frontiers in Microbiology

ORIGINAL RESEARCH
published: 30 November 2016
doi: 10.3389/fmicb.2016.01902

Genome Sequencing of *Sulfolobus* sp. A20 from Costa Rica and Comparative Analyses of the Putative Pathways of Carbon, Nitrogen, and Sulfur Metabolism in Various *Sulfolobus* Strains

*Xin Dai[1,2], Haina Wang[1,2], Zhenfeng Zhang[1], Kuan Li[3], Xiaoling Zhang[3], Marielos Mora-López[4], Chengying Jiang[1], Chang Liu[1,2], Li Wang[1], Yaxin Zhu[1], Walter Hernández-Ascencio[4], Zhiyang Dong[1] and Li Huang[1,2]**

[1] State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China, [2] College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China, [3] State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China, [4] Center for Research in Cell and Molecular Biology, Universidad de Costa Rica, San José, Costa Rica

OPEN ACCESS

Edited by:
Kian Mau Goh, Universiti Teknologi Malaysia, Malaysia

Reviewed by:
Yutaka Kawarabayasi, Kyushu University, Japan
Peter Redder, Paul Sabatier University, France

***Correspondence:**
Li Huang
huangl@sun.im.ac.cn

Specialty section:
This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology

Received: *20 August 2016*
Accepted: *14 November 2016*
Published: *30 November 2016*

Citation:
Dai X, Wang H, Zhang Z, Li K, Zhang X, Mora-López M, Jiang C, Liu C, Wang L, Zhu Y, Hernández-Ascencio W, Dong Z and Huang L (2016) Genome Sequencing of Sulfolobus sp. A20 from Costa Rica and Comparative Analyses of the Putative Pathways of Carbon, Nitrogen, and Sulfur Metabolism in Various Sulfolobus Strains. Front. Microbiol. 7:1902. doi: 10.3389/fmicb.2016.01902

The genome of *Sulfolobus* sp. A20 isolated from a hot spring in Costa Rica was sequenced. This circular genome of the strain is 2,688,317 bp in size and 34.8% in G+C content, and contains 2591 open reading frames (ORFs). Strain A20 shares ~95.6% identity at the 16S rRNA gene sequence level and <30% DNA-DNA hybridization (DDH) values with the most closely related known *Sulfolobus* species (i.e., *Sulfolobus islandicus* and *Sulfolobus solfataricus*), suggesting that it represents a novel *Sulfolobus* species. Comparison of the genome of strain A20 with those of the type strains of *S. solfataricus*, *Sulfolobus acidocaldarius*, *S. islandicus,* and *Sulfolobus tokodaii*, which were isolated from geographically separated areas, identified 1801 genes conserved among all *Sulfolobus* species analyzed (core genes). Comparative genome analyses show that central carbon metabolism in *Sulfolobus* is highly conserved, and enzymes involved in the Entner-Doudoroff pathway, the tricarboxylic acid cycle and the CO_2 fixation pathways are predominantly encoded by the core genes. All *Sulfolobus* species encode genes required for the conversion of ammonium into glutamate/glutamine. Some *Sulfolobus* strains have gained the ability to utilize additional nitrogen source such as nitrate (i.e., *S. islandicus* strain REY15A, LAL14/1, M14.25, and M16.27) or urea (i.e., *S. islandicus* HEV10/4, *S. tokodaii* strain7, and *S. metallicus* DSM 6482). The strategies for sulfur metabolism are most diverse and least understood. *S. tokodaii* encodes sulfur oxygenase/reductase (SOR), whereas both *S. islandicus* and *S. solfataricus* contain genes for sulfur reductase (SRE). However, neither SOR nor SRE genes exist in the genome of strain A20, raising the possibility that an unknown pathway for the utilization of elemental sulfur may be present in the strain. The ability of *Sulfolobus* to utilize nitrate

or sulfur is encoded by a gene cluster flanked by IS elements or their remnants. These clusters appear to have become fixed at a specific genomic site in some strains and lost in other strains during the course of evolution. The versatility in nitrogen and sulfur metabolism may represent adaptation of *Sulfolobus* to thriving in different habitats.

Keywords: *Sulfolobus*, strain A20, genome sequencing, comparative genomics, carbon metabolism, nitrogen metabolism, sulfur metabolism

INTRODUCTION

Archaea of genus *Sulfolobus* are widespread in solfataric fields around the globe. Known *Sulfolobus* species were mostly isolated from the Northern hemisphere (Brock et al., 1972; Grogan et al., 1990; Huber and Stetter, 1991; Jan et al., 1999; Suzuki et al., 2002; Xiang et al., 2003; Guo et al., 2011; Mao and Grogan, 2012; Zuo et al., 2015). These *Sulfolobus* isolates have been classified into nine species. Since *Sulfolobus* is readily grown and manipulated under laboratory conditions (Grogan, 1989), it has been used as a model for the study of Archaea. *Sulfolobus* also serves as a model for the study of eukaryotic genetic mechanisms because of the striking resemblance between Archaea and Eukarya in the flow of genetic information (Bell et al., 2002). In addition, *Sulfolobus* has been used as a host for the study of an increasing number of archaeal viruses and plasmids (Arnold et al., 2000; Rice et al., 2001; Xiang et al., 2003; Guo et al., 2011; Wang et al., 2015).

The complete genomes of 17 *Sulfolobus* strains belonging to four species have so far been deposited in GenBank. These include a *Sulfolobus tokodaii* strain (str.7) (Kawarabayasi et al., 2001), three *Sulfolobus solfataricus* strains (She et al., 2001; McCarthy et al., 2015), four *Sulfolobus acidocaldarius* strains (Chen et al., 2005; Mao and Grogan, 2012), and nine *Sulfolobus islandicus* strains (Reno et al., 2009; Guo et al., 2011; Zhang et al., 2013). Genomic comparisons show that *Sulfolobus* species are genetically diverged in relation to their geographic distance (Whitaker et al., 2003; Reno et al., 2009). Discontinuous and distantly separated habitats seem to be geographic barriers limiting gene flow among *Sulfolobus* populations. The variation in gene content among geographically diverse isolates is consistent with an isolation-by-distance model of diversification (Whitaker et al., 2003; Grogan et al., 2008; Reno et al., 2009). Apparently, genomic analyses of more geographically separated isolates would help shed more light on the genetic diversity and phylogenetic relationships of *Sulfolobus* strains.

All species of *Sulfolobus* are aerobic sulfur oxidizers, and many of them are initially described as autotrophs or mixotrophs (Brock et al., 1972). Two autotrophic carbon fixation cycles have been described in Crenarchaeota, i.e., the 3-hydroxypropionate/4-hydroxybutyrate (HP/HB) cycle and the dicarboxylate/4-hydroxybutyrate (DC/HB) cycle (Berg et al., 2007, 2010; Huber et al., 2008; Ramos-Vera et al., 2011). The HP/HB cycle was confirmed by biochemical assays in Sulfolobales including *Sulfolobus*, *Acidianus,* and *Metallosphaera* (Berg et al., 2007; Teufel et al., 2009; Estelmann et al., 2011; Demmer et al., 2013). H_2, hydrogen sulfide, sulfur, tetrathionate, and pyrite have been described as electron donors for autotrophically-grown *Sulfolobus* (Brock et al., 1972; Wood et al., 1987; Huber and Stetter, 1991; Huber et al., 1992). For the heterotrophical growth of *Sulfolobus*, the conversion of glucose to pyruvate was thought to rely on a non-phosphorylative Entner-Doudoroff (ED) pathway, as shown in *S. solfataricus* and *S. acidocaldarius* (Siebers et al., 1997). However, extensive *in vivo* and *in vitro* assays later indicated that both the semi-phosphorylative and the non-phosphorylative ED pathways might operate in *S. solfataricus* (Ahmed et al., 2005; Ettema et al., 2008). Genomic analyses of the metabolic pathways have been reported for several *Sulfolobus* strains (Sensen et al., 1998; Kawarabayasi et al., 2001; She et al., 2001; Chen et al., 2005; Guo et al., 2011; Jaubert et al., 2013). A further genomic comparison of metabolic pathways in various *Sulfolobus* strains will be of significance to the understanding of the strategies of the organisms to adapt to thriving in their environments. In the present study, we isolated a novel *Sulfolobus* species, denoted strain *Sulfolobus* sp. A20, from an acidic hot spring in Laguna Fumarólica, Costa Rica, and sequenced the genome of the strain. The 16S rRNA gene of strain A20 exhibits the highest sequence identity (∼95.6%) to those of *S. islandicus* and *S. solfataricus* isolates, but the significant differences suggest that strain A20 represents an independent *Sulfolobus* species. The genome of strain A20 was compared with all other available *Sulfolobus* genomes, and analyses of the pathways of carbon, nitrogen and sulfur metabolism in various *Sulfolobus* strains were performed.

MATERIALS AND METHODS

Isolation of Strain A20

A water sample FL1010-1 was collected in October 2010 from a hot spring, known as Laguna Fumarólica (10°46,365′ N and 85°20,646′ W, ∼85°C, pH 3–4), in the Las Palias hydrothermal field (Las Pailas sector), which is located in the southwest flank of the Rincón de la Vieja volcano crater. Rincón de la Vieja volcano (10°49′ N, 85°19′ W), an andesitic volcano in northwestern Costa Rica, belongs to the Circum Pacific Ring of Fire, which is a geothermal belt different from its nearest neighbors, the Yellowstone National Park and the Lassen Volcanic National Park. The sample was concentrated by tangential flow ultrafiltration through a hollow fiber membrane with a molecular mass cutoff of 6 kDa (Tianjin MOTIMO Membrane Technology, China). An enrichment culture was established by inoculating the concentrate in Zillig's medium (Zillig et al., 1994), which contained 0.3% $(NH_4)_2SO_4$, 0.05% $KH_2PO_4{\cdot}3H_2O$, 0.05% $MgSO_4{\cdot}7H_2O$, 0.01% KCl, 0.001% $Ca(NO_3)_2{\cdot}4H_2O$, 0.07% Glycine, 0.05% yeast extract, 0.2% sucrose, and 0.2% of a trace element solution (0.09% $MnCl_2{\cdot}4H_2O$, 0.225% $Na_2B_4O_7{\cdot}10H_2O$, 0.011% $ZnSO_4{\cdot}7H_2O$, 0.0025% $CuCl_2{\cdot}2H_2O$,

0.0015% $NaMoO_4 \cdot 2H_2O$, 0.0005% $CoSO_4 \cdot 7H_2O$). After incubation for 7–10 days at 75°C with shaking at 150 rpm, samples of the grown culture were spread on Zillig's medium plates solidified with 0.8% gelrite. The plates were incubated for 7 days at 75°C. Colonies were picked and purified by re-plating. Observation of the cells of strain A20 was carried out under a transmission electron microscope (JEM-1400, Jeol Ltd., Tokyo, Japan) at 80 kV by negatively staining with 2% uranyl acetate.

Genome Sequencing and Annotation

The genomic DNA of strain A20 was isolated and purified, as described (Chong, 2001), and sequenced on the Pacific Biosciences (PacBio) RS II and Illumina Hiseq 2000 systems at AnnoGenne, Beijing, China. The genome was assembled with SMRT analysis v2.3.0 and RS_HGAP_Assembly.3, and the genome assembly was improved by using the software Pilon (Walker et al., 2014). Identification of protein-coding open reading frames (ORFs) and annotation of the ORFs were performed by NCBI using the NCBI Prokaryotic Genome Annotation Pipeline (https://www.ncbi.nlm.nih.gov/genome/annotation_prok/). Genes were functionally annotated by BLAST search in COG, KEGG, Nr, and Pfam Databases (Camacho et al., 2009; Finn et al., 2011). Putative insertion sequence (IS) elements were identified by BLASTn search against the IS finder Database (http://www-is.biotoul.fr).

Comparative Genomics Analysis

The nucleotide sequences of all genome-sequenced *Sulfolobus* strains and the corresponding amino acid sequences were retrieved from the GenBank database and the NCBI Reference Sequence database (RefSeq) (**Table 1**). The dot plots of any two genomes for their genomic synteny were profiled with Mummer (Kurtz et al., 2004), and DNA-DNA hybridization (DDH) values *in silico* were computed using the Genome-to-Genome Distance Calculator (GGDC) version 2.0 (Meier-Kolthoff et al., 2013) by submitting the genome sequences to DSMZ (http://ggdc.dsmz.de) (Auch et al., 2010). All protein sequences derived from the *Sulfolobus* genomes were compared using all-by-all BLASTp with a threshold *E*-value 10^{-10}, and grouped into orthologous gene families by OrthoMCL (Li et al., 2003). Gene groups consisting of orthologous genes present in all genomes, in more than two but not all genomes or in only one genome were defined as core, variable, or individual gene groups, respectively. A Venn diagram of the orthologous analysis of gene families was built with R version 3.0.2.

Phylogenetic Analysis

The 16S rRNA gene sequences of *Sulfolobus* species were extracted from the genome sequences and aligned using the CLUSTAL X program (Thompson et al., 1997). Phylogenetic trees were constructed using the neighbor-joining, maximum-parsimony, and maximum-likelihood methods implemented in the software package MEGA version 5.0 (Tamura et al., 2011). Evolutionary distances were calculated using Kimura's two-parameter model. The resulting tree topologies were evaluated by bootstrap analysis with 1000 re-samplings.

Metabolic Pathway Assignments

The Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Ogata et al., 1999; Kanehisa and Goto, 2000) was used in the analysis of the metabolic pathways of *Sulfolobus* species. All amino acid sequences derived from the genomes of *Sulfolobus* were submitted to the KEGG database, and the metabolic functions of these sequences were annotated by kass (Moriya et al., 2007). The KO (KEGG Orthology) term and corresponding KEGG pathway for each ORF were automatically generated and provided.

Sequencing Data Accession Number

The genome data of *Sulfolobus* sp. A20 have been deposited in the Genbank database under accession number CP017006.

RESULTS

General Features of *Sulfolobus* sp. A20

Sulfolobus sp. A20 was isolated from a hot spring in Costa Rica. The cells of strain A20 were irregular cocci (0.8–1.0 μm in diameter) with flagella (**Figure 1**). Growth occurred at temperatures between 65 and 85°C, and pH between 2.0 and 4.5. The strain grew optimally at 75–85°C and pH 4.0. The doubling time of the strain was ∼14.3 h under the optimal growth conditions.

The Genome of *Sulfolobus* sp. A20

The genome of strain A20 was sequenced using a combination of PacBio RS II and Illumina Hiseq 2000 sequencing technologies with a 2 × 100 bp mode at a 150-fold and a 700-fold coverage, respectively. The genome consists of a single circular chromosome of 2,688,317 bp with 2591 ORFs, a single 16-23S rRNA cluster, a 5S rRNA gene, 45 tRNA genes and 5 miscellaneous RNA genes (misc RNAs). The average size of an ORF is ∼291 amino acids. No extra-chromosomal genetic elements were detected in the strain. The G+C content of the genome is 34.78%. BLASTp searches identified matches in the protein database at GenBank for ∼97.22% of the total ORFs of strain A20 (2519 ORFs). Among these ORFs, 2223 (∼85.80% of total ORFs) are most closely related to those from the genus *Sulfolobus*, and 227 are closely related to those from other genera of the Sulfolobales. The general features of the strain A20 genome are compared with those of the other sequenced Sulfolobus genomes in **Table 1**.

Strain A20 encodes a complete set of enzymes and proteins involved in DNA transactions, including DNA replication, DNA repair and recombination, and RNA transcription. These proteins are highly conserved among the *Sulfolobus* strains, whose genomes have been sequenced, and share the highest sequence identity with those from *S. islandicaus* or *S. solfataricus*. For example, DNA replication proteins, including ORC1-type DNA replication proteins (BFU36_RS04705, BFU36_RS02195, and BFU36_RS09865), mini-chromosome maintenance protein (MCM, BFU36_RS02210), primase subunits (BFU36_RS01270, BFU36_RS03220, and BFU36_RS03380), proliferating cell nuclear antigen subunits (PCNA, BFU36_RS01275, BFU36_RS03780, and BFU36_RS03820), replication factor

TABLE 1 | *Sulfolobus* strains with complete genome sequences.

Strains	GenBank accession no.	NCBI RefSeq no.	Genome size (Mb)	No. of ORFs	No. of rRNAs	No. of tRNAs	GC%	Habitat
Sulfolobus sp. A20	CP017006	NZ_CP017006	2.69	2591	3	45	34.8	Las Palias, Costa Rica
S. solfataricus P2	AE006641	NC_002754	2.99	2896	3	45	35.8	Naples, Italy
S. solfataricus P1	LT549890	NZ_LT549890	3.03	2967	3	45	35.8	Naples, Italy
S. solfataricus 98/2	CP001800	NC_017274	2.67	2605	3	45	35.8	Yellowstone, USA
S. islandicus REY15A	CP002425	NC_017276	2.52	2535	3	46	35.3	Reykjanes, Iceland
S. islandicus HVE10/4	CP002426	NC_017275	2.66	2692	3	44	35.1	Hvergaardi, Iceland
S. islandicus LAL14/1	CP003928	NC_021058	2.47	2505	3	45	35.1	Iceland
S. islandicus L.S.2.15	CP001399	NC_012589	2.74	2767	3	45	35.1	Lassen, USA
S. islandicus 14.25	CP001400	NC_012588	2.61	2682	3	45	35.1	Kamchatka, Russia
S. islandicus M16.4	CP001402	NC_012726	2.59	2678	3	45	35.0	Kamchatka, Russia
S. islandicus M16.27	CP001401	NC_012632	2.69	2766	3	45	35.0	Kamchatka, Russia
S. islandicus Y57.14	CP001403	NC_012622	2.7	2708	3	48	35.4	Yellowstone, USA
S. islandicus YN15.51	CP001404	NC_012623	2.81	2791	3	46	35.3	Yellowstone, USA
S. acidocaldarius DSM639	CP000077	NC_007181	2.23	2224	3	48	36.7	Yellowstone, USA
S. acidocaldarius N8	CP002817	NC_020246	2.18	2188	3	48	36.7	Hokkaido, Japan
S. acidocaldarius Ron121	CP002818	NC_020247	2.22	2227	3	30	36.7	Ronneburg, Germany
S. acidocaldarius SUSAZ	CP006977	NC_023069	2.06	2038	3	46	36.3	Los Azufres, Mexico
S. tokodaii str.7	BA000023	NC_003106	2.69	2764	3	46	32.8	Kyushu, Japan

C (RFC, BFU36_RS02175, and BFU36_RS02180) and DNA polymerases (BFU36_RS05445, BFU36_RS13105, and BFU36_RS03245), from strain A20 closely resemble their homologs at the amino acid sequence level from the other *Sulfolobus* strains. Strain A20 also encodes small, basic and nucleic acid-binding proteins, i.e., Cren7 (BFU36_RS01545), two Sul7d proteins (BFU36_RS09545 and BFU36_RS11200), and two members of the Sac10b family (BFU36_RS01605 and BFU36_RS01615).

Like other *Sulfolobus* strains, strain A20 carries integrative elements, CRISPR-based immune systems and antitoxin/toxin systems (Guo et al., 2011). About 13 ORFs are annotated as the homologs of transposase, and nine copies of putative insertion sequence (IS) elements are found. Among these IS elements, eight belong to the IS200/605 family and one to the IS607 family. Six CRISPR loci of the two subtypes (I-A and III-B) and cmr1-6 proteins are identified (Grissa et al., 2007). No apparent sequence homology was detected between the spacers and the known sequences of *Sulfolobus*/*Acidianus* viruses. Five copies of family II (VapBC) antitoxin-toxin gene pairs are found in the strain A20 genome.

Dot plot analysis reveals no genomic synteny between strain A20 and any of the genome-sequenced *Sulfolobus* strains. Pairwise DNA-DNA hybridization (DDH) *in silico* between strain A20 and one of the tested *Sulfolobus* strains, including *S. tokodaii* str.7, *S. acidocaldarius* DSM 639, three *S. solfataricus* strains, and four *S. islandicus* strains, produces DDH values between 16.7 and 23.1% (**Table 2**), which are far below the 70% threshold proposed for species definition (Tindall et al., 2009). These results suggest that strain A20 represents a novel *Sulfolobus* species.

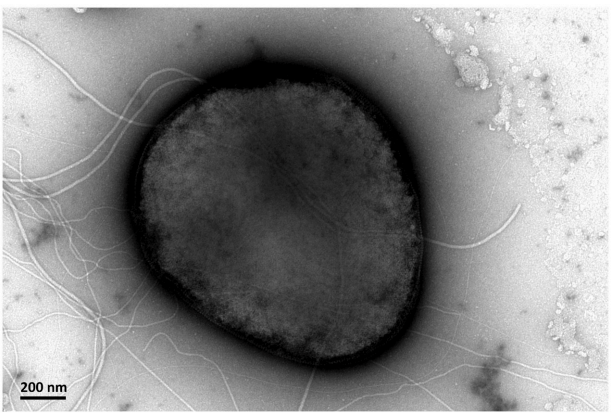

FIGURE 1 | A transmission electron micrograph showing the morphology of *Sulfolobus* sp. A20.

Phylogenetic Analysis of *Sulfolobus* Strains

The 16S rRNA gene sequence of strain A20 was retrieved from the genome sequence of the strain. BLAST searches show that it is most similar (∼95.6% identity) to those from several isolates of *S. islandicus* and *S. solfataricus*. The known *Sulfolobus* species appear to group into two main clades, as indicated by the phylogenetic analysis based on the 16S rRNA gene

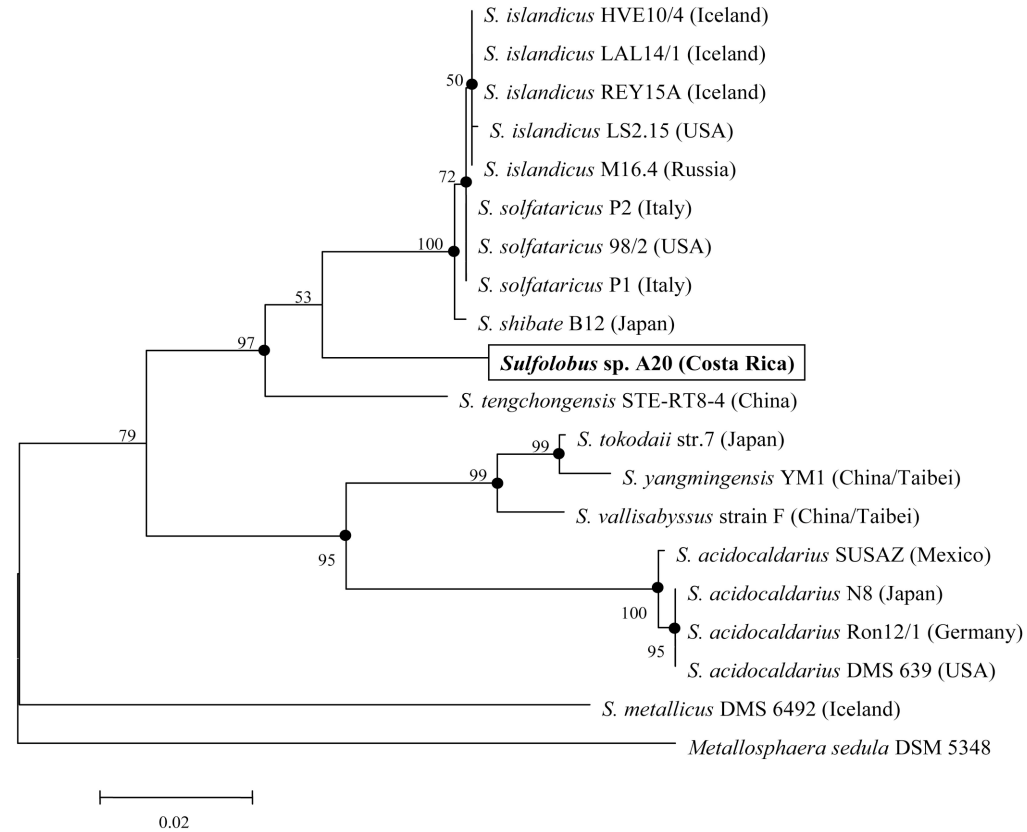

FIGURE 2 | Phylogenetic tree of genome-sequenced *Sulfolobus* strains based on the 16S rRNA gene sequences. *Metallosphaera sedula* DSM 5348 is used as the outgroup. Numbers denote the bootstrap percentages obtained with 1000 replicates.

TABLE 2 | *In silico* DNA-DNA hybridization (DDH) values (%) between *Sulfolobus* strains[a].

Sample	Strain	SSO			SIS				SAC	STO
	A20	P1	P2	98/2	LS2.15	REY15A	HVE10/4	LAL14/1	DSM639	str7
Strain A20	–	16.80	16.80	16.80	16.70	16.70	16.70	16.70	23.10	19.60
SSO-P1		–	94.80	91.50	38.00	37.50	37.30	37.40	18.80	24.00
SSO-P2			–	91.10	40.00	39.30	38.90	39.20	18.20	23.00
SSO-98/2				–	37.20	37.00	36.90	36.90	18.90	24.30
SIS-LS2.15					–	85.80	87.80	81.70	18.40	20.00
SIS-REY15A						–	94.90	94.10	18.10	21.10
SIS-HVE10/4							–	94.00	18.20	21.70
SIS-LAL14/1								–	18.10	21.20
SAC-DSM639									–	15.70
STO-str7										–

[a]*SSO, S. solfataricus; SAC, S. acidocaldarius; SIS, S. islandicus; STO, S. tokodaii.*

sequences (**Figure 2**). Strain A20, together with *S. islandicus*, *S. solfataricus, S. shibatae,* and *S. tengchongensis*, comprise one clade, while *S. acidocaldarius*, *S. tokodaii*, *S. vallisabyssus,* and *S. yangmingensis* make up the other. *S. metallicus* DSM6482, a strictly chemolithoautotrophic and ore-leaching *Sulfolobus* species, appears to be phylogenetically distant from the two main clades.

Core, Variable, and Individual Genes

A total of 18 *Sulfolobus* genomes, including the strain A20 genome, have been completely sequenced so far. To gain insight into the similarities and differences of the genomes from various *Sulfolobus* species, we compared the genome sequences available for the type strains of four *Sulfolobus* species, i.e., *S. acidocaldarius* DSM 639, *S. islandicus* REY15A, *S. solfataricus*

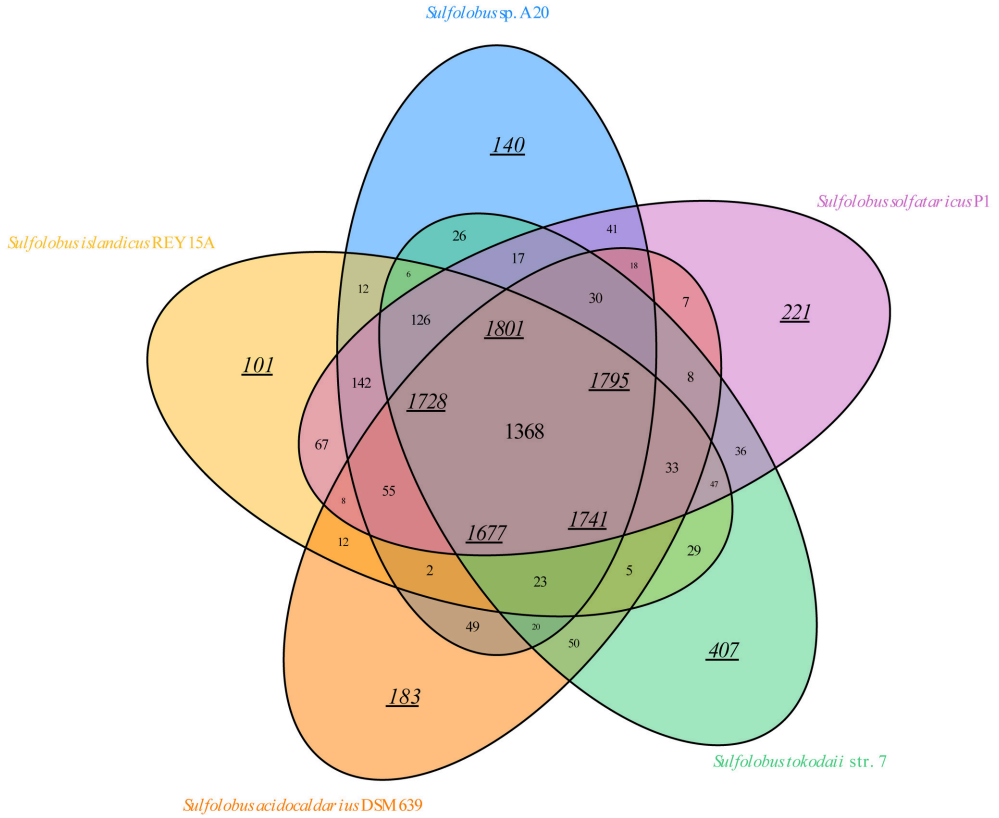

FIGURE 3 | Venn diagram of the conservation of protein-coding ORFs of the genome-sequenced type strains of *Sulfolobus* species. The overlaps between the ellipses show the gene groups shared by different strains with the number of shared gene groups indicated. The number of genes for each strain in a section of the diagram is shown by an underlined number in italics. Each underlined number in italics in the middle of the diagram indicates the number of core genes for a strain analyzed.

P1, and *S. tokodaii* str.7 as well as strain A20. The numbers of predicted ORFs for the five genomes are 2663 ± 439. The ORFs from these genomes are grouped into homologous groups. A total of 1368 gene groups form the core gene groups of the genus *Sulfolobus* (**Figure 3**). This number corresponds to 1801 genes (~69.51% of the total genes) in strain A20 (Table S1). Notably, the difference between these two numbers (i.e., 1368 gene groups vs. 1801 genes) is greater in strain A20 than in other *Sulfolobus* strains analyzed in this study, suggesting greater gene redundancy in A20 than in the other strains. Eight hundred and sixty nine gene groups are found in more than one, but not all, of the five genomes. These groups may constitute the variable parts of the *Sulfolobus* genomes. Strain A20 shares most gene groups with *S. solfataricus* P1 (1797), in agreement with their closest phylogenetic relationship. Moreover, the tested *Sulfolobus* genomes contain variable numbers of individual gene groups. In strain A20, 140 genes (~5.40% of the total ORFs) are not found in other four *Sulfolobus* strains. By comparison, *S. tokodaii* str.7 has the most individual genes (407, or ~14.72% of the total ORFs), whereas *S. islandicus* REY15A has the fewest individual genes (101, or ~3.98% of the total ORFs). Notably, the majority (>80%) of the individual genes encode hypothetical proteins. Conceivably, the exact numbers of core, variable and individual genes in *Sulfolobus* strains will change as the sample size increases but the general pattern of the distribution of these three groups of genes will likely remain.

Metabolic Pathways

KEGG analyses reveal that the genome of strain A20 contains 84, 3 and 10 genes encoding functions in central carbon metabolism, nitrogen metabolism and sulfur metabolism, respectively. As compared to other known *Sulfolobus* genomes, the A20 genome appears to have similar numbers of the genes encoding proteins or protein subunits involved in carbon and sulfur metabolism but fewer genes for nitrogen metabolism. In addition, a total of 15 different ATP-binding cassette (ABC) transporters are identified in the strain A20 genome. By comparison, the numbers of ABC transporters are 10–14 in various *S. islandicus* strains (Guo et al., 2011), 11 in *S. solfataricus* P2 (She et al., 2001), 6 in *S. tokodaii* str.7 (Kawarabayasi et al., 2001), and 3 in *S. acidocaldarius* DSM639 (Chen et al., 2005). The ABC transporters in strain A20 include those for the transportation of trehalose (BFU36_RS00560–BFU36_RS00575, 4 ORFs in all), arabinogalactan oligomer/maltooligosaccharide (BFU36_RS00855–BFU36_RS00870, 4 ORFs), and glucose/arabinose (BFU36_RS07440–BFU36_RS07455, 4 ORFs, and

TABLE 3 | Enzymes involved in the Entner-Doudoroff pathway in strain A20.

KO term	ORF(BFU36_RS)	Enzyme	Pathway[a]
K18125	06060	Glucose dehydrogenase	ED
K05308	06085	Gluconate dehydratase	ED
K18126	06095	2-keto-3-deoxygluconate kinase	sp ED
K11395	06090	2-keto-3-deoxy-6-phosphogluconate aldolase	sp ED
K18978	06100	Glyceraldehyde-3-phosphate dehydrogenase	sp ED
K15634	10260		ED
K15635	03725	Phosphglycorate mutase	
K01689	02015	Enoase	ED
K00873	01505	Pyruvate kinase	ED
K11395	06090	2-keto-3-deoxygluconate aldolase	np ED
K18020	09145		np ED
K18021	09155	Glyceraldehyde dehydrogenase	
K18022	09150		
K11529	02730	Glycerate kinase	np ED

[a]*sp, semi-phosphorylative pathway; np, non-phosphorylative pathway.*

BFU36_RS08120–BFU36_RS08130, 3 ORFs), suggesting the potential ability of strain A20 to utilize a wide range of sugars. There are 16 ORFs belonging to eight glycoside hydrolase (GHs) families, supporting the possibility that strain A20 uses a number of disaccharides and polysaccharides, e.g., cellobiose, maltotriose, mannan, and starch, for growth. A gene (BFU36_RS09315) encoding a putative trehalose glycosyltransferring synthase (TreT) exists in the genome of strain A20. TreT from *Thermoproteus tenax* has been shown to catalyze trehalose synthesis from NDP-glucose or glucose (Kouril et al., 2008). Therefore, it is possible that strain A20 is capable of trehalose synthesis. There is also a cluster of four putative carotenoid biosynthetic genes (BFU36_RS07010–BFU36_RS07025), encoding homologs of lycopene cyclase, phytoene synthase, beta-carotene hydroxylase and phytoene desaturase, respectively, in the strain A20 genome, and these genes are arranged in the same manner as those in *S. solfataricus* (Hemmi et al., 2003) (Table S2).

Central Carbon Metabolism

As revealed by the genome analysis of *S. solfataricus* P2, strain A20 lacks the classical Embden–Meyerhof–Parnas (EMP) and pentose phosphate pathways, since the genes encoding the homologs of the key enzymes in these pathways, i.e., phosphofructokinase in the former and glucose-6-phosphate dehydrogenase, 6-phosphogluconolactonase and 6-phosphogluconate dehydrogenase in the latter, are missing from the genomes (She et al., 2001; Ulas et al., 2012). Like other genome-sequenced *Sulfolobus* strains, strain A20 may utilize glucose through either the semi-phosphorylative or the non-phosphorylative-Entner-Doudoroff (ED) pathway, or both (**Table 3**). Like all other *Sulfolobus* species, strain A20 contains all genes involved in the tricarboxylic acid (TCA) cycle, except for those encoding the alpha-ketoglutarate dehydrogenase complex. The genes for the alpha-ketoglutarate dehydrogenase complex are replaced by those encoding the two subunits of 2-oxoacid:ferredoxin oxidoreductase, an enzyme catalyzing coenzyme A-dependent oxidative decarboxylation of 2-oxoacids (Zillig, 1991; Nishizawa et al., 2005). Intriguingly, the copy number of the genes for 2-oxoacid:ferredoxin oxidoreductase varies among *Sulfolobus* species. A single copy of the genes are present in strain A20, *S. solfataricus* and *S. islandicus*, whereas two copies of the genes are found in *S. acidocaldarius* and *S. tokodaii*, in apparent agreement with the phylogenetic relationship among these species (**Figure 2**).

All tested *Sulfolobus* strains are mixotrophs capable of growing chemolithotrophically on CO_2 with inorganic sulfur compounds (RISCs) as an energy source or heterotrophically on organic compounds (Brock et al., 1972; Keeling et al., 1998; Jan et al., 1999). Two CO_2 fixation pathways, i.e., the 3-hydroxypropionate/4-hydroxybutyrate (HP/HB) cycle and the dicarboxylate/4-hydroxybutyrate (DC/HB) cycle, have been reported to exist in (hyper)thermophilic autotrophic Crenarchaeota (Berg et al., 2010). Like the other 17 *Sulfolobus* genomes, the strain A20 genome contains all of the genes encoding homologs of the enzymes of the two cycles (**Figure 4**).

Nitrogen Metabolism

Like all other *Sulfolobus* genomes, the A20 genome contains genes encoding putative glutamate dehydrogenase (BFU36_RS08195), glutamine synthetase (BFU36_RS04000, BFU36_RS09525, and BFU36_RS10890) and the two subunits of carbamoylphosphate synthase (BFU36_RS02825 and BFU36_RS02830) (**Table 4**). It seems that all *Sulfolobus* strains employ a common strategy in the utilization of ammonium as a universal nitrogen source for the synthesis of glutamate, glutamine and carbamoyl-phosphate.

It is worth noting that four of the *S. islandicus* strains (i.e., REY15A, LAL14/1, M14.25, and M16.27) isolated from Iceland and Russia carry the *narGHJI* operon encoding a nitrate reductase and a nitrate transporter (*narK*) (**Table 4**), and, therefore, are potentially capable of utilizing nitrate. An operon encoding the subunits of urease (UreAB and UreC) and its accessory proteins (UreE, UreF, and UreG) is found in the genomes of *S. islandicus* HEV10/4, *S. tokodaii* str.7 and *S. metallicus* DSM 6482, suggesting that these strains are probably able to hydrolyze urea. Besides, genes for a putative cyanate lyase and a formamidase are found in the genomes of *S. tokodaii* str.7 and *S. islandicus* HEV10/4, respectively, suggesting a broader spectrum of nitrogen sources for these *Sulfolobus* strains.

Sulfur Metabolism

All sequenced *Sulfolobus* genomes contain a gene cluster (BFU36_RS07995–BFU36_RS08005 in strain A20) coding for sulfite reductase, phosphoadenosine phosphosulfate reductase, and sulfate adenylyltransferase (**Tables 4**, **5**). These enzymes probably catalyze the conversion of hydrogen sulfide into

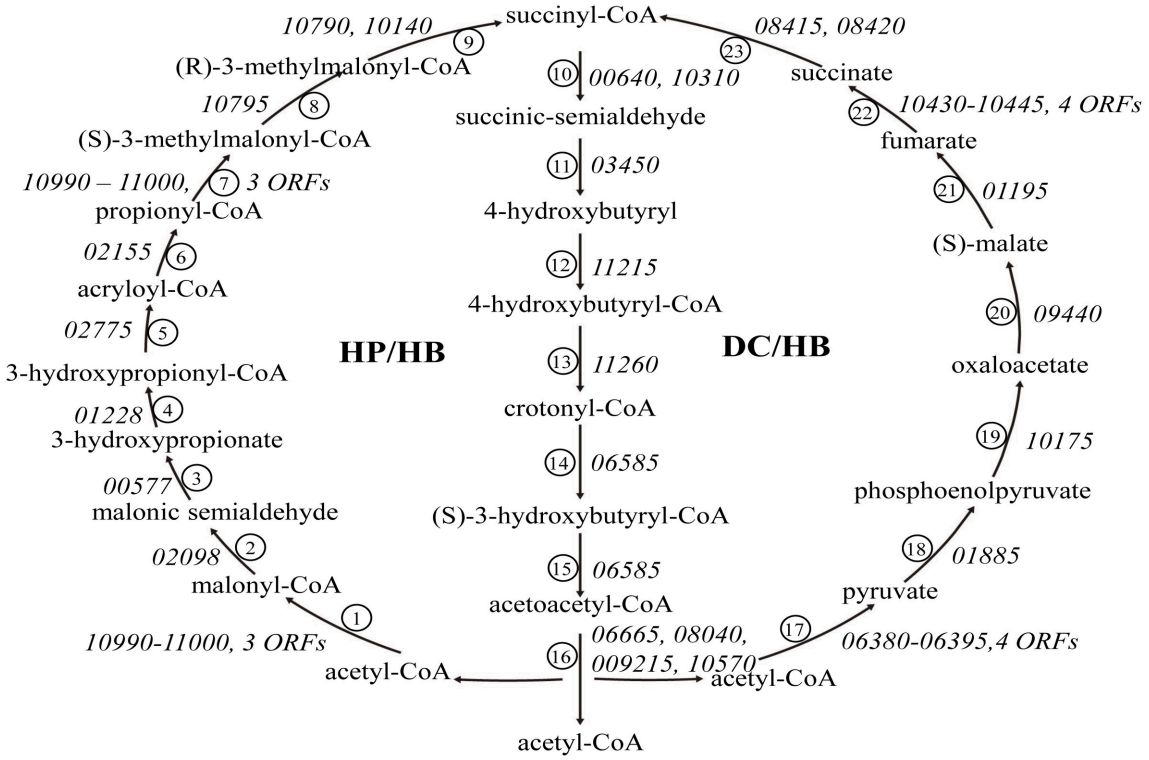

FIGURE 4 | The 3-hydroxypropionate/4-hydroxybutyrate (HP/HB) cycle and the dicarboxylate/ 4-hydroxybutyrate (DC/HB) cycle in *Sulfolobus*. Homologs of the enzymes in the two pathways in strain A20 are indicated by ORF numbers for the strain: ① acetyl-CoA caroxylase, ② malonyl-CoA reductase (NADPH), ③ malonate semialdehyde reductase (NADPH), ④ 3-hydroxypropionate-CoA ligase (AMP-forming), ⑤ 3-hydroxypropionyl-CoA dehydratase, ⑥ acrryloyl-CoA reductase (NADPH), ⑦ propionyl-CoA carboxylase, ⑧ methylmalonyl-CoA epimerase, ⑨ methylmalonyl-CoA mutase, ⑩ succinyl-CoA reductase, ⑪ succinic semialdehyde reductse (NADPH), ⑫ 4-hydroxybutyrate-CoA ligase (AMP forming), ⑬ 4-hydroxybutyryl-CoA dehydratase, ⑭ crotonyl-CoA hydratase, ⑮ (S)-3-hydroxybutyryl-CoA dehydrogenase (NAD$^+$), ⑯ acetoacetyl-CoA beta-ketothiolase, ⑰ pyruvate synthase, ⑱ pyruvate:water dikinase, ⑲ PEP carboxylase, ⑳ malate dehydrogenase (NAD), ㉑ fumarate hydratase, ㉒ fumarate reductase, ㉓ succinyl-CoA synthetase (ADP-forming).

sulfite, and the subsequent transformation of sulfite into sulfate, with concomitant generation of ATP through substrate level phosphorylation (Kappler and Dahl, 2001; Rohwerder and Sand, 2007). A sulfide:quinine oxidoreductase (SQR) gene also exists in all *Sulfolobus* genomes (BFU36_RS09190 in strain A20). SQR may catalyze the oxidation of hydrogen sulfide into polysulfide (Rohwerder and Sand, 2007; Brito et al., 2009). Intriguingly, no homologs of sulfur oxygenase/reductase (SOR), a key enzyme for archaeal sulfur oxidation (Kletzin, 1992; Urich et al., 2006), are found in the genomes of *Sulfolobus* except for that of *S. tokodaii* str.7 (Kawarabayasi et al., 2001). The mechanism of elemental sulfur oxidization in *Sulfolobus* strains lacking SOR remains unknown. Putative genes for sulfur reductase (SRE) and thiosulfate:quinine oxidoreductase (TQO), which serve key roles in the reduction of elemental sulfur into hydrogen sulfide and the transformation of thiosulfate into tetrathionate, respectively (Laska et al., 2003; Guiral et al., 2005; Liu et al., 2012), are also found variably in *Sulfolobus* genomes (**Tables 4**, **5**). Strain A20 and *S. tokodaii* str.7 carry *doxDA* (BFU36_RS07850–BFU36_RS07855), which encode a TQO homolog. *S. islandicus* and *S. solfataricus* have an SRE-encoding gene cluster (*sreABC*) and *doxDA*. *S. acidocaldarius* contains neither of the genes.

DISCUSSION

Sulfolobus sp. A20 was isolated from a hot spring in Costa Rica and the genomic DNA of the strain was completely sequenced. The addition of strain A20 to the growing list of the members of the genus *Sulfolobus* would aid further biogeographic comparison and evolutionary studies of this interesting group of archaea.

Sequence analysis indicates that strain A20 might be a mixotroph. The strain appears to be able to fix CO_2 via the HP/HB cycle. It is also capable of metabolizing glucose through a branched-ED pathway and the TCA cycle, as are other *Sulfolobus* strains. In general, genes involved in central carbon metabolism are conserved in all sequenced *Sulfolobus* genomes. Some of the genes may exist in different numbers of copies and/or be arranged differently among different species, and the differences are in apparent agreement with the phylogenetic relationship rather than the geographical separation of the species (**Figure 2**). It is of interest that genes encoding enzymes for CO_2 fixation through both HP/HB and DC/HB cycles are found in strain A20 and other sequenced *Sulfolobus* genomes. A similar finding has been reported for the genome of *Acidianus hospitalis* W1, a

TABLE 4 | Patterns of the distribution of genes encoding putative enzymes in nitrogen and sulfur metabolism in various *Sulfolobus* strains[a].

KO terms	*S. solfataricus*			*S. islandicus*								*S. acidocaldarius*				
	A20	P2	P1	98/2	HEV10/4	REY15A	LAL14.1	M16.4	M14.25	M16.27	LS2.1.5	YG57.14	YN15.51	DSM639	N8	Ron121	SUSAZ	*S. tokodaii* str.7
NITROGEN METABOLISM[b]																		
K00261 (*gdhA*)	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
K01915 (*glnA*)	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
K01955 (*carB*)	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
K01956 (*carA*)	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
K00370 (*narG*)						■	■		■	■								
K00371 (*narH*)						■	■		■	■								
K00373 (*narJ*)						■	■		■	■								
K00374 (*narI*)						■	■		■	■								
K02575 (*narK*)						■	■		■	■								
K01725 (*cynS*)																		■
K01455 (*for*)					■													
SULFUR METABOLISM[c]																		
K01011 (*tst*)	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
K00381 (*cysI*)	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
K00390 (*cysH*)	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
K00958 (*sat*)	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
K01738 (*cysK*)	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
K01739 (*metB*)	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
K17218 (*sqr*)	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
K16936 (*doxA*)	■	■	■	■	■	■	■	■	■	■	■	■	■					■
K16937 (*doxD*)	■	■	■	■	■	■	■	■	■	■	■	■	■				■	■
K16952 (*sor*)																		■
K17219 (*sreA*)		■	■	■	■	■	■	■	■	■	■	■	■					
K17220 (*sreB*)		■	■	■	■	■	■	■	■	■	■	■	■					
K17221 (*sreC*)		■	■	■	■	■	■	■	■	■	■	■	■					

[a] *The presence of KO terms in nitrogen and sulfur metabolism is shown in gray.*

[b] *Nitrogen metabolism: gdhA, glutamate dehydrogenase (NAD(P)$^+$) [EC 1.4.1.3]; glnA, glutamine synthetase [EC 6.3.1.2]; carB, carbamoyl-phosphate synthase large subunit [EC 6.3.5.5]; carA, carbamoyl-phosphate synthase small subunit [EC 6.3.5.5]; narG, nitrate reductase/nitrite oxidoreductase, alpha subunit [EC 1.7.5.1; 1.7.99.4]; narH, nitrate reductase/nitrite oxidoreductase, beta subunit [EC 1.7.5.1; 1.7.99.4]; narJ, nitrate reductase delta subunit; narI, nitrate reductase gamma subunit [EC 1.7.5.1; 1.7.99.4]; narK, nitrate/nitrite transporter; cynS, cyanate lyase [EC 4.2.1.104]; for, formamidase [EC 3.5.1.49].*

[c] *Sulfur metabolism: tst, thiosulfate sulfurtransferase [EC 2.8.1.1]; cysI, sulfite reductase (NADPH) hemoprotein [EC 1.8.1.2]; cysH, phosphoadenosine phosphosulfate reductase [EC 1.8.4.8; 1.8.4.10]; sat, sulfate adenylyltransferase [EC 2.7.7.4]; cysK, cysteine synthase A [EC 2.5.1.47]; metB, cystathionine gamma-synthase [EC 2.5.1.48]; sqr, sulfide:quinone oxidoreductase [EC 1.8.5.4]; doxA, thiosulfate dehydrogenase [quinone] small subunit [EC 1.8.5.2]; doxD, thiosulfate dehydrogenase [quinone] large subunit [EC 1.8.5.2]; sor, sulfur oxygenase/reductase [EC 1.13.11.55]; sreA, sulfur reductase molybdopterin subunit; sreB, sulfur reductase FeS subunit; sreC, sulfur reductase membrane anchor.*

facultative anaerobe of the *Sulfolobales* (You et al., 2014). The two pathways differ in their sensitivity to oxygen, although they share many enzymes and intermediates in common (Ramos-Vera et al., 2011). The HP/HB cycle is more oxygen-tolerant than the DC/HB cycle since pyruvate synthase, one of key enzymes in the latter cycle, is oxygen sensitive (Jahn et al., 2007; Huber et al., 2008). As aerobes or microaerobes, members of the Sulfolobales have been shown to fix CO_2 through the HP/HB cycle. However, genes coding for putative pyruvate synthase, pyruvate:water dikinase and PEP carboxylase in the DC/HB cycle were found to be expressed, although at a low level, in *Metallosphaera sedula*, an aerobe closely related to *Sulfolobus* strains (Berg et al., 2010). Therefore, we infer that the DC/HB pathway may also be employed by *Sulfolobus* to fix CO_2 under certain conditions.

Similarly, genes involved in the two ED pathways, i.e., the semi-phosphorylated pathway and the non-phosphorylated pathway, are also conserved in all the sequenced *Sulfolobus* genomes. The two ED pathways were named as the archaeal branched ED pathway (Sato and Atomi, 2011), and their functions were verified in *S. solfataricus* (Ahmed et al., 2005). The redundancy of the pathways for central carbon metabolism in *Sulfolobus* may contribute to the adaption of the organisms to thriving in the extreme and oligotrophic habitats.

TABLE 5 | Predicted reactions in sulfur metabolism in *Sulfolobus*[a].

Strains	A20	*S. acidocaldarius*	*S. solfataricus*	*S. islandicus*	*S. tokodaii*
thiosulfate + cyanide ⇋ **sulfite** + thiocyanate (*tst*)	+	+	+	+	+
hydrogen sulfide + $3NADP^+ + 3H_2O$ ⇋ **sulfite** + $3NADPH + 3H^+$ (*sqr*)	+	+	+	+	+
PAP + **sulfite** + thioredoxin disulfide ⇋ **PAPS** + thioredoxin (*cysH*)	+	+	+	+	+
APS + diphosphate ⇋ **sulfate** + ATP (*sat*)	+	+	+	+	+
Sulfur → **hydrogen sulfide** + **thiosulfate** (*sor*)	−	−	−	−	+
Thiosulfate + 6-decylubiquinone ⇋ **tetrathionate** + 6-decylubiquinol (*doxAD*)	+	−	+	+	+
Sulfur + hydrogen ⇋ **hydrogen sulfide** (*sre*)	−	−	+	+	−

[a]*APS, adenylyl sulfate; PAPS, 3′-phosphoadenylyl sulfate; PAP, adenosine 3′, 5′-bisphosphate.*

All *Sulfolobus* genomes contained a complete pathway for ammonium assimilation, which is similar to that found in heterotrophic bacteria (Zalkin, 1993; Guo et al., 2011; Wang et al., 2016), suggesting that *Sulfolobus* prefers to use ammonia as the nitrogen source. Strain A20 is probably unable to use other inorganic nitrogen sources for growth, while several of the *S. islandicus* strains and *S. tokodaii* str.7 might be able to use nitrate, urea, cyanate or formamide as their nitrogen source. These results point to the diversity of nitrogen utilization by *Sulfolobus*. It remains to be determined if the difference in the ability of *Sulfolobus* strains to use inorganic nitrogen compounds correlates with the availability of the nitrogen sources in the habitats of the strains.

Genomic analyses reveal the presence of transposase genes and repeating sequences near the *nar* gene cluster, suggesting the potential mobility of the cluster. The *nar* cluster was found at either of the two genomic sites in four *S. islandicus* strains containing the cluster. In the two *S. islandicus* strains from Iceland (i.e., REY15A and LAL14/1), the *nar* cluster resides on the complementary strand downstream of a sequence encoding a GntR family transcriptional regulator, a CoA ester lyase and an esterase (SIRE_RS02235–SIRE_RS02245 in REY15A and SIL_RS02325–SIL_RS02335 in LAL14/1). This site of potential *nar* insertion is termed insertion site A. On the other hand, in the two strains from Kamchatka (i.e., M16.27 and M14.25), the cluster is located downstream of a sequence encoding a 3-hydroxyacyl-CoA dehydrogenase, an AMP-dependence synthetase and an acetyl-CoA synthetase (M1627_RS04095–M1627_RS04105 in M16.27 and M1425_RS04080–M1425_RS04090 in M14.25). We denote this potential location for the insertion of the *nar* cluster insertion site B. Although only two strains were found to contain the *nar* cluster at insertion site A, this insertion site is present in all *Sulfolobus* strains analyzed in this study. Variation occurs downstream of the site. There are seven types of gene organization downstream of insertion site A in the 18 strains (Tables S3, S4). The tandem array of the three genes at insertion site B is found only in *S. islandicus* strains isolated form Kamchatka, Yellowstone National Park (YNP), and Lassen in USA (Tables S3, S4). Three general patterns of gene arrangement were identified at insertion site B. The two *S. islandicus* strains from USA (i.e., L.S 2.15 and Y57.14) are of one type, and the two Kamchatka *S. islandicus* strains (i.e., M16.27 and M14.25) belong to the other type. Remarkable variation in gene arrangement indicates that the two sites are where active transposition has taken place. The biogeographical difference in genomic location of the *nar* gene cluster presumably resulted from the transposition of the cluster. Since the presence of the *nar* cluster is restricted to *S. islandicus* and some of the strains in this species lack the gene cluster, we hypothesize that the species originally carried the cluster. When it spread to various geographical locations, loss or transposition of the gene cluster occurred, producing variants that thrive in various parts of the globe today. Whether the *nar* cluster was originally acquired through horizontal gene transfer is unclear. However, no significant difference in GC content between the gene cluster and the genome was detected.

Elemental sulfur metabolism is complex in *Sulfolobus*, and relatively low conservation in sulfur metabolism exists among the sequenced genomes. Strain A20 is likely capable of utilizing hydrogen sulfide because of the presence in its genome a conserved gene cluster for sulfur metabolism (Kawarabayasi et al., 2001; Chen et al., 2005). Although most *Sulfolobus* strains have been described as sulfur-oxidizing microbes (Brock et al., 1972), the biochemical process of elemental sulfur oxidation has yet to be fully understood. The *sor* gene encoding the classical sulfur oxygenase/reductase required for the initial step in the archaeal sulfur oxidation pathway (Urich et al., 2006) is present in none of the sequenced *Sulfolobus* genomes except for the genome of *S. tokodaii* str.7 (Kawarabayasi et al., 2001; She et al., 2001; Chen et al., 2005; Guo et al., 2011; Jaubert et al., 2013). Instead, there is a gene cluster encoding sulfur reductase (SRE), which reduces S^0 with the help of a hydrogenase in anaerobically grown *Acidianus ambivalens* (Laska et al., 2003), in the genomes of *S. solfataricus* and *S. islandicus*. However, no hydrogenase genes have been identified in the two species. So, whether and how the sulfur reductase catalyzes sulfur reduction in the absence of a hydrogenase under aerobic conditions remains to be determined. It has been reported that *Sulfolobus tokodaii* str.7 grows poorly in the presence of elemental sulfur under the facultatively chemolithotrophic conditions (Suzuki et al., 2002), although it encodes a homolog of the classical sulfur oxygenase/reductase. However, the strain was able to oxidize hydrogen sulfide into sulfate (Kawarabayasi et al., 2001), suggesting the possibility of functional divergence of the homologs of sulfur oxygenase/reductase in *Sulfolobus*. Therefore,

further investigation is needed to understand the mechanisms underlining elemental sulfur metabolism in *Sulfolobus*.

The *sre* gene cluster is flanked upstream by a hypothetical protein and a 4Fe-4S ferredoxin and downstream by another 4Fe-4S ferredoxin and two hypothetical proteins (Tables S3, S4). This entire sequence is located downstream of a *cupin* gene. Based on the presence of genes between *cupin* and the *sre* cluster, three types of gene arrangement were identified at this site. A transposase gene is located between *cupin* and the *sre* cluster in *S. solfataricus* strains P1 and P2, both of which were isolated from Naples, Italy. However, no transposase gene at this site was found in *S. solfataricus* strain 98/2 or *S. islandicus* strains from YNP. Instead, a gene for the large subunit of nitricoxide redutase is present at this site in these strains. By comparison, a pseudogene is in the place of the transposase gene in *S. islandicus* strain 14.25 from Kamchatka. The two other *S. islandicus* strains (i.e., M16.4 and M16.27) from Kamchatka contain multiple transposase genes as well as hypothetical proteins at the site. Patterns of gene arrangement upstream of the *sre* gene cluster appear to carry distinct geographical markers, since they exhibit similarity among closely located strains of the same species. Whether the function of the *sre* gene cluster is affected by its genomic environment is unclear.

A putative *tusA-dsrE2-dsrE3A* gene cluster is linked to the *hdr* cluster (*hdrC1-hdrB1A-hyp-hdrC2-hdrB2*) in all *Sulfolobus* genomes. The *hdr* cluster encodes a heterodisulfide-reductase complex, which may be involved in sulfur transfer and reversible reduction of the disulfide bond X-S-S-X in *Acidithiobacillus ferrrooxidans* (Quatrini et al., 2009; Liu et al., 2014), while the *tusA-dsrE2-dsrE3A* gene cluster may encode functions in the transformation of tetrathionate into thiosulfate in *Metallosphaera cuprina* (Liu et al., 2014). How the two genomically linked gene clusters function in sulfur metabolism remains to be understood.

Taken together, our genomic analyses reveal that these *Sulfolobus* species are conserved in central carbon metabolism, but differ in the ability to use inorganic nitrogen and sulfur sources. The ability of *Sulfolobus* to utilize nitrate or sulfur is encoded by a gene cluster flanked by IS elements or their remnants. These clusters appear to have become fixed at a specific genomic site in some strains and lost in other strains during the course of evolution.

AUTHOR CONTRIBUTIONS

XD and LH designed the project. XD and ZZ analyzed the data. HW, LW, YZ, ZD, MM-L, and WH-A collected sample, purified the strain and prepared the genomic DNA for sequencing. KL and XZ performed bioinformatic analysis of the genome sequences. CJ and CL analyzed the pathways of sulfur metabolism. LH, XD and ZZ wrote the manuscript.

ACKNOWLEDGMENTS

We thank Drs. Hailiang Dong and Yong Tao for their valuable comments. This work was supported by National Natural Science Foundation of China grant 31130003. Sampling was partially supported by Grant VI 801–B0–530 from Vicerrectoría de Investigación, Universidad de Costa Rica (San José, Costa Rica). Access to the site and collecting permits were respectively granted by Biodiversity Institutional Commission (University of Costa Rica) (Resolution No. 011, 2010) and Guanacaste Conservation Area (Resolution No. ACG-PI-018-2012), Ministry of Environment, Energy and Telecommunications, Costa Rica.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2016.01902/full#supplementary-material

REFERENCES

Ahmed, H., Ettema, T. J., Tjaden, B., Geerling, A. C., van der Oost, J., and Siebers, B. (2005). The semi-phosphorylative Entner-Doudoroff pathway in hyperthermophilic archaea: a re-evaluation. *Biochem. J.* 390(Pt 2), 529–540. doi: 10.1042/BJ20041711

Arnold, H. P., Ziese, U., and Zillig, W. (2000). SNDV, a novel virus of the extremely thermophilic and acidophilic archaeon *Sulfolobus. Virology* 272, 409–416. doi: 10.1006/viro.2000.0375

Auch, A. F., von Jan, M., Klenk, H. P., and Göker, M. (2010). Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand. Genomic Sci.* 2, 117–134. doi: 10.4056/sigs.531120

Bell, S. D., Botting, C. H., Wardleworth, B. N., Jackson, S. P., and White, M. F. (2002). The interaction of Alba, a conserved archaeal, chromatin protein, with Sir2 and its regulation by acetylation. *Science* 296, 148–151. doi: 10.1126/science.1070506

Berg, I. A., Kockelkorn, D., Buckel, W., and Fuchs, G. (2007). A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in *Archaea. Science* 318, 1782–1786. doi: 10.1126/science.1149976

Berg, I. A., Ramos-Vera, W. H., Petri, A., Huber, H., and Fuchs, G. (2010). Study of the distribution of autotrophic CO_2 fixation cycles in *Crenarchaeota. Microbiology* 156, 256–269. doi: 10.1099/mic.0.034298-0

Brito, J. A., Sousa, F. L., Stelter, M., Bandeiras, T. M., Vonrhein, C., Teixeira, M., et al. (2009). Structural and functional insights into sulfide:quinone oxidoreductase. *Biochemistry* 48, 5613–5622. doi: 10.1021/bi9003827

Brock, T. D., Brock, K. M., Belly, R. T., and Weiss, R. L. (1972). *Sulfolobus:* a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Arch. Mikrobiol.* 84, 54–68. doi: 10.1007/BF00408082

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST plus: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421

Chen, L. M., Brügger, K., Skovgaard, M., Redder, P., She, Q. X., Torarinsson, E., et al. (2005). The genome of *Sulfolobus acidocaldarius*, a model organism of the *Crenarchaeota. J. Bacteriol.* 187, 4992–4999. doi: 10.1128/JB.187.14.4992-4999.2005

Chong, L. (2001). Molecular cloning – A laboratory manual, 3rd edition. *Science* 292, 446–446. doi: 10.1126/science.1060677

Demmer, U., Warkentin, E., Srivastava, A., Kockelkorn, D., Pötter, M., Marx, A., et al. (2013). Structural basis for a bispecific $NADP^+$ and CoA binding site in an archaeal malonyl-coenzyme A reductase. *J. Biol. Chem.* 288, 6363–6370. doi: 10.1074/jbc.M112.421263

Estelmann, S., Hügler, M., Eisenreich, W., Werner, K., Berg, I. A., Ramos-Vera, W. H., et al. (2011). Labeling and enzyme studies of the central carbon metabolism in *Metallosphaera sedula*. *J. Bacteriol.* 193, 1191–1200. doi: 10.1128/JB.01155-10

Ettema, T. J., Ahmed, H., Geerling, A. C., van der Oost, J., and Siebers, B. (2008). The non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase (GAPN) of *Sulfolobus solfataricus*: a key-enzyme of the semi-phosphorylative branch of the Entner-Doudoroff pathway. *Extremophiles* 12, 75–88. doi: 10.1007/s00792-007-0082-1

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37. doi: 10.1093/nar/gkr367

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35, W52–W57. doi: 10.1093/nar/gkm360

Grogan, D., Palm, P., and Zillig, W. (1990). Isolate B12, which harbors a virus-like element, represents a new species of the archaebacterial genus *Sulfolobus, Sulfolobus shibatae*, sp. nov. *Arch. Microbiol.* 154, 594–599. doi: 10.1007/BF00248842

Grogan, D. W. (1989). Phenotypic characterization of the archaebacterial genus *Sulfolobus:* comparison of five wild-type strains. *J. Bacteriol.* 171, 6710–6719. doi: 10.1128/jb.171.12.6710-6719.1989

Grogan, D. W., Ozarzak, M. A., and Bernander, R. (2008). Variation in gene content among geographically diverse *Sulfolobus* isolates. *Environ. Microbiol.* 10, 137–146. doi: 10.1111/j.1462-2920.2007.01439.x

Guiral, M., Tron, P., Aubert, C., Gloter, A., Iobbi-Nivol, C., and Giudici-Orticoni, M. T. (2005). A membrane-bound multienzyme, hydrogen-oxidizing, and sulfur-reducing complex from the hyperthermophilic bacterium *Aquifex aeolicus*. *J. Biol. Chem.* 280, 42004–42015. doi: 10.1074/jbc.M508034200

Guo, L., Brügger, K., Liu, C., Shah, S. A., Zheng, H. J., Zhu, Y. Q., et al. (2011). Genome analyses of Icelandic strains of *Sulfolobus islandicus*, model organisms for genetic and virus-host interaction studies. *J. Bacteriol.* 193, 1672–1680. doi: 10.1128/JB.01487-10

Hemmi, H., Ikejiri, S., Nakayama, T., and Nishino, T. (2003). Fusion-type lycopene beta-cyclase from a thermoacidophilic archaeon *Sulfolobus solfataricus*. *Biochem. Biophys. Res. Commun.* 305, 586–591. doi: 10.1016/S0006-291X(03)00798-8

Huber, G., Drobner, E., Huber, H., and Stetter, K. O. (1992). Growth by aerobic oxidation of molecular hydrogen in *Archaea*—a metabolic property so far unknown for this domain. *Syst. Appl. Microbiol.* 15, 502–504. doi: 10.1016/S0723-2020(11)80108-6

Huber, G., and Stetter, K. O. (1991). *Sulfolobus metallicus,* sp. nov, a novel strictly chemolithoautotrophic thermophilic archaeal species of metal-mobilizers. *Syst. Appl. Microbiol.* 14, 372–378. doi: 10.1016/S0723-2020(11)80312-7

Huber, H., Gallenberger, M., Jahn, U., Eylert, E., Berg, I. A., Kockelkorn, D., et al. (2008). A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation cycle in the hyperthermophilic Archaeum *Ignicoccus hospitalis*. *Proc. Natl. Acad. Sci. U.S.A.* 105, 7851–7856. doi: 10.1073/pnas.0801043105

Jahn, U., Huber, H., Eisenreich, W., Hügler, M., and Fuchs, G. (2007). Insights into the autotrophic CO^2 fixation pathway of the archaeon *Ignicoccus hospitalis*: comprehensive analysis of the central carbon metabolism. *J. Bacteriol.* 189, 4108–4119. doi: 10.1128/JB.00047-07

Jan, R. L., Wu, J., Chaw, S. M., Tsai, C. W., and Tsen, S. D. (1999). A novel species of thermoacidophilic archaeon, *Sulfolobus yangmingensis* sp. nov. *Int. J. Syst. Bacteriol.* 49(Pt 4), 1809–1816. doi: 10.1099/00207713-49-4-1809

Jaubert, C., Danioux, C., Oberto, J., Cortez, D., Bize, A., Krupovic, M., et al. (2013). Genomics and genetics of *Sulfolobus islandicus* LAL14/1, a model hyperthermophilic archaeon. *Open Biol.* 3:130010. doi: 10.1098/rsob.130010

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27

Kappler, U., and Dahl, C. (2001). Enzymology and molecular biology of prokaryotic sulfite oxidation. *FEMS Microbiol. Lett.* 203, 1–9. doi: 10.1111/j.1574-6968.2001.tb10813.x

Kawarabayasi, Y., Hino, Y., Horikawa, H., Jin-no, K., Takahashi, M., Sekine, M., et al. (2001). Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain7. *DNA Res.* 8, 123–140. doi: 10.1093/dnares/8.4.123

Keeling, P. J., Klenk, H. P., Singh, R. K., Schenk, M. E., Sensen, C. W., Zillig, W., et al. (1998). *Sulfolobus islandicus* plasmids pRN1 and pRN2 share distant but common evolutionary ancestry. *Extremophiles* 2, 391–393. doi: 10.1007/s007920050083

Kletzin, A. (1992). Molecular characterization of the sor gene, which encodes the sulfur oxygenase reductase of the thermoacidophilic archaeum *Desulfurolobus ambivalens*. *J. Bacteriol.* 174, 5854–5859. doi: 10.1128/jb.174.18.5854-5859.1992

Kouril, T., Zaparty, M., Marrero, J., Brinkmann, H., and Siebers, B. (2008). A novel trehalose synthesizing pathway in the hyperthermophilic crenarchaeon *Thermoproteus tenax*: the unidirectional TreT pathway. *Arch. Microbiol.* 190, 355–369. doi: 10.1007/s00203-008-0377-3

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12. doi: 10.1186/gb-2004-5-2-r12

Laska, S., Lottspeich, F., and Kletzin, A. (2003). Membrane-bound hydrogenase and sulfur reductase of the hyperthermophilic and acidophilic archaeon *Acidianus ambivalens*. *Microbiology* 149, 2357–2371. doi: 10.1099/mic.0.26455-0

Li, L., Stoeckert, C. J., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503

Liu, L. J., Stockdreher, Y., Koch, T., Sun, S. T., Fan, Z., Josten, M., et al. (2014). Thiosulfate transfer mediated by DsrE/TusA homologs from acidothermophilic sulfur-oxidizing archaeon *Metallosphaera cuprina*. *J. Biol. Chem.* 289, 26949–26959. doi: 10.1074/jbc.M114.591669

Liu, Y., Beer, L. L., and Whitman, W. B. (2012). Sulfur metabolism in archaea reveals novel processes. *Environ. Microbiol.* 14, 2632–2644. doi: 10.1111/j.1462-2920.2012.02783.x

Mao, D., and Grogan, D. (2012). Genomic evidence of rapid, global-scale gene flow in a *Sulfolobus* species. *ISME J.* 6, 1613–1616. doi: 10.1038/ismej.2012.20

McCarthy, S., Gradnigo, J., Johnson, T., Payne, S., Lipzen, A., Martin, J., et al. (2015). Complete genome sequence of *Sulfolobus solfataricus* strain 98/2 and evolved derivatives. *Genome Announc.* 3:e00549-15. doi: 10.1128/genomeA.00549-15

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P., and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14:60. doi: 10.1186/1471-2105-14-60

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35, W182–W185. doi: 10.1093/nar/gkm321

Nishizawa, Y., Yabuki, T., Fukuda, E., and Wakagi, T. (2005). Gene expression and characterization of two 2-oxoacid:ferredoxin oxidoreductases from *Aeropyrum pernix* K1. *FEBS Lett.* 579, 2319–2322. doi: 10.1016/j.febslet.2004.11.117

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 27, 29–34. doi: 10.1093/nar/27.1.29

Quatrini, R., Appia-Ayme, C., Denis, Y., Jedlicki, E., Holmes, D. S., and Bonnefoy, V. (2009). Extending the models for iron and sulfur oxidation in the extreme Acidophile *Acidithiobacillus ferrooxidans*. *BMC Genomics* 10:394. doi: 10.1186/1471-2164-10-394

Ramos-Vera, W. H., Weiss, M., Strittmatter, E., Kockelkorn, D., and Fuchs, G. (2011). Identification of missing genes and enzymes for autotrophic carbon fixation in *Crenarchaeota*. *J. Bacteriol.* 193, 1201–1211. doi: 10.1128/JB.01156-10

Reno, M. L., Held, N. L., Fields, C. J., Burke, P. V., and Whitaker, R. J. (2009). Biogeography of the *Sulfolobus islandicus* pan-genome. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8605–8610. doi: 10.1073/pnas.0808945106

Rice, G., Stedman, K., Snyder, J., Wiedenheft, B., Willits, D., Brumfield, S., et al. (2001). Viruses from extreme thermal environments. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13341–13345. doi: 10.1073/pnas.231170198

Rohwerder, T., and Sand, W. (2007). Oxidation of inorganic sulfur compounds in acidophilic prokaryotes. *Eng. Life Sci.* 7, 301–309. doi: 10.1002/elsc.200720204

Sato, T., and Atomi, H. (2011). Novel metabolic pathways in *Archaea*. *Curr. Opin. Microbiol.* 14, 307–314. doi: 10.1016/j.mib.2011.04.014

Sensen, C. W., Charlebois, R. L., Chow, C., Clausen, I. G., Curtis, B., Doolittle, W. F., et al. (1998). Completing the sequence of the *Sulfolobus solfataricus* P2 genome. *Extremophiles* 2, 305–312. doi: 10.1007/s007920050073

She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., et al. (2001). The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc. Natl. Acad. Sci. U.S.A.* 98, 7835–7840. doi: 10.1073/pnas.141222098

Siebers, B., Wendisch, V. F., and Hensel, R. (1997). Carbohydrate metabolism in *Thermoproteus tenax*: *in vivo* utilization of the non-phosphorylative

Entner-Doudoroff pathway and characterization of its first enzyme, glucose dehydrogenase. *Arch. Microbiol.* 168, 120–127. doi: 10.1007/s002030050477

Suzuki, T., Iwasaki, T., Uzawa, T., Hara, K., Nemoto, N., Kon, T., et al. (2002). *Sulfolobus tokodaii* sp. nov. (f. *Sulfolobus* sp. strain 7), a new member of the genus *Sulfolobus* isolated from Beppu Hot Springs, Japan. *Extremophiles* 6, 39–44. doi: 10.1007/s007920100221

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121

Teufel, R., Kung, J. W., Kockelkorn, D., Alber, B. E., and Fuchs, G. (2009). 3-hydroxypropionyl-coenzyme A dehydratase and acryloyl-coenzyme A reductase, enzymes of the autotrophic 3-hydroxypropionate/4-hydroxybutyrate cycle in the Sulfolobales. *J. Bacteriol.* 191, 4572–4581. doi: 10.1128/JB.00068-09

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882. doi: 10.1093/nar/25.24.4876

Tindall, B. J., Rosselló-Móra, R., Busse, H. J., Ludwig, W., and Kampfer, P. (2009). Notes on the characterization of prokaryote strains for taxonomic purposes. *Int. J. Syst. Evol. Microbiol.* 60, 249–266. doi: 10.1099/ijs.0.016949-0

Ulas, T., Riemer, S. A., Zaparty, M., Siebers, B., and Schomburg, D. (2012). Genome-scale reconstruction and analysis of the metabolic network in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *PLoS ONE* 7:e43401. doi: 10.1371/journal.pone.0043401

Urich, T., Gomes, C. M., Kletzin, A., and Frazao, C. (2006). X-ray structure of a self-compartmentalizing sulfur cycle metalloenzyme. *Science* 311, 996–1000. doi: 10.1126/science.1120306

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* 9:e112963. doi: 10.1371/journal.pone.0112963

Wang, H., Peng, N., Shah, S. A., Huang, L., and She, Q. X. (2015). *Archaeal* extrachromosomal genetic elements. *Microbiol. Mol. Biol. Rev.* 79, 117–152. doi: 10.1128/MMBR.00042-14

Wang, J., Yan, D., Dixon, R., and Wang, Y. P. (2016). Deciphering the principles of bacterial nitrogen dietary preferences: a strategy for nutrient containment. *MBio* 7:e00792-16. doi: 10.1128/mBio.00792-16

Whitaker, R. J., Grogan, D. W., and Taylor, J. W. (2003). Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* 301, 976–978. doi: 10.1126/science.1086909

Wood, A. P., Kelly, D. P., and Norris, P. R. (1987). Autotrophic growth of four *Sulfolobus* strains on tetrathionate and the effect of organic nutrients. *Arch. Microbiol.* 146, 382–389. doi: 10.1007/BF00410940

Xiang, X., Dong, X., and Huang, L. (2003). *Sulfolobus tengchongensis* sp. nov., a novel thermoacidophilic archaeon isolated from a hot spring in Tengchong, China. *Extremophiles* 7, 493–498. doi: 10.1007/s00792-003-0355-2

You, X., Zhang, Q., Zheng, H., and Jiang, C. (2014). [Adaptation of *Acidianus hospitalis* W1 to oligotrophic and acidic hot spring environments]. *Wei Sheng Wu Xue Bao* 54, 1193–1203. doi: 10.13343/j.cnki.wsxb.2014.10.012

Zalkin, H. (1993). The amidotransferases. *Adv. Enzymol. Relat. Areas Mol. Biol.* 66, 203–309. doi: 10.1002/9780470123126.ch5

Zhang, C., Krause, D. J., and Whitaker, R. J. (2013). *Sulfolobus islandicus*: a model system for evolutionary genomics. *Biochem. Soc. Trans.* 41, 458–462. doi: 10.1042/BST20120338

Zillig, W. (1991). Comparative biochemistry of *Archaea* and *Bacteria*. *Curr. Opin. Genet. Dev.* 1, 544–551. doi: 10.1016/S0959-437X(05)80206-0

Zillig, W., Kletzin, A., Schleper, C., Holz, I., Janekovic, D., Hain, J., et al. (1994). Screening for Sulfolobales, their plasmids and their viruses in icelandic solfataras. *Syst. Appl. Microbiol.* 16, 609–628. doi: 10.1016/S0723-2020(11)80333-4

Zuo, G., Xu, Z., and Hao, B. (2015). Phylogeny and taxonomy of *Archaea*: a comparison of the whole-genome-based CVTree approach with 16S rRNA sequence analysis. *Life (Basel)* 5, 949–968. doi: 10.3390/life5010949

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Dai, Wang, Zhang, Li, Zhang, Mora-López, Jiang, Liu, Wang, Zhu, Hernández-Ascencio, Dong and Huang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.