



Gene Turnover Contributes to the Evolutionary Adaptation of *Acidithiobacillus caldus*: Insights from Comparative Genomics

Xian Zhang^{1,2}, Xueduan Liu^{1,2}, Qiang He³, Weiling Dong¹, Xiaoxia Zhang⁴, Fenliang Fan⁵, Deliang Peng⁶, Wenkun Huang⁶ and Huaqun Yin^{1,2*}

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, ³ Department of Civil and Environmental Engineering, the University of Tennessee, Knoxville, TN, USA, ⁴ Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing, China, ⁵ Key Laboratory of Plant Nutrition and Fertilizer, Chinese Academy of Agricultural Sciences, Beijing, China, ⁶ State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, China

OPEN ACCESS

Edited by:

Edgardo Donati,
National University of La Plata,
Argentina

Reviewed by:

Om V. Singh,
University of Pittsburgh, USA
Jorge Hernan Valdes,
Center for Genomics
and Bioinformatics/Universidad
Mayor, Chile

*Correspondence:

Huaqun Yin
yinhuaqun@gmail.com

Specialty section:

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

Received: 27 June 2016

Accepted: 22 November 2016

Published: 06 December 2016

Citation:

Zhang X, Liu X, He Q, Dong W,
Zhang X, Fan F, Peng D, Huang W
and Yin H (2016) Gene Turnover
Contributes to the Evolutionary
Adaptation of *Acidithiobacillus caldus*:
Insights from Comparative Genomics.
Front. Microbiol. 7:1960.
doi: 10.3389/fmicb.2016.01960

Acidithiobacillus caldus is an extremely acidophilic sulfur-oxidizer with specialized characteristics, such as tolerance to low pH and heavy metal resistance. To gain novel insights into its genetic complexity, we chosen six *A. caldus* strains for comparative survey. All strains analyzed in this study differ in geographic origins as well as in ecological preferences. Based on phylogenomic analysis, we clustered the six *A. caldus* strains isolated from various ecological niches into two groups: group 1 strains with smaller genomes and group 2 strains with larger genomes. We found no obvious intraspecific divergence with respect to predicted genes that are related to central metabolism and stress management strategies between these two groups. Although numerous highly homogeneous genes were observed, high genetic diversity was also detected. Preliminary inspection provided a first glimpse of the potential correlation between intraspecific diversity at the genome level and environmental variation, especially geochemical conditions. Evolutionary genetic analyses further showed evidence that the difference in environmental conditions might be a crucial factor to drive the divergent evolution of *A. caldus* species. We identified a diverse pool of mobile genetic elements including insertion sequences and genomic islands, which suggests a high frequency of genetic exchange in these harsh habitats. Comprehensive analysis revealed that gene gains and losses were both dominant evolutionary forces that directed the genomic diversification of *A. caldus* species. For instance, horizontal gene transfer and gene duplication events in group 2 strains might contribute to an increase in microbial DNA content and novel functions. Moreover, genomes undergo extensive changes in group 1 strains such as removal of potential non-functional DNA, which results in the formation of compact and streamlined genomes. Taken together, the findings presented herein show highly frequent gene turnover of *A. caldus* species that inhabit extremely acidic environments, and shed new light on the contribution of gene turnover to the evolutionary adaptation of acidophiles.

Keywords: *Acidithiobacillus caldus*, comparative genomics, intraspecific diversity, gene turnover, evolutionary adaptation

INTRODUCTION

Acidithiobacillus caldus (formerly *Thiobacillus caldus*), a moderately thermophilic, obligately chemolithoautotrophic, and extremely acidophilic sulfur-oxidizing bacterium (Hallberg and Lindström, 1994, 1996), is of interest for its potential role in industrial bioleaching (Rawlings, 1998; Dopson and Lindström, 1999). *A. caldus* exploits elemental sulfur and a wide range of reduced inorganic sulfur compounds at moderately high temperatures to support autotrophic growth (Mangold et al., 2011; Chen et al., 2012). It is the primary member of a consortium of sulfur oxidizers in different toxic-laden acidic environments, which are termed “extreme environments,” including coal pile and spoil, gold-bearing reactor operation, as well as low-grade copper bioleaching heap (Valdes et al., 2009; You et al., 2011; Zhang et al., 2016c). Considering that *A. caldus* inhabits harsh environments for prolonged periods and accommodates both sudden stress changes and long-term stress conditions in various habitats, gene flow and genetic drift might frequently occur. As such, the flexible gene repertoire generated by gene exchange has imparted *A. caldus* with extensive genetic material for diversification of function and phenotype. Therefore, research focusing on the correlation between genomic changes and evolutionary adaptation is of great interest.

The accumulation of genomic changes underlying evolutionary adaptation has often been viewed as a complex process, and has been subject to many influences and complications (Barrick et al., 2009). As stated by Carretero-Paulet et al. (2015), homologous genes derived from newly formed subgenomes might undergo asymmetric fractionation via mutational events, which include nucleotide substitutions, gene gains and losses, and changes in genomic structure and organization (Librado et al., 2014). In terms of neutral mutation theory, mutations underlying gene and genome evolution, though not necessarily beneficial, should accumulate at a constant rate by drift (Kimura, 1984). Another view is that the substitution rates for beneficial and deleterious mutations depend on environmental selection, as well as population size and structure (Gillespie, 1991; Ohta, 1992). For many years, the crucial role of gene and genome duplications (namely, neofunctionalization and subfunctionalization) in governing organismal evolution has been acknowledged (Ohno, 1970; Force et al., 1999; Innan and Kondrashov, 2010; Kulmuni et al., 2013). Only in recent decades has great attention been paid to the molecular mechanisms of gene loss (deletion or pseudogenization) as a pervasive source of genetic change, which is believed to be another key evolutionary event that causes adaptive phenotypic diversity (Albalat and Cañestro, 2016). In recent years, a number of analytical methods for population genomics and molecular evolution have provided substantial evidence to determine the relative contribution of diverse evolutionary forces, which shape genome organization, architecture, and diversity in response to environmental perturbations (Librado et al., 2014). In eukaryotes, gene family evolution has often been modeled after a phylogenetic birth-and-death (BD) process (Nei and Rooney, 2005). This BD model, though suitable to account for single-gene duplications, might

not be appropriate for calculating gene turnover rates given that horizontal gene transfer (HGT) events occur in certain organisms (Librado et al., 2014). However, an alternative gain-and-death (GD) stochastic model in a maximum-likelihood statistical framework was applied to circumvent this limitation (Librado et al., 2012). Unlike the birth process, gains in the developed GD model can accommodate all kinds of gene acquisitions, irrespective of their original source, even including HGT (Librado et al., 2014). In this study, we are interested in whether the aforementioned theoretical and analytical approaches can be applied to explain the relationship between genetic change and adaptive evolution of *A. caldus* inhabiting extraordinarily extreme environments.

Members of *A. caldus* species are ubiquitous throughout many sulfur-rich acidic environments worldwide (Table 1), indicating their adaptation to various niches with high concentrations of toxic substrates, such as coal spoil, gold-bearing bioleaching reactor, and copper mine tailing. In recent years, revolutionary technologies and tools have allowed for the rapid characterization of microbial genome sequences (MacLean et al., 2009; Metzker, 2010). Accurate analyses of gene family evolution have been made possible owing to the increasing availability of closely related genomes (Hahn et al., 2007; Sánchez-Gracia et al., 2009; Vieira and Rozas, 2011). Furthermore, acquisition of numerous additional genomes has fuelled a new field termed comparative genomics (Jacobsen et al., 2011), which is useful for investigating microbial genome evolution and even mechanisms for speciation (González et al., 2014; Justice et al., 2014; Ullrich et al., 2016). Comparative surveys based on the available genomes of the two *A. caldus* strains ATCC 51756 and SM-1 have revealed that both strains harbor a relatively high proportion of unique gene complements (Acuña et al., 2013). These gene complements represent a diverse pool of mobile genetic elements, including insertion sequences (ISs), genomic islands (GIs), and integrative conjugative and mobilizable elements. Yet, limited information is available on the contribution of diverse evolutionary forces to the genomic diversification of *A. caldus*. Given this knowledge gap, we have isolated and sequenced four new *A. caldus* strains from different geographic origins (Table 1).

In this study, we estimated the phylogenetic relationships of *A. caldus* strains based on their genomic sequences (four newly sequenced genomes and two existing genomes from a public database), and performed an exhaustive study of the GD dynamics, with special focus on genetic exchange underlying evolutionary adaptation. These findings, to some extent, highlight the role of gene turnover in the evolutionary diversification of *A. caldus* and adaptation to specific lifestyles and environmental niches.

MATERIALS AND METHODS

DNA Sequencing and Bioinformatics Analysis

Genome sequences for six strains were retrieved in this study, including *A. caldus* ATCC 51756, SM-1, DX, S1, ZBY, and ZJ.

TABLE 1 | General features of sequenced chromosomes in *A. caldus* strains.

Organism	<i>A. caldus</i> SM-1	<i>A. caldus</i> ATCC 51756	<i>A. caldus</i> S1	<i>A. caldus</i> DX	<i>A. caldus</i> ZBY	<i>A. caldus</i> ZJ
Geographic origin	Gold-bearing bioleaching reactor, China	Coal spoil at the Kingsbury Mine, UK	Coal heap drainage, Jiangxi, China	Copper mine tailings, Jiangxi, China	Copper mine tailings, Chambishi, Zambia	Copper mine tailings, Fujian, China
Status	Complete	Complete	Draft	Draft	Draft	Draft
Accession number	NC_015850	NZ_CP005986	LZYH00000000	LZYE00000000	LZYF00000000	LZYG00000000
Total bases (bp)	2,932,225	2,777,717	2,792,792	3,122,206	3,160,074	3,143,077
Completeness*			89.75	98.76	98.76	98.14
Coverage	38×	120×	92×	95×	89×	76×
GC content (%)	61.32	61.72	60.90	61.01	60.98	61.00
Number of contigs	1	1	1,208	390	414	386
Maximum sequence length	2,932,225	2,777,717	26,396	102,019	77,380	74,790
Minimum sequence length	2,932,225	2,777,717	200	207	201	206
N50 (bp)	2,932,225	2,777,717	4,735	22,157	18,983	18,308
N90 (bp)	2,932,225	2,777,717	617	4,321	4,123	2,291
Number of rRNA operon (5s-16s-23s)	2	2	1	1	1	1
Number of tRNA	47	49	32	46	47	46
Number of coding sequences	2,833	2,699	2,874	2,942	3,017	2,984
Proteins with predicted function	2,042	2,008	1,860	2,109	2,161	2,144
Reference	You et al. (2011)	Valdes et al. (2009)	This study	This study	This study	This study

*Genome completeness was estimated using the CheckM. Strains SM-1 and ATCC 51756 with complete genome were excluded.

Of these bacteria, the type strain ATCC 51756 was isolated from a coal spoil in Kingsbury, UK (Marsh and Norris, 1983), strain SM-1 was from an industrial reactor used in bioleaching operation (Liu et al., 2007), and the other strains (DX, S1, ZBY, and ZJ) were obtained from the China Center for Type Culture Collection. More details for geographic origins of these four new strains were shown in **Table 1**. Genome sequences of strains ATCC 51756 and SM-1, including chromosomal and plasmid sequences, were downloaded from the GenBank database. For strains DX, S1, ZBY, and ZJ, chromosomal DNA was sequenced by an Illumina MiSeq sequencer (Illumina, Inc., USA), using the paired-end sequencing approach with an average DNA insert size of 300 bp and typical read-length of 150 bp. Subsequently, bioinformatics analysis of raw sequences was performed as described previously (Yin et al., 2014), primarily including quality control, genome assembly, computational prediction of coding sequences (CDS) and other genome features such as rRNA and tRNA, as well as functional assignments against public databases (NCBI-nr and COG). Genome completeness of each strain was also estimated using the program CheckM (Parks et al., 2015). Additionally, circular maps showing chromosome architecture were drawn using the Circos software (Krzywinski et al., 2009).

Intergenomic distance scores were calculated using the web service Genome-to-Genome Distance Calculator (GGDC) 2.1 (Meier-Kolthoff et al., 2013). The distance $d(X, Y)$

between genome X and Y was calculated according to the formula:

$$d(X, Y) = 1 - \frac{2 \cdot I_{XY}}{H_{XY} + H_{YX}} \quad (1)$$

in which, I_{XY} denotes the sum of identical base pairs over all high-scoring segment pairs (HSPs, which are intergenomic matches), while H_{XY} and/or H_{YX} denote the total length of all HSPs. Heatmap was shown using the software HemI (Deng et al., 2014).

16S Ribosomal RNA (rRNA) Gene-Based and Whole Genome-Based Phylogenetic Tree

Phylogenetic relationship based on 16S rRNA sequences of *Acidithiobacillus* strains was analyzed using MEGA v5.05 with neighbor-joining method. The robustness of clustering was evaluated by 1,000 bootstrap replicates. Additionally, the phylogenetic relationships between complete and draft genomes from *A. caldus* strains were estimated. We employed an online platform CVTree3 (Zuo and Hao, 2015) to construct the whole-genome based phylogenetic tree using a composition vector approach. This whole-genome-based and alignment-free prokaryotic phylogeny was validated by directly comparing our result with the taxonomy of these strains, as opposed to performing statistical resampling tests such as bootstrap or

jackknife. The genome sequence of *Acidithiobacillus ferrooxidans* ATCC 23270 was chosen as an outgroup. Subsequently, visualization of phylogenetic tree was executed using the MEGA v5.05 (Tamura et al., 2011).

Pan-Genome Analysis

Species diversity could be identified by analyzing gene repertoire across all strains of a species, i.e., the pan-genome (Tettelin et al., 2008). PanOCT v3.18 (Fouts et al., 2012) with a BLASTP all-against-all comparison of entire proteins (E -value $\leq 1e^{-5}$; sequence identity $\geq 50\%$) was used to identify shared and unique gene content. Subsequently, annotation of core genome and strain-specific genes was implemented using BLAST against the extended COG database (Franceschini et al., 2013).

Gene Family Evolution

Groups of orthologous sequences (orthogroups, herein referred to as gene families) in all six *A. caldus* strains were classified by clustering with OrthoFinder v0.4 (Emms and Kelly, 2015), using a Markov cluster algorithm. Transposable elements were excluded, given that these gene sequences might interfere with our analyses owing to lineage-specific expansions (Carretero-Paulet et al., 2015).

To analyze the evolutionary rates of gene families, we applied the developed computational program BadiRate v1.35 using a GD stochastic model (Librado et al., 2012). The gain (γ) and death (δ) rates of gene families were estimated using a branch-specific rates (GD-BR-ML) model assuming that each phylogenetic branch had its own specific turnover rate.

Mobile Gene Elements, Insertion Sequence Elements, Transposable Elements, and Genomic Islands

IS family annotation and transposase inspection was done by BLAST comparison (E -value $\leq 1e^{-5}$) against the ISFinder database with manual detection of the surrounding significant search hits (Siguiet et al., 2006). The program SeqWord Genomic Island Sniffer (Bezuidt et al., 2009) was implemented to identify the putative horizontally transferred elements distributed in the chromosome of *A. caldus* ATCC 51756. Then, the prediction of genes in the putative horizontally transferred elements was performed using the MetaGeneAnnotator (Noguchi et al., 2008). For the other chromosomes, the computational tool IslandViewer 3 (Dhillon et al., 2015), which integrates three different prediction methods including IslandPick (Langille et al., 2008), IslandPath-DIMOB (Hsiao et al., 2003), and SIGI-HMM (Waack et al., 2006), was used to predict GIs. The GC content of GI sequences was calculated using the NGS QC Toolkit (Patel and Jain, 2012). Due to the high number of contigs, *A. caldus* S1 was excluded from the GI prediction.

Availability of Supporting Data

The data sets supporting our results in this study are available in the GenBank repository. These Whole Genome Shotgun projects of four newly sequenced *A. caldus* strains have been

deposited at the DDBJ/ENA/GenBank under the accession numbers LZYE00000000 (DX), LZYP00000000 (ZBY), LZYH00000000 (S1), and LZYG00000000 (ZJ). Additionally, the versions described in this paper are version LZYE01000000, LZYP01000000, LZYH01000000, and LZYG01000000, respectively.

RESULTS AND DISCUSSION

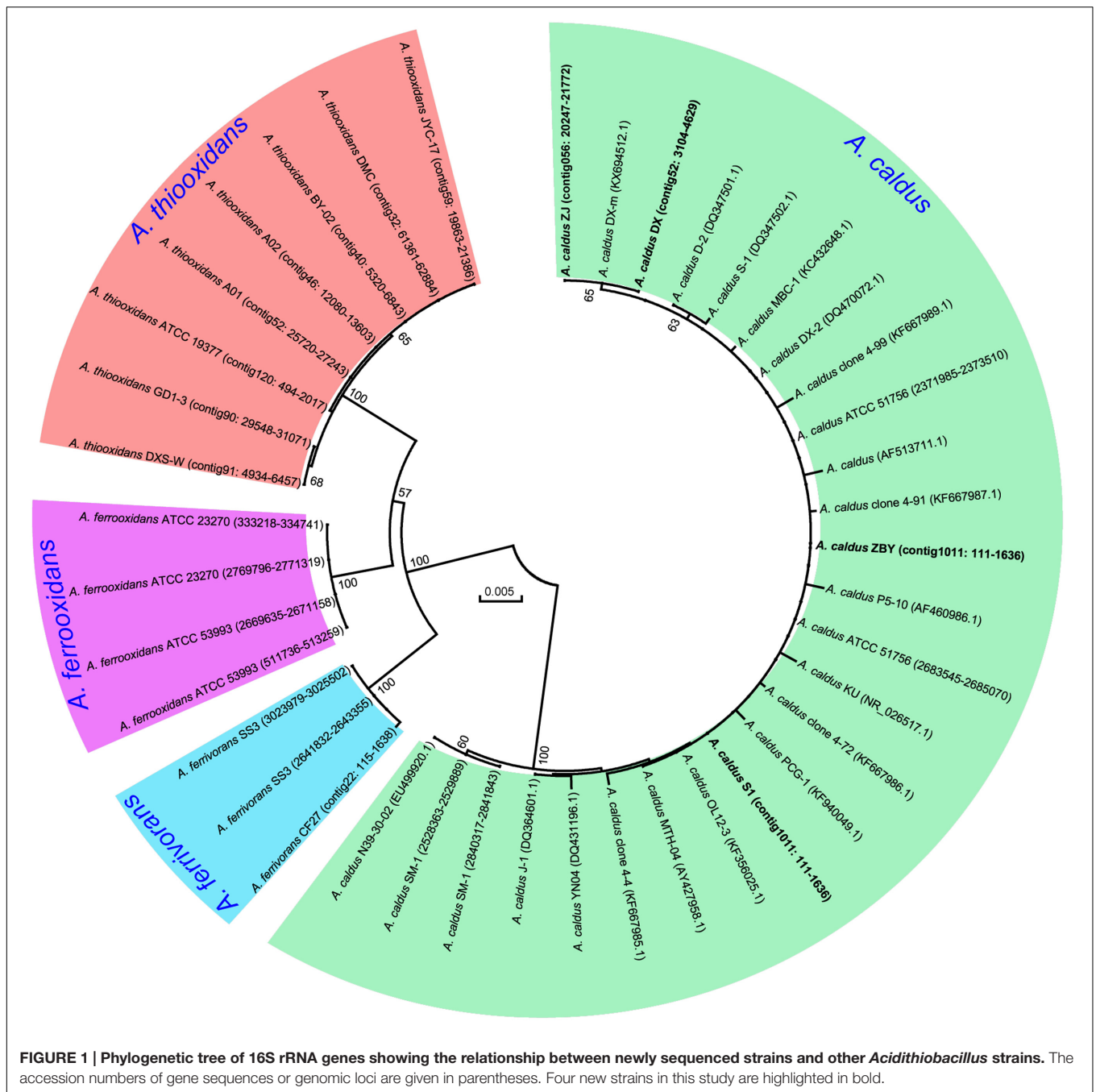
Overview of the *A. caldus* Chromosomes

The circular chromosomes of *A. caldus* strains varied from 2.78 to 3.16 Mb (Table 1). *A. caldus* strains DX, ZBY, and ZJ, which were isolated from a copper mine, possess larger chromosomes than the other strains inhabiting the divergent habitats. Genome-size variations in bacteria correspond to variations in gene number as bacterial genomes are tightly packed, and most sequences are functional protein-coding regions (Mira et al., 2001). Accordingly, strains with larger genome were predicted to harbor more CDSs compared to other strains in this study. Additionally, the evaluation of quality and completeness of genome assemblies supported the reliability of pan-genome analysis, although strain S1 had relatively low genome completeness in comparison with its closely related counterparts (Table 1).

In all *A. caldus* strains, the mean percentage GC content of these chromosomal DNAs (60.90–61.72% for all six strains) was much higher than that observed for other recognized *Acidithiobacillus* spp., e.g., *A. ferrooxidans*, *A. thiooxidans*, and *A. ferrivorans*. It might be reasonable considering that *A. caldus* species was known as the only known mesothermophile within the Acidithiobacillales (Acuña et al., 2013), and GC content of prokaryotic genomes was positively correlated with optimal growth temperature (Musto et al., 2004, 2006).

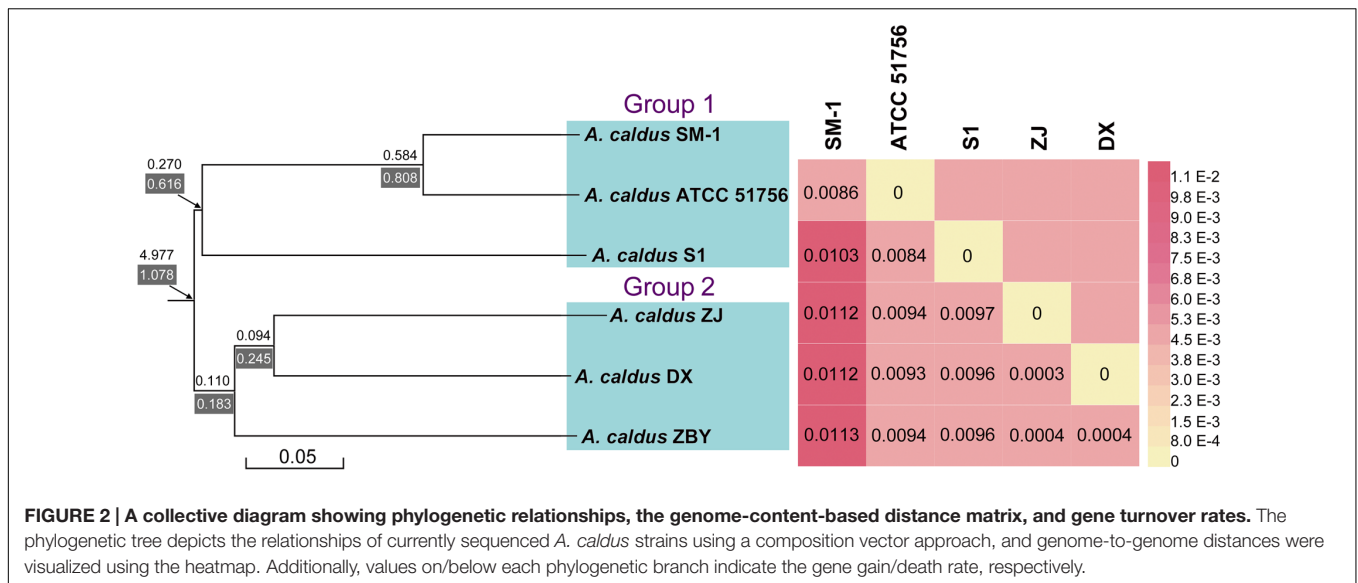
Evolutionary Relationship of *A. caldus* Strains

A phylogenetic tree based on 16S rRNA genes of *Acidithiobacillus* strains preliminarily demonstrated that these four newly sequenced strains in this study were taxonomically affiliated with *A. caldus* (Figure 1). To further identify the evolutionary relationships of *A. caldus* strains, a whole-genome-based and alignment-free phylogenetic tree was constructed (Figure 2). Additionally, GGDC analyses were employed to support the phylogenetic relationship. This phylogenomic tree showed that three strains isolated from the copper mine (namely, ZJ, DX, and ZBY) were clustered together (group 2 in Figure 2). Similarly, an earlier study reported that taxonomic clustering of six strains belonging to the genus *Novosphingobium* was generally influenced by their respective source of isolation (Gan et al., 2013). Further inspection revealed that the geographic distribution of strain ZBY was distinctively differed from those of the other two strains (ZJ and DX), and the genome-content-based distance matrix implied a slight evolutionary divergence (Figure 2). The correlation between intraspecific divergence and geographic distribution was also observed within the closely related *A. thiooxidans* species by comparative genomic



analysis (Zhang et al., 2016a). In the group 1 (Figure 2), interestingly, *A. caldus* SM-1 was obtained from a bioleaching reactor used for low grade gold-bearing minerals (Acuña et al., 2013), and the strain ATCC 51756 was isolated from a coal spoil; moreover, phylogenetic analysis revealed that these two strains were more closely related to each other than to the other four strains examined in this study. We therefore suspect that strain SM-1 might originally be isolated from an acidic setting similar to the habitat for ATCC 51756.

Ji et al. (2014) showed that the differences in adaptive evolution were attributable to different ecoiniche by genetically analyzing the marine and freshwater magnetospirilla. Accordingly, we propose that environmental variation, particularly geochemical conditions, might be a determinant of genomic diversity of *A. caldus* strains. From an alternative perspective, it appears that geographic distribution has less of an influence on hereditary variation in comparison with ecoiniche difference. The findings were consistent with an earlier study showing that environmental heterogeneity has relatively more



influence on microbial biogeography compared to geographic distance (Lin et al., 2013).

Gene Contents in *A. caldus* Strains

Gene prediction showed that the chromosomes of *A. caldus* strains contained 2,699 (ATCC 51756), 2,833 (SM-1), 2,874 (S1), 2,942 (DX), 3,017 (ZBY), and 2,984 (ZJ) predicted CDS. Functional analysis based on COG categories (Supplementary Table S1) revealed that the four most abundant functional categories within all *A. caldus* strains were “function unknown [S],” “replication, recombination, and repair [L],” “cell wall/membrane/envelope biogenesis [M],” and “energy production and conversion [C].” As reported by Silver and Phung (1996), high concentrations of toxic substrates such as heavy metals might cause a high rate of DNA damage. Thus, it was expected that CDS involved in COG category [L] would be abundant in *A. caldus* strains. Additionally, these data can also explain why this finding was distinct from previous studies analyzing the COG classification of other organisms such as marine magnetospirillum *Magnetospira* sp. QH-2 (Ji et al., 2014), given that the concentrations of potential toxic substrates in the extreme environment were much higher than those in the marine environment.

A previous study based on four genomes of “*Ferrovum*” strains highlighted the most distinct differences in interspecific metabolisms (Ullrich et al., 2016). However, in our study the assignment of CDS to the COG classification revealed that no significant differences in the number of assigned CDS were observed between the six genomes (Supplementary Table S1), probably suggesting few group-specific metabolic traits.

Comparison of Inferred Metabolic Traits and Niche Adaptation

Comparison of the Central Metabolism

In light of COG assignment aforementioned, we further observed CDS related to the predicted metabolic profiles. Compared with

other metabolic models reported in the literature, including carbon metabolism (You et al., 2011; Zhang et al., 2016b), nitrogen uptake (Levicán et al., 2008; Justice et al., 2014), and sulfur oxidation (Mangold et al., 2011; Chen et al., 2012; Yin et al., 2014), all strains in our study were predicted to contain numerous genes involved in central metabolism (Supplementary Table S2). The metabolic potentials of all strains were reconstructed and compared to each other for the identification of shared metabolic features as well as group- or strain-specific traits (Figure 3). Comprehensive analysis of these metabolism-related genes focuses on the main differences between the six *A. caldus* strains. As depicted in Figure 3, however, the evidence showed low intraspecific genetic diversity in the predicted metabolic profiles between *A. caldus* strains. A suite of genes involved in carbon assimilation were found in all strains. *A. caldus* fixes carbon dioxide via the classical Calvin–Benson–Bassham (CBB) cycle, and harbors a gene cluster predicted to encode carbon dioxide-concentrating protein (CcmK) with various copies, carboxysome shell protein (CsoS), carboxysomal shell carbonic anhydrase (CsoSCA), and ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO; Supplementary Table S2). Moreover, *A. caldus* operates a complete Embden–Meyerhof pathway (EMP) or glycolysis, pentose phosphate pathway (PPP), and incomplete tricarboxylic acid (TCA) cycle, which lacks the 2-oxoglutarate dehydrogenase complex (Valdés et al., 2008b).

With respect to nitrogen uptake, although *A. caldus* lacks nitrogenase directing the fixation of molecular nitrogen (Valdés et al., 2008a), assimilation of nitrate, nitrite, and ammonia plays a critical role in meeting nitrogen requirements. *A. caldus* utilizes nitrate or nitrite via nitrate transporter (NRT) and nitrate/nitrite transporter (Nrt). However, NRT was not present in strain ATCC 51756. Genes associated with dissimilatory nitrate reduction were identified, while a gene involved in assimilatory nitrate reduction (*nasA*) was absent in strain ATCC 51756. Though absent, it appears that the non-existence of those genes had little

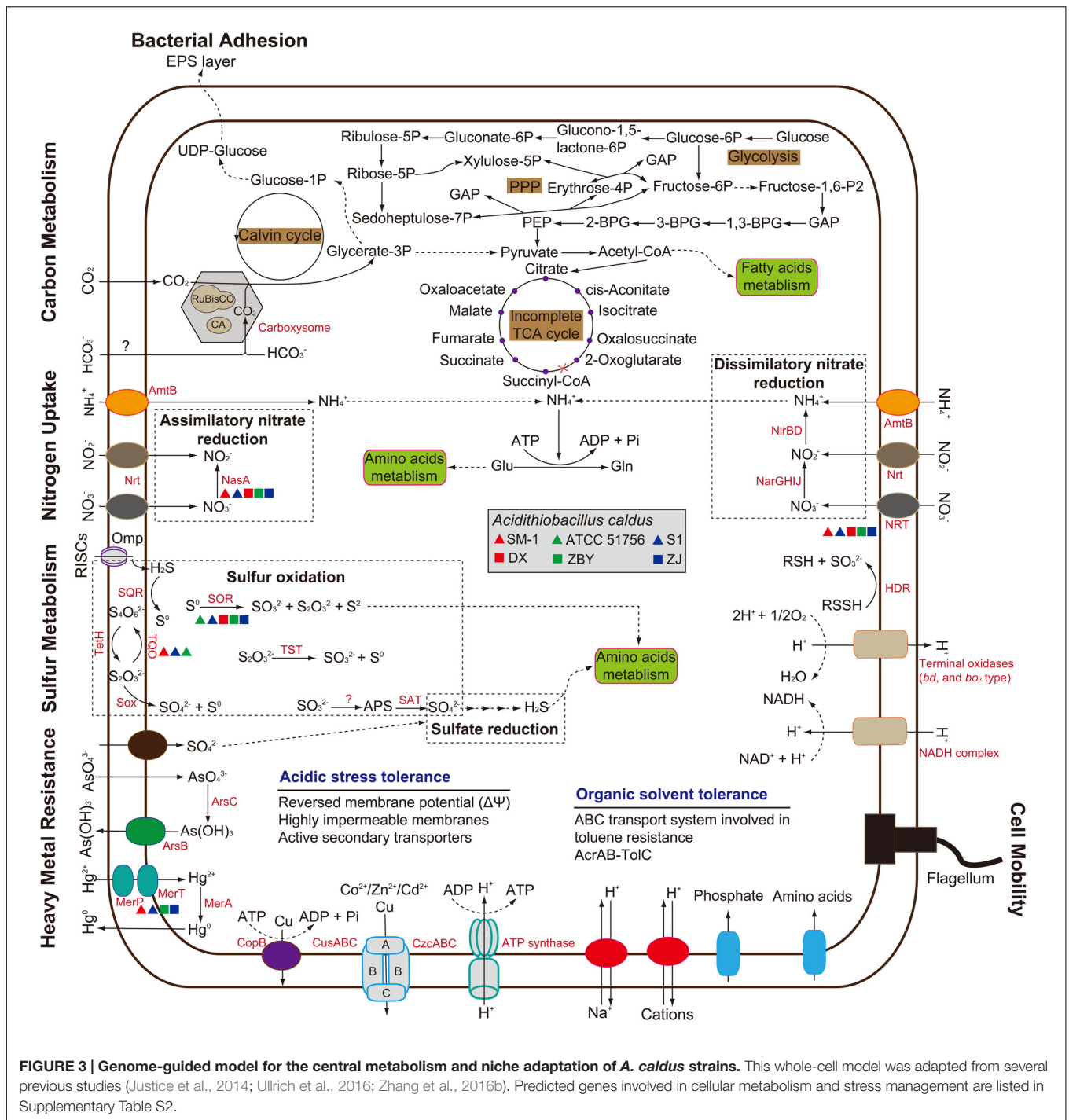


FIGURE 3 | Genome-guided model for the central metabolism and niche adaptation of *A. caldus* strains. This whole-cell model was adapted from several previous studies (Justice et al., 2014; Ullrich et al., 2016; Zhang et al., 2016b). Predicted genes involved in cellular metabolism and stress management are listed in Supplementary Table S2.

influence on the assimilation of nitrate. Additionally, all strains share the potential to take up extracellular ammonia into the cell via AmtB transporter (Figure 3) under low nitrogen levels (Levicán et al., 2008), and to convert it to glutamine via glutamine synthetase.

In recent years, the sulfur oxidation system in *A. caldus* has been well studied (Mangold et al., 2011; Chen et al., 2012). According to reported sequences, numerous genes related to

sulfur oxidation were found. Additionally, all *A. caldus* strains harbor genes predicted to be involved in sulfate reduction (Supplementary Table S2). Of note, the *sor* gene encoding sulfur oxygenase reductase, an important enzyme catalyzing a disproportionation reaction of cytoplasmic sulfur (Zhang et al., 2015), was absent in strain SM-1 (Figure 3). Group 2 strains lack the gene encoding the putative thiosulfate:quinone oxidoreductase. Thus, whether other alternative genes exist

in these strains needs to be studied further. Similar to the well-studied model for electron transfer of *A. ferrooxidans* (Valdés et al., 2008a), *A. caldus* potentially employs the electron transfer pathway from sulfur oxidation to (1) various types of terminal oxidases to generate a proton gradient or (2) to NADH complex to produce reducing power (Chen et al., 2012).

To some extent, investigation of genes involved in central metabolism supported the results of COG assignment that there were no obvious intraspecific differences. In other words, comparison of intraspecific genomes showed that only slight differences were observed in metabolic profiles, at least in central metabolism.

Response to Environmental Stress

Microbial response to environmental stresses is always a critical issue in ecological fields (Yin et al., 2015). A long-term experiment with *Escherichia coli* revealed complex coupling between organismal adaptation and genome evolution, which occurred even in a constant environment (Barrick et al., 2009). In the context of the six *A. caldus* strains, bacterial adhesion, motility, heavy metal resistance, and organic solvent tolerance were taken into account (Figure 3). All strains share a core set of genes potentially related to environmental adaptation (Supplementary Table S2). The presence of genes encoding extracellular polymeric substances precursors and type IV pili in *A. caldus* suggests a cell adhesion on mineral surface. This trait provides a reaction space between cell and mineral surface, thereby increasing the dissolution of metal sulfides (Watling, 2006; González et al., 2013). Genes assigned to COG category [N] (cell mobility) and [T] (signal transduction) were also observed, but there were few differences between these two groups (Supplementary Table S1). A full suite of genes associated with flagellar assembly were found in all strains, suggesting that *A. caldus* strains had the capacity to swim across environmental gradients and to colonize new sites.

Extremely acidic environments, especially bioleaching systems, are regarded as having extremely high concentrations of soluble and potentially toxic substrates such as heavy metals, including arsenic, mercury, copper, and cadmium (Valdés et al., 2008a) and organic extractants, such as Lix984n (Zhou et al., 2012). A series of gene clusters potentially encoding functional enzymes were identified, suggesting that *A. caldus* has the ability to cope with high concentrations of heavy metal ions. As for organic solvent tolerance, a six-gene cluster, encoding ABC transporter ATP-binding protein, hypothetical protein, toluene tolerance protein, mce-related protein, toluene tolerance protein Ttg2B, and toluene ABC transporter ATP-binding protein, was found in all strains. Additionally, an *acrAB-tolC* operon potentially encoding AcrB (transporter AcrB/AcrD/AcrF family protein), AcrA (RND family efflux transporter MFP subunit), and TolC (outer membrane efflux protein) in each genome indicated that *A. caldus* can utilize the pumps associated with resistance-nodulation-cell division protein to transfer these organic substrates.

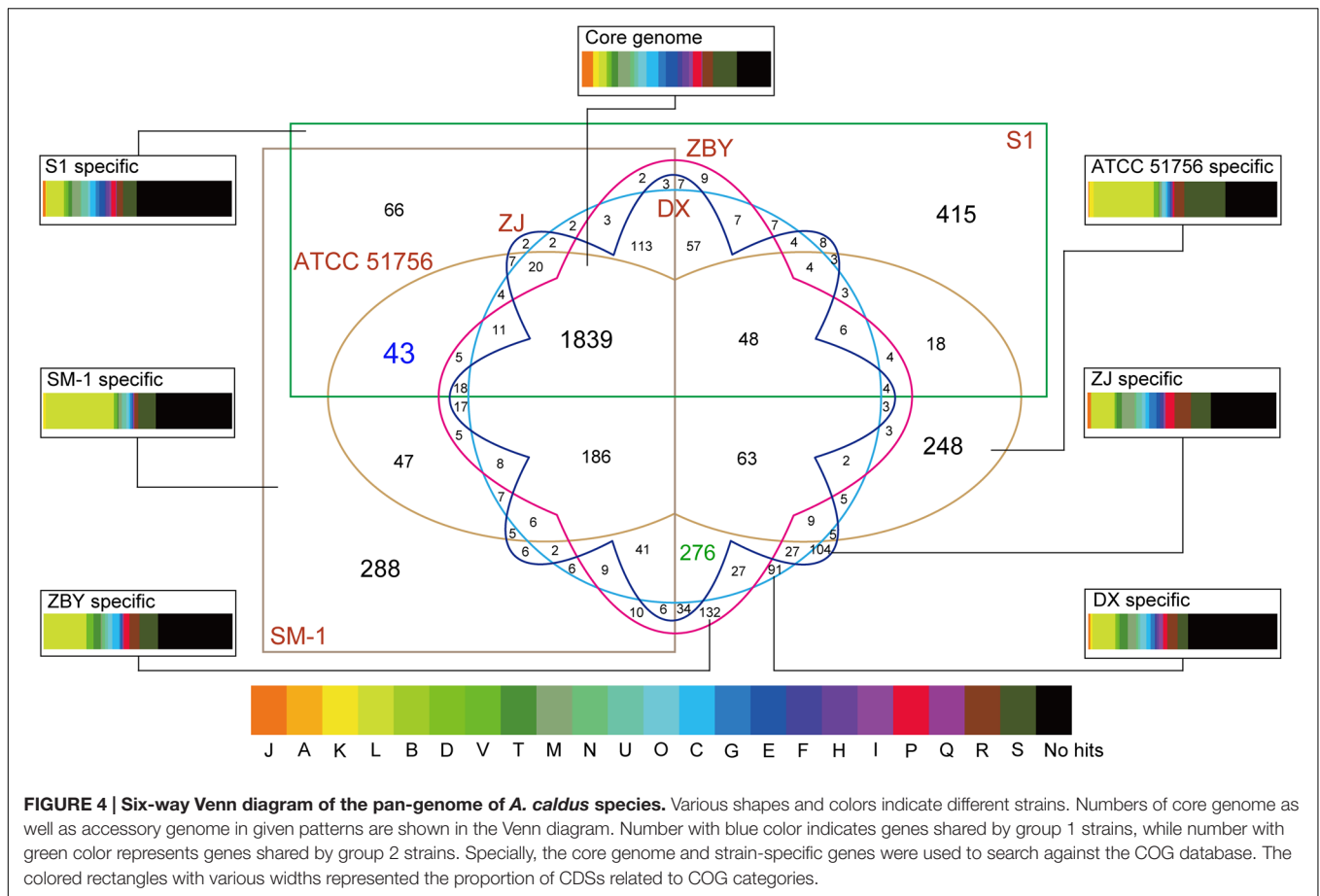
Pan-Genome Analysis

As shown above, numerous homologous genes associated with metabolic pathways as well as environmental adaptation were observed. To gain a deeper understanding of group- and strain-specific features, pan-genome analysis of *A. caldus* species was performed. A total of 4,424 CDS acquired from the four newly sequenced chromosomes plus two available chromosomes in the public database were clustered using the PanOCT. Pairwise BLAST comparisons indicated that 1,839 orthologs (41.57%) with a high percentage across all six strains were identified as the *A. caldus* core genome (Figure 4). The remaining variable 1,307 clusters were classified as the *A. caldus* accessory genome. Furthermore, strain-specific clusters were observed among the six *A. caldus* strains.

Functional assignment based on the core genome was employed to investigate the proportion of proteins in each COG category. As depicted in Figure 4, the core genome in *A. caldus* strains was commonly enriched in the COG category [M] (cell wall/membrane/envelope biogenesis; 6.36%). Additionally, our results showed that CDSs involving COG categories [C] (energy production and conversion; 6.30%) and [E] (amino acid transport and metabolism; 6.25%) were abundant. The large proportion of these genes indicated that energy utilization and uptake of nutrients in these strains might be more efficient to better adapt to the challenging environment. In other words, these findings were in line with an earlier report detailing that core genes provided functions that were essential to the basic lifestyle of the species (Medini et al., 2005).

Persistent genes encoding essential functions are stably maintained in genomes under constant selection (Nuñez et al., 2013), while dispensable or accessory genes are frequently gained or lost (Medini et al., 2005). Therefore, the accessory genome contributes to intraspecific diversity (Tettelin et al., 2008). Here, we identified many transposases by alignment of accessory genes against the NCBI-nr database (Supplementary Table S3), suggesting roles in shaping the evolution of protein families. Similarly, previously studies based on available genomes revealed that plentiful accessory genes were probably acquired by HGT (Tian et al., 2012; Sugawara et al., 2013). Additionally, it is particularly noteworthy that strain-specific genes were found to be enriched in the COG category [L] (replication, recombination and repair; Figure 4), thus supporting the view that the accessory genome confers selective advantages such as niche adaptation.

In particular, a total of 43 and 276 group-specific genes shared by group 1 and group 2 strains, respectively, were detected (Figure 4). Functional profiling based on COGs revealed that most of these predicted CDS were assigned to no COG category, probably indicating the existence of many group-specific CDS with unidentified function (Supplementary Table S4). Further inspection underscored that the abundant genes involved in certain COG categories, including [L] (replication, recombination, and repair), [M] (cell wall/membrane/envelope biogenesis), and [P] (inorganic ion transport and metabolism), might be necessary for the group 2 strains. A reasonable explanation is that copper bioleaching heap, the habitat for group 2 strains, has high concentrations of toxic metals (Zhang et al., 2016c). Microbes in such an extreme environment might



harbor potential strategies to cope with the chemical constraints of their natural functions. Additionally, COG categories [S] (general function prediction only) and [R] (function unknown) were relatively abundant in all groups, further highlighting the role of these unknown functional CDSs in genomic differentiation.

Mobile and Transposable Elements

Prediction and classification of transposable elements using ISFinder indicated that a large number of IS elements, which accounted for various proportions of the total CDS in each chromosome (ranging from 1.8 to 5.8%), were randomly distributed over the chromosomes of the *A. caldus* strains (Supplementary Table S5). Although the types of IS families were similar to each other, their distribution and relative abundance varied with each strain. Among them, some of these IS elements were identified to cluster in flexible chromosomal regions that did not satisfy the criteria of other putative mobile elements such as GIs; these findings were consistent with those from an earlier study (Acuña et al., 2013). As stated by Bentley and Parkhill (2004), the progressive loss of gene order in a prokaryotic genome might be attributed to several events including gene deletion, IS and repeat expansion, as well as recombination or rearrangement. Given this, *A. caldus* SM-1 as well as ATCC 51756 might have

higher genome plasticities compared with other closely related strains, mainly because of the acquisition of IS elements during evolution.

Aside from IS elements, the putative GI elements in all *A. caldus* strains were also identified. Results showed that several GIs ranging from 4 to 58 kb were widespread in the chromosomes of *A. caldus* strains (Supplementary Table S6). Additionally, most CDS in the GIs were annotated as hypothetical proteins. Further analyses showed the presence of integrases or mobile genetic elements such as transposase, thereby indicating that various putative GIs might be acquired via HGT. In light of the view that underscores the contribution of horizontal (lateral) gene transfer (HGT) in the expansion of gene repertoires of prokaryotes (Ochman et al., 2000; Gogarten et al., 2002; Treangen and Rocha, 2011), we inferred that the frequency of HGT was high in group 2 strains with larger genomes, conferring a predominant role in shaping their evolution and allowing the acquisition of novel adaptive functions. We emphasized the role of GIs in adaptation to specific lifestyles and environmental niches, considering that many GIs were highly relevant for niche-specific adaptation (Wu et al., 2011). Furthermore, numerous genes in *A. caldus* species might be obtained by genetic exchange as suggested by the presence of a large load of mobile genetic elements including IS elements, transposases, and GIs. Consequently,

TABLE 2 | Summary of genes assigned or unassigned to orthogroups in six *A. caldus* strains.

Number	SM-1	ATCC 51756	S1	ZJ	DX	ZBY
Orthogroups	2,514	2,422	2,395	2,818	2,800	2,828
Genes in orthogroups	2,709	2,538	2,477	2,918	2,886	2,942
Genes unassigned to any orthogroup	124	161	397	66	56	75

changes in genome structure and gene copy number might provide *A. caldus* strains with a survival advantage for rapid adaptation and survival in highly acidic and metal-laden environments.

Probabilistic Analysis of Gene Family Turnover

Gene families in *A. caldus* strains were classified as orthogroups using OrthoFinder (Table 2). Our classification identified up to 3,109 orthogroups (containing two or more genes in all selected strains), which included 16,470 sequences. There were fewer genes in group 1 strains with small genomes clustered into multigene orthogroups than in group 2 strains. However, these smaller genomes contained more unassigned genes than any orthogroups compared to the others. These results appear to be explained in part by intense fractionation pressure (Carretero-Paulet et al., 2015). In other words, multigene families in smaller genomes might be under continuous deletion pressure and, as a result, these genomes tend to be smaller in comparison with their counterparts in larger genomes.

BadiRate analysis, using a full likelihood method, was applied to examine the evolutionary dynamics of gene families across the *A. caldus* species, and to characterize the expansion and/or contraction of genomes. The statistical framework not only estimates GD rates in a decoupled manner, using two independent parameters (γ and δ), but also explicitly takes into account certain key features in prokaryotic evolution, such as HGT. A stochastic GD-BR-ML model statistically evaluating the turnover rates demonstrated that a large number of orthologous genes frequently undergoing high gain and/or death events have evolved from ancestral genes (Figure 2). Particularly, gene families in *A. caldus* species rapidly expand through gene gain (duplication) and slowly contract through gene death (deletion or pseudogenization), indicating that the extensive recruitment of genes involved in long-term evolution confers an ecological advantage for survival and proliferation under extremely acidic conditions. Moreover, the phylogenetic branch with the higher death rate ($\delta = 0.616$ and/or 0.808) indicated that group 1 strains with smaller genomes might be derived from free-living ancestors by the genome-reductive evolutionary process (Figure 2). Given that genome reduction coincided with the increase in frequency of mobile elements and repeated sequences (Moran, 2003), multiple IS elements identified in strain SM-1 and ATCC 51756 (Supplementary Table S5) might play a key role in mediating intrachromosomal recombination, thereby leading to rearrangements and gene loss. However, the dispensable genes in the above-mentioned microorganisms might suffer

extensive loss and non-functionalization. The compact genomes in the given organisms can perform essential functions for cellular survival and replication, as the loss of dispensable genes has little effect on bacterial fitness, at least under certain environmental conditions (Albalat and Cañestro, 2016). Despite their smaller or near-minimal size, all reduced genomes still retain the essential gene set, and are thereby able to support cellular life both in stable and changing circumstances (Moya et al., 2009). Therefore, small genomes in group 1 strains would be more tightly packed by selective reduction, and are thus more streamlined than their larger genome counterparts.

Gene turnover in group 2 strains was also estimated. As illustrated in Figure 2, phylogenetic branches showed a lower gene turnover rate in group 2 strains compared to that in group 1 strains. Additionally, we found that the rates of gene death were slightly higher than the gain turnover rates. There were two possible explanations for these results. The number of genes gained from HGT as well as gene duplication events might be significant enough to account for the increase of microbial DNA content and novel functions, and play a key role in evolution (Mira et al., 2001; Navarre et al., 2006). This hypothesis may also be supported by an earlier genetic study on the evolution of *Bacillus anthracis* virulence, which revealed that key genes that cause anthrax in this bacterium were identified as acquired by HGT (Zwick et al., 2012). However, a conceivable explanation underlying environment-dependent conditional dispensability indicates that genes in a given species would be dispensable if they were related to certain processes that were only required in a specific untested environments (Albalat and Cañestro, 2016). Of note, it is challenging to assess which genes are regarded as dispensable or essential components by coupling genotypes with phenotypes. In view of the complexity of environmental conditions in copper mines, low deletion pressures might provide microbes with a major fitness advantage for growth in adverse environments. Furthermore, large genomes in bacteria correspond to species that have the ability to tackle various environmental stimuli (Schneiker et al., 2007). Accordingly, large bacterial genomes might have an adaptive role in the evolution of group 2 strains.

CONCLUSION

Six chromosomes of the extreme acidophile *A. caldus* were valuable resources for the investigation of genetic diversity and evolutionary adaptation. A phylogenetic tree based on chromosomal sequences of *A. caldus* species showed a potential correlation between genomic diversity

and geochemical characteristics. Further analysis revealed that chemical constraint in respective natural habitat might be a determinant contributing to genetic diversification. Apparently, genetic analyses indicated that gene gain and loss were both dominant evolutionary forces in the adaptive evolution of *A. caldus* species. During adaptation to these adverse environmental conditions, GD rates varied in different settings, resulting in genomic differentiation and speciation. The compact and streamlined genomes might undergo selective deletion pressure, whereas large genomes had been extensively recruited by intraspecific or interspecific genetic exchange. These genome-guided findings in our study, to some extent, provide novel insights into the evolutionary adaptation of *A. caldus* species.

AUTHOR CONTRIBUTIONS

XnZ, XL, and QH conceived and designed the experiments. XnZ and WD performed the experiments. XnZ analyzed the data. XnZ wrote the paper. XoZ, FF, DP, WH, and HY revised the manuscript.

REFERENCES

- Acuña, L. G., Cárdenas, J. P., Covarrubias, P. C., Haristoy, J. J., Flores, R., Nuñez, H., et al. (2013). Architecture and gene repertoire of the flexible genome of the extreme acidophile *Acidithiobacillus caldus*. *PLoS ONE* 8:e78237. doi: 10.1371/journal.pone.0078237
- Albalat, R., and Cañestro, C. (2016). Evolution by gene loss. *Nat. Rev. Genet.* 17, 379–391. doi: 10.1038/nrg.2016.39
- Barrick, J. E., Yu, D. S., Yoon, S. H., Jeong, H., Oh, T. K., Schneider, D., et al. (2009). Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 461, 1243–1247. doi: 10.1038/nature08480
- Bentley, S. D., and Parkhill, J. (2004). Comparative genomic structure of prokaryotes. *Annu. Rev. Genet.* 38, 771–791. doi: 10.1146/annurev.genet.38.072902.094318
- Bezuidt, O., Lima-Mendez, G., and Reva, O. N. (2009). SeqWord gene island sniffer: a program to study the lateral genetic exchange among bacteria. *W. Acad. Sci. Eng. Techn.* 58, 410–415.
- Carretero-Paulet, L., Librado, P., Chang, T., Ibarra-Laclette, E., Herrera-Estrella, L., Rozas, J., et al. (2015). High gene family turnover rates and gene space adaptation in the compact genome of the carnivorous plant *Utricularia gibba*. *Mol. Biol. Evol.* 32, 1284–1295. doi: 10.1093/molbev/msv020
- Chen, L., Ren, Y., Lin, J., Liu, X., Pang, X., and Lin, J. (2012). *Acidithiobacillus caldus* sulfur oxidation model based on transcriptome analysis between the wild type and sulfur oxygenase reductase defective mutant. *PLoS ONE* 7:e39470. doi: 10.1371/journal.pone.0039470
- Deng, W., Wang, Y., Liu, Z., Cheng, H., and Xue, Y. (2014). HemI: a toolkit for illustrating heatmaps. *PLoS ONE* 9:e111988. doi: 10.1371/journal.pone.0111988
- Dhillon, B. K., Laird, M. R., Shay, J. A., Winsor, G. L., Lo, R., Nizam, F., et al. (2015). IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. *Nucleic Acids Res.* 43, W104–W108. doi: 10.1093/nar/gkv401
- Dopson, M., and Lindström, E. B. (1999). Potential role of *Thiobacillus caldus* in arsenopyrite bioleaching. *Appl. Environ. Microbiol.* 65, 36–40.
- Emms, D. M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves ortholog inference accuracy. *Genome Biol.* 16, 157. doi: 10.1186/s13059-015-0721-2
- Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531–1545.

FUNDING

This work was supported by the National Natural Science Foundation of China (No. 31570113 and No. 41573072) and the Fundamental Research Funds for the Central Universities of Central South University (No. 2016zzts102).

ACKNOWLEDGMENTS

We thank Dr. Qichao Tu in Zhejiang University and Dr. Guanyun Wei in Nanjing Normal University for helpful discussion and suggestions. Also, we thank the National Center for Biotechnology Information (NCBI) for providing the genomic sequences of *A. caldus* strains ATCC 51756 and SM-1.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.01960/full#supplementary-material>

- Fouts, D. E., Brinkac, L., Beck, E., Inman, J., and Sutton, G. (2012). PanOCT: automated clustering of orthologs using conserved gene neighborhood for pan-genomic analysis of bacterial strains and closely related species. *Nucleic Acids Res.* 40, e172. doi: 10.1093/nar/gks757
- Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., et al. (2013). STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 41, D808–D815. doi: 10.1093/nar/gks1094
- Gan, H. M., Hudson, A. O., Rahman, A. Y. A., Chan, K. G., and Savka, M. A. (2013). Comparative genomic analysis of six bacteria belonging to the genus *Novosphingobium*: insights into marine adaptation, cell-cell signaling and bioremediation. *BMC Genomics* 14:431. doi: 10.1186/1471-2164-14-431
- Gillespie, J. H. (1991). *The Causes of Molecular Evolution*. Oxford: Oxford University Press.
- Gogarten, J. P., Doolittle, W. F., and Lawrence, J. G. (2002). Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* 19, 2226–2238. doi: 10.1093/oxfordjournals.molbev.a004046
- González, A., Bellenberg, S., Mamani, S., Ruiz, L., Echeverría, A., Souleré, L., et al. (2013). AHL signaling molecules with a large acyl chain enhance biofilm formation on sulfur and metal sulfides by the bioleaching bacterium *Acidithiobacillus ferrooxidans*. *Appl. Microbiol. Biotechnol.* 97, 3729–3737. doi: 10.1007/s00253-012-4229-3
- González, C., Yanquepe, M., Cardenas, J. P., Valdes, J., Quatrini, R., Holmes, D. S., et al. (2014). Genetic variability of psychrotolerant *Acidithiobacillus ferrivorans* revealed by (meta)genomic analysis. *Res. Microbiol.* 165, 726–734. doi: 10.1016/j.resmic.2014.08.005
- Hahn, M. W., Han, M. V., and Han, S. (2007). Gene family evolution across 12 *Drosophila* genomes. *PLoS Genet.* 3:e197. doi: 10.1371/journal.pgen.0030197
- Hallberg, K. B., and Lindström, E. B. (1994). Characterization of *Thiobacillus caldus* sp. nov., a moderately thermophilic acidophile. *Microbiology* 140, 3451–3456. doi: 10.1099/13500872-140-12-3451
- Hallberg, K. B., and Lindström, E. B. (1996). Multiple serotypes of the moderate thermophile *Thiobacillus caldus*, a limitation of immunological assays for biomining microorganisms. *Appl. Environ. Microbiol.* 62, 4243–4246.
- Hsiao, W., Wan, I., Jones, S. J., and Brinkman, F. S. L. (2003). IslandPath: aiding detection of genomic islands in prokaryotes. *Bioinformatics* 19, 418–420. doi: 10.1093/bioinformatics/btg004

- Innan, H., and Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. *Nat. Rev. Genet.* 11, 97–108. doi: 10.1038/nrg2689
- Jacobsen, A., Hendriksen, R. S., Aaresturp, F. M., Ussery, D. W., and Friis, C. (2011). The *Salmonella enterica* Pan-genome. *Microb. Ecol.* 62, 487–504. doi: 10.1007/s00248-011-9880-1
- Ji, B., Zhang, S., Arnoux, P., Rouy, Z., Alberto, F., Philippe, N., et al. (2014). Comparative genomic analysis provides insights into the evolution and niche adaptation of marine *Magnetospira* sp. QH-2 strain. *Environ. Microbiol.* 16, 525–544. doi: 10.1111/1462-2920.12180
- Justice, N. B., Norman, A., Brown, C. T., Singh, A., Thomas, B. C., and Banfield, J. F. (2014). Comparison of environmental and isolate *Sulfobacillus* genomes reveals diverse carbon, sulfur, nitrogen, and hydrogen metabolisms. *BMC Genomics* 15:1107. doi: 10.1186/1471-2164-15-1107
- Kimura, M. (1984). *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.
- Krzywinski, M., Schein, J., Biro, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109
- Kulmuni, J., Wurm, Y., and Pamilo, P. (2013). Comparative genomics of chemosensory protein genes reveals rapid evolution and positive selection in ant-specific duplicates. *Heredity* 110, 538–547. doi: 10.1038/hdy.2012.122
- Langille, M. G., Hsiao, W. W., and Brinkman, F. S. (2008). Evaluation of genomic island predictors using a comparative genomics approach. *BMC Bioinformatics* 9:329. doi: 10.1186/1471-2105-9-329
- Levicán, G., Ugalde, J. A., Ehrenfeld, N., Maass, A., and Parada, P. (2008). Comparative genomic analysis of carbon and nitrogen assimilation mechanisms in three indigenous bioleaching bacteria: predictions and validations. *BMC Genomics* 9:581. doi: 10.1186/1471-2164-9-581
- Librado, P., Vieira, F. G., and Rozas, J. (2012). BadiRate: estimating family turnover rates by likelihood-based methods. *Bioinformatics* 28, 279–281. doi: 10.1093/bioinformatics/btr623
- Librado, P., Vieira, F. G., Sánchez-Gracia, A., Kolokotronis, S., and Rozas, J. (2014). Mycobacterial phylogenomics: an enhanced method for gene turnover analysis reveals uneven levels of gene gain and loss among species and gene families. *Genome Biol. Evol.* 6, 1454–1465. doi: 10.1093/gbe/evu117
- Lin, W., Wang, Y., Gorby, Y., Nealson, K., and Pan, Y. (2013). Integrating niche-based process and spatial process in biogeography of magnetotactic bacteria. *Sci. Rep.* 3, 1643. doi: 10.1038/srep01643
- Liu, X., Lin, J., Zhang, Z., Bian, J., Zhao, Q., Liu, Y., et al. (2007). Construction of conjugative gene transfer system between *E. coli* and moderately thermophilic, extremely acidophilic *Acidithiobacillus caldus* MTH-04. *J. Microbiol. Biotechnol.* 17, 162–167.
- MacLean, D., Jones, J. D. G., and Studholme, D. J. (2009). Application of 'next-generation' sequencing technologies to microbial genetics. *Nat. Rev. Microbiol.* 7, 287–296. doi: 10.1038/nrmicro2122
- Mangold, S., Valdés, J., Holmes, D. S., and Dopson, M. (2011). Sulfur metabolism in the extreme acidophile *Acidithiobacillus caldus*. *Front. Microbiol.* 2:17. doi: 10.3389/fmicb.2011.00017
- Marsh, R. M., and Norris, P. R. (1983). The isolation of some thermophilic, autotrophic, iron- and sulfur-oxidizing bacteria. *FEMS Microbiol. Lett.* 17, 311–315. doi: 10.1111/j.1574-6968.1983.tb00426.x
- Medini, D., Donati, C., Tettelin, H., Massignani, V., and Rappuoli, R. (2005). The microbial pan-genome. *Curr. Opin. Genet. Dev.* 15, 589–594. doi: 10.1016/j.gde.2005.09.006
- Meier-Kolthoff, J. P., Auch, A. F., Klenk, H., 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
- Metzker, M. L. (2010). Sequencing technologies—the next generation. *Nat. Rev. Genet.* 11, 31–46. doi: 10.1038/nrg2626
- Mira, A., Ochman, H., and Moran, N. A. (2001). Deletional bias and the evolution of bacterial genomes. *Trends Genet.* 17, 589–596. doi: 10.1016/S0168-9525(01)02447-7
- Moran, N. A. (2003). Tracing the evolution of gene loss in obligate bacterial symbionts. *Curr. Opin. Microbiol.* 6, 512–518. doi: 10.1016/j.mib.2003.08.001
- Moya, A., Gil, R., Latorre, A., Peretó, J., Garcillán-Barcia, M. P., and De La Cruz, F. (2009). Toward minimal bacterial cells: evolution vs. design. *FEMS Microbiol. Rev.* 33, 225–235. doi: 10.1111/j.1574-6976.2008.00151.x
- Musto, H., Naya, H., Zavala, A., Romero, H., Alvarez-Valín, F., and Bernardi, G. (2006). Genomic GC level, optimal growth temperature, and genome size in prokaryotes. *Biochem. Biophys. Res. Commun.* 347, 1–3. doi: 10.1016/j.bbrc.2006.06.054
- Musto, H., Naya, H., Zavala, A., Romero, H., Alvarez-Valín, F., and Bernardi, G. (2004). Correlations between genomic GC levels and optimal growth temperatures in prokaryotes. *FEBS Lett.* 573, 73–77. doi: 10.1016/j.febslet.2004.07.056
- Navarre, W. W., Porwollik, S., Wang, Y., McClelland, M., Rosen, H., Libby, S. J., et al. (2006). Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. *Science* 313, 236–238. doi: 10.1126/science.1128794
- Nei, M., and Rooney, A. P. (2005). Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genet.* 39, 121–152. doi: 10.1146/annurev.genet.39.073003.112240
- Noguchi, H., Taniguchi, T., and Itoh, T. (2008). MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res.* 15, 387–396. doi: 10.1093/dnares/dsn027
- Nuñez, P. A., Romero, H., Farber, M. D., and Rocha, E. P. C. (2013). Natural selection for operons depends on genome size. *Genome Biol. Evol.* 5, 2242–2254. doi: 10.1093/gbe/evt174
- Ochman, H., Lawrence, J. G., and Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299–304. doi: 10.1038/35012500
- Ohno, S. (1970). *Evolution by Gene Duplication*. Berlin: Springer.
- Ohta, T. (1992). The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Syst.* 23, 263–286. doi: 10.1146/annurev.es.23.110192.001403
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114
- Patel, R. K., and Jain, M. (2012). NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS ONE* 7:e30619. doi: 10.1371/journal.pone.0030619
- Rawlings, D. E. (1998). Industrial practice and the biology of leaching of metals from ores. *J. Ind. Microbiol. Biotechnol.* 20, 268–274. doi: 10.1038/sj.jim.2900522
- Sánchez-Gracia, A., Vieira, F. G., and Rozas, J. (2009). Molecular evolution of the major chemosensory gene families in insects. *Heredity* 103, 208–216. doi: 10.1038/hdy.2009.55
- Schneiker, S., Perlova, O., Kaiser, O., Gerth, K., Alici, A., Altmeyer, M. O., et al. (2007). Complete genome sequence of the myxobacterium *Sorangium cellulosum*. *Nat. Biotechnol.* 25, 1281–1289. doi: 10.1038/nbt1354
- Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., and Chandler, M. (2006). ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* 34, D32–D36. doi: 10.1093/nar/gkj014
- Silver, S., and Phung, L. T. (1996). Bacterial heavy metal resistance: new surprises. *Annu. Rev. Microbiol.* 50, 753–789. doi: 10.1146/annurev.micro.50.1.753
- Sugawara, M., Epstein, B., Badgley, B. D., Unno, T., Xu, L., Reese, J., et al. (2013). Comparative genomics of the core and accessory genomes of 48 *Sinorhizobium* strains comprising five genospecies. *Genome Biol.* 14, R17. doi: 10.1186/gb-2013-14-2-r17
- 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
- Tettelin, H., Riley, D., Cattuto, C., and Medini, D. (2008). Comparative genomics: the bacterial pan-genome. *Curr. Opin. Microbiol.* 11, 472–477. doi: 10.1016/j.mib.2008.09.006
- Tian, C. F., Zhou, Y. J., Zhang, Y. M., Li, Q. Q., Zhang, Y. Z., Li, D. F., et al. (2012). Comparative genomics of rhizobia nodulating soybean suggests extensive recruitment of lineage-specific genes in adaptations. *Proc. Natl. Acad. Sci. U.S.A.* 109, 8629–8634. doi: 10.1073/pnas.1120436109
- Treangen, T. J., and Rocha, E. P. C. (2011). Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet.* 7:e1001284. doi: 10.1371/journal.pgen.1001284
- Ullrich, S. R., González, C., Poehlein, A., Tischler, J. S., Daniel, R., Schlömann, M., et al. (2016). Gene loss and horizontal gene transfer contributed to the genome

- evolution of the extreme acidophile “Ferrovum”. *Front. Microbiol.* 7:797. doi: 10.3389/fmicb.2016.00797
- Valdés, J., Pedroso, I., Quatrini, R., Dodson, R. J., Tettelin, H., Blake, R., et al. (2008a). *Acidithiobacillus ferrooxidans* metabolism: from genome sequence to industrial applications. *BMC Genomics* 9:597. doi: 10.1186/1471-2164-9-597
- Valdés, J., Pedroso, I., Quatrini, R., and Holmes, D. S. (2008b). Comparative genome analysis of *Acidithiobacillus ferrooxidans*, *A. thiooxidans* and *A. caldus*: Insights into their metabolism and ecophysiology. *Hydrometallurgy* 94, 180–184. doi: 10.1016/j.hydromet.2008.05.039
- Valdes, J., Quatrini, R., Hallberg, K., Dopson, M., Valenzuela, P. D. T., and Holmes, D. S. (2009). Draft genome sequence of the extremely acidophilic bacterium *Acidithiobacillus caldus* ATCC 51756 reveals metabolic versatility in the genus *Acidithiobacillus*. *J. Bacteriol.* 191, 5877–5878. doi: 10.1128/JB.00843-09
- Vieira, F. G., and Rozas, J. (2011). Comparative genomics of the odorant-binding and chemosensory protein gene families across the Arthropoda: origin and evolutionary history of the chemosensory system. *Genome Biol. Evol.* 3, 476–490. doi: 10.1093/gbe/evr033
- Waack, S., Keller, O., Asper, R., Brodag, T., Damm, C., Fricke, W. F., et al. (2006). Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC Bioinformatics* 7:142. doi: 10.1186/1471-2105-7-142
- Watling, H. R. (2006). The bioleaching of sulphide minerals with emphasis on copper sulphides—a review. *Hydrometallurgy* 84, 81–108. doi: 10.1016/j.hydromet.2006.05.001
- Wu, X., Monchy, S., Taghavi, S., Zhu, W., Ramos, J., and van der Lelie, D. (2011). Comparative genomics and functional analysis of niche-specific adaptation in *Pseudomonas putida*. *FEMS Microbiol. Rev.* 35, 299–323. doi: 10.1111/j.1574-6976.2010.00249.x
- Yin, H., Niu, J., Ren, Y., Cong, J., Zhang, X., Fan, F., et al. (2015). An integrated insight into the response of sedimentary microbial communities to heavy metal contamination. *Sci. Rep.* 5, 14266. doi: 10.1038/srep14266
- Yin, H., Zhang, X., Li, X., He, Z., Liang, Y., Guo, X., et al. (2014). Whole-genome sequencing reveals novel insights into sulfur oxidation in the extremophile *Acidithiobacillus thiooxidans*. *BMC Microbiol.* 14:179. doi: 10.1186/1471-2180-14-179
- You, X. Y., Guo, X., Zheng, H. J., Zhang, M. J., Liu, L. J., Zhu, Y. Q., et al. (2011). Unraveling the *Acidithiobacillus caldus* complete genome and its central metabolisms for carbon assimilation. *J. Genet. Genomics* 38, 243–252. doi: 10.1016/j.jgg.2011.04.006
- Zhang, X., Feng, X., Tao, J., Ma, L., Xiao, Y., Liang, Y., et al. (2016a). Comparative genomics of the extreme acidophile *Acidithiobacillus thiooxidans* reveals intraspecific divergence and niche adaptation. *Int. J. Mol. Sci.* 17, 1355. doi: 10.3390/ijms17081355
- Zhang, X., Liu, X., Liang, Y., Fan, F., Zhang, X., and Yin, H. (2016b). Metabolic diversity and adaptive mechanisms of iron- and/or sulfur-oxidizing autotrophic acidophiles in extremely acidic environments. *Environ. Microbiol. Rep.* 8, 738–751. doi: 10.1111/1758-2229.12435
- Zhang, X., Niu, J., Liang, Y., Liu, X., and Yin, H. (2016c). Metagenome-scale analysis yields insights into the structure and function of microbial communities in a copper bioleaching heap. *BMC Genet.* 17:21. doi: 10.1186/s12863-016-0330-4
- Zhang, X., Huaqun, Y., Yili, L., Qiu, G., and Liu, X. (2015). Theoretical model of the structure and the reaction mechanisms of sulfur oxygenase reductase in *Acidithiobacillus thiooxidans*. *Adv. Mater. Res.* 1130, 67–70. doi: 10.4028/www.scientific.net/AMR.1130.67
- Zhou, Z., Fang, Y., Li, Q., Yin, H., Qin, W., Liang, Y., et al. (2012). Global transcriptional analysis of stress-response strategies in *Acidithiobacillus ferrooxidans* ATCC 23270 exposed to organic extractant-Lix984n. *World J. Microbiol. Biotechnol.* 28, 1045–1055. doi: 10.1007/s11274-011-0903-3
- Zuo, G., and Hao, B. (2015). CVTree3 web server for whole-genome-based and alignment-free prokaryotic phylogeny and taxonomy. *Genomics Proteomics Bioinformatics* 13, 321–331. doi: 10.1016/j.gpb.2015.08.004
- Zwick, M. E., Joseph, S. J., Didelot, X., Chen, P. E., Bishop-Lilly, K. A., Stewart, A. C., et al. (2012). Genomic characterization of the *Bacillus cereus* sensu lato species: backdrop to the evolution of *Bacillus anthracis*. *Genome Res.* 22, 1512–1524. doi: 10.1101/gr.134437.111

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 Zhang, Liu, He, Dong, Zhang, Fan, Peng, Huang and Yin. 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.