Comparative Genomic Analysis Reveals Distribution, Organization and Evolution of Metal Resistance Genes in Genus *Acidithiobacillus*

Liangzhi Li,a,b Zhenghua Liu,a,b Delong Meng,a,b,* Xueduan Liu,a,b Xing Li,a,b Ming 1
Zhang,a,b Jiemeng Tao,a,b Yabing Gu,a,b Shuiping Zhongc,d and Huaqun Yin,a,b,*

*aSchool of Minerals Processing and Bioengineering, Central South University, Changsha, China
bKey Laboratory of Biometallurgy of Ministry of Education, Central South University,
Changsha, China
cCollege of Zijin Mining, Fuzhou University, Fuzhou, Fujian, China
dNational Key Laboratory of Comprehensive Utilization of Low-Grade Refractory Gold Ores, Shanghang, China

* Correspondence:*
Huaqun Yin
yinhuaqun_cs@sina.com;
Delong Meng
delong.meng@csu.edu.cn

Abstract

Members of genus *Acidithiobacillus* which can adapt up to extremely high concentration of
heavy metals are universally found at acid mine drainage (AMD) sites. We here perform comparative genomic analysis of 37 strains within the genus *Acidithiobacillus* to answer the untouched questions as to the mechanisms and the evolutionary history of metal resistance genes in *Acidithiobacillus* spp. Results showed that the evolutionary history of metal resistance genes in *Acidithiobacillus* spp. involved a combination of gene gains and losses, HGT and gene duplication. Phylogenetic analyses revealed that metal resistance genes in *Acidithiobacillus* spp. were acquired by early horizontal gene transfer (HGT) events from species that shared habitats with *Acidithiobacillus* spp. such as *Acidihalobacter, Thiobacillus, Acidiferrobacter* and *Thiomonas*. Multi-copper oxidase genes involved in copper detoxification were lost in iron-oxidizing *A. ferridurans, A. ferrivorans* and *A. ferrooxidans*, and were replaced by rusticyanin during evolution. In addition, widespread purifying selection and predicted high expression levels emphasized the indispensable roles of metal resistance genes for *Acidithiobacillus* spp. to adapt harsh environments. Taking together, the results suggested that *Acidithiobacillus* spp. recruited and consolidated additional novel functionalities in adaption to challenging environments via HGT, gene duplication and purifying selection. This study shed light on the distribution, organization, functionality and the complex evolutionary history of metal resistance genes in *Acidithiobacillus* spp.

**IMPORTANCE** Horizontal gene transfer (HGT), together with natural selection and gene duplication, are three main engines that drive adaptive evolution of microbial genomes. Previous studies indicated that HGT was a main adaptive mechanism for acidophile to cope with heavy
metal-rich environments. However, evidences of HGT in *Acidithiobacillus* species in response to challenging metal-rich environments and mechanisms addressing how metal resistance genes originated and evolved in *Acidithiobacillus* are still lacking. The findings of this study revealed fascinating phenomena of putative cross-phylum HGT, suggested that *Acidithiobacillus* spp. recruited and consolidated additional novel functionalities in adaption to challenging environments via HGT, gene duplication and purifying selection. Taken together, the insights gained in this study have improved our understanding on the metal resistance strategies of *Acidithiobacillus* spp..

**Keywords:** comparative genomics, *Acidithiobacillus* spp., metal resistance, evolution, horizontal gene transfer

1 **Introduction**

The genus *Acidithiobacillus* is comprised of a group of obligatory acidophilic chemolithotrophic Gram-negative bacteria that conduct energy conservation by oxidizing reduced inorganic sulfur compounds (RISCs), such as thiosulfate (S$_2$O$_3^{2-}$), hydrogen sulfide (H$_2$S) and elemental sulfur (S$^0$). This attribute has been exploited in bioleaching of ores and has also been the main cause of accelerating formation of polluting acid mine drainage (AMD) water which bring disruptive consequences for terrestrial and aquatic ecosystems (1-5). Currently, 8 species have been reported in genus *Acidithiobacillus* including *Acidithiobacillus albertensis, Acidithiobacillus*
caldus, Acidithiobacillus cuprithermicus, Acidithiobacillus ferridurans, Acidithiobacillus ferriphilus, Acidithiobacillus ferrivorans, Acidithiobacillus ferrooxidans and Acidithiobacillus thiooxidans. Acidithiobacillus species are universally found at AMD sites where highly toxic inorganic forms of heavy metals and metalloids such as arsenic, mercury, copper, cadmium and cobalt often dominate (1, 6-9). As revealed in previous studies, strains of A. ferrooxidans could adapt up to 800 mM Cu^{2+}, 84 mM As^{3+}, 1071 mM Zn^{2+}, 500 mM Cd^{2+}, 1000 mM Ni^{2+} and 50 mM Hg^{2+} (10, 11), while strains of A. thiooxidans and A. ferrooxidans were able to adapt up to 120 and 100 mM As^{5+} (12). It is obvious that Acidithiobacillus species surviving in acidic metal-rich leaching environments must possess certain highly efficient and advanced metal resistance mechanisms, rendering them ideal research models to improve our understanding of metal resistance. There are four main strategies that acidophile generally utilizes to increase resistance to metals including chelation and precipitation of toxic metals, efflux of the toxic metals out of the cell by molecular pumps, enzymic conversion of metal ions to less toxic forms and/or enhanced cell wall and membrane that are impermeable for toxic metal ions (13). These strategies are closely related to the genetic properties of the bacteria. The genes conferring toxic metal resistance are prone to cluster together as gene operons in Acidithiobacillus genomes such as the mer operons that confer mercury resistance (14). However, the mechanisms driving the genes to cluster together as gene operons are far from clear. For example, the mer operon generally contains merA, merD, merR, merP, merC, and merT, although their organization may be different. Reported genes related to mercury, arsenate, cadmium, zinc, cobalt and copper
Horizontal gene transfer (HGT), together with natural selection and gene duplication, are three main engines that drive adaptive evolution of microbial genomes, whereas their relative importance is still ambiguous (15, 16). HGT events generally occur via mobile genetic elements (MGEs) carrying functional genes that can enhance the adaptability and versatility of the acceptors. Recent studies revealed that over 20% of the microbial genes may be acquired through HGT after analyzing gene-transfer webs amongst a vast number of sequenced genomes (17, 18). Mobile genetic elements (MGEs) such as plasmids, phages, genomic islands (GIs), transposons and insertion sequences are indicative of HGT events. GIs represent a versatile gene pool which are large chromosomal regions usually found adjacent to or integrated within tRNA genes (19). Deviant G+C contents, as well as incongruences between phylogenies of functional genes and species can also be used as indicatives of HGT events (20-23). Some previous studies have described overall metal resistance mechanisms in acidophile and indicated that horizontal gene transfer (HGT) was a main adaptive mechanism for acidophile to cope with heavy metal-rich environments (13, 24-27), for example, it was reported that *A. ferrooxidans* ATCC 53993 contained GIs inserted after tRNA metabolism related gene (MiaB type) carrying genes encoding mercury detoxification determinants (*merA, merC* and *merR*) and copper translocating P-type ATPase (13).
However, direct evidences of HGT in other *Acidithiobacillus* species in response to challenging metal-rich environments and mechanisms addressing how metal resistance genes originated and evolved in *Acidithiobacillus* are still lacking. The rapid development of next-generation sequencing technologies has enriched the public database with numerous sequenced genomes. Exploiting these resources, large-scale comparative genomic analyses are improving our understanding of extensive genetic diversity and evolutionary history. To date, 8 species have been reported in genus *Acidithiobacillus* and strains of 6 species (*A. thiooxidans, A. ferrooxidans, A. caldus, A. ferrivorans, A. albertensis* and *A. ferridurans*) have been sequenced. In an attempt to reveal the genetic traits and phylogenetic history of metal resistance in genus *Acidithiobacillus*, comparative genomic analysis of 37 strains within the genus *Acidithiobacillus* was carried out. In addition, molecular phylogenetic analysis, gene contents, molecular conservation, linear analysis along with selective pressure and codon usage analysis were used to explore the evolution of toxic metal resistance genes in genus *Acidithiobacillus* (28-31). Our results provided explicit evidences that HGT played a crucial role in driving the evolution of heavy metal resistance in *Acidithiobacillus* species surviving extreme environments. The insights gained in this study improved our understanding on the complex evolutionary history of toxic metal resistance genes in *Acidithiobacillus* spp. and its roles in the biogeochemical cycling of different metals.

2 Results
2.1 Genomic Features

A summary of features for 38 available sequenced genomes of *Acidithiobacillus* spp. are listed in Table 2. The G+C contents of the 38 genomes range from 52.6 to 61.5%. These genomes vary in size by approximately 2.48 mega-bases (ranging from 1.70 to 4.18 Mb) with coding sequence (CDS) numbers ranging from 1617 to 4359, suggesting substantial strain-to-strain variations.

2.2 Phylogenetic Analyses and Pan-genome Analyses

To associate the distribution of metal resistance genes in *Acidithiobacillus* spp. with their phylogenetic affiliation, we constructed the phylogenetic tree of the *Acidithiobacillus* spp. with closely related species based on 16S rRNA gene sequences using NJ methods (Supplementary Fig. S1). Results showed that *Acidithiobacillus* species were phylogenetically close to *Thermithiobacillus* spp.. Since no 16S rRNA gene sequence could be retrieved from the genome of *Acidithiobacillus* sp. NORP59, we constructed phylogenetic tree of the 38 *Acidithiobacillus* genomes with their phylogenetic affiliations based on whole genome sequences (Supplementary Fig. S2). Results showed that *Acidithiobacillus* sp. NORP59 form a clade with *Thioglobus* sp. MED-G25, distant from standard strains of *Acidithiobacillus* spp., in addition, *Acidithiobacillus* sp. NORP59 showed low values of BLASTN-based average nucleotide identity (ANI; average 75%) and tetranucleotide composition regression (Tetra; average 0.758) in comparison with other available *Acidithiobacillus* strains (Supplementary Table S1), indicating that strain NORP59 was
not a member of genus *Acidithiobacillus*, thus it was excluded from any further analyses.

To reveal the genomic features specific to each strain, we identified orthologous groups among the 37 *Acidithiobacillus* genomes using BPGA pipeline (32). Our analysis of the total of 37 genomes revealed a pan genome containing 141,937 putative protein-coding genes in the genus *Acidithiobacillus*. Out of these 141,937 genes, 16,317 (11.5%) were clustered into the core genome of *Acidithiobacillus* spp., 95704 (67.4%) were represented in the accessory genome and 5332 (3.8%) were identified as strain-specific genes. The number of specific genes ranges from 0 to 811 (Fig. 1A). Result of BacMet database annotation revealed that accessory genome had a higher proportion (8.2%, namely 7846 genes) of genes involved in heavy metal resistance activities (Fig. 1C) than core genome (5.2%, namely 851 genes) and unique genes (0.3%, namely 16 genes). We constructed robust phylogenetic tree of the 37 genomes based on concatenation of the 110 single-copy core genes in each genome using the NJ method (Supplementary Fig. S3). The phylogenetic trees, inferred using ML and UPGMA methods (Supplementary Fig. S4, S5), were congruent with the NJ phylogenetic tree, showing that the phylogenetic tree was highly reliable. A chronogram for NJ phylogenetic tree was constructed with calibration point to explore the divergence times of *Acidithiobacillus* (Fig. 2). The most recent common ancestor (MRCA) of *Acidithiobacillus* (i.e., the emergence time of *A. caldus*) was estimated to have emerged around 800 MYA (Fig. 2), whereas the most recent common ancestor (MRCA) of *A. thiooxidans*, *A. ferrooxidans*, *A. ferrivorans* and *A. ferridurans* was estimated to have emerged around 439 MYA.
These well-supported core gene trees showed that some strains in different reported species were grouped together, such as, *A. ferridurans* JCM 18981 and *A. ferrooxidans* IO-2C, *Acidithiobacillus* sp. GGI-221 and strains of *A. ferrooxidans*, *A. albertensis* DSM 14366 and strains of *A. thiooxidans*. These results suggested that there were mistakes in the classification of *Acidithiobacillus* spp.. ANI approach was further applied to evaluate the genomic similarity of microorganisms (33). The ANI results (Supplementary Table S1) justified the conclusion of phylogenetic analysis. All 37 strains could be classified into 9 species based on ANI ≥ 96%.

Comparison of strains JCM 18981 and IO-2C resulted in high ANI (98.9%), suggesting that they belong to the same species (*A. ferridurans*). Comparison between strains *A. albertensis* DSM 14366 and *A. thiooxidans* spp. resulted in high ANI (97.3%), suggesting that they belong to the same species (*A. thiooxidans*). Comparison between strains GGI-221 and *A. ferrooxidans* spp. resulted in a high ANI (99.5%), suggesting that they belong to the same species (*A. ferrooxidans*).

Thus in this article below we renamed strain IO-2C to *A. ferridurans* IO-2C (*A. ferrooxidans* IO-2C), strain DSM 14366 to *A. thiooxidans* DSM 14366 (*A. albertensis* DSM 14366), strain GGI-221 to *A. ferrooxidans* GGI-221 (*Acidithiobacillus* sp. GGI-221).

### 2.3 Models Extrapolation of Pan-genome of *Acidithiobacillus*

Vernikos et al. (34) stated that mathematical extrapolation of pan-genome would be highly reliable provided that sufficient genomes (more than five) are involved. As the deduced power
law regression function \[ P_s(n) = 4451.42n^{0.36465} \] shown (Fig. 1B), the pan-genome of *Acidithiobacillus* had a parameter (\( \gamma \)) of 0.365 falling into the range \( 0 < \gamma < 1 \), suggesting that the pan-genome was open (35). The deduced exponential regression \[ F_c(n) = 2052.65e^{-0.0714494n} \] revealed that the extrapolated curve of core genome followed a steep slope, reaching a minimum of 441 gene families after the 37th genome was added (Fig. 1B). The number of core genes was relatively constant that an extra genome added would not significantly affect the size of core genome, which was consistent with the notion that core genes were conserved genes universally present in all strains (36).

### 2.4 Identification of Putative Mobile Genetic Elements

Mobile genetic elements (MGEs) are specific genome segments capable of intra- and extracellular movements, carrying gene(s) encoding for putative functions, which are considered as indicatives of HGT events (37). In this study, MGEs including transposases, integrases, genomic islands (GIs) and phage-associated genes were identified and compared in all genomes. Each *Acidithiobacillus* genome in this study contained a number of transposons (Supplementary Table S3). The number of transposon copies per genome ranged from 143 (*A. thiooxidans* CLST) to 334 (*A. caldus* SM-1). Members of the IS3, IS5, IS21 and IS110 families were the most common. The number of GIs copies per genome ranged from 25 (*A. ferrivorans* SS3) to 66 (*A. thiooxidans* CLST) (Supplementary Table S4). The number of prophages and/or prophage
remnants ranged from 0 to 6 (Supplementary Table S5). BacMet annotation results revealed that genomes of *Acidithiobacillus* spp. contained a vast repertoire of metal resistance genes with an average 160 BacMet hits in each genome. Visualization of chromosome of representative strains of different species of *Acidithiobacillus* (Supplementary Fig. S6) revealed that many of these genes were distributed within genomic island regions or flanking with other mobile genetic elements such as transposons, indicating that putative HGT events contributed greatly to the adaptive survival of *Acidithiobacillus* spp. in toxic metal-rich niches.

### 2.5 Estimation of Gene Family Gains and Losses

We further investigated how the metal resistance related gene families might have been gained and lost during the evolution of *A. thiooxidans*, *A. ferroxidans*, *A. ferrivorans*, *A. ferridurans* and *A. caldus* (Supplementary Fig. S7). Gene family gains were observed on the branches from *A. ferridurans*, *A. caldus* and the the most recent common ancestor (MRCA) of *A. ferridurans* and *A. ferroxidans*. In all, new gene families arose over time probably via HGT, with only a few gene families being lost.

### 2.6 Case Study of Hg-Related Genes

The overall distribution and organization of Hg resistance genes in 37 *Acidithiobacillus* strains are summarized in Supplementary Fig. S8. Hg resistance genes *merA*, *merC*, and *merR* were
detected in most Acidithiobacillus genomes, while merD, merP, and merT, were only detected in A. ferrooxidans, A. thiooxidans and A. ferrivorans genomes. The gene merB that encodes organomercury lyase was missing in Acidithiobacillus genomes. The mer genes in a genome were prone to group together as mer clusters in Acidithiobacillus genomes, which could be categorized into six sub-groups (merCAD, merRTPA, merTAR, merACR, merAR and merA). Some strains in this study possessed two different mer clusters in their genomes (Supplementary Fig. S8). However, the mer clusters were not detected in Acidithiobacillus sp. SH, possibly resulting from incomplete sequencing.

2.6.1 The Origin and Evolution of mer Clusters

The divergent distribution and organization of mer genes in Acidithiobacillus raised a question as to their origin and evolution. The deviant G+C content can be used as a detect method of HGT (20, 21). Herein, we detected the G+C contents of mer clusters and their corresponding genomes. The results showed that the G+C contents of the merCAD clusters were higher than those of the genomes in A. ferrooxidans strains (62.8 vs. 57.6–61.0); the G+C contents of respective merAPTR clusters were higher than those of the genomes in A. thiooxidans strains (60.5–61.1 vs. 52.4–53.2), A. ferrooxidans strains (62.6–62. vs. 58.6) and A. ferridurans strains (63.7–64.1 vs. 58.4–61.3). The merTAR and merACR gene clusters also showed variations of G+C contents between mer gene clusters and their corresponding genomes (Supplementary Fig. S9). To further
To elucidate the origin of *mer* genes clusters in *Acidithiobacillus*, a NJ phylogenetic tree was constructed based on the MerA protein sequences. As shown in Fig. 3, the strains could be separated into three groups, of which the group I included *merCAD* and *merRTP A* clusters, the group II represented *merACR* clusters and group III contained *merRTAC, merTA and merAR* clusters and stand-alone *merA* genes. The phylogenetic tree revealed that the group I and *mer* genes of Burkholderiales members were sister groups, while the group II were more similar to *mer* genes from *Acidihalobacter* spp., the group III were more similar to *mer* genes of *Thiobacillus* spp. and *Acidiferrobacter thiooxydans*, implying that the group I *mer* cluster may be
acquired via HGT from members of Burkholderiales, and group II mer cluster from Acidihalobacter spp. and group III mer cluster from Thiobacillus spp. or Acidiferrobacter spp. in early evolutionary history. In addition, A. thiooxidans Licanantay containing cluster merRTAC branches near the base of the merTA and merAR clade. We inferred that among the group III, gene clusters merRTAC, merTA and merAR were likely originated in the A. thiooxidans followed by gene loss in A. thiooxidans and HGT to the A. ferrivorans. Besides, only stand-alone merA genes were detected in strains of A. caldus. The phylogenetic tree revealed that merA1 and merA2 genes of A. caldus form separate groups, in which merA2 genes of A. caldus were more similar to those from A. ferrivorans. The genes in the upstream and genes in the downstream were conserved for merA1 and merA2 genes, respectively (Supplementary Fig. S10). Several genes encoding mobile element protein which are indicative of HGT were found near the merA2 genes. All these indicated that merA2 genes of A. caldus arose from early HGT events, most likely from members within genus Acidithiobacillus.

2.7 Case Study of As-Related Genes

The overall distribution and organization of As resistance genes in Acidithiobacillus spp. are summarized in Supplementary Fig. S11. As resistance genes (ars genes) including arsR, arsB, and arsC were detected in all Acidithiobacillus genomes, while arsA and arsD were only detected in A. ferrivorans genomes and some strains from other species including A.
ferrooxidans DLC-5 and A. ferridurans IO-2C (A. ferrooxidans IO-2C), arsH genes were only detected in genomes of A. ferrooxidans (within arsCRBH cluster) and A. thiooxidans (stand-alone) (Supplementary Fig. S11). Stand-alone arsM genes were also detected in all species. However, aio, arr involved in arsenite oxidation and respiratory reduction of arsenate were not found in Acidithiobacillus genomes, suggesting that cytoplasmic As(V) reduction and As(III) extrusion are the main As resistance strategies used in genus Acidithiobacillus spp.. The ars genes in Acidithiobacillus genomes were prone to group together as ars clusters including four sub-groups, arsADCR, arsRB, arsBRC and arsBRCH. The arsB gene of A. caldus MTH-04 was intercepted by an exogenous insertion sequence which may result in malfuction of this arsB gene, rendering A. caldus MTH-04 vulnerable to toxic arsenic ions (Supplementary Fig. S11).

2.7.1 The Origin and Evolution of ars Clusters

The distribution and organization of ars genes in Acidithiobacillus raise a question as to their evolution. The results showed that the G+C contents of the arsADCR clusters were higher than those of the genomes in A. ferrivorans strains (60.0–63.3 vs. 56.4–58.6); the G+C contents of respective arsBRC clusters were slightly higher than those of the genomes in A. caldus strains (62.3–63.1 vs. 60.9–61.5) and A. thiooxidans strains (53.1–55.3 vs. 52.4–54.3), showing variation of G+C contents between clusters and the corresponding genomes (Supplementary Fig. S12). Comparison of the chromosomal regions flanking the ars gene clusters among...
Acidithiobacillus strains revealed that the genes in the upstream and downstream regions were conserved among strains of the same species (Supplementary Fig. S13). There were several genes encoding mobile element protein (transposases or recombinases) locating near the arsADCR and arsBRC clusters, and there were tRNA genes flanking the arsBRC clusters of *A. caldus*, indicating putative HGT events (Supplementary Fig. S13). Strains of the same species share the same insertion sites, whereas strains of different species resulted in different insertion sites, suggesting that ars clusters may be acquired more than once via different HGT events. The flanking regions of the arsBRC gene clusters in strain *A. thiooxidans* were homologous to the corresponding regions of strain *A. ferrivorans* SS3 and *A. ferrooxidans* ATCC 23270 that have no arsBRC; the same phenomena were found in arsBR gene clusters and arsH genes. These results suggested that ars genes may be acquired by different HGT events followed by gene loss in some strains (Supplementary Fig. S13).

To gain insights into the origin of ars genes clusters, NJ phylogenetic trees were constructed based on the concatenated ArsBRC protein sequences and ArsAD protein sequences. As shown in Fig. 4, phylogenetic analysis revealed that the strains possessing arsBRC clusters formed monophyletic clade. *Acidithiobacillus* spp. likely obtained the arsBRC cluster via HGT from *Acidihalobacter* followed by rearrangement and gene loss since ars gene clusters of *Acidihalobacter prosperus* branch near the base of the clade of *Acidithiobacillus*. NJ phylogenetic trees based on concatenated ArsAD protein sequences show that the ars clusters of
Acidithiobacillus and ars clusters of Alcaligenes faecalis, Pseudomonas hunanensis and Sulfuriferula spp. were sister groups (Fig. 5), implying that Alcaligenes faecalis, Pseudomonas hunanensis and Sulfuriferula spp. are likely donors of arsADCR cluster.

Stand-alone arsM genes encoding As(III) S-adenosylmethyltransferase were found located distant from canonical ars gene clusters. NJ phylogenetic trees constructed based on the ArsM protein sequences revealed (Supplementary Fig. S14) that the Acidithiobacillus strains formed three separate sub-groups, of which sub-group I comprises ArsM from A. thiooxidans, A. ferrooxidans, A. ferridurans, A. ferrivorans and Acidithiobacillus sp. SH, grouping with ArsM from Thiomonas. Sub-group II comprises two deviant ArsM from A. ferrooxidans and A. thiooxidans (GenBank accession: WP_11352344401; WP_080707752.1) grouping with ArsM from Melaminivora and Pseudomonas. Sub-group III comprises arsM from A. caldus grouping with ArsM from members of Rhodospirillales. Incongruences between phylogenies of ArsM and species imply that the sub-group I arsM genes may be acquired via HGT from members of Thiomonas, sub-group II arsM genes from members of genus Melaminivora and Pseudomonas and sub-group III arsM genes from members of Rhodospirillales in early evolutionary history.

2.8 Case Study of Cd, Zn, Co, Cu-Related Genes

Gene clusters encoding Czc/CusABC homologs which are involved in the detoxification of
divalent cations (cadmium and zinc and cobalt)(38, 39) and monovalent cations (copper and silver)(13, 40, 41) were detected in all *Acidithiobacillus* genomes (Supplementary Fig. S15). In addition, genes encoding CzcD and genes encoding copper-translocating P-type ATPase (Cop) were detected in all *Acidithiobacillus* genomes, most of which located next to *czcABC* gene clusters. However, genes that encode multi-copper oxidase were only detected in strains of *A. thiooxidans* and *A. caldus*. The above-mentioned genes represent the dominant genetic traits that contribute resistance to Cd, Zn, Co and Cu in *Acidithiobacillus* genomes. Considerable numbers of additional copies of *czc*-like and *cop*-like genes were also present. The *czc* clusters could be divided into two sub-groups, *czcABC* and *czcCAB* (Supplementary Fig. S15). The gene clusters in the organization of *czcABC* were observed in strains across all species of *Acidithiobacillus*, of which strains of *A. ferrooxidans* and *A. ferridurans* possess another copies of *czcABC* located on different sites suggesting that *czc* clusters may be acquired more than once. There are also tRNA genes and genes encoding mobile element protein flanking *czcABC* clusters (Supplementary Fig. S15). While each *A. caldus* strain possess a *czcCAB* cluster in addition to existing *czcABC* cluster (Supplementary Fig. S15). The *czc* gene clusters exhibited more than 90% identity within each sub-group and over 80% identity between two sub-groups. Actually, the *czc* genes were detected in all strains of *Acidithiobacillus*. The genes in the upstream and downstream regions of a *czcABC* cluster were conserved among strains of all species of *Acidithiobacillus* (Supplementary Fig. S15A), which indicated that this *czcABC* gene cluster was acquired before the divergence of genus *Acidithiobacillus*. Notably, *cop* genes encoding copper-translocating
P-type ATPase located in the upstream the \textit{czcABC} clusters are conserved across all species of \textit{Acidithiobacillus} (Supplementary Fig. S15A).

### 2.8.1 The Origin and Evolution of \textit{czc} Clusters

The G+C contents of the \textit{czcABC} clusters are higher than those of the genomes in \textit{Acidithiobacillus} spp. (62.2–64.2 vs. 52.9–61.5) except for \textit{A. caldus}, which is lower than those of the \textit{A. caldus} genomes (59.4–59.5 vs. 60.9–61.4); the G+C contents of respective \textit{czcBAC} clusters are higher than those of the genomes in \textit{A. caldus} strains (64.9–66.0 vs. 60.9–61.5), presenting variations of G+C contents between clusters and the corresponding genomes (Supplementary Fig. S16). To gain insights into the origin of \textit{czc} genes clusters in \textit{Acidithiobacillus}, a NJ phylogenetic tree was constructed based on concatanated CzcABC protein sequences. As shown in Fig. 6, the strains possessing \textit{czcABC} and \textit{czcBAC} clusters form separate groups. Notably, the phylogeny reveals that the \textit{czcABC} of \textit{Acidithiobacillus} and \textit{czc} clusters of \textit{Thiomonas} spp. and \textit{Acidihalobacter} spp. are sister groups, whereas \textit{czcBAC} of \textit{Acidithiobacillus} and \textit{czc} clusters of another clade of \textit{Thiomonas} spp. are sister groups. These results implied that the \textit{czcABC} cluster may be acquired via HGT from members of \textit{Acidihalobacter} and \textit{Thiomonas}, while \textit{czcBAC} may be acquired from members of \textit{Thiomonas}.

### 2.8.2 The \textit{mco} Genes in \textit{Acidithiobacillus} spp.


Stand-alone *mco* genes that encode multi-copper oxidase which oxidized Cu\(^+\) to the less toxic Cu\(^{2+}\) form were found in all strains of *A. thiooxidans* and *A. caldus* in this study. NJ phylogenetic tree based on the multi-copper oxidase protein sequences reveals that the multi-copper oxidases of *Acidithiobacillus* were adjacent to those of *Acidihalobacter* spp., implying that *Acidithiobacillus* and *Acidihalobacter* have a common multi-copper oxidases ancestor (Supplementary Fig. S17). Synteny of the chromosomal regions flanking the *mco* genes showed that multi-copper oxidase was probably lost in strains of *A. ferrivorans*, *A. ferridurans* and *A. ferrooxidans* (Supplementary Fig. S18).

### 2.8.3 The *czc*-like and *cop*-like Genes in *Acidithiobacillus*

Our studies revealed that there are multiple *czc*-like genes and *cop*-like genes present in the genomes including 1–5 pairs of *czcA*-like genes, 1–5 pairs of *czcB*-like genes, 1–4 pairs of *czcC*-like genes, 1–4 pairs of *czcD*-like genes and 1–3 pairs of *cop*-like genes (Supplementary Fig. S20, S21, S22, S23 and S24). Alignment of residues of CzcA and CzcA-like proteins reveals that the “DDE” motifs essential for CzcA function (42) are highly conserved (Supplementary Fig. S25). Alignments of residues of Cop and Cop-like proteins reveals that the metal binding domains (CASC....CASC), translocation domains (CPCAMGLA) and phosphorylation domains (FDKTGTLT) (43) are conserved, indicating that these genes may be functional (Supplementary Fig. S26). Phylogenetic analysis revealed that *czc*-like and *cop*-like genes from the same species...
clustered together and there were incongruences between phylogenies of functional genes and species, which suggested that multiple cze-like and cop-like genes may arise from gene duplication and HGT (Supplementary Fig. S20, S21, S22, S23 and S24).

2.9 Assessment of Functionality of Metal Resistance Genes

We also assessed indirectly the functionality of above-mentioned metal resistance related genes by assessing the strength of natural selection acting on these genes, as well as CAI (Codon Adaption Index) of these genes. Results showed that 83.3% of above-mentioned metal resistance genes had a rate of nonsynonymous substitutions that was lower than the rate of synonymous substitutions (dN/dS < 1), suggesting purifying selection on these genes (Fig. 7 and Supplementary Table S6). Only a merA gene from Acidithiobacillus thiooxidans A02 (dN/dS = 1.51), a cop gene from Acidithiobacillus ferrooxidans CCM 4253 (dN/dS = 1.39) and an arsH gene from Acidithiobacillus ferridurans JCM 18981 (dN/dS = 1.82) showed dN/dS ratios >1, indicating that they might be under positive selection. The lowest dN/dS ratio was observed for arsB genes (average dN/dS = 0.06) and arsM genes (average dN/dS = 0.04), showing strong purifying selection. All these metal resistance genes with dN/dS ratio < 1 were supposed to be most essential for the growth of Acidithiobacillus spp.. However, merA genes from A. thiooxidans, arsR genes from A. thiooxidans, merC genes from A. caldus, merP genes from A. thiooxidans and thiooxidans and A. ferridurans, merC genes from A. caldus, mco genes from A. thiooxidans and
some cop-like and czc-like genes show dN/dS ≈ 1, indicating that selection force had relaxed on them (Fig. 7 and Supplementary Table S6). Highly expressed genes in bacteria often have a stronger codon bias than genes expressed at lower levels, due to translational selection (44-46).

In this study, expression levels of metal resistance related genes in the Acidithiobacillus spp. were predicted using CAI (Codon Adaptation Index) as a numerical estimator (47, 48). Approximately 53.2% of the metal resistance genes were predicted to be highly expressed genes using CAI cutoff values: greater than 0.73 in A. thiooxidans, greater than 0.75 in A. caldus and greater than 0.76 in A. ferrooxidans, A. ferrivorans and A. ferridurans (Fig. 8 and Supplementary Table S7). The cutoff values were indicated with average CAI values of Tu genes in each species.

3 Discussion

In the present work, we characterized the genomic features of Acidithiobacillus spp. focusing on the distribution, organization, the complex evolutionary history and functionality of metal resistance genes. With the rapid advances in genomic analysis technology, new measurements such as ANI are being developed to evaluate the genomic similarity between bacteria (33). In this study, we redetermined the phylogenetic status of 38 Acidithiobacillus spp. using ANI and phylogenetic trees. Our results showed that strains DSM 14366, GGI-221 and IO-2C were misnamed. We therefore reclassified strain IO-2C to A. ferridurans IO-2C (A. ferrooxidans), DSM 14366 to A. thiooxidans DSM 14366 (A. albertensis DSM 14366), GGI-221 to A.
ferrooxidans GGI-221 (Acidithiobacillus sp. GGI-221). Low ANI values of Acidithiobacillus sp. NORP59 in comparison with other Acidithiobacillus strains and distant phylogenetic relation indicated that strain NORP59 is not a member of genus Acidithiobacillus. Our study revealed complex taxonomy of genus Acidithiobacillus and demonstrated the usefulness and accuracy of advanced bioinformatics measurements such as ANI in rectifying inaccurate identifications from previous studies.

There is generally a positive correlation between the abundance of mobile genetic elements and the frequency of HGT (49). We found that Acidithiobacillus spp. contain 0–6 prophages and/or prophage remnants, 143–334 IS elements and 25–66 GIs per genome, suggesting that Acidithiobacillus spp. were rich in mobile genetic elements. The existence of transposable elements and prophage near the heavy metal resistance genes and gene clusters suggested that they may be involved in HGT of heavy metal resistance genes, which was in consistence with previous studies illustrating that HGT was a major mechanism that confers bacteria crucial traits such as metal resistance in adaptation to toxic metal-containing habitats (15, 50-52). In addition, BacMet database annotation revealed that accessory genome had a higher proportion (8.2%, namely 7846 genes) of genes related to heavy metal resistance activity (Fig. 1C) and models extrapolation revealed that pan-genome of 37 Acidithiobacillus spp. was ‘open’, which was in line with results that metal resistance associated gene family gain events overcome loss events. We inferred that during evolutionary niche adaption, Acidithiobacillus spp. had accessorized
their genome armaments with genetic elements transferred from members outside their species or
genus, with the help of MGEs. The “accessory-genome” was therefore considered as a massive
resource that provided *Acidithiobacillus* spp. with unprecedented elasticity to improve their
fitness adapting harsh conditions (53-55). Our results also showed that transformation of heavy
metal to less toxic forms followed by efflux out of cells was an often-used metal resistance
strategy of *Acidithiobacillus* spp., which is also a process contributing to biogeochemical cycling
of metals.

We implemented a combination of several approaches to discover putative HGT events including
deviant G+C content detection, phylogenetic analysis, mobile genetic element detection and
chromosomal synteny analysis since it was difficult to identify HGT events via deviant G+C
contents alone if HGT happened to occur between the organisms with similar G+C contents (56).
The different gene contents and organization of metal resistance genes suggested complicated
evolutionary history of these genes, and also suggested different genetic requirements for
effective cellular metal detoxification and regulation of metal resistance operons. We supposed
*Thiomonas* spp. to be the donors of *arsM* gene and *czcABC, czcBAC* gene clusters in
*Acidithiobacillus* spp., which was reasonable since members of *Thiomonas* spp. were found
ubiquitous in heavy metal-contaminated environments, especially in AMD sites, the same habitat
as those of *Acidithiobacillus* species and usually contained a vast flexible gene pool, for instance,
the genome of *T. arsenitoxydans* contained a circular chromosome and a plasmid (pTHI) where
more than 20 GIs (ThGEI-A–ThGEI-S) resided, which were associated with heavy metal resistance and conjugation (57-59). In addition, the median G+C content of *Thiomonas* spp. genomes was 64.0% which was close to those of *czcABC* (64.2%), *czcBAC* (64.9%) gene clusters in *Acidithiobacillus* spp. All these suppose *Thiomonas* spp. to be the most possible sources of multiple heavy metal resistance operons that were acquired by *Acidithiobacillus* spp. via cross-phylum HGT. Similar traits were also shown in case of *Acidihalobacter* spp., the putative cross-phylum donors of *merACR, arsBRC* and multi-copper oxidase encoding gene which was a group of halophilic, mesophilic, heavy metal tolerant, extreme acidophiles that shared same habitats with *Acidithiobacillus* spp., with an average genome G+C content (61.9%) similar to those of *merACR* (60.0%), *arsBRC* (61.7%) gene clusters in *Acidithiobacillus* spp. (60, 61). This could also be applied to *Acidiferrobacter* spp. that inhabit AMD sites, which were predicted to be the cross-phylum donors of *merAPTR* gene cluster and rusticyanin encoding gene (62). Multiple *czc*-like genes and *cop*-like genes were present in the genomes of *Acidithiobacillus* spp. which may arise from gene duplication. Metal resistance proteins encoded by genes such as *czc* and *cop* are in high demand for *Acidithiobacillus* spp. to cope with harsh metal-rich environments. The presence of duplicate metal resistance genes may be beneficial by enhancing regulatory elasticity of *Acidithiobacillus* spp. simply because extra amounts of protein or RNA products are generated (63).

The gene contexts of metal resistance genes in different *Acidithiobacillus* species were not.
conserved which indicated that these genes may be acquired more than once via different HGT events. Synteny of the chromosomal regions flanking the mco genes showed that multi-copper oxidases were possibly lost in strains of *A. ferridurans*, *A. ferrivorans* and *A. ferrooxidans* during evolution since *mco* genes in *A. thiooxidans* experienced relaxed selection pressure (dN/dS ≈ 1) and *A. thiooxidans* arose before the speciation of *A. ferridurans*, *A. ferrivorans* and *A. ferrooxidans* in evolutionary history (Supplementary Fig. S7). We supposed that the function of this enzyme was replaced in *A. ferridurans*, *A. ferrivorans* and *A. ferrooxidans* by rusticyanin, a periplasmic protein with affinity for Cu$^+$ forming part of an iron-oxidizing supercomplex and was supposed to confer a higher level of copper resistance (64, 65). Rusticyanin was absent in species incapable of oxidizing iron such as *A. thiooxidans* and *A. caldus* and was acquired via HGT from *Acidiferrobacter* spp. to *A. ferridurans*, *A. ferrivorans* and *A. ferrooxidans*, as revealed by phylogenetic analysis (Supplementary Fig. S19). Most non-synonymous substitutions are deleterious especially for functionally important genes, since the functionality of important proteins might be substantially affected even if small changes happens in its sequences, resulting in dN/dS < 1 of these genes considering that mutants of these genes are prone to be eliminated by purifying (negative) selection. In the contrary, when new mutations are beneficial conferring adaptive advantages, these genes might be under diversifying (directional or positive) selection, reflected by dN/dS > 1, namely there are more non-synonymous than synonymous nucleotide substitutions. Nonsynonymous nucleotide substitutions and synonymous substitutions will accumulate at a similar rate as (dN/dS ≈ 1) if genes become non-functional in the population,
since there is no selection pressure acting to maintain the adaptive variant (neutral evolution), and these genes are more prone to lost during evolution (66). Results showed that 83.3% of metal resistance genes in *Acidithiobacillus* species had a lower rate of nonsynonymous substitution than that of synonymous substitution (dN/dS < 1), which is in consistent with the assumption of predominant purifying selection on functionally important metal resistance genes. Besides, approximately 53.2% of the metal resistance genes were predicted to be highly expressed with CAI (Codon Adaption Index). There is a close correlation between codon usage bias and gene expressivity which results from selection for efficient and accurate translation of highly expressed genes (67, 68). CAI was used to measure the synonymous codon usage bias for a given DNA sequence with ever-improving algorithm by comparing the similarity between the synonymous codon usage of a gene and the synonymous codon frequency of a reference datasets of more than 30 highly expressed genes derived from prokaryotes and lower eukaryotes, which can serve as a reliable indicator of codon usage bias as well as numerical estimator of gene expressivity for genes of prokaryotes and lower eukaryotes (65, 69). Widespread low dN/dS ratios and relative high CAI values across metal resistance related genes emphasized their indispansible functions in supporting the growth of these extremophiles in the face of harsh environments. Taken together, our research revealed that *Acidithiobacillus* spp. recruited and consolidated additional novel resistant functionalities in adaption to challenging metal-rich environments via HGT, gene duplication and purifying selection. However, further experimental confirmations of the functionalities and expression levels of metal resistance related genes in
Acidithiobacillus spp. are still necessary for advanced understanding of these genes.

4 Concluding Remarks

We performed comparative genomic analysis of 37 strains within the genus Acidithiobacillus to explore the distribution, organization, functionality and the complex evolutionary history of metal resistance genes in Acidithiobacillus spp. Results indicated that the evolutionary history of toxic metal resistance genes in Acidithiobacillus spp. involved a combination of gene gain and loss, HGT, gene duplication and gene family expansion. Gene clusters involved in toxic metal resistance such as mer, ars, cze/cus were found to be acquired by early horizontal gene transfer (HGT) events from sources related to Burkholderiales, Rhodospirillales, Acidihalobacter, Thiobacillus, Acidiferrobacter, Alcaligenes, Pseudomonas, Sulfuriferula, Melaminivora and Thiomonas, many of which share habitats with Acidithiobacillus spp. Whereas multi-copper oxidase gene of iron-oxidizing A. ferridurans, A. ferrivorans and A. ferrooxidans involved in copper detoxification was lost, and was replaced by rusticyanin during evolution. Assessment of functionality of metal resistance genes showed that 83.3% of metal resistance genes in Acidithiobacillus spp. were under purifying selection (dN/dS < 1) and 53.2% of them were highly expressed. The insights gained in this study have greatly improved our understanding on the metal resistance strategies of Acidithiobacillus spp.

5 Materials and Methods
A total of 38 sequenced strains belonging to genus *Acidithiobacillus* were collected from NCBI database, containing one to three sequenced standard strains in each species and four unidentified strains. General features of bacterial genomes used in this study were summarized in Table 2.

### 5.1 Average Nucleotide Identity (ANI)

Comparisons of average nucleotide identity (ANI) based on BLAST algorithm (70) and tetranucleotide frequency correlation coefficient (Tetra) (71) were conducted using the web server JSpeciesWS (33, 72) running with default parameters.

### 5.2 Phylogenetic Analyses

To associate the distribution of metal resistance genes in *Acidithiobacillus* spp. with their phylogenetic affiliation, we constructed the phylogenetic tree of the *Acidithiobacillus* spp. with closed related species based on 16S rRNA gene sequences using Neighbor-joining method in MEGA v7.0 (73) with rRNA genes predicted using the RNAmmer v1.2 (74). Since no 16S rRNA gene sequence could be retrieved from the genome of *Acidithiobacillus* sp. NORP59, we also constructed phylogenetic tree of the 38 *Acidithiobacillus* spp. genomes with their phylogenetic affiliation based on whole genome sequences with CVTree 3 (75). Phylogenetic tree of the 37 *Acidithiobacillus* spp. genomes based on concatenation of the 110 single-copy core genes in a genome was constructed with the maximum-likelihood (ML) method using MEGA v7.0 (73).
with Jones-Taylor-Thornton (JTT) model (76) and 1000 bootstraps, rooted by *Escherichia coli* and *Salmonella enterica*. The ambiguous areas of alignment were removed by Gblocks 0.91b (77) before tree constructing. We also constructed Neighbor-joining tree and UPGMA tree with MEGA v7.0 (73) both under p-distance model with 1000 bootstraps of above-mention core genes. A chronogram for NJ phylogenetic tree of the 37 *Acidithiobacillus* species with branch lengths reflecting divergence times was inferred using the divergence time between *Escherichia coli* and *Salmonella enterica* (140 MYA, millions of years ago) (78) as calibration points with Timetree Wizard, a built-in program of MEGA v7.0 (73) with the JTT matrix-based model (76) and the Realtime method (79).

5.3 Genes for Toxic Metal Resistance

We identified genes involved in heavy metal resistance in *Acidithiobacillus* genomes by performing BLASTP search against the BacMet database (80) which contains genes with experimentally confirmed metal resistance functions. Genes meeting the following cutoffs: e-value $<1e^{-50}$, sequence similarity $>35\%$ were considered as reliable hits. The evolutionary history of genes for toxic metal resistance was inferred using MEGA v7.0 (73) with the Neighbor-Joining method with 1000 bootstraps.

5.4 Pan-Genome Analysis and Models Extrapolation of *Acidithiobacillus*
Bacterial Pan Genome Analysis tool (BPGA) pipeline (32) was used to identify orthologous groups among *Acidithiobacillus* testing genomes and to extrapolate the pan-genome models of *Acidithiobacillus* spp. applying default parameters. We defined the set of genes shared among all testing strains as their core genome, the set of genes shared with more than two but not all testing strains as their accessory genome, the set of genes not shared with other strains in each testing strain as the unique genes and the set of non-homologous genes with all testing genomes as the pan genome. The details of all testing strains used are listed in Table 2. We identified genes involved in heavy metal resistance in pan genome, core genome, accessory genome and unique genes by performing BLASTP search against the BacMet database (80) with cutoffs: e-value <1e−50, sequence similarity >35%.

In this study, the size of *Acidithiobacillus* pan-genome was extrapolated implementing an power law regression function $P_s = \kappa n^\gamma$ using a built-in program of BPGA pipeline (32), in which $P_s$ represents the total number of non-orthologous gene families within its pan-genome, $n$ represents the number of tested strains, and both $\kappa$ and $\gamma$ are free parameters (35). The exponent $\gamma < 0$ suggests the pan-genome is ‘closed’ considering the size of the pan-genome is reaching a constant as extra genomes adding in. Conversely, species is predicted to harbor an ‘open’ pan-genome for $0 < \gamma < 1$. In addition, the size of the core-genome was extrapolated fitting into an exponential decay function $F_c = \kappa_c \exp(-n/\tau_c) + \Omega$, with a built-in program of BPGA pipeline (32), where $F_c$ is the number of core gene families, while $\kappa_c$, $\tau_c$, and $\Omega$ are free parameters (81).
5.5 Estimation of Gene Family Gains and Losses

We used BadiRate (82) with ‘BDI-FR-CWP’ method to estimate the number of metal resistance related gene family gains and losses during the evolution of Acidithiobacillus. We conducted the analysis only on 5 representative Acidithiobacillus strains including A. thiooxidans CLST, A. ferrooxidans ATCC 53993, A. ferrivorans PQ33, A. ferridurans JCM 18981 and A. caldus SM-1, because the estimation required complete sets of testing gene families which are only available in species with considerable phylogenetic distance and have high quality (complete or nearly complete) genome databases. We inferred the orthologous groups of metal resistance genes based on reciprocal best hits within each Acidithiobacillus species using the CD-HIT software (global sequence identity cut-off is 0.9 and bandwidth of alignment cutoff is 20aa) (83). These data were used to construct a gene family size file [a matrix (column = species, row = number of orthologous groups)]. We used the topology of the core-gene tree as the reference tree. The number of gene families related to metal resistance in each internal node was inferred using those numbers in branch species and the phylogenetic branch lengths. Gene family gains and losses in each phylogenetic branch were then estimated using Wagner parsimony algorithm under the BDI stochastic model and Free Rates (FR) model assuming that each branch has its own turnover rates.
5.6 Prediction of Mobile Genetic Elements

We used the ISFinder platform (84) to predict and classify insertion sequences (IS) and transposases distributed over *Acidithiobacillus* genomes, with default criteria. We used the IslandViewer 4 platform (85), which integrates two prediction method that make use of sequence composition, i.e., IslandPath-DIMOB (86) and SIGI-HMM (87), and a comparative genomic islands prediction method IslandPick (88), to identify putative genomic islands (GIs). We used PHASTER (Phage Search Tool Enhanced Release) (89) for identification and annotation of prophage sequences within *Acidithiobacillus* genomes. In addition, tRNA genes were predicted using the tRNAscan-SE v2.0 (90). The genes involved in the heavy metal resistance were visualized, along with the tRNAs, insertion sequences, prophage sequences and putative genomic islands, as well as the core genes, accessory genes and unique genes using BRIG-0.95 (91).

5.7 Selective Pressure Analysis and Expressivity Prediction

Strength of natural selection was assessed by estimating the relative rates of non-synonymous (dN) to synonymous (dS) nucleotide substitutions of orthologous metal resistance genes with the HyPhy package (92) using the adaptive branch-site random effects likelihood (aBSREL) method (93) via datamonkey web server (94). A gene in the node or tip of a given tree was considered under positive (diversifying) selection (dN/dS > 1) or negative (purifying) selection (dN/dS < 1) if the P value was <0.05 using the Likelihood Ratio Test after correcting for multiple testing. We
used CAI (Codon Adaption Index) as a numerical estimator of gene expression level (47, 48), a method previously described by G Wu et al. (95). We used CAIcal (96) to calculate CAI values of above-mentioned metal resistance genes. CAI value ranges from 0 to 1.0, with higher CAI value indicating higher expression level. Putative highly expressed (PHX) genes related to toxic metal resistance in the *Acidithiobacillus* spp. were inferred, with *Tu* genes as references which encode elongation factors which are supposed to be highly expressed across most organisms (97).

6 Acknowledgments

We thank the NCBI for providing the genome sequences of *Acidithiobacillus* spp. strains. This work was funded by the National Natural Science Foundation of China (grants 31570113, 41807345 and 41573072) and open funding of state key laboratory of comprehensive Utilization of Low-Grade Refractory Gold Ores (Zijin Mining Group Co., Ltd) (no. 738010212). We would also like to thank the Hunan International Scientific and Technological Cooperation Base of Environmental Microbiology and Application, China. We declare no competing interests.

7 References


Table 1. Genes involved in toxic metal resistance and transformation.

<table>
<thead>
<tr>
<th>Category</th>
<th>Gene</th>
<th>Protein</th>
<th>Function</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mercury Detoxification</td>
<td>merA</td>
<td>Mercuric reductase</td>
<td>Reduce Hg(II) to gaseous mercury Hg(0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>merD</td>
<td>Regulatory protein</td>
<td>Downregulate mer operon expression</td>
<td></td>
</tr>
<tr>
<td></td>
<td>merR</td>
<td>Transcriptional activator/repressor</td>
<td>Regulate the overall expression level of mer operon</td>
<td>(98)</td>
</tr>
<tr>
<td></td>
<td>merP</td>
<td>Periplasmic protein</td>
<td>Hg(II) scavenging</td>
<td></td>
</tr>
<tr>
<td></td>
<td>merC</td>
<td>Inner membrane spanning protein</td>
<td>Transport Hg(II) to the cytoplasm</td>
<td></td>
</tr>
<tr>
<td></td>
<td>merT</td>
<td>Inner membrane spanning protein</td>
<td>Transport Hg(II) to the cytoplasm</td>
<td></td>
</tr>
<tr>
<td><strong>gene</strong></td>
<td><strong>function</strong></td>
<td><strong>role</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td>---------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>merB</td>
<td>Organomercury lyase</td>
<td>Break down organomercury</td>
<td></td>
<td></td>
</tr>
<tr>
<td>arsR</td>
<td>Transcriptional regulator ArsR</td>
<td>Repress the basal expression level of the <em>ars</em> operon</td>
<td></td>
<td></td>
</tr>
<tr>
<td>arsB</td>
<td>As(III) efflux pump protein</td>
<td>Pump out As(III) out of the cell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>arsC</td>
<td>As(V) reductase ArsC</td>
<td>Transform As(V) to As(III)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>arsH</td>
<td>Organoarsenical oxidase</td>
<td>Impacting the oxidation of arsenic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>arsA</td>
<td>As(III) active ATPase</td>
<td>Convert membrane potential to ATP, provide from ATP hydrolysis to ArsB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>arsD</td>
<td>As metallochaperone</td>
<td>Binds and transfers As(III) to ArsA; repress the upper expression levels of the <em>ars</em> operon</td>
<td></td>
<td></td>
</tr>
<tr>
<td>arrA</td>
<td>Large subunit of respiratory As(V) reductase</td>
<td>Provide energy by reducing As(V) to As(III)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>arrB</td>
<td>Small subunit of respiratory As(V) reductase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene</td>
<td>Function</td>
<td>Description</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>----------</td>
<td>-------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>aioA</strong></td>
<td>Large subunit of As(III) oxidase</td>
<td>Provide energy by oxidizing As(III) to As(V)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>aioB</strong></td>
<td>Small subunit of As(III) oxidase</td>
<td>Catalyzes the formation of volatile As(III) S-adenosyl-methyltransferase from As(III) trimethylarsine</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>arsM</strong></td>
<td>As(III) S-adenosyl-methyltransferase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>czc/cusA</strong></td>
<td>Large inner membrane protein</td>
<td>Form a tripartite complex Czc/CusABC that pump out cadmium, zinc, cobalt and copper cations via the mechanism of cation-proton antiporter</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>czc/cusB</strong></td>
<td>Periplasmic coupling protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>czc/cusC</strong></td>
<td>Smaller outer membrane protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>czcD</strong></td>
<td>Cation diffusion facilitator</td>
<td>Efflux of cadmium ions; regulation of the Czc system</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>cop</strong></td>
<td>Copper -translocating P-type ATPase</td>
<td>Pump out Cu(I) ions from the cytosol to the periplasm</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>mco</strong></td>
<td>Multi-copper oxidase</td>
<td>Oxidize Cu(I) to the less toxic form Cu(II)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2. General features of bacterial genomes used in this study.

<table>
<thead>
<tr>
<th>Organism/Name</th>
<th>Strain</th>
<th>Genbank Acc. no.</th>
<th>Level</th>
<th>Size (Mb)</th>
<th>GC%</th>
<th>Gene</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. thiooxidans</td>
<td>A01</td>
<td>GCA_000559045.1</td>
<td>contig</td>
<td>3.8</td>
<td>53.1</td>
<td>4131</td>
<td>3719</td>
</tr>
<tr>
<td>A. thiooxidans</td>
<td>Licanantay</td>
<td>GCA_000709715.1</td>
<td>contig</td>
<td>3.9</td>
<td>52.8</td>
<td>4359</td>
<td>3774</td>
</tr>
<tr>
<td>A. thiooxidans</td>
<td>DMC</td>
<td>GCA_001705625.1</td>
<td>contig</td>
<td>3.9</td>
<td>53.1</td>
<td>4196</td>
<td>3772</td>
</tr>
<tr>
<td>A. thiooxidans</td>
<td>A02</td>
<td>GCA_001705645.1</td>
<td>contig</td>
<td>3.7</td>
<td>53.0</td>
<td>4034</td>
<td>3639</td>
</tr>
<tr>
<td>A. thiooxidans</td>
<td>GD1-3</td>
<td>GCA_001705695.1</td>
<td>contig</td>
<td>3.9</td>
<td>52.9</td>
<td>4225</td>
<td>3824</td>
</tr>
<tr>
<td>A. thiooxidans</td>
<td>BY-02</td>
<td>GCA_001705725.1</td>
<td>contig</td>
<td>3.8</td>
<td>53.1</td>
<td>4196</td>
<td>3683</td>
</tr>
<tr>
<td>A. thiooxidans</td>
<td>JYC-17</td>
<td>GCA_001705755.1</td>
<td>contig</td>
<td>3.8</td>
<td>53.1</td>
<td>4176</td>
<td>3737</td>
</tr>
<tr>
<td>A. thiooxidans</td>
<td>DXS-W</td>
<td>GCA_001705805.1</td>
<td>contig</td>
<td>3.9</td>
<td>52.9</td>
<td>4229</td>
<td>3833</td>
</tr>
<tr>
<td>A. thiooxidans</td>
<td>ZBY</td>
<td>GCA_001756595.1</td>
<td>contig</td>
<td>3.8</td>
<td>53.2</td>
<td>4106</td>
<td>3714</td>
</tr>
<tr>
<td>A. thiooxidans</td>
<td>CLST</td>
<td>GCA_002079865.1</td>
<td>scaffold</td>
<td>4.0</td>
<td>52.4</td>
<td>4045</td>
<td>3619</td>
</tr>
<tr>
<td>A. thiooxidans</td>
<td>ATCC 19377</td>
<td>GCA_000227215.2</td>
<td>contig</td>
<td>3.0</td>
<td>53.2</td>
<td>3106</td>
<td>2847</td>
</tr>
<tr>
<td>A. ferrooxidans</td>
<td>BY0502</td>
<td>GCA_001652185.1</td>
<td>contig</td>
<td>3.0</td>
<td>56.8</td>
<td>3186</td>
<td>2821</td>
</tr>
<tr>
<td>A. ferrooxidans</td>
<td>CCM 4253</td>
<td>GCA_003233765.1</td>
<td>contig</td>
<td>3.2</td>
<td>58.6</td>
<td>3278</td>
<td>3049</td>
</tr>
<tr>
<td>A. ferrooxidans</td>
<td>IO-2C</td>
<td>GCA_003309025.1</td>
<td>contig</td>
<td>2.7</td>
<td>58.7</td>
<td>2822</td>
<td>2620</td>
</tr>
<tr>
<td>A. ferrooxidans</td>
<td>YQH-1</td>
<td>GCA_001418795.1</td>
<td>scaffold</td>
<td>3.1</td>
<td>58.6</td>
<td>3189</td>
<td>2956</td>
</tr>
<tr>
<td>Organism</td>
<td>Accession</td>
<td>GenBank ID</td>
<td>Component</td>
<td>Length</td>
<td>Start</td>
<td>End</td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------</td>
<td>-------------------</td>
<td>-----------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td><em>A. ferrooxidans</em></td>
<td>GCA_000732185.1</td>
<td>contig</td>
<td>4.2</td>
<td>57.6</td>
<td>4289</td>
<td>4006</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CP001219.1</td>
<td>compete</td>
<td>3.0</td>
<td>58.8</td>
<td>3087</td>
<td>2909</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CP001132.1</td>
<td>complete</td>
<td>2.9</td>
<td>58.9</td>
<td>2926</td>
<td>2779</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GCA_001559335.1</td>
<td>contig</td>
<td>3.1</td>
<td>58.6</td>
<td>3179</td>
<td>2955</td>
<td></td>
</tr>
<tr>
<td><em>Acidithiobacillus</em> sp.</td>
<td>GCA_002847505.1</td>
<td>contig</td>
<td>2.9</td>
<td>54.3</td>
<td>2987</td>
<td>2799</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GCA_002733395.1</td>
<td>contig</td>
<td>3.6</td>
<td>52.6</td>
<td>3525</td>
<td>3260</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GCA_00179815.2</td>
<td>contig</td>
<td>3.2</td>
<td>58.6</td>
<td>4079</td>
<td>4007</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GCA_002341825.1</td>
<td>contig</td>
<td>1.7</td>
<td>58.5</td>
<td>1617</td>
<td>1596</td>
<td></td>
</tr>
<tr>
<td><em>A. albertensis</em></td>
<td>GCA_001931655.1</td>
<td>contig</td>
<td>3.5</td>
<td>52.5</td>
<td>3622</td>
<td>3354</td>
<td></td>
</tr>
<tr>
<td><em>A. ferrivorans</em></td>
<td>CP002985.1</td>
<td>complete</td>
<td>3.2</td>
<td>56.6</td>
<td>3206</td>
<td>3016</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GCA_900174455.1</td>
<td>scaffold</td>
<td>3.5</td>
<td>56.5</td>
<td>3555</td>
<td>3296</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GCA_000750615.1</td>
<td>contig</td>
<td>3.4</td>
<td>56.4</td>
<td>3557</td>
<td>3236</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GCA_001685225.1</td>
<td>contig</td>
<td>3.0</td>
<td>56.6</td>
<td>3087</td>
<td>2778</td>
<td></td>
</tr>
<tr>
<td>Strain</td>
<td>Accession</td>
<td>Type</td>
<td>Length</td>
<td>Completion</td>
<td>GC%</td>
<td>Start</td>
<td>End</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------</td>
<td>----------</td>
<td>--------</td>
<td>------------</td>
<td>------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td><em>A. ferrivorans</em></td>
<td>PQ33</td>
<td>contig</td>
<td>3.3</td>
<td>56.6</td>
<td>3382</td>
<td>3077</td>
<td></td>
</tr>
<tr>
<td><em>A. ferrivorans</em></td>
<td>21-59-9</td>
<td>scaffold</td>
<td>2.7</td>
<td>59.0</td>
<td>3208</td>
<td>2909</td>
<td></td>
</tr>
<tr>
<td><em>A. caldus</em></td>
<td>ATCC 51756</td>
<td>complete</td>
<td>3.0</td>
<td>61.4</td>
<td>3005</td>
<td>2849</td>
<td></td>
</tr>
<tr>
<td><em>A. caldus</em></td>
<td>SM-1</td>
<td>complete</td>
<td>3.2</td>
<td>60.9</td>
<td>3254</td>
<td>3106</td>
<td></td>
</tr>
<tr>
<td><em>A. caldus</em></td>
<td>MTH-04</td>
<td>complete</td>
<td>3.0</td>
<td>61.5</td>
<td>3102</td>
<td>2713</td>
<td></td>
</tr>
<tr>
<td><em>A. caldus</em></td>
<td>DX</td>
<td>contig</td>
<td>3.1</td>
<td>61.0</td>
<td>3179</td>
<td>2861</td>
<td></td>
</tr>
<tr>
<td><em>A. caldus</em></td>
<td>ZBY</td>
<td>contig</td>
<td>3.2</td>
<td>61.0</td>
<td>3210</td>
<td>2902</td>
<td></td>
</tr>
<tr>
<td><em>A. caldus</em></td>
<td>ZJ</td>
<td>contig</td>
<td>3.1</td>
<td>61.0</td>
<td>3223</td>
<td>2884</td>
<td></td>
</tr>
<tr>
<td><em>A. caldus</em></td>
<td>S1</td>
<td>contig</td>
<td>2.8</td>
<td>60.9</td>
<td>2906</td>
<td>2396</td>
<td></td>
</tr>
<tr>
<td><em>A. ferridurans</em></td>
<td>JCM 18981</td>
<td>compete</td>
<td>2.9</td>
<td>58.4</td>
<td>3043</td>
<td>2802</td>
<td></td>
</tr>
</tbody>
</table>
FIG 1. Pan-genome analysis of strains in the genus *Acidithiobacillus*. 

(A) Petal diagram of Pan-genome. Each strain is represented by an oval that is colored: *A. ferrivorans* (navy blue), *A. caldus* (indigo), *A. albertensis* (yellow), *A. ferridurans* (orange), *A. ferrooxidans* (green), *A. thiooxidans* (pink), *A. ferrivorans* (navy blue), unclassified (purple). The center is the number of orthologous coding sequences shared by all strains (i.e., the core genome). Numbers in non-overlapping portions of each oval show the number of CDSs unique to each strain. The total number of protein coding genes within each genome is listed below the strain name. (B) Mathematical modeling of pan genome and core genome of *Acidithiobacillus*. (C) Portions of genes involved in heavy metal resistance in unique genes, accessory genome, core genome and pan genome according BacMet database annotation.
**FIG 2.** Chronogram of the 37 *Acidithiobacillus* strains. The value near each internal branch is the estimated emerge time for that branch. Nodes with fossil record corrections are indicated with a red asterisk.
FIG 3. Neighbor-joining (NJ) phylogenetic tree of the MerA protein sequences derived from *Acidithiobacillus* spp. strains and other representative species. Bootstrap values are indicated at each node based on a total of 1,000 bootstrap replicates. Branches represent Group I *mer* clusters were marked in green; Branches represent Group II *mer* clusters were marked in blue;
Branches represent Group III mer clusters. They were marked in red and pink.

**FIG 4.** Neighbor joining phylogenetic tree of concatenated ArsBRC protein sequences derived from *Acidithiobacillus* spp. strains and other representative species. Bootstrap values are indicated at each node based on a total of 1,000 bootstrap replicates. Branches represent ars clusters of *Acidithiobacillus* spp. were marked in blue.
FIG 5. Neighbor joining phylogenetic tree of concatenated ArsAD protein sequences derived from *Acidithiobacillus* spp. strains and other representative species. Bootstrap values are indicated at each node based on a total of 1,000 bootstrap replicates. Branches represent ars clusters of *Acidithiobacillus* spp. were marked in red.
FIG 6. Neighbor joining phylogenetic tree of concatenated CzcA, CzcB and CzcC protein sequences derived from Acidithiobacillus spp. strains and other representative species. Bootstrap values are indicated at each node based on a total of 1,000 bootstrap replicates. Branches represent czcABC clusters were marked in green; czcBAC clusters formed in another group were marked in blue.
**FIG 7.** Distribution and range of selection pressures on different metal resistance genes of *Acidithiobacillus* spp.

**FIG 8.** Range of CAI values of different metal resistance genes of *Acidithiobacillus* spp. with *Tu* gene as reference.