus and T. marianensis contain fewer spacer-repeat units. There are only 7 repeat/spacer sequences in S. acidophilus [3], and 17 repeat/spacer sequences in T. marianensis [35]. Unexpectedly, only 2 spacers in S. thermosulfidooxidans show similarities to the spacers in T. marianensis genome, and none of spacers show similarity to the spacers in S. acidophilus genome.

Energy Metabolism

Sulfobacillus thermosulfidooxidans can grow mixotrophically by aerobic oxidation of ferrous iron, sulfur, and sulfide in the presence of organic compounds and concomitant fixation of inorganic carbon [36]. The oxidation and electron transfer pathways of ISC are very complicated and various in different microbes, making their prediction and elucidation difficult [37,38]. Besides, some steps occur spontaneously, without enzymatic catalysis in the oxidation of ISC. Previous studies of S. thermosulfidooxidans detected several enzymatic activities involved in the oxidation of ISC, but the specific genes related with these activities were not identified [39,40]. Based on genome analysis, more than 30 genes encoding enzymes and electron transfer proteins predicted to be involved in the oxidation of inorganic sulfur compounds (ISCs) were detected in the genome (Table S4 in File S1). Genome-based model for ISC metabolism in S. thermosulfidooxidans is proposed (Fig. 2). Furthermore, semi-quantitative RT-PCR also indicated that all analyzed genes related with sulfur oxidations are expressed during growth on elemental sulfur.

**The oxidation of elemental sulfur.** Sulfur oxydgenase reductase (SOR) is considered a cytoplasmic enzyme oxidizing elemental sulfur in the cytoplasm in many sulfur-oxidizing bacteria, although it is largely unclear how sulfur is transferred into cytoplasm. It can catalyze substrate sulfur into hydrogen sulfide, sulfite and thiosulfate [41,42,43]. In this reaction, elemental sulfur is both the electron donor and one of the two know acceptors, the other being oxygen. The enzyme is different from sulfur dioxygenase [44] and sulfur reductase [45], in that both activities are found together. However, recent study shows that the reaction catalyzed by SOR doesn’t couple with the electron transfer chain or substrate-level phosphorylation in A. caldus. The predicted SOR protein in S. thermosulfidooxidans shows a characteristic domain of the SOR family (Pfam: PF07682) and all activity-required residues [46]. And the predicted protein shows 72% similarity with the SOR in S. acidophilus. Semi-quantitative RT-PCR indicated that the expression of Sor gene is very high during growth on elemental sulfur, which suggests that the SOR in S. thermosulfidooxidans is also involved in the oxidation of elemental sulfur. Surprisingly, the SORs in Sulfobacillus genus are assigned into the archaeal cluster and most closely related to a homologue in archaeal Ferroplasma acidarmanus which also survives in extremely acidic environments (Fig. 3A). These suggest that Sulfobacillus genus may exchange the sulfur-oxidizing gene with other extremophiles sharing the similar niche.

One of products from the SOR catalyzed reactions, hydrogen sulfide, is considered to be oxidized by sulfide quinone reductase (SQR). In this reaction, two electrons are transferred to the electron transfer chain by the quinone. In Gram-negative A. ferrooxidans, hydrogen sulfide is oxidized by SQR located in the cytoplasmic membrane [37]. One copy of Sqr gene was identified in the genome of S. thermosulfidooxidans. Meanwhile, an orthologous gene is very high SQR similarity was also identified in S. acidophilus. The conservation of sqr gene context in the genus Sulfobacillus is observed with A. ferrooxidans, which strongly suggests that sulfide quinone reductase also has the similar functional properties in Gram-positive Sulfobacillus genus.

Another product from the SOR catalyzed reactions, thiosulfate, may be catalyzed by thiosulfate sulfurrtransferase (TST) or rhodanese. Five genes encoding thiosulfate sulfurtransferase (TST) or rhodanese are dispersed in S. thermosulfidooxidans genome. The presence of the rhodanese motif associated with ubiquitin C-terminal hydrolases and phosphatases makes these enzymes possess the potential of sulfur oxidation [47]. They are proposed to transfer a sulfur atom from thiosulfate to sulfur acceptors like thiol proteins (RSH) with the production of sulfite in A. caldus [48]. Subsequently, the product from the TST catalyzed reactions, sulfane sulfide (RSSH), is proposed to be the substrate of heterodisulfide reductase (HDR). Heterodisulfide reductase is three subunit complex HdrABC that can catalyze the oxidation of RSSH to regenerate RSH with coupling with the electron transfer chain. Three copies of HdrABC operon have been identified in S. thermosulfidooxidans (Table S4 in File S1). One copy of HdrABC operon has a high similarity with that of S. acidophilus, while the
The sulfite oxidation pathway. The known periplasmic enzymes [sorAB or soxCD] involved in the direct oxidation of sulfite [30,51] were not identified in both S. thermosulfidooxidans and S. acidophilus. An oxidoreductase molybdopterin binding protein was identified as a possible sulfite oxidase, but the putative protein only harbors a molybdopterin domain and a twin-arginine translocation pathway signal sequence without dimerization domain and N-terminal heme domain (Table S4 in File S1). Furthermore, no additional subunit containing heme domain that is required for sulfite oxidase was identified in S. acidophilus and S. thermosulfidooxidans. Therefore, subsequent oxidation of sulfite is most likely to occur in the cytoplasm (Fig. 2). The most possible way for sulfite oxidation is that sulfite is converted to adenosine-5'-phosphosulfate (APS) and then oxidized to sulfate via an indirect pathway controlled by APS reductase and sulfate adenylyltransferase (SAT). Electrons from APS reductase and sulfate adenylyltransferase indeed catalyzes APS to sulfate. In A. ferrooxidans, the missing function is postulated to be accomplished by the hypothetical gene embedded in the hdr locus of sulfur oxidizers [37]. The conserved hypothetical gene was also found in the hdr locus of S. thermosulfidooxidans genome (Table S4 in File S1).

The S4I pathway. The gene encoding tetrathionate hydrolase (TTH) was also identified in S. thermosulfidooxidans, which is thought to be involved in the hydrolysis of tetrathionate to generate sulfur, sulfate and thiosulfate. And the activity of this enzyme has been studied in Acidithiobacillus genus [52,53]. The TTH of S. thermosulfidooxidans shows 57% similarity with that of A. caldus and has a conserved pyrrolo-quinoline quinone domain. Previous experimental data showed that the A. caldus TTH is a soluble periplasmic protein with maximum activity at pH 3.0 [54]. Furthermore, doxDA genes present in S. thermosulfidooxidans genome are predicted to encode a thiosulfate/quinone oxidoreductase (TQR). Orthologous genes of doxDA were also detected in S. acidophilus, which supports that the enzyme in Sulfolobus genus has the similar functional properties. In our model, the TQR is proposed to catalyze thiosulfate to tetrathionate and transfer two electrons to the quinone. The consecutive reactions catalyzed by TTH and TQR promote the sulfur oxidation in the periplasm of S. thermosulfidooxidans.

The iron(II) oxidation. In extremely acidic environment, the most detailed account of iron(II) oxidation pathways is available for the Gram-negative bacterium A. ferrooxidans [9]. The model for iron oxidation in A. ferrooxidans is related to two transcriptional units, the petI and rus operons [55]. However, the genome of S. thermosulfidooxidans doesn’t contain petI and rus operons, and these genes are not discovered in S. acidophilus either. Unexpectedly, a gene encoding the blue copper protein sulfocyanin was found in the S. thermosulfidooxidans genome. The predicted protein in S. thermosulfidooxidans shows a characteristic domain of sulfocyanin (Pfam: PF06525). Sulfocyanin, sharing sequence characteristics with A. ferrooxidans rusticyanins, is proposed to transfer electrons during iron oxidation in acidiphilic archaea Ferroplasma spp. [56,57]. Besides, semi-quantitative RT-PCR indicated that the sulfocyanin gene in S. thermosulfidooxidans is
Figure 3. Phylogenetic analysis of key genes related with the oxidation of inorganic sulfur compounds and ferrous iron. (A) Phylogenetic tree of sulfur oxidase reductase based on an alignment of 334 amino acid positions with the neighbor-joining method. The numbers associated with the branches refer to bootstrap values (confidence limits) resulting from 1,000 replicate resamplings. The scale represents the number of amino acid substitutions per site. *S. thermosulfidooxidans* are shown in bold. (B) Phylogenetic tree of sulfocyanin/rusticyanin proteins of archaea and bacteria based on an alignment of 263 amino acid positions with the neighbor-joining method. The numbers associated with the branches refer to bootstrap values (confidence limits) resulting from 1,000 replicate resamplings. The scale represents the number of amino acid substitutions per site. *S. thermosulfidooxidans* are shown in bold.

doi:10.1371/journal.pone.0099417.g003
highly expressed during growth on ferrous sulfate. Thus, it is possible that the predicted sulfocyanin is a component of the electron transport chain in iron oxidation of *S. thermosulfidooxidans*, although more details about iron(II) oxidation pathways are still unknown. Interestingly, phylogenetic analysis revealed that the protein is assigned to the archaeal cluster (Fig. 3B), which makes the origin of sulfocyanin in *S. thermosulfidooxidans* elusive.

Electron Transfer Chain

*S. thermosulfidooxidans* encodes a fairly complete respiratory chain consisting of complexes 1–5, which is necessary for energy generation and reverse electron transport (Table S4 in File S1). Three *cytCAB* copies encoding subunits of cytochrome *bd* complex and four gene clusters that code for *aa3*-type terminal oxidase were detected in the draft genome. The analysis also revealed that *S. thermosulfidooxidans* has 16 genes encoding all subunits of type I NADH dehydrogenase (subunit F of which has three copies). In consistent with other members of the *Sulfobacillus* genus, *S. thermosulfidooxidans* also lacks most components of electron transfer chain in the iron oxidation, only Cyt *aa3*, a subunit of the cytochrome complex *bc1* and cytochrome *c* present. However, numerous other genes putatively involved in electron transfer chain were found in *S. thermosulfidooxidans* genome (Table S4 in File S1). The gene redundancy may provide regulatory flexibility to confront environmental changes such as nutrient deficiency and different substrate phosphorylations.

Central Carbon Metabolism

*S. thermosulfidooxidans* encodes all key genes of the Calvin cycle carbon fixation pathway (Fig. 4; Table S4 in File S1). Comparing with *S. acidophilus* genomes [3], *S. thermosulfidooxidans* only contains a gene cluster encoding form I ribulose bisphosphate carboxylase and lacks homologs of form II ribulose bisphosphate carboxylase. Form I and II ribulose bisphosphate carboxylases are regulated to adapt to environmental conditions with different levels of CO₂ in *Hydrogenovibrio marinus* [58]. It was reported that *S. thermosulfidooxidans* could grow autotrophically at the CO₂ content of the supplied air to 5–10% [59]. Thus, the complete Calvin cycle confirms that carbon dioxide can be the major source of carbon for *S. thermosulfidooxidans*.

In many organisms, the 3-phosphoglyceraldehyde generated by CO₂ fixation via the Calvin cycle enters the glycolysis/glucconeogenesis pathways [47,60]. The genes identified for the pathways in *S. thermosulfidooxidans* with their reactions and potential interconnections with other biosynthetic pathways are presented (Fig. 4). Fixed carbon can be channeled in either of two directions: for glycogen biosynthesis, or to provide carbon backbones for anabolic reactions. For glycogen biosynthesis, *S. thermosulfidooxidans* contains a gene that are predicted to encode fructose biphosphate aldolase (EC: 4.1.2.13), which catalyzes the formation of fructose-1, 6-biphosphate. And a fructose biphosphatase gene was identified in *S. thermosulfidooxidans* genome. However, in accordance with all published *Sulfobacillus* genomes, these genomes lack orthologous genes encoding three key enzymes for glycolgen biosynthesis: (i) glucose-1P-adenylyltransferase; (ii) glycolgen synthase; (iii) 1, 4-alpha-glucan-branching protein. Moreover, *S. thermosulfidooxidans* also lacks glycerol phosphorylase that is involved in decomposing glycerol, thus regenerating glucose-1P from the non-reducing terminus of glycerol [61,62]. These results suggested that glycerol may not be the main carbon-stored substance in *S. thermosulfidooxidans*.

*S. thermosulfidooxidans* harbours candidate enzymes for all steps of an oxidative TCA (Fig. 4 and Table S4 in File S1). The conversion of 2-oxoglutarate to succinyl-CoA is flexible in *S. thermosulfidooxidans*. It can be catalyzed directly by 2-oxoacid: ferredoxin oxidoreductase (EC: 1.2.7.3) or continuously catalyzed in two steps by 2-oxoglutarate dehydrogenase complex. Moreover, it is notable that *S. thermosulfidooxidans* seems more versatile in the production and the conversion of central metabolite pyruvate (Table S4 in File S1). Except for one copy of malate dehydrogenase (EC: 1.1.1.37), the genome also has two genes with homology to malic enzyme (EC: 1.1.1.38) that catalyzes the reversible conversion of malate to pyruvate and thus is involved in glucogenesis and anaerobiosis. Furthermore, *S. thermosulfidooxidans* encodes an alanine dehydrogenase (EC: 1.4.1.1) that catalyzes the reversible deamination of alanine to pyruvate (Table S4 in File S1). Alternatively, alanine formation from pyruvate catalyzed by this enzyme might not only be important for protein biosynthesis but could also have a function in ammonia storage and ammonia toxicity alleviation. On the other hand, alanine might thus represent an important source of pyruvate for *S. thermosulfidooxidans*, as it possesses various amino acid and oligo/dipeptide transporters (Table S5 in File S1).

In *S. thermosulfidooxidans*, the process of pentose sugar synthesis is very flexible. *S. thermosulfidooxidans* possesses a functional oxidative pentose phosphate pathway (Fig. 4 and Table S4 in File S1), which can be used for the generation of pentose sugars in many bacteria and archaea [63,64,65]. Meanwhile, homologues of key genes for the non-oxidative pentose phosphate pathway could also be identified in *S. thermosulfidooxidans* as well. Specifically, *S. thermosulfidooxidans* encodes a 6-phosphoglucuronolactonase and two glucose-6-phosphate 1-dehydrogenases that catalyze the interconversion of hexose and pentose (Table S4 in File S1). Furthermore, two genes coding for 6-phosphogluconate dehydrogenase, a key enzyme that bridges the oxidative and non-oxidative part in pentose phosphate pathway, were identified, one of which is located in proximity to two other genes putatively involved in the pentose phosphate pathway, a transketolase and a bifunctional transaldolase/phosphoglucose isomerase. The bifunctional transaldolase/phosphoglucose isomerase is found as a fused protein harboring one transaldolase domain and one glucose-6-phosphate isomerase domain in many bacteria [66]. It is noteworthy that a gene encodes a non-functional glucose-6-phosphate isomerase is also identified in the genome (Table S4 in File S1).

Transport and Resistance

The genome of *S. thermosulfidooxidans* encodes at least 230 putative transporter proteins (Table S5 in File S1), which are the structural elements of approximately 90 transport systems (some of them consisting of several proteins) and represent 50 transporter families [22,67,68]. Furthermore, all important components of the general secretion (Sec) and twin-arginine translocation (Tat) pathways have been identified in the genome (Table S4 in File S1). The complement of transport systems in *S. thermosulfidooxidans* is roughly reminiscent of other Clostridiales family XVII *incertae sedis* but more than 80 transport proteins have best hits without *Clostridiales* family XVII *incertae sedis*.

Among the *S. thermosulfidooxidans* transport proteins, those belonging to the ATP-binding cassette (ABC) superfamily are the most represented (n = 87), which might be involved in the uptake of organic molecules. Besides, more than 15 transport systems composed of ATP-binding cassette (ABC) superfamily proteins responsible for special organic molecules are also identified. These complement the higher flexibility of *S. thermosulfidooxidans* in the central carbon metabolism. In addition, *S. thermosulfidooxidans* also harbors an oligo/dipeptide transport system as well as several aminopeptidases (leucyl- and methionyl-aminopeptidase), which could be used to release amino acid from
the imported peptides. The uptake system of amino acids or peptides may reduce the energy cost of protein synthesis, and could also serve for acquisition of substrates for anaplerotic reactions.

Although the mechanism of transport of extracellular S0 to the cytoplasm is not clear, several candidates have been proposed to play important roles for the S0 transport in green sulfur bacteria [41,69]. One possibility is that the thioredoxin SoxW acts together with thiol-disulfide interchange protein DsbD within the periplasm in transferring S0 across the inner membrane [70]. One gene is identified that potentially codes for a DsbD protein in the draft genome (Table S5 in File S1). No significant homology of SoxW has been found in S. thermosulfidooxidans, however, numerous thioredoxin genes are identified in S. thermosulfidooxidans and they possibly can perform the same function as SoxW. Furthermore, S. thermosulfidooxidans also possesses at least four sulfonate/nitrate/taurine transport systems. These systems can transport sulfonate compounds into the cytoplasm, and the sulfur in sulfonate compounds may finally be oxidized as energy source.

S. thermosulfidooxidans also possesses, like other Sulfobacillus members, various resistance systems including more than 30 putative metal ion efflux proteins belonging to 13 different transporter families (Table S5 in File S1). Most of transport systems responsible for metal ion efflux belong to ABC transporter superfamily. Except for these, we also find some oxidoreductases that are closely related with metal resistances (Table S5 in File S1). For example, S. thermosulfidooxidans encodes a mercuric reductase, which belongs to a FAD-containing flavoprotein and can reduce Hg\(^{2+}\) to Hg\(^0\) utilizing NADPH [71,72]. The reduction of Hg\(^{2+}\) is an important step of mercuric resistance. As for arsenic resistance, a remarkable feature is the presence of two arS genes coding for arsenate reductase not previously described in S. thermosulfidooxidans [73,74]. The enzyme arsenate reductase is required to confer resistance to As(V) for organisms, since the non-enzymatic reduction of As(V) is too slow to be physiologically significant [75]. And the resulting As(III) can be pumped out of the cells by the ArsA/ArsB ATPase [76]. Surprisingly, phylogenetic analysis showed that one of arsenate reductase is most closely related to a homologue in the thermophilic bacterium Thermaerobacter subterraneus but the other is most closely related to a homologue in the acidophilic thermophilic bacterium Alicyclobacillus acidocaldarius (Table S5 and Fig. S3 in File S1). These indicate that S. thermosulfidooxidans may obtain new arsenic resistance capacities from other extremophiles sharing a similar niche. In addition, S. thermosulfidooxidans has three antibiotic transporter systems (Table S5 in File S1), which is absent from other Sulfobacillus genomes. It is tempting to speculate that the unique ecological niche makes S. thermosulfidooxidans obtain new antibiotic resistances.

Conclusions

Comparative genome analysis of S. thermosulfidooxidans genome revealed that the gene content for sulfur oxidation is similar to other sulfur-oxidizing acidophiles, but also revealed some features not yet found in other acidophiles. A novel sulfur oxygenase
reduce is suggested to play a key role in the sulfur oxidation of \textit{S. thermosulfidooxidans}. It can catalyze substrate sulfur into hydrogen sulfide, sulfite and thiosulfate \cite{41,43}. Although the iron oxidation is still unclear, the predicted sulfochlorin is proposed to be an important component of the electron transport chain in the iron oxidation process of \textit{S. thermosulfidooxidans}, as it happens in acidophilic archaebacterium \textit{Ferrobacillus sp.} \cite{56}. In addition, \textit{S. thermosulfidooxidans} has more flexibility in the central carbon metabolism including two pentose phosphate pathways, flexible conversion of the central metabolite pyruvate and the ability to metabolize various organic compounds. However, glycogen may not be used as a substance of energy source in \textit{S. thermosulfidooxidans}. It also possesses numerous transport systems of organic compounds including multiple sugars, oligopeptide/dipeptide, malic acid, and various amino acids. These transport systems complement the higher flexibility of \textit{S. thermosulfidooxidans} in the central carbon metabolism. Furthermore, it also encodes an impressive collection of resistance proteins that will provide a surviving advantage for living in the acid hot spring containing high concentration of various heavy metals. The physiological versatility of \textit{S. thermosulfidooxidans} might be an essential factor for the competitive success in the extreme acidic environment.

**Supporting Information**

File S1  Figures S1, S2, and S3 and Tables S1, S2, S3, S4 and S5. **Figure S1.** A. Phylogeny of \textit{Clostridiales} family XVII \textit{incertae sedis} including \textit{S. thermosulfidooxidans}. Phylogenetic tree based on the 16S rRNA gene was constructed with the neighbor-joining method. The scale bar represents the number of nucleotide substitutions per site. B. Electron micrograph of \textit{S. thermosulfidooxidans} shows its morphology. **Figure S2.** Venn diagrams showing the numbers of COGs shared between the predicted proteomes of \textit{S. thermosulfidooxidans}, \textit{S. acidiphilus}, and \textit{T. marianensis}. Proteins that were not grouped into COGs are represented as specific proteins for each organism. Numbers in brackets behind species names indicate total number of predicted proteins. **Figure S3.** Phylogenetic tree of arsenate reductase. The phylogenetic tree was constructed based on a manually corrected alignment of 159 amino acid positions with the neighbor-joining method. The numbers associated with the branches refer to bootstrap values (confidence limits) resulting from 1,000 replicate resamplings. The scale represents the number of amino acid substitutions per site. The two sequences in \textit{S. thermosulfidooxidans} are shown in bold. **Table S1.** Primers used for semi-quantitative RT-PCR. **Table S2.** Selecting genes encoding informational processing proteins. **Table S3.** Genes encoding transposase and other enzymes related with gene transposition. **Table S4.** Genes encoding proteins of diverse metabolic pathways. **Table S5.** Genes encoding transporter proteins. (ZIP)

**Acknowledgments**

The authors would like to thank Dr. Zhih He in University of Oklahoma for helpful discussion and advices.

**Author Contributions**

Conceived and designed the experiments: XL HY GQ. Performed the experiments: XG YL QH XY LM X.Zhang. Analyzed the data: XG QH X.Zhou. Contributed reagents/materials/analysis tools: XG QH. Wrote the paper: XG.
