Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Complete genome sequencing and comparative genomic analyses of *Bacillus* sp. S3, a novel hyper Sb(III)-oxidizing bacterium

Jiaokun Li
Central South University
Tianyuan Gu
Central South University
Weimin Zeng (zengweimin1024@126.com)
Central South University
Runlan Yu
Central South University
Yuandong Liu
Central South University
Xueling Wu
Central South University
Li Shen
Central South University
Guanzhou Qiu
Central South University
Liang Zhi Li
Central South University

Research article

Keywords: Bacillus sp. S3, Sb(III)-resistance, genome sequencing, comparative genome, heavy metal(loid)

Posted Date: December 2nd, 2019

DOI: https://doi.org/10.21203/rs.2.17919/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Version of Record: A version of this preprint was published at BMC Microbiology on May 1st, 2020. See the published version at https://doi.org/10.1186/s12866-020-01737-3.

Abstract

Background: Antimonite [Sb(III)]-oxidizing bacterium have great potential in the environmental bioremediation of Sb-polluted sites. Bacillus sp. S3 isolated from antimony-contaminated soil showed high Sb(III) resistance and Sb(III) oxidation efficiency. However, very little genomic information and evolutionary relationships that bacterial oxidation of Sb(III) is available.

Results: Here, we identified a 5579638 bp chromosome with 40.30% GC content and a 241339 bp plasmid with 36.74% GC content in the complete genome of Bacillus sp. S3. Genomic annotation showed that Bacillus sp. S3 contains key aioB gene potentially encoding As(III)/Sb(III) oxidase, is not shared with other Bacillus strains. Further, a series of genes associated with Sb(III) and other heavy metal(loid) were also ascertained, reflecting adaptive advantage for growth in the harsh eco-environment. It is noteworthy that Bacillus sp. S3 is a novel species within the Bacillus genus as indicated by phylogenetic relationship and the average nucleotide identities (ANI) analysis. The presence of genomic plasticity demonstrated a high number of mobile genetic elements (MGEs) that were mainly distributed on chromosomes within the Bacillus genus. The core genome contained 554 core genes and many unique genes were dissected in analyzed genomes, indicating a conserved core but adaptive pan repertoire. Whole genomic alignment indicates that frequently genomic reshuffling and rearrangements, genetic gain and loss, and other recombination events occurred during the evolutionary history of Bacillus genus. In addition, the origin and evolution analysis of Sb(III)-resistance genes revealed that evolutionary relationships and events of horizontal gene transfer (HGT) among the Bacillus genus. The assessment of functionality of heavy metal(loid) resistance genes emphasized its indispensable roles in the harsh eco-environment of Bacillus genus. The real-time Quantitative PCR(RT-qPCR) results of Sb(III)-related genes indicated that the Sb(III) resistance was constantly increased under the Sb(III) stress.

Conclusions: The insights provided in this study shed light on the molecular details of Bacillus sp. S3 coping with Sb(III), which extended our understanding on the evolutionary relationship between Bacillus sp. S3 and other closely related species and will enrich the Sb(III) resistance genetic data sources.

Background

The hazardous heavy metal(loid)s such as Antimony (Sb), Arsenic (As), Cadmium (Cd), Chromium (Cr) and Lead (Pb) has exerted a serious threat to the natural environments in many parts of the world [1–3]. Natural biogeochemical cycle and anthropogenic activities such as mining activities, rapid urbanization, and industrialization have all contributed to elevated levels of heavy metal(loid) in soils in recent decades [4–5]. Heavy metal(loid) pollution has become an environmental issue of great public concern, which leads to heavy metal(loid) accumulation in the human food chain. In recent years, conventional remediation technologies have been developed to remove heavy metal(loid) from contaminated surroundings, such as ion exchange, membrane separation, congulation/flocculation, electrochemical methods, extraction and adsorption [6–7]. Most of the physiochemical remediation methods cannot be applied to large-scale fields because of the high cost, generation of secondary pollution and their unsustainable nature [8]. Alternatively, the approaches of bioremediation is an emerging technology, which utilizes inherent biological mechanisms to eradicate pollutants from heavy metal contaminated sites due to its low-cost and environmentally friendly advantages [9]. Microorganisms are able to alleviate the toxicity of heavy metal(loid) using various resistance strategies, such as converting, chelation, adsorption and accumulation [10].

Antimony (Sb), one of the later known toxic metalloid, is also used as a flame retardants, Pb-Sb alloy, and a catalyst polyethylene glycol terephthalate[11]. Sb and its compounds are listed as priority pollutants by the United states Environmental Protection Agency (USEPA, 1979) and the European Union (CEC, 1976) [7, 12]. The maximum acceptable concentration of Sb in drinking water has been set by the World Health Organization (WHO) at 6 µg/liter [13]. The main Sb species include antimonite [(Sb III)] and antimonate [Sb (V)] in natural system, which can interconvert via biogeochemical

processes. Sb(V) is more stable in aerobic environments than Sb (III), and Sb (III) is more toxic than Sb(V) due to Sb(III) is high affinity with thiol-containing proteins [14]. Consequently, microbial Sb(III) oxidation, which transforms the toxic Sb(III) to the less toxic Sb(V), has a significant value of environmental Sb bioremediation [15].

Our previous study have confirmed that *Bacillus* sp. S3 could aerobically oxidize 100 μM Sb(III) in 2 days (50 $\mu M \cdot d^{-1}$) with high Sb(III) resistance (5.5mM) [16]. Subsequently, the bacterium has the ability to cope with multiple heavy metal(loid)s via various adaptive strategies [17]. Many studies have reported on new isolated bacterial strains (more than 60 aerobic Sb(III) oxidizers) including *Shinella* sp. strain NLS1, *Ensifer* sp. strain NLS4, *Acinetobacter* sp. JL7, *Comamonas* sp. JL25, *Comamonas* sp. JL40, *Comamonas* sp. S44, *Stenotrophomonas* sp. JL9, and *Bosea*sp. AS–1, which isolated from antimony-contaminated soil and mine on a global scale in recent years [7, 11, 14, 18]. Although increasing numbers of studies have focused on microbial Sb oxidation, microbially medited Sb transformation have not been well characterized thus far, especially in the *Bacillus* genus.

It is generally noted that *Bacillus* is a large genus of the Gram-positive, heterotrophic, endospore-forming bacteria and belongs to Bacilli, Phylum [19]. Members of genus *Bacillus* provides a model system for the study of metal ions and exhibit broad resistance to heavy metal [20]. With the extending of the third generation sequencing platform PacBio RSII, a large number of the *Bacillus* strains (more than 4551) have been genomic sequenced. In contrast, there is little literatures on Sb (III) oxidation and resistance mechanisms in genus *Bacillus,* for example, arsenite oxidase AioBA responsible for As(III) oxidation in *A. tumefaciens* 5A was reported to also function as a Sb(III) oxidase [13], the arsenite oxidase AioAB is composed of a large (AioA) and a small (AioB) subunit [21], a novel Sb(III) oxidase AnoA was discovered to catalyze Sb(III) oxidation in *A. tumefaciens* GW4 with $NADP^+$ as the co-factor [22], the cellular H_2O_2 catalyzed bacterial Sb(III) oxidation as an abiotic oxidant [23]. Although these studies provide great advance, there has never been a comprehensive research focused on the effects on of Sb in terms of whole genome and comparative genomics. To further understand the molecular details of *Bacillus* sp. S3 in response to Sb(III), determination of the genomic information of Sb(III)-resistance strains is crucial.Here, we applied genomes sequence and the comparative genomic analysis to study the Sb resistance mechanism and evolutionary relationship of *Bacillus* sp. S3.

Methods

DNA Sequencing, assembly and annotation

The genomic DNA of *Bacillus* sp. S3 was extracted from 5mL exponential phase cell using E.Z.B.A Bacterial DNA Kit (Omega) according to the manufacturer's instructions. The genomic information was obtained by whole-genome shotgun sequencing using the Pacific Biosciences (PacBio) RSII platform (Pacific Biosciences, Menlo Park, CA, USA) at the Institute of Microbiology, Chinese Academy of Sciences (Beijing, China). DNA was sheared with a Covaris instrument (Covaris Inc., Woburn, MA, USA) to 500 bases, and fragmented DNAs were checked by an Agilent Bioanalyzer DNA 7500 kit. Genomic DNA prepared from *Bacillus* sp. S3 was used to produce whole shotgun libraries, a total of 1,596 Gb raw data comprising 88,313 reads encompassing 874,728,983 bases was generated using reference based approach (Table S1). The de novo assemblies were performed using the SPAdes (version 2.6), final genome assembly results are listed in Table S2. Protein-coding regions in the assembled sequences were predicted using Prodigal, tRNA, rRNA, and ncRNA were separately identified by tRNA-scan, barrnap and RNAmmer [24]. Genome Annotation was carried out using the Rapid Annotation using Subsystem Technology (RAST) server [25]. The functional category the genes were determined against the retrieved, National Center for Biotechnology Information (NCBI), non-redundant (NR), Cluster of Orthologous Groups(COG), Kyoto Encyclopedia of Genes and Genomes(KEGG) database with E-value cut-off set to $1e^{-5}$ and subsequent filtering for the best hit [26-27].

Genes for heavy metal(loid)s resistance

The gene related to heavy metal(loid)s in the *Bacillus* sp. S3 genome and other comparative *Bacillus* genomes was identified using Prodigal and RAST with just 25% homology.

Phylogenetic analysis and average nucleotide identity (ANI)

Phylogenetic analysis was reconstructed based on the 16S rRNA gene sequences, the 16S rRNA sequence of type strains of *Bacillus* spp. were fetched from the NCBI GenBank database. Additionally, we constructed the phylogenetic tree based on the the alignment of nucleotide sequences for the 554 single-copy genes shared by all strains in the current research. To further explore the genomic similarities among strains, the phylogenetic tree based on whole-genomes composition vector with a K value of 6 was visualized using the web server CVTree3 [28], which one *Encephalitozoon cuniculi* GB M1 strain was designed as an outgroup. Finally, the most closely related strains were retrieved from NCBI and the consensus tree was constructed using the neighbor-joining method with 1000 bootstrap consensus values by using MEGA 7.0.26 [29]. The ANI values and tetramer usage patterns were calculated using the web server JSpecies1.2.1 [30] based on a BLAST algorithm and tetranucleotide frequency correlation coefficient (Tetra) with default parameters [31].

Comparative genomics

Bacterial Pan Genome Analysis tool (BPGA) was used to extrapolate pipeline pan-genome models with default parameters, and all of orthologous groups among testing genus *Bacillus* were identified [32]. The core genomes is the common set of shared genes among all testing strains, the pan-genomes is the entire set of genes within test genomes, the accessory genes is the set of genes shared with more than two but not all testing strains and unique genes is the set of genes in each strain not shared with other strains [27]. The details of the strains used are listed in Table1. Furthermore, synteny maps were generated to unravel the degree of rearrangements (insertions, deletions, duplications) by identifying conserved locally collinear blocks (LCBs) among genomes, followed by multiple alignment of conserved genomic sequence using Mauve version 20150226 [33].

Prediction of mobile genetic elements (MGEs)

Genomic islands (GIs) were detected using the web server IslandViewer4[8] using three prediction methods including IslandPath-DIMOB, SIGI-HMM, and IslandPick with default parameters [34, 35]. Insertion sequences (ISs) were predicted and classified using the ISFinder platform against the ISfinder database with default criteria [36]. CRISPR arrays were detected using the CRISPR Finder online server to perform BLAST searches against dbCRISPR (CRISPR database) [37]. PHAST [38] was used to scan prophages by BLASTing against the NCBI and the prophage databases.

Selective pressure analysis and expressivity prediction

The CAI (codon adaptation index) values of selected genes were analyzed using Codon W1.4.2 (http://codonw.sourceforge.net//) and CAI Calculator 2 (http://www.evolvingcode.net/codon/CalculateCAIs.php). The mode and strength of natural selection in protein sequences was estimated by evaluated the ratio of nonsynonymous (*dN*) to synonymous (*dS*) nucleotide substitutions rates by using the online software Datamonkey, the HyPhy package with Single-Likelihood Ancestor Counting method was used to detect the selection sites [39].

Nucleotide sequence accession number

The complete genome sequence of *Bacillus* sp. S3 were posted in the NCBI GenBank database with accession number of CP039727.1.

Results And Discussion

Characterize *Bacillus* sp. S3 with different heavy metal(loid)s

Bacillus sp. S3 showed high tolerance to multiple heavy metal(loid)s in the previous study [17]. As unambiguously shown in Fig. 1, the SEM images of *Bacillus* sp. S3 intuitively showed that the cell walls were enveloped by filaments, possibly as a result of the presence of extracellular polymeric substances. Intriguingly, after cultivation of *Bacillus* sp. S3 for exponential phase under the Sb(III) stress, the cell surface became smoother than control. As shown in Fig. S1, no physical Sb(III) adsorption was detected on *Bacillus* sp. S3 cell surfaces for the energy dispersive spectrometer (EDS) spectra measurements, the result was the same as other studies [9]. When the initial concentration was 1mM Pb(II), the peak value of EDS was significant from that of the control group and other treatment groups. The SEM results showed that the a smaller cell size, lesser wrinkled cell wall and the occurrence of intracellular dissolution in present of heavy metal(loid), implying that the elevated level of heavy metal(loid) ions may suppress the secretion of extracellular polymers substance and normal metabolism. The smallness of bacterial cells provided a large contact interface, which would facilitate the interaction between heavy metal(loid) and biosorption process of *Bacillus* sp. S3. The EDS spectra results revealed that *Bacillus* sp. S3 might further absorb Pb(II) to a certain extent compared with other heavy metal(loid)s, such as extracellular adsorption and surface complexation [40].

General genome feature of *Bacillus* sp. S3

The draft genome of *Bacillus* sp. S3 was preliminary analyzed, consisting of 269 contigs (Fig. 2A). Furthermore, the complete genome of *Bacillus* sp. S3 contains a single circular chromosome of 5579638 bp with an average GC content of 40.30% and one plasmid (denoted as pBS-032) of 241339 bp with 36.74% GC content (Fig. 2B, C). A total of 88,313 reads were acquired by Pacific Biosciences (PacBio) RSII platform, the 5344 protein-coding sequences cover 83.13% of the genome and the average length of the coding sequences is 904 bp, as well as 36 rRNAs, 104 tRNAs, and 14 miscRNAs were detected (Table 1). The genomic features of *Bacillus* sp. S3 and other 44 closely related strains are summarized in Table 2, the average chromosome length of the 45 *Bacillus* genomes was 4.99 Mb with ranging of 2.2-7.08 Mb and the average GC content was 40.31% with ranging from 35.4% to 47.8%, indicating substantial species-to-species or strain-to-strain variation and similar genomic characteristics. Among these investigated strains, *Bacillus* sp. OxB-1 showed the highest GC content (47.8%), and *B. thuringiensis* and *B. cereus* SJ1 showed lower GC contents (35.4%).

Notably, large numbers of heavy metal(loid) resistance genes were located on chromosome in *Bacillus* sp. S3 rather than plasmid pBS-032 (Table 2), and these genes could be used for further analysis of genetic diversity and evolutionary. But unlike chromosome, the pBS-032 was found to contain large numbers of hypothetical proteins. The *aioAB* genes encoding As(III) oxidase is responsible for Sb(III)/As(III) oxidation and relieved that Sb(III) was toxic to strain [13]. As resistance genes (*aioB*, *arsB* and *arsC*) were detected in all comparative *Bacillus* genus. The *aioB* gene was involved in the *Bacillus* sp. S3, however, was not found other *Bacillus* genomes. The result suggesting that cytoplasmic As(V) reduction and As(III) extrusion are the main As/Sb resistance strategies used in *Bacillus* genus. Stand-alone *aioB* genes encoding As(III) oxidase was found located chromosome conferring the *Bacillus* sp. S3 with As(III)/Sb(III) oxidizer. The *ars* operon were located on chromosome in *Bacillus* sp. S3, three *arsB* genes encoding Sb(III)/As(III) efflux pump membrane proteins and *arsC* gene encoding As(V) reductase, which conferring As/Sb resistance and pumping out intracellular Sb(III). The phosphate (Pi) related genes were also identified, such as *phoB*, *pstS*, and *phoR*, are preferentially co-regulated with As(III) oxidation and can be induced by Sb(III) in previous report [41].

As shown in Fig. S2, 237 Carbohydrate-active Enzymes (CAZymes) were of *Bacillus* sp. S3 fall into 6 classes. The heavy metal(loid) ions may bind non-specifically to regions of theses enzymes to induce inhibition [42].This interactions would be the detoxification of heavy metal(loid) needs a great deal of energy (e.g. for pumping out intracellular metal ions by ATPases, or for the strengthened expression of metal(loid) resistance proteins, etc.) through increased conversion of carbohydrates for cellular demand. It implied that these enzymes might play an important role in coping with the metal(loid) ions.

Based on BLASTP searches (e value<1e-5), there were 4371 and 3610 predicted genes involved in COG database (Fig. S3, Table S3) and GO database (Fig. S4, Table S4), respectively. A high proportion of genes in COG database were assigned to the R (general function prediction only), E (Amino acid transport and metabolism), G (Carbohydrate transport and metabolism), C (Energy production and conversion), K (Transcription), and L (Replication, recombination, and repair) categories. Compared to other bacteria, enrichment profiles of *Bacillus* sp. S3 genes assigned to COG functional categories show an overabundance of genes involved in cell motility, energy production and conversion, DNA recombination and repair, defense mechanisms, and secondary metabolite biosynthesis and transport. These resistance genes involved in defense and repair mechanisms for dealing with heavy metals propelled by harsh eco-environments, such as nitrate and heavy metals [10]. These resistance genes include ABC antiporters and $Cd^{2+}/Zn^{2+}/Co^{2+}$ efflux components (CzcABC, CzcD) is probably important for *Bacillus* sp. S3 in the adaption of specific niche.

3691 predicted genes were involved in 180 KEGG pathways (Fig. S5, Table S5) which were divided into “Metabolism” (637), “Biosynthesis of secondary metabolites” (285), “Microbial metabolism in diverse environments” (215), “Carbon metabolism” (135), “Biosynthesis of amino acids” (134) and “ABC transporters” (122). Previous study has shown that the DNA damage and oxidative stress were induced to reduce cell damage under copper exposure in *Mytilus coruscus*, four KEGG pathways (DNA replication, apoptosis, aminoacyl-tRNA biosynthesis and mismatch repair) shown differentially expressed genes between control and copper treated group by KEGG pathway enrichment analysis [42]. The arsenite oxidase (AioB) is assigined to metabolic pathway (ko01100), the phosphate-binding protein (PstS), sn-glycerol-3-phosphate-binding periplasmic protein (UgpB) and phosphate import ATP-binding protein (PstB) are assigned to ABC transporters pathway (ko02010), the copper-ion-binding protein CopZ is assigned to mineral absorption pathways (ko04978), these pathways may play an important role in response to the metal(loid) toxicity.

Phylogenetic analyses

To evaluate the phylogenetic relationship, we downloaded 44 *Bacillus* genome sequences (including 16 complete genomes) and their annotations from the NCBI GenBank database and RAST sever as shown in Table 2. The phylogenetic tree revealed that the representative *Bacillus* sp. S3 based on 16S rRNA gene sequences belonged to *Bacillus* genus and the phylogenetic relationship with other genus *Bacillus* (L75, G8 and G18) was grouped together (Fig. 3A). 16S rRNA gene, while has conventionally been used in assessing bacterial taxonomy and phylogeny, proved undefined and controversial in the past [43]. At the same time, the phylogenetic tree between 554 single-copy core genes and whole-genome-based CV are congruent with that of 16S rRNA gene sequences ((Fig. 3B, C), indicating that *Bacillus* sp. S3 may belong to the species of *Bacillus bataviensis*. The difference of phylogenetic trees between whole genome and core genes suggested that the flexible genes could be crucial in altering the genome content and shaping the topology of the trees. These results showed that *Bacillus* sp. S3 in the same species reported was grouped together, however, the topologies of the three phylogenetic trees presented some differences.

ANI, a method can be applied to evaluate the genomic distance and delineate species in evolutionary progress and overcome the difficulty of conventional deviations caused by evolutionary mutation rate and HGT events [31]. As show in Table 3, the closest ANI values 81.51% between *Bacillus* sp. S3 and other reference strains considerably lower than threshold value of 95–96% of the boundary for species circumscription [44], notably represented S3 is a novel species [45]. The ANI results justify the findings from the phylogenetic analysis.

Core and pan genomes analysis

To clarify the genomic features specific to each *Bacillus* strain, all genes from tested *Bacillus* strain were described by MP method in the pan-genome analysis pipeline with a 50% cut-off for protein sequence identity. The pan-genome is the complete gene inventory of a species. There were 39,933 orthologs constituting the pan genome for the *Bacillus* genus (Fig. 4A). Out of these 39,933 orthologs, 554 them (1.38% of total pan genome) were identified as core conserved genome,

and 16234 them were identified as strain-specific genes. Accessory gene number varied from 1189 to 4,431 (mean: 3693), 3,893were represented in the accessory genomes of *Bacillus* sp. S3. Accessory gene is known as indispensable orthologs, the variability of accessory gene indicating the flexibility of genome structure. After comparing strain-specific genes, the variability in the total number of strain-specific genes ranges from 0 to 1560 genes (mean 360). *Bacillus* sp. OxB-1 had the highest amount of these (n = 1560), reflecting the greatest difference with other tested genomes. As shown in Fig. 4B, the more new orthologs were intuitively observed after addition of more genomes to the group, implying a stabilized core structure and an open pan-genome of *Bacillus* strains. These core genes shared with all *Bacillus* genomes also could be classified into different COG categories (Fig. 4C), which is agreement with previous reports that larger prokaryotic genomes tend to pile up genes directly or indirectly involved in different metabolism [46]. The KEGG annotation of 385 specific genes of *Bacillus* sp. S3 showed that 10 genes involved in the Environmental Information Processing including one cobalt transport system protein-encoding gene *corA* and one As(III) efflux pump membrane proteins-encoding gene *arsB*. Further, small number of strain-specific genes (<40%) were assigned to the COG categories for the *Bacillus* sp. S3, these specific genes was mainly found to be enriched in phosphotransferase system and ABC-type metal ion transport system. The result revealed specific adaptive strategies for *Bacillus* sp. S3 in response to harsh eco-environment.

Mobile gene elements in *Bacillus* genomes

It is well recognized that bacteria genomes have notable genome plasticity by several elements of HGT, known as mobile gene elements(GEIs, IS, Prophages) and CRISPRs. The presence of the majority of GEIs in *Bacillus* sp. S3 and other comparative strains renders clues about the genomic plasticity of these isolates (Table S6). We identified 20/5/15 GEIs in *Bacillus* sp. S3 through three methods, these GEIs mediated genetic rearrangement and HGT are possibly conducive to the integrated pool of transposase. Analysis of transposable elements showed that numerous insertion sequence (IS) elements were distributed over the genomes and plasmid of *Bacillus* strains, harboring IS1, IS2, IS3, IS4, IS5, IS21 and IS256. The numbers of insertion sequence (IS) elements could magnify the size of genome, and result in frequently genomic exchange with other community members [47].

As shown in Table S6, all *Bacillus* genomes could be served as feasible targets of phage infections is explained by these discovered phage-associated genes. A total of 5 intact (100% score) prophage region was predicted in the genome of *Bacillus* sp. S3, their information containing size: 11485 bp, 5841 bp, 8630 bp, 8468 bp, 7959 bp; coding sequence (CDS): 11, 7, 6, 10, 8; GC content: 42.84%, 38.02%, 35.25%, 36.25%, 33.84% (Figure S6). It is interesting note that the number of CRISPRs varied between 0 and 15 per strain and CRISPR loci absolutely scattered the chromosome, 4 confirmed CRISPR with 39, 20, 60, 4 spacer was detected in *Bacillus* sp. S3. Our findings clearly suggested that the *Bacillus* genus are able to trigger various defense mechanisms against invasion of exogenous DNA for maintaining the stability of their genetic architecture in procedure of evolution.

Signatures of HGT events generally could be facilitated by MGEs, such as integration sites, anomalous GC contents, or varied codon usage, suggesting that the potential contributor of the observed genomic rearrangements between the genomes of genus *Bacillus* []47. MGEs is regarded as a key driving forces of genome evolution and play a pivotal role in HGT events [11], indicating that high genomic plasticity in *Bacillus* genus was extended to potential strategies to cope with high metal(loid) ion concentrations of their natural habitats. Genomic island (GEIs), which have been committed to provide antibiotic resistance to the host bacteria, are now generally divided into 4 categories based on their function, including resistance island (RIs), virulence genes, metabolic islands, and symbiotic island(SIs), these islands promoting symbiotic integration of the host with other microorganisms [48, 49]. CRISPR-Cas systems are a type of adaptive immunity detected in bacteria and archaea, which protect them against invading genetic elements [50-51]. *Bacillus* sp. S3 may have acquired specific functional gene clusters related-metal(loid) from other genera via HGT and genomic reshuffling, which was integrated into its genome, thereby putatively enhancing the adaptability response to high concentration of heavy metal(loid) ion.

Comparative genomic analysis of *Bacillus* genus

Mauve is used to construct multiple genome alignments in large-scale evolutionary events such as genome rearrangement, inversion, and other recombination. To prove the extent of genomic shuffling, the whole genome sequence of *Bacillus* sp. S3 was also compared individually with selected strains using mauve with default settings. As shown in Fig. 5, synteny comparisons of *Bacillus* sp. S3 with other 10 complete genomes were generated. Synteny maps showed the presence of large-scale blocks of inversions and several crisscrossing locally collinear blocks (LCBs) were inspected in relation to strains *Bacillus* sp. S3. 615 LCBs with a minimum weight of 45 were exhibited between S3 and *B. bataviensis* LMG21833, other comparative information were distinctly observed. Conserved structural synteny and lack of inversions and rearrangements among members of *Bacillus* group suggesting *Bacillus* sp. S3 underwent intricate and significant genome rearrangement relative to tested *Bacillus* genomes occurring as a function of genome distance. Mauve analyses showed the presence of large-scale evolutionary events, suggesting the wide and sophisticated rearrangement profiles occurring at the genus level [52].

The origin and evolution of heavy Sb(III)-related genes

To examined the origin and evolution of Sb(III)-related genes in *Bacillus* sp. S3, allow *aioB*, *arsB* and *arsC*, these genes that contribute to Sb(III) resistance and operon transcriptional regulator in the closely related *Bacillus* genus were selected. Genome evolution can be driven by the acquisition and loss of genes, which is often conducted by HGT, genomic reshuffling and natural selection [53]. Comprehensive analysis of HGT events was performed, since it is difficult to identify via either phylogenetic analysis or deviant G+C content. As shown in Fig. S7, the deviant G+C content between genes and the genome can be used as a detect method of HGT [44].

The overall origin and evolution of As/Sb resistance genes in *Bacillus* genus were analyzed by phylogenetic trees. We speculated that the *aioB* gene in *Bacillus* sp. S3 from the evolution of 2Fe-2S ferredoxin (Fig. S8). To further elucidate the evolution of *arsB* cluster/*arsC*, NJ inferred phylogenetic trees constructed based on ArsB and ArsC proteins showed that *arsB* clusters of *Bacillus* sp. S3 formed separate 3 groups and a monophyletic clade, suggesting that the *arsB* cluster/*arsC* of *Bacillus* sp. S3 may originate from a common ancestor with *B. bataviensis* LMG21833. *Bacillus* sp. S3 likely obtained *arsB* cluster from *B. bataviensis* LMG21833, *B. vireti* and *B. drentensis*, since prosperus branch of *arsB* cluster of *Bacillus* sp. S3 near the base of the clade (Fig. S9, Fig. S10). As shown in Fig. S10, the *arsC* gene of *Bacillus drentensis* may be acquired via HGT events from *Rhodococcus qingshengii*. The results showed that *Bacillus* sp. S3 acquired As/Sb resistance genes via HGT. These data suggest that the As/Sb resistance genes were horizontally transferred between the *Bacillus* sp. S3 and other species. The results was found to consistent with recent study showing that bacterial As resistance and detoxification acquired via HGT, driven by adaption to habitats containing As [54].

3.9 Assessment of functionality of heavy metal(loid)s-related genes

The CAI has been used for a numerical calculator of gene expression level described by the previous delineation [55], as a result of highly expressed genes in bacteria are prone to magnify stronger codon bias. The CAI value varies from 0 to 1.0, with higher CAI value indicating a higher expression level [56]. In order to assess indirectly the functionality of metal(loid) resistance genes, the appraisal of the strength of natural selection was performed, along with the CAIs of these genes. Putative highly expressed (PHX) genes associated to metal(loid) resistance in the *Bacillus* genus were inferred, using *gerd* gene encoding spore germination protein as references, which encode elongation factors that can be deduced to be highly expressed across most organisms. As shown in Fig. 6A, the CAI values of these above-mentioned metal(loid)s resistance genes were calculated. The cutoff values were indicated with average CAI values of *Gerd* genes in each species. Our result showed that only about 8% of the metal(loid) resistance genes were predicted to be PHX genes that greater than 0.75, while lots of genes have CAI values range from 0.4 to 0.8. The PHX genes leads to stronger codon bias than those of low expressed level genes, resulting from codon translational selection.

A gene in the node or tip of a given tree was considered under diversifying selection ($dN/dS > 1$), evolve neutrally ($dN/dS \approx 1$), or purifying selection ($dN/dS > 1$) using the likelihood ratio test after adjusting for multiple testing (P value < 0.1). As shown in Fig. 6B, 96.3% of 10 selected genes associated metal(loid) had a ratio of nonsynonymous substitutions (*dN/dS*<1), implying that these genes is indispensable factor for the above-mentioned *Bacillus* species under the pressure of purifying selection. Only an *arsC* gene from *B. bataviensis* LMG21833 (*dN/dS*=1.63), double *chrA* genes from *B. niacin* DSM 2923 and *B. liceniformis* ATCC 14580 (*dN/dS*=2.08, *dN/dS*=1.28), a *copZ* gene from *B. firmus* NCTC 10335 (*dN/dS*=1.28), a *corA* gene from *Bacillus* sp. LF1 (*dN/dS*=2.80), double *znuB* genes from *B. soli* 15604 and *B. bataviensis* LMG21833 (*dN/dS*=2.02, *dN/dS*=1.1) showed that *dN/dS*>1, suggesting that they might be under diversifying selection. Furthermore, the lowest *dN/dS* ratio was remarked for *copA* gene (average *dN/dS*=0.08) and the *zur* gene (average *dN/dS*=0.09), demonstrating strong purifying selection. Additionally, these genes including *arsB* from *Bacillus mesonae* H20-5, *chrR* from *Bacillus mesonae* H20-5 and *Bacillus* sp. S3, *copZ* from *B. thuringiensis* 97-27, *Bacillus glycinifermentans* BGLY and *B. oceanisediminis strain* Bhandara28, *corA* from *B. firmus strain* 14_TX, showed *dN/dS* ratio=1.0, indicating that selection force have little effect on them.

Transcription Analysis in *Bacillus* sp. S3 with or without Sb(III)

To gain the insights into the role of *Bacillus* sp. S3-specific genes annotated as “arsenate oxidase(*aio*)” and “protein of resistance function (*ars*)” in presence of Sb(III), the transcript level of *aioB*, *arsB* cluster (*arsB_1*, *arsB_2* and *arsB_3*), *arsC* and *psts_1* was investigated by qRT-PCR. Primers used are listed in Table S7, where 16S rRNA gene was used as an internal reference. As shown in Fig. 7, the results showed that the expression level of most of genes were up-regulated induced by Sb(III) except for *arsC*. Although the exposure of 0.5h Sb(III) decreased the expression level of *aioB* and *arsB* cluster, the expression level of *aioB* and *arsB cluster* were up-regulated after 1h, 2h and 4h. The transcript level of *aioB* gene was 15.8, 4.4, 2.6-fold change higher with 100μM Sb(III) from 1 to 4h compared to uninduced culture. Nevertheless, when the *Bacillus* sp. S3 was exposed to high Sb(III) treatment (200 and 300 μM), the transcription level of *aioB* gene was up-regulated 2.6, 2.1, 1.3-fold change and 2.5, 2.8, 2.1 from 1 to 4h. Thus, the transcription level of *aioB* have nothing to do with the concentration and stress length of Sb(III). Additionally, the transcription level of *arsB* cluster (2h and 4h) were remarkably enhanced, moreover, the transcription of *arsB* cluster and *psts_1* were improved along with increasing the Sb(III) concentration. The increased transcription level of *aioB* and *arsB* in the presence of Sb(III) suggest that Sb(III) could stimulate the expression levels of As(III)/Sb(III) resistance genes, which may act synergistically to release the toxicity of Sb(III) in *Bacillus* sp. S3.

It has been reported that *aioA* expression was induced by Sb(III) [12]. Our results showed that expression level of *aioB* was up-regulated from 1 to 4h compared to uninduced cell, indicated that the *aioB* gene is induced by Sb(III) and play a putative part in oxidizing Sb(III). Nevertheless, 0.5h Sb(III) exposure suggesting that the *aioB* genes were not up-regulated during the earlier time points, the expression *aioB* of higher Sb(III) concentration (200 and 300 μM) notably lower than 100μM Sb(III). These results well-documented demonstrate the higher Sb(III) could to some extent inhabit *aioB* expression, which is basically consistent with the previous reports that the excessive As(III) treatment could inhibit the *aioB* expression [13]. In addition, the transcription of *psts_1* genes involved in phosphate metabolism and co-regulated the genes *aioBA* was induced by Sb(III), suggesting that *Bacillus* sp. S3 required more DNA repair and amino acids synthesis processes through increased the production of PRPP in response to Sb(III) [22].

Conclusion

In this study, we sequenced a hyper Sb(III) oxidation strain *Bacillus* sp.S3 and performed comparative genomic study of the *Bacillus* group, representing substantial improvements over previously published results. The majority of genes encoding metal(loid) resistance proteins and MGEs were discovered in *Bacillus* sp.S3, which could adapt to a metal(loid)-contaminated environment. Meanwhile, there is an arsenate oxidase AioB in the *Bacillus* sp. S3 with not shared with other *Bacillus* genus, which could play a key role in the process of Sb(III)-oxidizing. Furthermore, *Bacillus* sp. S3 was identified as

a new species by phylogenetic trees and ANI analysis. 6 selected genes related to Sb(III) in the *Bacillus* sp. S3 were all induced Sb(III) using RT-qPCR, indicating these genes involved in alleviating the toxicity of Sb(III). The comparative genome within the 45 *Bacillus* species and origins and evolution of Sb(III)-resistance genes were explored. Hence, our study revealed the molecular basis in response to Sb(III), for *Bacillus* sp.S3, and represented a first glimpse into the understanding of the genomic plasticity and evolutionary relationship among the *Bacillus* genus. As a consequence of the lack of comprehensive analysis with respect to genetic expression and regulation by *Bacillus* sp. S3, the molecular basis of microorganism-Sb(III) needs further clarify in the near future via gene knockouts and protein characterization.

Declarations

Acknowledgements

We gratefully acknowledge the Institute of Microbiology, Chinese Academy of Sciences (Beijing, China) providing the Pacific Biosciences (PacBio) RSII. We would like to thank Chenbing Ai, Xiaoyan Wu, and Fei Liu for their research assistance.

Funding

This work was supported by the National Natural Science Foundation of China (No. 31470230, 51320105006, 51604308), Key Research and Development Projects in Hunan Province (2018WK2012), Natural Science Foundation of Hunan Province of China (No.2018JJ2486), the Youth Talent Foundation of Hunan Province of China(No.2017RS3003), Fundamental Research Funds for the Central University of Central South University (No. 2019zzts687).

Availability of date and materials

All sequences involved in this study are available from the DDBJ/ENA/GenBank under accession number CP039727 and CP039728, respectively.

Authors' contributions

TG analyzed the date and wrote the manuscript. LL and LJ assisted in data analysis and revised the manuscript. LJ, YR, and ZW conceived the project. SL, LW and WX helped revised the manuscript. All authors discussed the manuscript draft and agreed to the final content.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing financial interest.

References

[1] Cheung, K. H., and Gu, J.-D. (2007). Mechanism of hexavalent chromium detoxification by microorganisms and bioremediation application potential: A review. *International Biodeterioration & Biodegradation.* 59(1), 8–15. doi: 10.1016/j.ibiod.2006.05.002.

[2] Guo, H., Luo, S., Chen, L., Xiao, X., Xi, Q., Wei, W., et al. (2010). Bioremediation of heavy metals by growing hyperaccumulaor endophytic bacterium *Bacillus* sp. L14. *Bioresour Technol.* 101(22), 8599–8605. doi: 10.1016/j.biortech.2010.06.085.

[3] Nguyen, T. A., Ngo, H. H., Guo, W. S., Zhang, J., Liang, S., Yue, Q. Y., et al. (2013). Applicability of agricultural waste and by-products for adsorptive removal of heavy metals from wastewater. *Bioresour Technol.* 148, 574–585. doi: 10.1016/j.biortech.2013.08.124.

[4] Ma, Y., Rajkumar, M., Zhang, C., and Freitas, H. (2016). Beneficial role of bacterial endophytes in heavy metal phytoremediation. *J Environ Manage.* 174, 14–25. doi: 10.1016/j.jenvman.2016.02.047.

[5] Essa, A.M. M., Al Abboud, M. A., and Khatib, S. I. (2018). Metal transformation as a strategy for bacterial detoxification of heavy metals. *J Basic Microbiol.* 58(1), 17–29. doi: 10.1002/jobm.201700143.

[6] Manasi, Rajesh, V., Santhana Krishna Kumar, A., and Rajesh, N. (2014). Biosorption of cadmium using a novel bacterium isolated from an electronic industry effluent. *Chemical Engineering Journal.* 235, 176–185. doi: 10.1016/j.cej.2013.09.016.

[7] Nguyen, V. K., Choi, W., Yu, J., and Lee, T. (2017). Microbial oxidation of antimonite and arsenite by bacteria isolated from antimony-contaminated soils. *International Journal of Hydrogen Energy.* 42(45), 27832–27842. doi: 10.1016/j.ijhydene.2017.08.056.

[8] Sheng, P. X., Ting, Y.-P., Chen, J. P. J. I., and Research, E. C. (2007). Biosorption of heavy metal ions (Pb, Cu, and Cd) from aqueous solutions by the marine alga *Sargassum* sp. in single-and multiple-metal systems. 46(8), 2438–2444. doi: 10.1021/ie0615786.

[9] He, M., Li, X., Liu, H., Miller, S. J., Wang, G., and Rensing, C. (2011). Characterization and genomic analysis of a highly chromate resistant and reducing bacterial strain *Lysinibacillus fusiformis* ZC1. *J Hazard Mater* 185(2–3), 682–688. doi: 10.1016/j.jhazmat.2010.09.072

[10] Hemme, C. L., Deng, Y., Gentry, T. J., Fields, M. W., Wu, L., Barua, S., et al. (2010). Metagenomic insights into evolution of a heavy metal-contaminated groundwater microbial community. *ISME J.* 4(5), 660–672. doi: 10.1038/ismej.2009.154.

[11] Lu, X., Zhang, Y., Liu, C., Wu, M., and Wang, H. (2018). Characterization of the antimonite- and arsenite-oxidizing bacterium *Bosea* sp. AS–1 and its potential application in arsenic removal. *J Hazard Mater* 359, 527–534. doi: 10.1016/j.jhazmat.2018.07.112.

[12] Li, J., Yang, B., Shi, M., Yuan, K., Guo, W., Li, M., et al. (2017). Effects upon metabolic pathways and energy production by Sb(III) and As(III)/Sb(III)-oxidase gene *aioA* in *Agrobacterium tumefaciens* GW4. *PLoS One.* 12(2), e0172823. doi: 10.1371/journal.pone.0172823.

[13] Wang, Q., Warelow, T. P., Kang, Y. S., Romano, C., Osborne, T. H., Lehr, C. R., et al. (2015). Arsenite oxidase also functions as an antimonite oxidase. *Appl Environ Microbiol* 81(6), 1959–1965. doi: 10.1128/AEM.02981–14.

[14] Li, J., Wang, Q., Zhang, S., Qin, D., and Wang, G. (2013). Phylogenetic and genome analyses of antimony-oxidizing bacteria isolated from antimony mined soil. *International Biodeterioration & Biodegradation.* 76, 76–80. doi: 10.1016/j.ibiod.2012.06.009.

[15] Terry, L. R., Kulp, T. R., Wiatrowski, H., Miller, L. G., and Oremland, R. S. (2015). Microbiological oxidation of antimony(III) with oxygen or nitrate by bacteria isolated from contaminated mine sediments. *Appl Environ Microbiol.* 81(24), 8478–8488. doi: 10.1128/AEM.01970–15.

[16] Li, J., Yu, H., Wu, X., Shen, L., Liu, Y., Qiu, G., et al. (2018). Novel Hyper Antimony-Oxidizing Bacteria Isolated from Contaminated Mine Soils in China. *Geomicrobiology Journal.* 35(8), 713–720. doi: 10.1080/01490451.2018.1454556.

[17] Zeng, W., Li, F., Wu, C., Yu, R., Wu, X., Shen, L., et al. (2019). Role of extracellular polymeric substance (EPS) in toxicity response of soil bacteria *Bacillus* sp. S3 to multiple heavy metals. *Bioprocess Biosyst Eng.* 2019, 1–15. doi: 10.1007/s00449–019–02213–7.

[18] Xiong, J., Li, D., Li, H., He, M., Miller, S. J., Yu, L., et al. (2011). Genome analysis and characterization of zinc efflux systems of a highly zinc-resistant bacterium, *Comamonas testosteroni* S44. *Res Microbiol.* 162(7), 671–679. doi: 10.1016/j.resmic.2011.06.002.

[19] Çolak, F., Atar, N., Yazıcıoğlu, D., and Olgun, A. (2011). Biosorption of lead from aqueous solutions by *Bacillus* strains possessing heavy-metal resistance. *Chemical Engineering Journal.* 173(2), 422–428. doi: 10.1016/j.cej.2011.07.084.

[20] Moore, C. M., and Helmann, J. D. (2005). Metal ion homeostasis in *Bacillus subtilis. Curr Opin Microbiol.* 8(2), 188–195. doi: 10.1016/j.mib.2005.02.007.

[21] Lehr, C. R., Kashyap, D. R., and McDermott, T. R. (2007). New insights into microbial oxidation of antimony and arsenic. *Appl Environ Microbiol.* 73(7), 2386–2389. doi: 10.1128/AEM.02789–06.

[22] Li, J., Wang, Q., Li, M., Yang, B., Shi, M., Guo, W., et al. (2015). Proteomics and Genetics for Identification of a Bacterial Antimonite Oxidase in *Agrobacterium tumefaciens. Environ Sci Technol.* 49(10), 5980–5989. doi: 10.1021/es506318b.

[23] Li, J., Zhang, Y., Zheng, S., Liu, F., and Wang, G. (2019). Anaerobic Bacterial Immobilization and Removal of Toxic Sb(III) Coupled With Fe(II)/Sb(III) Oxidation and Denitrification. *Front. Microbiol.* 10*:* 360. doi: 10.3389/fmicb.2019.00360.

[24] Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L.J (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* 11(1)*:* 119. doi: 10.1016/j.chemosphere.2014.01.066. doi:10.1186/1471–2105–11–119

[25] Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 42, D206–214. doi: 10.1093/nar/gkt1226.

[26] Sun, Z., Harris, H. M., McCann, A., Guo, C., Argimon, S., Zhang, W., et al. (2015). Expanding the biotechnology potential of *lactobacilli* through comparative genomics of 213 strains and associated genera. *Nat Commun.* 6*:* 8322. doi: 10.1038/ncomms9322.

[27] Tian, X., Zhang, Z., Yang, T., Chen, M., Li, J., Chen, F., et al. (2016). Comparative Genomics Analysis of *Streptomyces* Species Reveals Their Adaptation to the Marine Environment and Their Diversity at the Genomic Level. *Front. Microbiol.* 7*:* 998. doi: 10.3389/fmicb.2016.00998.

[28] Qi, J., Luo, H., and Hao, B. (2004). CVTree: a phylogenetic tree reconstruction tool based on whole genomes. *Nucleic Acids Res.* 32, W45–47. doi: 10.1093/nar/gkh362.

[29] Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 33(7), 1870–1874. doi: 10.1093/molbev/msw054.

[30] Richter, M., Rossello-Mora, R., Oliver Glockner, F., and Peplies, J. (2016). JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics.* 32(6), 929–931. doi: 10.1093/bioinformatics/btv68.

[31] Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* 106(45), 19126–19131. doi: 10.1073/pnas.0906412106.

[32] Chaudhari, N. M., Gupta, V. K., and Dutta, C. (2016). BPGA-an ultra-fast pan-genome analysis pipeline. *Sci Rep.* 6, 24373. doi: 10.1038/srep24373.

[33] Darling, A. E., Mau, B., and Perna, N. T. (2010). ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One.* 5(6), e11147. doi: 10.1371/journal.pone.0011147.

[34] Soares, S. C., Oliveira, L. C., Jaiswal, A. K., and Azevedo, V. (2016). Genomic Islands: an overview of current software and future improvements. *J Integr Bioinform.* 13(1), 301. doi: 10.2390/biecoll-jib–2016–301.

[35] Bertelli, C., Brinkman, F. S. L., and Valencia, A. (2018). Improved genomic island predictions with IslandPath-DIMOB. *Bioinformatics.* 34(13), 2161–2167. doi: 10.1093/bioinformatics/bty095.

[36] Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., and Chandler, M. (2006). ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* 34, D32–36. doi: 10.1093/nar/gkj014.

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

[38]Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., and Wishart, D. S. (2011). PHAST: a fast phage search tool. *Nucleic Acids Res.* 39, W347–352. doi: 10.1093/nar/gkr485.

[39] Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V., and Kosakovsky Pond, S. L. (2018). Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol Biol Evol.* doi: 10.1093/molbev/msx335.

[40] Huang, F., Guo, C.-L., Lu, G.-N., Yi, X.-Y., Zhu, L.-D., and Dang, Z. (2014). Bioaccumulation characterization of cadmium by growing Bacillus cereus RC–1 and its mechanism. *Chemosphere.* 109, 134–142. doi: 10.1016/j.chemosphere.2014.01.066.

[41] Chen, F., Cao, Y., Wei, S., Li, Y., Li, X., Wang, Q., et al. (2015). Regulation of arsenite oxidation by the phosphate two-component system PhoBR in *Halomonas* sp. HAL1. *Front. Microbiol.* 6: 923. doi: 10.3389/fmicb.2015.00923.

[42] Paulo, R. H., Caroline, H., Rosane, M. P., Adelar, B. R.d.C. G. S., Jose, L.d.C. S., Maria, d.L. T. M. P., et al. (2014). Xylanase from Fusarium heterosporum: Properties and influence of thiol compounds on xylanase activity. *African Journal of Biotechnology.* 13(9), 1047–1055. doi: 10.5897/ajb2013.13282.

[43] Hahnke, R. L., Meier-Kolthoff, J. P., Garcia-Lopez, M., Mukherjee, S., Huntemann, M., Ivanova, N. N., et al. (2016). Genome-Based Taxonomic Classification of Bacteroidetes. *Front. Microbiol.* 7: 2003. doi: 10.3389/fmicb.2016.02003.

[44] Xie, J. B., Du, Z., Bai, L., Tian, C., Zhang, Y., Xie, J. Y., et al. (2014). Comparative genomic analysis of N_2-fixing and non-N_2-fixing *Paenibacillus* spp.: organization, evolution and expression of the nitrogen fixation genes. *PLoS Genet.* 10(3), e1004231. doi: 10.1371/journal.pgen.1004231.

[45] Belbahri, L., Chenari Bouket, A., Rekik, I., Alenezi, F. N., Vallat, A., Luptakova, L., et al. (2017). Comparative Genomics of *Bacillus amyloliquefaciens* Strains Reveals a Core Genome with Traits for Habitat Adaptation and a Secondary Metabolites Rich Accessory Genome. *Front. Microbiol.* 8:1438. doi: 10.3389/fmicb.2017.01438.

[46] Smarda, P., Bures, P., Horova, L., Leitch, I. J., Mucina, L., Pacini, E., et al. (2014). Ecological and evolutionary significance of genomic GC content diversity in monocots. *Proc. Natl. Acad. Sci. U.S.A.* 111(39), E4096–4102. doi: 10.1073/pnas.1321152111.

[47] Zhang, X., Liu, X., Li, L., Wei, G., Zhang, D., Liang, Y., et al. (2019). Phylogeny, Divergent Evolution, and Speciation of Sulfur-Oxidizing Acidithiobacillus Populations. *BMC Genomics.* 20(1)*:* 438. doi: 10.1186/s12864–019–5827–6.

[48] Juhas, M., van der Meer, J. R., Gaillard, M., Harding, R. M., Hood, D. W., and Crook, D. W. (2009). Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev* 33(2), 376–393. doi: 10.1111/j.1574–6976.2008.00136.x.

[49] Tumapa, S., Holden, M. T., Vesaratchavest, M., Wuthiekanun, V., Limmathurotsakul, D., Chierakul, W., et al. (2008). *Burkholderia pseudomallei* genome plasticity associated with genomic island variation. *BMC Genomics.* 9: 190. doi: 10.1186/1471–2164–9–190.

[50] Sorek, R., Kunin, V., and Hugenholtz, P. (2008). CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nature Reviews Microbiology.* 6(3), 181–186. doi: 10.1038/nrmicro1793.

[51] Zhang Q, Rho M, Tang H, et al. (2013). CRISPR-Cas systems target a diverse collection of invasive mobile genetic elements in human microbiomes. *Genome biol.* 14(4), R40. doi:10.1186/gb–2013–14–4-r40

[52] Pathak, A., Chauhan, A., Blom, J., Indest, K. J., Jung, C. M., Stothard, P., et al. (2016). Comparative Genomics and Metabolic Analysis Reveals Peculiar Characteristics of *Rhodococcus opacus* Strain M213 Particularly for Naphthalene Degradation. *PLoS One.* 11(8), e0161032. doi: 10.1371/journal.pone.0161032.

[53] Ochman, H., Lawrence, J. G., and Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature.* 405(6784), 299–304. doi: 10.1038/35012500.

[54] Wang, L., Wang, J., and Jing, C. (2017). Comparative Genomic Analysis Reveals Organization, Function and Evolution of ars Genes in *Pantoea* spp. *Front. Microbiol.* 8, 471. doi: 10.3389/fmicb.2017.00471.

[55] Moury, B., and Simon, V. (2011). *dN/dS*-based methods detect positive selection linked to trade-offs between different fitness traits in the coat protein of potato virus Y. *Mol Biol Evol.* 28(9), 2707–2717. doi: 10.1093/molbev/msr105.

[56] Wu, G., Culley, D. E., and Zhang, W. (2005). Predicted highly expressed genes in the genomes of *Streptomyces coelicolor* and *Streptomyces avermitilis* and the implications for their metabolism. *Microbiology.* 151, 2175–2187. doi: 10.1099/mic.0.27833–0.

Tables

Table 1 Genomic features of the chromosome and plasmid of *Bacillus* sp. S3

Features	Chromosome	Plasmid
Genome size (bp)	5579,638	241,339
Total number of genes	5,131	234
Gene Length (bp):	4638,424	203,412
Gene Average Length (bp):	904	869
Gene Length /Genome (%):	83.13%	84.28%
GC Content in gene region (%):	41.17%	37.63%
G+C% numbers	40.30%	36.74%
Protein conding genes	5,344	234
tRNAs	104	0
rRNAs	36	0
Protein-coding genes (CDS)	5,344	235
GIs	12	3
CRISPRs	4	0
Prophages	5	0
Transposase genes of IS elements	115	48

Table 2 Genomes statistical information of the 45 strains used in this study.

No.	Organism	NCBI accession NO.	Level	Size (Mb)	GC%	No. of Genes	No. of Proteins	rRNA	tRNA
1	*Bacillus asahii* OM18	NZ_CP026095	Complete	4.89	37.45	4,824	4,419	54	150
2	*B. bataviensis* LMG 21833	AJLS00000000.1	Contig	5.37	39.6	5,238	5,207	-	23
3	*B. cereus* SJ1	ADFM00000000.1	Contig	5.16	35.40	2,192	2,184	-	8
4	*B. cucumis Strain* V32-6	PGVE00000000.1	Contig	5.71	38.6	5,552	5,309	10	129
5	*Bacillus dielmoensis* FF4(T)	NZ_CCAD000000000	Complete	4.57	40.9	4,433	4,184	20	136
6	*B. drentensis* NBRC 102427	BCUX00000000.1	Contig	5.16	38.7	4,998	4,851	-	28
7	*B. drentensis strain* FJAT-10044	LUUU00000000.1	Scaffold	5.3	38.9	5,163	4,890	51	125
8	*B. firmus* NBRC 15306	BCUY00000000.1	Contig	4.42	41.7	4,515	4,170	-	26
9	*B. firmus* NCTC10335	UFTC00000000.1	Contig	4.8	41.70	4,948	4,436	36	108
10	*B. firmus strain* 14_TX	QNSF00000000.1	Scaffold	5.84	40.8	6,009	5,834	23	102
11	*B. glycinifermentans* BGLY	LT603683.1	Complete	4.61	46.1	4,735	4,381	24	82
12	*B. glycinifermentans* SRCM103574	CP035232.1	Complete	4.81	46.01	4,948	4,572	25	82
13	*B. licheniformis* ATCC 14580	CP000002.3	Complete	4.22	46.20	4,382	4,219	21	72
14	*B. licheniformis strain* YNP1-TSU	NZ_CM007615.1	Chromosome	4.25	45.9	4,388	4,227	4	66
15	*B. mesonae strain* FJAT-13985	LUUQ00000000.1	Scaffold	5.81	40.3	5,553	5,309	40	104
16	*B. mesonae strain* H20-5	CP022572.1	Complete	5.84	40.4	5,600	5,330	41	109
17	*B. methanolicus* MGA3	CP007739.1	Complete	3.34	38.7	3,355	3,092	27	91
18	*B. niacini strain* DSM 2923	JRYQ00000000.1	Scaffold	2.2	38.3	2,167	1,922	3	35
19	*B. novalis* NBRC 102450	NZ_BCVP00000000.1	Contig	5.57	39.9	5,387	5,212	-	31
20	*B. novalis strain* FJAT-14227	LUUR00000000.1	Scaffold	5.67	40.0	5,517	5,247	37	118
21	*B. oceanisediminis strain* Bhandara28	MBRJ00000000.1	Contig	5.88	40.8	5,925	5,705	33	106
22	*B. soli* NBRC 102451	BCVI00000000.1	Contig	5.46	39.5	5,289	5,109	3	18
23	*B. soli strain* DSM 15604	NISV00000000.1	Scaffold	5.58	39.7	5,451	5,148	43	92
24	*B. subtilis subsp. spizizenii str.* W23	CP002183.1	Complete	4.03	43.9	4,116	3,912	24	77
25	*B. subtilis subsp. subtilis str.* 168	AL009126.3	Complete	4.22	43.5	4,536	4,237	30	86
26	*B. thuringiensis serovar konkukian str.* 97-27	AE017355.1	Complete	5.24	35.4	5,263	5,117	41	105
27	*B. thuringiensis* YBT-1518	CP005935.1	Complete	6	35.4	6,371	5,837	45	90
28	*B. vireti* LMG 21834	ALAN00000000.1	Contig	5.28	39.7	5,106	5,084	-	21
29	*B. vireti* Strain DSM 15602	LDNB00000000.1	Scaffold	5.31	39.8	5,118	4,794	13	89
30	*B. velezensis* FZB42	CP000560.1	Complete	3.91	46.50	3,892	3,687	29	88
31	*Bacillus* sp. AFS006103	NTXX00000000.	Scaffold	5.18	38.6	5,018	4,840	5	92
32	*Bacillus* sp. OK048	FNHN00000000.1	Scaffold	5.17	38.0	5,225	5,058	25	115
33	*Bacillus* sp. OV166	FXWM00000000.1	Contig	7.08	38.3	7,182	6,561	60	152
34	*Bacillus* sp. UNC41MFS5	JMLP00000000.1	Scaffold	3.27	38.6	3,107	3,014	6	25
35	*Bacillus* sp. UNC438CL73TsuS30	AXVA00000000.1	Scaffold	3.06	39.0	2,983	2,867	6	41
36	*Bacillus* sp. LF1	CVRB00000000.1	Contig	5.6	38.1	5,524	5,277	17	102
37	*Bacillus* sp. FJAT-18017	CP012602.1	Complete	5.27	42.4	5,018	4,825	30	85
38	*Bacillus* sp. FJAT-29814	LMTJ00000000.1	Scaffold	5.89	41.9	5,791	5,596	11	85
39	*Bacillus* sp. X1	CP008855.1	Complete	3.42	38.10	3,433	3,103	36	122
40	*Bacillus* sp. MUM 116	MLYR00000000.1	Contig	5.72	38.4	5,600	5,273	25	165
41	*Bacillus* sp. OxB-1	AP013294.1	Complete	3.59	47.80	3,604	3,438	22	83
42	*Bacillus* sp. WN066	SMYO00000000.1	Contig	6.21	38.60	6,131	5,757	36	151
43	*Bacillus* sp. 7884-1	NPDD00000000.1	Contig	6.00	38.00	5,848	5,597	-	46
44	*Bacillus* sp. MRMR6	MSLS00000000.1	Contig	5.44	38.80	5,276	4,978	50	108
45	*Bacillus* sp. S3	CP039727	Complete	5.58	40.30	5,131	5,344	104	36

"-": unpublished

Table 3 Genes associated with putative heavy metal(loid)s resistance in *Bacillus* sp. S3

Category	Gene_ID	Gene	Protein	Function
Arsenate/arsenite detoxification	TBCP-5282_g_1112	*aioB*	Small subunit of arsenite oxidase	As(III) oxidation
	TBCP-5282_g_0536	*arsB_123*	Putative arsenical pump membrane protein	As(III) efflux pump
	TBCP-5282_g_1915			
	TBCP-5282_g_2525			
	TBCP-5282_g_1916	*arsC*	Arsenate reductase	As(V) reduction
Copper detoxification	TBCP-5282_g_1542 TBCP-5282_g_3015 TBCP-5282_g_3334 TBCP-5282_g_3337	*copA_1234*	Copper-exporting P-type ATPase; Lead, cadmium, zinc and mercury transporting ATPase	Cation translocation P-type ATPase
	TBCP-5282_g_2193 TBCP-5282_g_3338	*copZ_12*	Copper chaperone	
	TBCP-5282_g_1004	*cutC*	Copper homeostasis protein	
Chromate detoxification	TBCP-5282_g_4443	*chrR*	Chromate reductase	
	TBCP-5282_g_4179 TBCP-5282_g_4178 TBCP-5282_g_2372	*chrA_123*	Chromate transport protein	
Cadmium, zinc, cobalt, mercury detoxification	TBCP-5282_g_1905	*cadC*	Cadmium resistance transcriptional regulatory protein	
	TBCP-5282_g_1906	*cadA*	putative cadmium-transporting ATPase	
	TBCP-5282_g_4163	*zupT*	Zinc transporter	
	TBCP-5282_g_1166	*zosA*	Zinc-transporting ATPase	
	TBCP-5282_g_1794	*yeiR*	Zinc-binding GTPase	
	TBCP-5282_g_3393	*zur*	Zinc uptake regulation protein	
	TBCP-5282_g_4163	*znuA*	Zinc ABC transporter, substrate-binding protein	
	TBCP-5282_g_3965	*znuB*	Zinc ABC transporter, permease protein	
	TBCP-5282_g_1573	*znuC*	Zinc ABC transporter, ATP-binding protein	
	TBCP-5282_g_4038	*czcD*	Cadmium, cobalt and zinc/H(+)-K(+) antiporter	Cation efflux system protein
	TBCP-5282_g_3865	*corC*	Magnesium and cobalt efflux protein	
	TBCP-5282_g_4915	*corA*	Cobalt/magnesium transport	

			protein	
		merR	Mercuric resistance operon regulatory protein	
Nickel, molybdenum, detoxification	TBCP-5282_g_4042	*nikMN*	nickel transport protein	
	TBCP-5282_g_4137	*modA*	Molybdate-binding periplasmic protein	
	TBCP-5282_g_5286	*modB*	Molybdenum transport system permease protein	
Other detoxification		*mntH*	Divalent metal cation transporter	

Table 4 Average nucleotide identities (ANI) analysis of *Bacillus* sp. S3 and other *Bacillus* species.

	S3	LMG 21833	LMG 21834	DSM 15602	NBRC 102451	DSM 15604	NBRC 102450	FJAT-14227	ATCC 14580	YNP1-TSU	OM18	LF1	W23	168	OK048
Bacillus sp. S3	*	81.68	78.48	78.47	78.34	78.37	78.25	78.3	67.49	66.8	69.62	74.47	67.72	67.8	74.02
B. bataviensis LMG 21833	81.34	*	78.7	78.68	78.84	78.84	78.99	78.98	66.61	66.49	69.28	74.06	66.92	66.99	73.69
B. vireti LMG 21834	78.25	78.78	*	99.99	79.6	79.61	89.59	89.59	66.72	66.61	68.86	74.48	66.77	66.86	73.94
B. vireti strain DSM 15602	78.15	78.7	99.98	*	79.7	79.7	89.6	89.61	66.75	66.65	68.79	74.42	66.85	66.91	73.99
B. soli NBRC 102451	78.06	78.73	79.74	79.74	*	99.99	80.58	80.6	66.51	66.45	69.01	74.81	66.75	66.79	74.28
B. soli strain DSM 15604	78.39	79.05	79.94	79.92	99.99	*	80.76	80.78	67.61	66.9	69.99	75.11	67.9	67.92	74.83
B. novalis NBRC 102450	78	78.97	89.46	89.46	80.6	80.61	*	100	66.7	66.6	68.54	74.53	66.69	66.71	74.08
B. novalis FJAT-14227	78.42	79.31	89.58	89.56	80.66	80.7	99.99	*	67.63	67.04	69.49	74.92	67.55	67.76	74.43
B. licheniformis ATCC 14580	67.35	67.09	67.24	67.29	67.08	67.16	67.32	67.33	*	99.57	66.99	67.12	72.15	72.16	67.13
B. licheniformis YNP1-TSU	66.76	66.55	66.71	66.73	66.54	66.58	66.86	66.88	99.41	*	66.6	66.58	71.83	71.75	66.74
B. asahii strain OM18	70.12	70.21	69.86	69.8	70	70.05	69.79	69.85	67.96	67.09	*	70.07	68.52	68.46	69.93
Bacillus sp. LF1	74.3	74.32	74.64	74.66	75.01	75.04	74.75	74.8	66.83	66.64	69.47	*	67.14	67.19	73.34
B. subtilis subsp. str. W23	67.61	67.52	67.49	67.44	67.38	67.43	67.48	67.54	72.12	71.88	67.83	67.47	*	92.5	67.29
B. subtilis subsp. str. 168	67.67	67.62	67.53	67.47	67.46	67.45	67.52	67.53	72.14	71.82	67.87	67.53	92.23	*	67.57
Bacillus sp. OK048	74.04	73.97	74.26	74.3	74.56	74.59	74.21	74.26	66.58	66.53	69.16	73.35	66.86	66.85	*

Figures

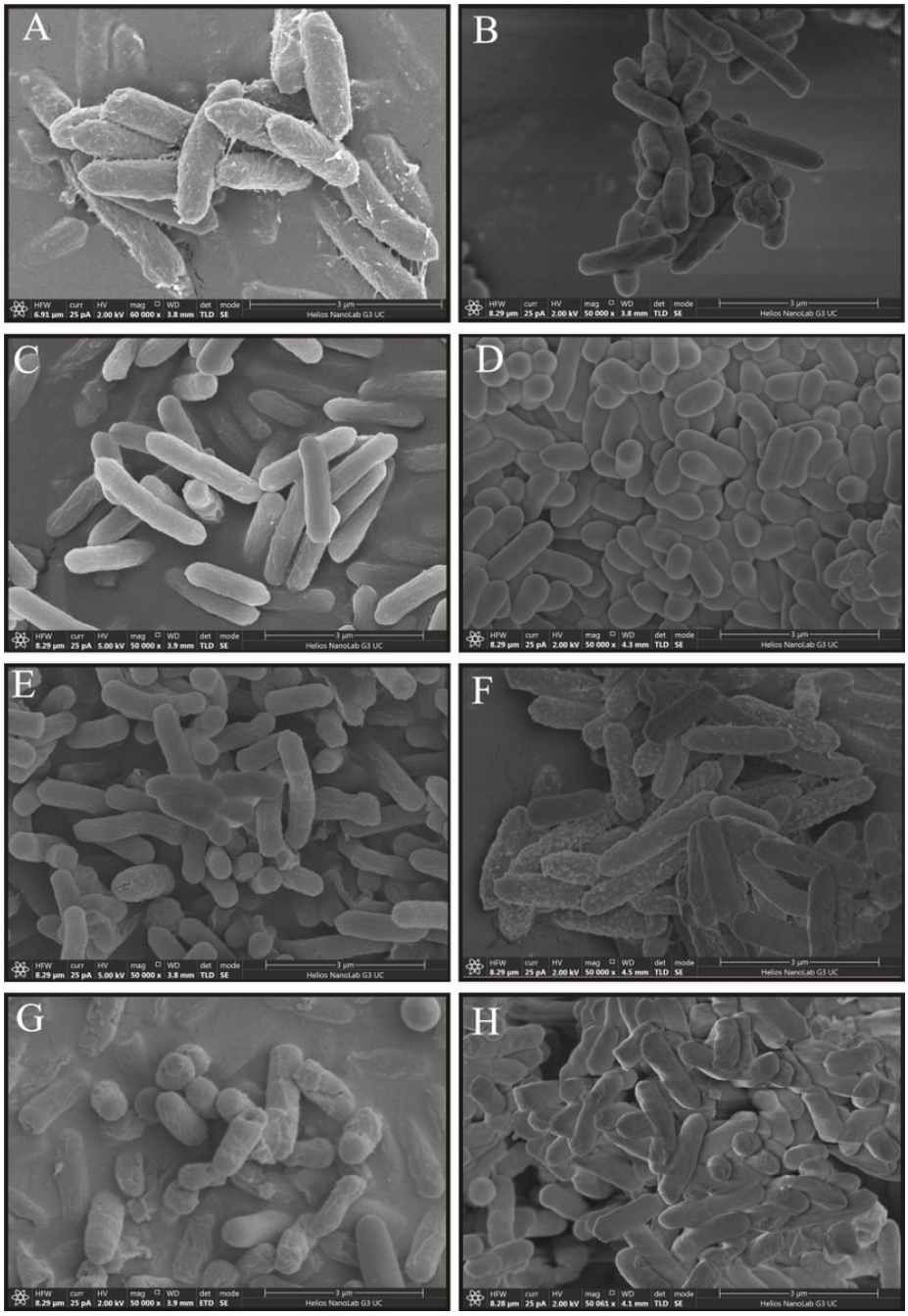

Figure 1

Scanning electron microscope (SEM) micrograph of Bacillus sp. S3 before and after the various heavy metal ions exposure: CK (A); Sb(III) (B); As(III) (C); Cd(II) (D); Cr(VI) (E); Pb(II) (F); Cu(II) (G); Zn(II) (H).

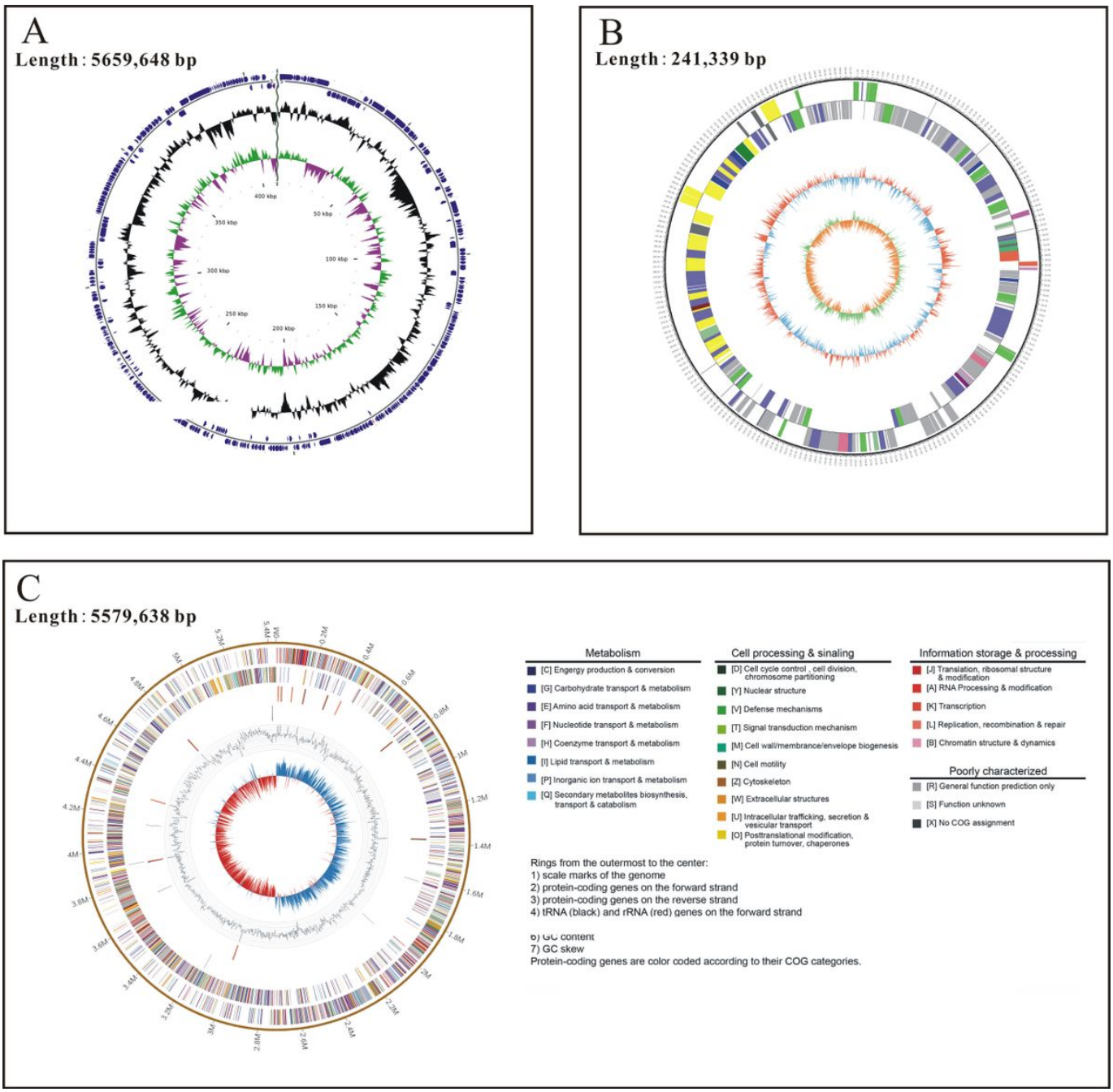

Figure 2

Circular diagrams of the Bacillus sp. S3 displaying relevant genome features. Genomic map of Bacillus sp. S3 with the draft genome sequence (A); Circular genome map of Bacillus sp. S3 chromosome (B) and plasmid (C) with the following information is provided from outside to inside: (1) scale marks of genomes; (2) assigned COG classes of protein-coding genes (CDSs) on the forward strand as indicated by relevant colors; (3) forward strand CDSs; (4) tRNA(black) and rRNA(red) genes on the forward strand; (5) tRNA(black) and rRNA(red) genes on the reversed strand; (6) GC content(swell outward/inward indicates higher/lower G+C compared with the average G+C content); (7) GC skew (cyan/red indicate positive/negative values).

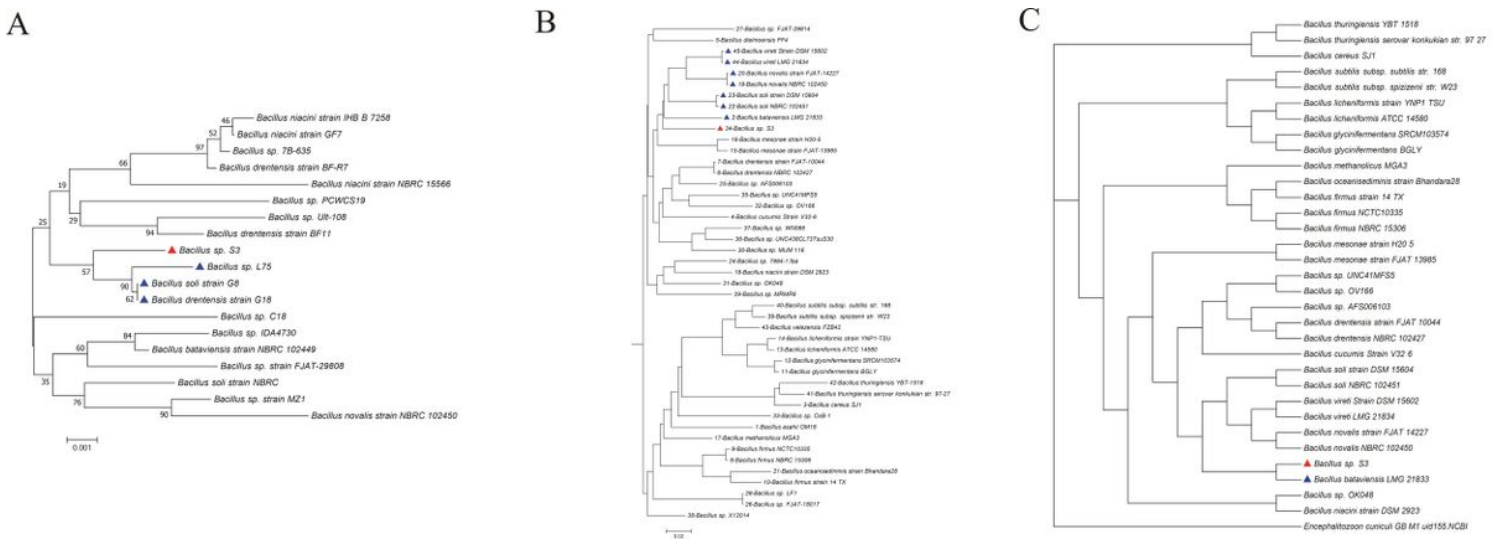

Figure 3

Phylogenetic relationships of 45 Bacillus strains. Phylogenetic trees based on (A)16s rRNA genes derived from Bacillus sp. S3 and other close-related strains. (B) 554 single-copy core genes. Bootstrap values are indicated at each node based on a total of 1,000 bootstrap replicates. (C) Whole-genome-based phylogeny trees using a composition vector (CV) approach. Bacillus sp. S3 with other closely related strains formed a group were marked in blue and S3 was marked in red blot. Encephalitozoon cuniculi GB M1 is regarded as a out group.

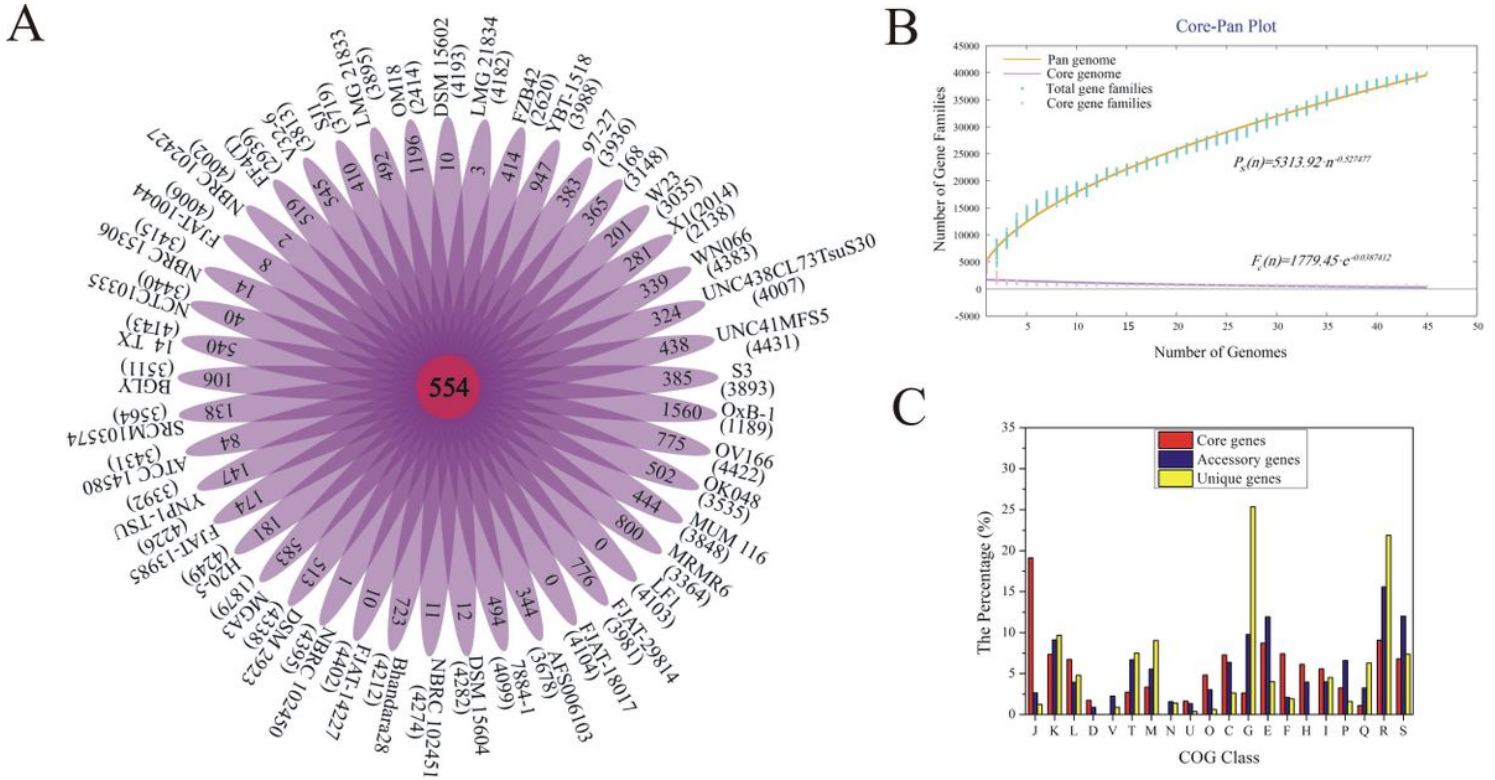

Figure 4

Pan genome analysis of strains within the Bacillus genus. (A) Venn diagram displaying the numbers of core gene and flexible genes for each of the 45 Bacillus strains. (B) Mathematical modeling of the pan-genome and core genome of Bacillus. (C) Proportion of genes enriched in the clusters of orthologous groups (COG) categories in unique genes, accessory genome, and pan-genome according to COG database.

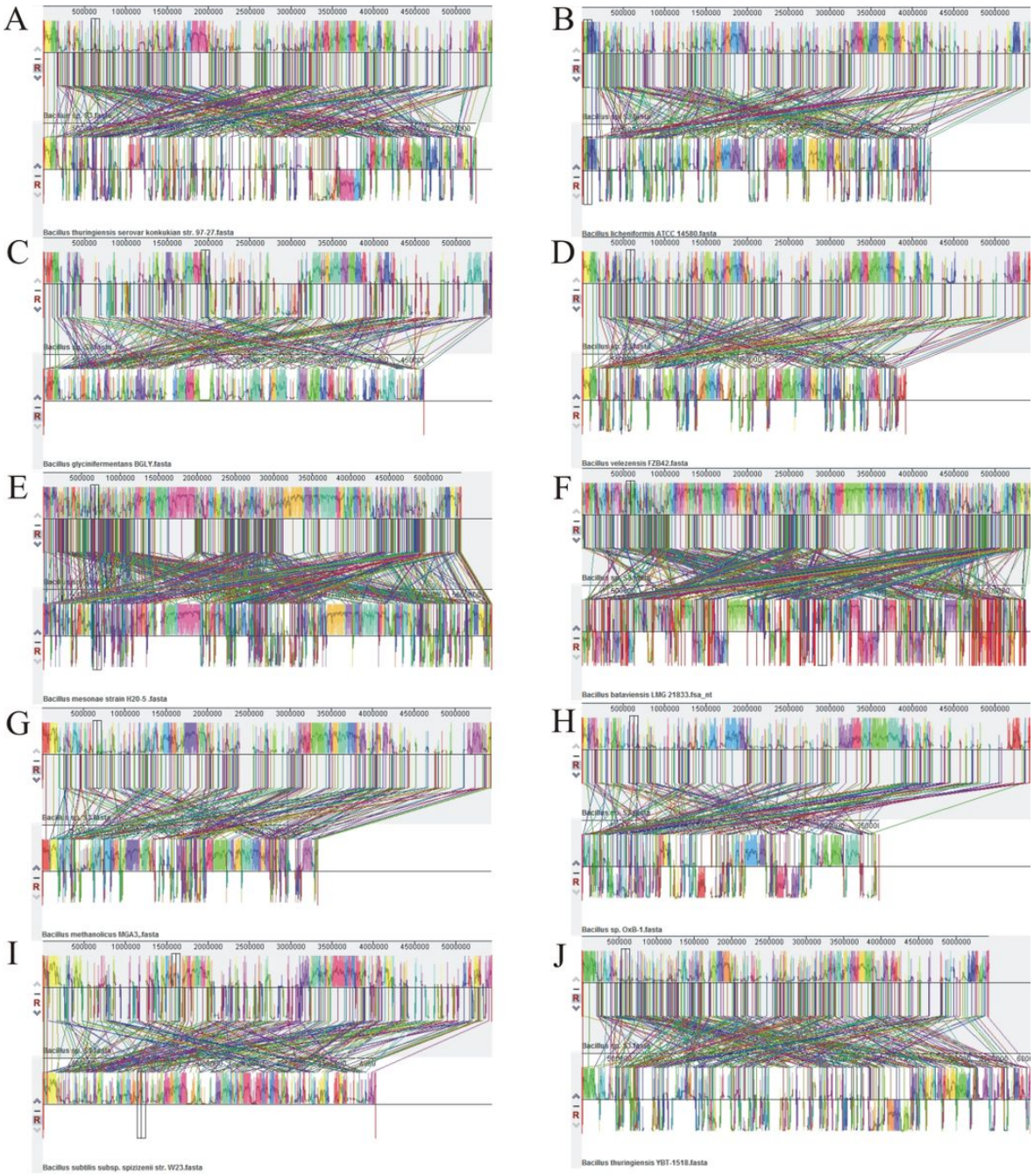

Figure 5

Whole genomes comparative alignment of Bacillus sp. S3 with 10 selected Bacillus strain. Bacillus thuringiensis serovar konkukian str. 97-27 (A); Bacillus licheniformis ATCC 14580 (B); Bacillus glycinifermentans BGLY (C); Bacillus velezensis FZB42 (D); Bacillus mesonae strain H20-5 (E); Bacillus bataviensis LMG 21833 (F); Bacillus methanolicus MGA3 (G); Bacillus sp. OxB-1 (H); Bacillus subtilis subsp. spizizenii str. W23 (I); Bacillus thuringiensis YBT-1518 (J). Boxes of different colors demonstrate the sequence coordinates and the conserved segments represented locally collinear blocks (LCBs) (or locally conserved regions). The LCBs above and below the reference line of the consistent color represent the orientation of the LCBs relative to the reference sequence for each genome. White areas represent possibly contain genome-specific sequence elements and those genomic positions that did not adequately align between the selected genomes.

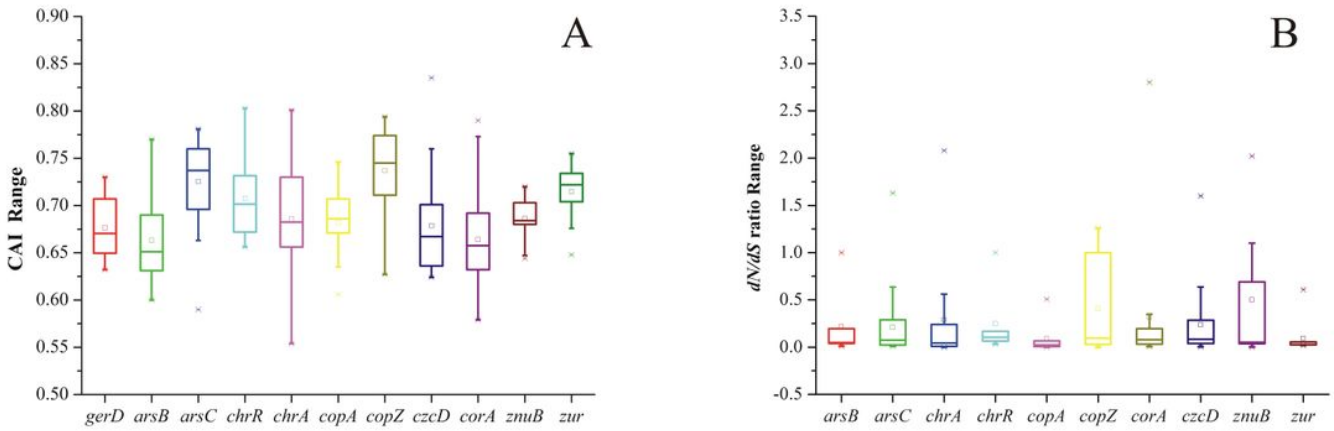

Figure 6

Distributions and range of selection pressure on various metal resistance genes within the genus Bacillus (A), Range of CAI values of different metal resistance genes of Bacillus with gerD gene as a reference(B).

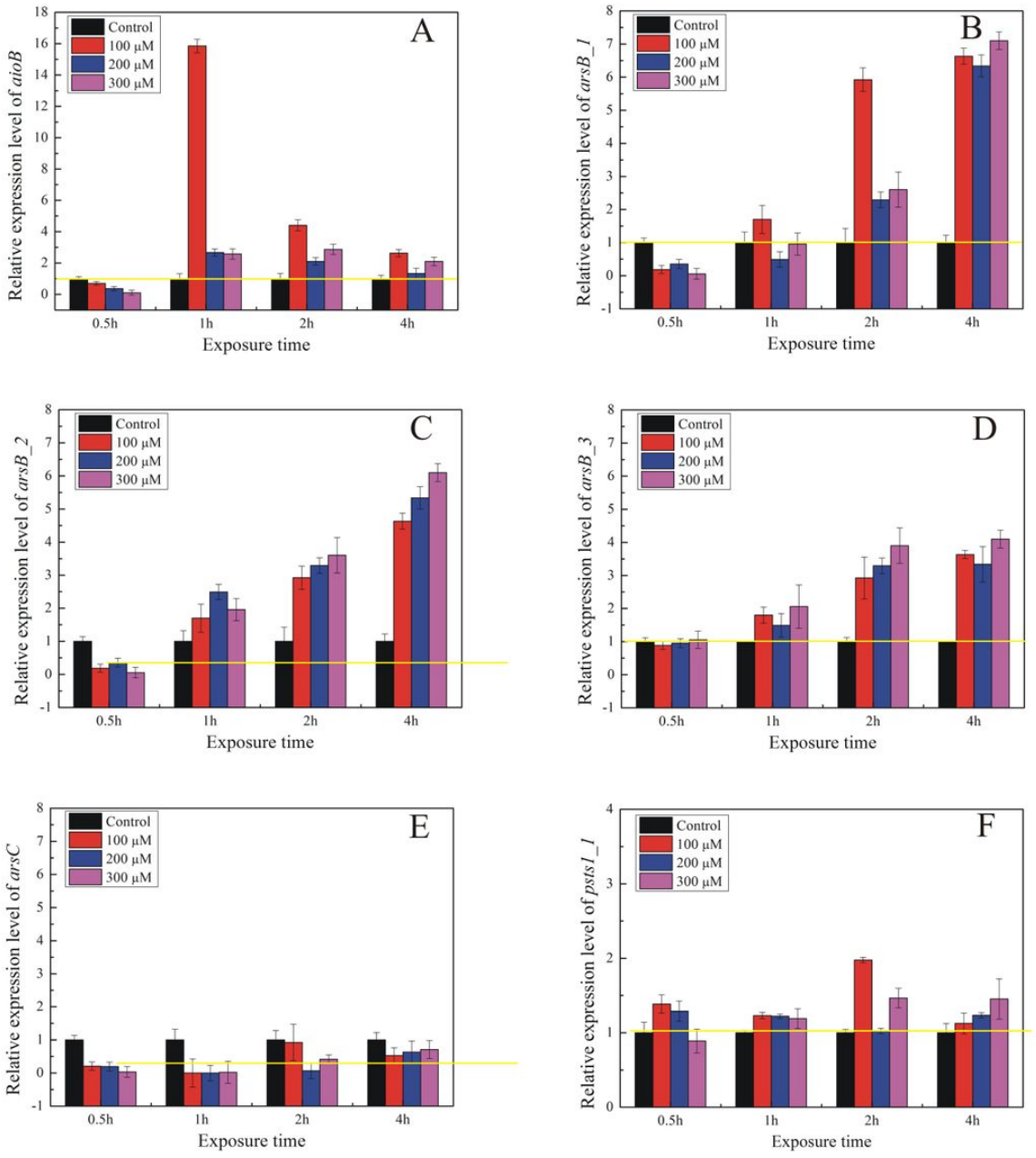

Figure 7

Quantitative reverse transcriptase-PCR analysis of the genes encoding proteins involved in antimonite/arsenate oxidation (A), antimonite/arsenite resistance (B, C, D), antimonate/arsenate reductase (E) and phosphate metabolism (F). Total RNA was isolated from strain S3 at different times (0.5 h, 1 h, 2 h and 4 h) cultured with control, 100 μM, 200 μM, 300 μM Sb(III) in CDM medium, respectively. Date shown as the mean of three replicates, with the error bars representing ± SD.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS5TKEGGdatabases.xlsx
- Supplementarymaterial.docx

- TableS4GOdatabases.xlsx
- TableS6Mobilegeneticelements.xlsx