



Article

Comparative Genomics of the Extreme Acidophile *Acidithiobacillus thiooxidans* Reveals Intraspecific Divergence and Niche Adaptation

Xian Zhang ^{1,2}, Xue Feng ¹, Jiemeng Tao ¹, Liyuan Ma ¹, Yunhua Xiao ¹, Yili Liang ^{1,2},
Xueduan Liu ^{1,2} and Huaqun Yin ^{1,2,*}

¹ School of Minerals Processing and Bioengineering, Central South University, Changsha 410083, China; zixuange2010@126.com (X.Z.); fengxue@csu.edu.cn (X.F.); taojiemeng@csu.edu.cn (J.T.); maliyuan2008@csu.edu.cn (L.M.); huazipiaoling.123@163.com (Y.X.); liangyili@hotmail.com (Y.L.); xueduanliu@yahoo.com (X.L.)

² Key Laboratory of Biometallurgy of Ministry of Education, Central South University, Changsha 410083, China

* Correspondence: yinhuaqun@gmail.com; Tel.: +86-731-8883-0546

Academic Editor: Patrick C. Y. Woo

Received: 3 July 2016; Accepted: 11 August 2016; Published: 19 August 2016

Abstract: *Acidithiobacillus thiooxidans* known for its ubiquity in diverse acidic and sulfur-bearing environments worldwide was used as the research subject in this study. To explore the genomic fluidity and intraspecific diversity of *Acidithiobacillus thiooxidans* (*A. thiooxidans*) species, comparative genomics based on nine draft genomes was performed. Phylogenomic scrutiny provided first insights into the multiple groupings of these strains, suggesting that genetic diversity might be potentially correlated with their geographic distribution as well as geochemical conditions. While these strains shared a large number of common genes, they displayed differences in gene content. Functional assignment indicated that the core genome was essential for microbial basic activities such as energy acquisition and uptake of nutrients, whereas the accessory genome was thought to be involved in niche adaptation. Comprehensive analysis of their predicted central metabolism revealed that few differences were observed among these strains. Further analyses showed evidences of relevance between environmental conditions and genomic diversification. Furthermore, a diverse pool of mobile genetic elements including insertion sequences and genomic islands in all *A. thiooxidans* strains probably demonstrated the frequent genetic flow (such as lateral gene transfer) in the extremely acidic environments. From another perspective, these elements might endow *A. thiooxidans* species with capacities to withstand the chemical constraints of their natural habitats. Taken together, our findings bring some valuable data to better understand the genomic diversity and econiche adaptation within *A. thiooxidans* strains.

Keywords: *Acidithiobacillus thiooxidans*; comparative genomics; intraspecific diversity; niche adaptation

1. Introduction

Extraordinarily extreme environments [1–3] are the habitats for extremophiles, although they were previously thought of as almost insurmountable physical and chemical barriers to life [4]. These extreme environmental conditions are inhospitable to the growth of most life [5,6]. However, acidophilic microorganisms, especially prokaryotic acidophiles (eubacteria and archaea), are considerably diverse in natural and man-made acidic environments (pH < 3) [7]. Thus, it is of great interest to identify the potential mechanisms that ensure microorganisms survive and proliferate in these extreme environments.

With the advance of high-throughput sequencing, numerous genomes derived from a wide range of organisms were sequenced continuously, thereby fueling the development of comparative genomics [8]. In contrast to standard genetic researches, which have inherent limits to elucidate the hereditary traits of species, acquisition of genomic sequences generated by high-throughput sequencing provides plentiful gene contents and further enables future studies to explore the primary issues that investigators are interested in. Acidophiles such as *Acidithiobacillus*, *Sulfobacillus*, and *Leptospirillum* were isolated from the extremely acidic environments, and their sequenced genomes were used for comparative survey [6,9,10]. Nevertheless, limited information is currently known about intraspecific variability of acidophiles at the genome level. Comparative genomics determining genomic differences among multiple strains of individual species can unravel the extent of intraspecific diversity [11]. The gene repertoire represented across all strains reveals the genomic diversity of a species, i.e., pan-genome, which consists of “core genome” (common genes to all strains of a species) and “dispensable or accessory genome” (genes shared by some but not all strains of the species as well as strain-specific genes) [12]. The core genome includes all common genes that are essential for its basic lifestyle and major phenotypic traits, while dispensable genome confers selective advantages such as niche adaptation, antibiotic resistance, and colonization of new hosts [12,13]. However, it remains unclear whether these intriguing findings concerning the genomic analyses of other organisms could be applied to acidophiles isolated from the harsh environments that are physico-chemically and ecologically distinct from the “normal” environments.

Acidithiobacillus, the acidophilic and obligately chemolithoautotrophic bacterium, is widely found in various acidic environments worldwide [7,14]. *Acidithiobacillus thiooxidans* (*A. thiooxidans*) is an aerobic mesophilic microorganism belonging to the genus *Acidithiobacillus* and is considered to play an important role in industrial bioleaching [10]. Until recently, three genomes of *A. thiooxidans* strains including ATCC 19377, A01, and Licanantay from various habitats (Table 1) have been sequenced and submitted to National Center for Biotechnology Information (NCBI) [15–17]. When the draft genome sequences were released, numerous genes were annotated and public, thus providing a wealth of valuable information. As a result, much research could be invested to identify novel insights into the genotypic characteristics.

Table 1. Strains of *Acidithiobacillus thiooxidans* used for comparison survey in this study.

Strain	Geographic Origin	Reference
Licanantay	Copper mine, Atacama, Chile	[17]
ATCC 19377	Kimmeridge clay, UK	[15]
GD1-3	Copper Mine, Shaoguan, Guangdong, China	This study
DXS-W	Copper Mine, Dongxiang Mountain, Hami, Xinjiang, China	This study
A02	Coal heap drainage, Pingxiang, Jiangxi, China	This study
A01	Coal dump, Pingxiang, Jiangxi, China	[16]
BY-02	Copper Mine, Baiyin, Gansu, China	This study
DMC	Coal heap drainage, Chenzhou, Hunan, China	This study
TYC-17	Copper Mine, Baiyin, Gansu, China	This study

In this study, six new genomic DNA of *A. thiooxidans* strains isolated from different acidic environments in China (Table 1) were extracted and sequenced. Together with three aforementioned genomes publicly available in the GenBank database, a global genomic comparison was executed. Our work showed a data-driven approach to elucidate the similarities and differences among *A. thiooxidans* genomes, aiming to explore the genetic diversity and niche adaptation within *A. thiooxidans* strains.

2. Results and Discussion

2.1. General Features of *Acidithiobacillus thiooxidans* (*A. thiooxidans*) Genomes

Six new genomic DNA were subjected to Illumina MiSeq sequencing platform, and an average of 240 Mb raw reads (short DNA sequences) in each genome was yielded. After quality control using NGS QC Toolkit, high quality (HQ) reads (85.3% to 87.80%) were retained for subsequent analyses. All HQ reads aforementioned were used for sequence assembly, and an in-house Perl script was then employed to filter the assembled sequences under 200 bp, resulting in the draft genome assemblies. Genome characteristics were summarized in Table S1. Of note, the draft genome of *A. thiooxidans* ATCC 19377 sequenced previously [15] was much smaller than these of the others, preliminarily inferring that the low-coverage genome sequencing (9.6-fold) might contribute to the missing of large fragments genomic DNA. Nevertheless, high quality and completeness of genome assemblies was acquired in this strain (Table S1). Thus, pan-genome analysis could be reliable as the high quality of genome completeness was estimated in all strains.

As listed in Table S1, total genome sizes varied among all strains (3.02 to 3.95 Mb). As stated by Nuñez et al. [18], genome size in prokaryotes is related to metabolic diversity, effective population size, regulatory complexity, and horizontal transfer rates. And larger genomes might have a high adaptive plasticity compared to smaller genomes [17]. Previous studies showed that the genome of *A. ferrooxidans* DSM 16786 isolated from mining processes [19] was much larger than that of *A. ferrooxidans* ATCC 23270, which was from the bituminous effluent of coal mine [17]. Likewise, the larger genome size of *A. thiooxidans* strains GD1-3, DXS-W, and Licanantay obtained from copper mining implied that putative gene acquisition might enable them to adapt to the high concentration of metals in these metal mines. Also, the number of putative coding sequence (CDS) (3087 to 4200) and mean mole percentage GC content (52.84% to 53.17%) varied among all sequenced genomes (Figure 1 and Table S1). There is slight difference about mean percentage GC content among strains, while the number of CDS varied greatly. The CDS counts of strain Licanantay (4200) and GD1-3 (4171) were slightly more than those of the others.

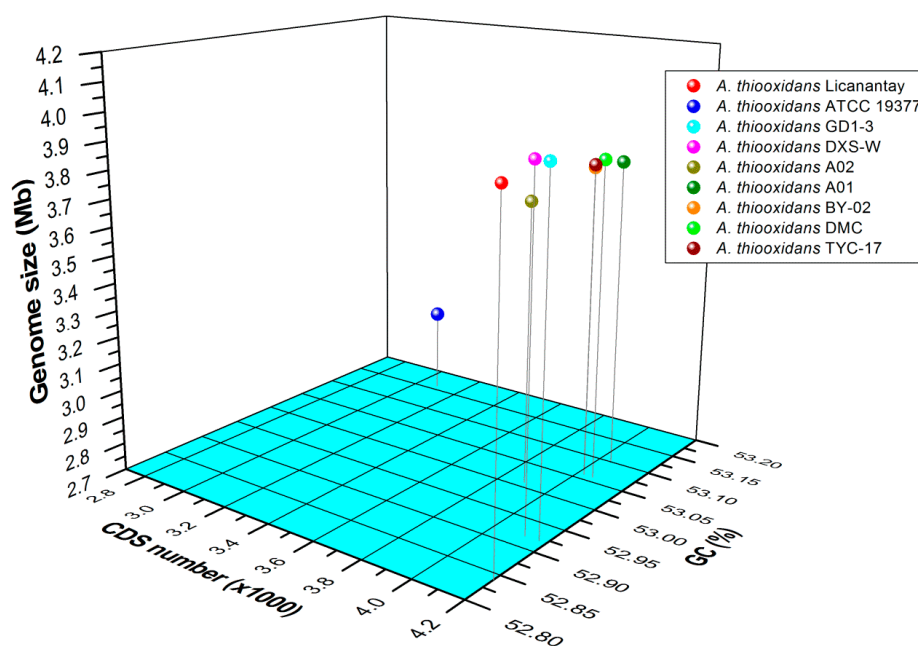


Figure 1. Three-dimensional plots of genome size, coding sequence (CDS) number, and GC content of the nine *Acidithiobacillus thiooxidans* (*A. thiooxidans*) strains sequenced in this study. The available genomes from strains Licanantay, ATCC 19377, and A01 were acquired from the public database, and the others were sequenced in this study.

2.2. Pan-Genome Analysis of *A. thiooxidans* Strains

To understand the pan-genome of *A. thiooxidans* more deeply, 7186 protein CDSs obtained from the six newly sequenced genomes plus three available genomes from the public database were clustered using the program PanOCT (Developer, Location) with a 50% sequence identity cut-off. Herein, 2043 (28.31%) orthologs were identified as the *A. thiooxidans* core genome, and the remaining variable genes were defined as the accessory genome of *A. thiooxidans* (Figure 2A). In particular, our results showed that Licanantay had the largest number of unique genes (1001), followed by ATCC 19377 (421). These were similar to what Travisany et al. [17] reported previously. In their study, comparative analysis showed that strain-specific genes in the Licanantay might be involved in adaptation to its specific biomining environment and several genetic motility genes likely acquired by horizontal gene transfer in mining environments.

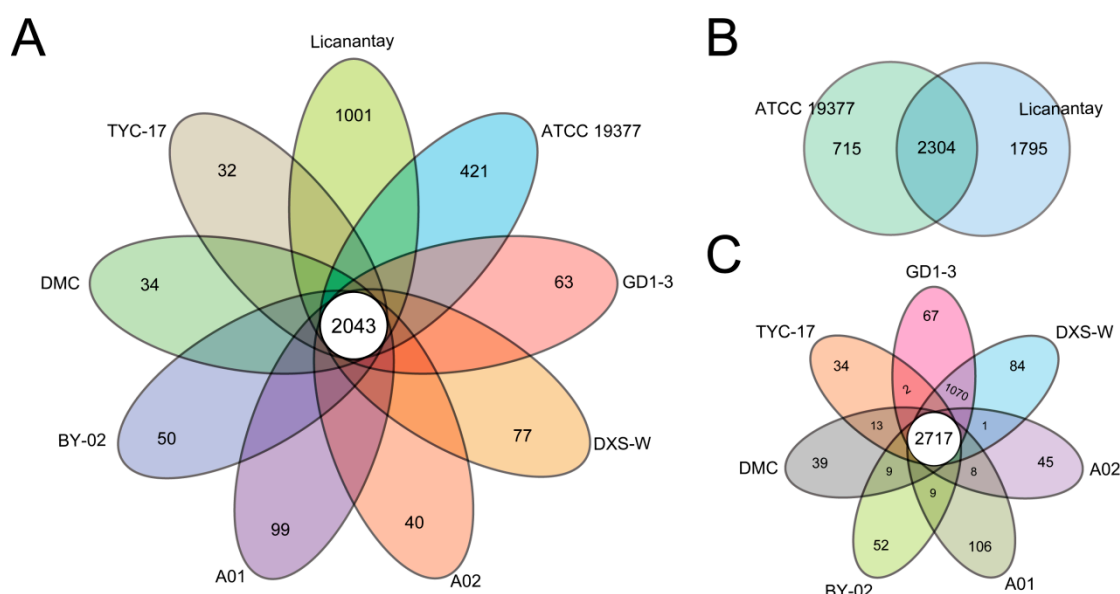


Figure 2. The pan-genome of *Acidithiobacillus thiooxidans* strains. The flower plots and Venn diagram demonstrate the number of shared, accessory and strain-specific genes among *A. thiooxidans* strains. Each strain was represented by an oval or circle that was colored. (A) Flower plot showing the numbers (in the petals) correspond to the unique genes of each strains, and the number of core genome common to all *A. thiooxidans* strains (in the center); (B) Venn diagram showing the numbers of unique genes and core orthologous genes between *A. thiooxidans* ATCC 19377 and Licanantay; (C) flower plot showing the numbers of CDSs among all *A. thiooxidans* strains in this study except for ATCC 19377 and Licanantay.

Considering the closely relation between Licanantay and ATCC 19377 [17], we further identified their shared genes. 2304 orthologous genes were present, and the unique genes in Licanantay and ATCC 19377 were 1795 and 715, respectively (Figure 2B), further indicating that Licanantay with much more strain-specific genes had the advantage to adapt the environmental conditions. Additionally, core orthologous genes and unique genes within seven other strains were examined. 2717 orthologs were identified, and the number of strain-specific genes in each strain varied from 34 to 106 (Figure 2C). In particular, GD1-3 and DXS-W shared a large number of genes (1070) only between the two of them, suggesting a closely correlation with each other.

2.3. Phylogenomic Tree Based on Core Genome

Since the 16S rRNA gene sequences among each pair of *A. thiooxidans* strains are highly similar, we could not assess the phylogenetic distance between strains using these sequences alone. In this study, a phylogenomic tree based on their core genome (Figure 3) showed that nine strains were

apparently divided into three main groupings. As depicted in Figure 3, *A. thiooxidans* Licanantay was genetically distinct from other strains included in this study, and strains GD1-3, DXS-W, A02, A01, BY-02, DMC, and TYC-17 were closely related to each other. In fact, the strains ATCC 19377 and Licanantay were originally isolated from Kimmeridge clay and Chilean copper mine respectively [15,17], and the others from various acidic environments in China (Table 1). Thus, a hypothesis was proposed that hereditary difference might be related to the geographic distribution.

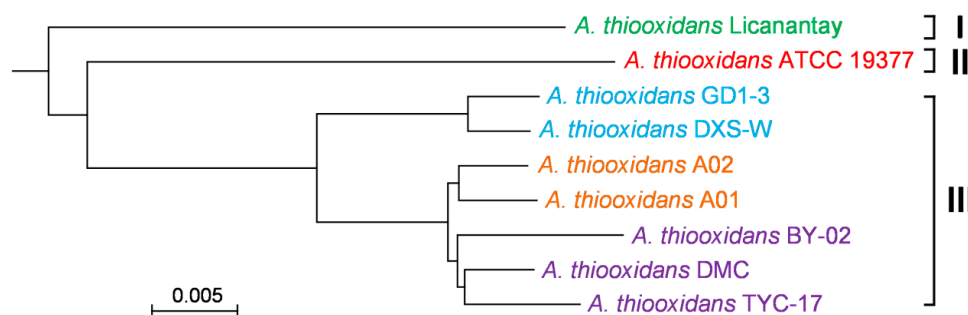


Figure 3. Phylogenomic tree of sequenced *Acidithiobacillus thiooxidans* strains based on their core genome. These strains from various geographic origins were clustered into three distinct classes. Classe I represents strain Licanantay isolated from Kimmeridge clay, class II represents strain ATCC 19377 from Chilean copper mine, and class III represents certain strains isolated from different acidic environments in China.

Further inspection showed that these six strains from China were classified into three clusters. As reported by Douillard et al. [20], phylogenomic analysis showed that multiple groupings of *Lactobacillus rhamnosus* partly be related to their ecological niches. In our study, *A. thiooxidans* strains GD1-3 and DXS-W were isolated from the similar environments, and strains A01 and A02 from the same sampling points. While strains BY-02 and TYC-17 were isolated from copper mine, and DMC from coal heap drainage, these three strains gathered in a cluster. Unfortunately, the detailed geochemical conditions of these six bacteria, at that time, were not measured. Thus, the limited samples and experimental data restricted our further analysis to determine whether genetic difference was correlated with the geochemical characteristics in these acidic environments.

2.4. Functional Features of the Pan-Genome

To identify possible intraspecific diversification in functions, the functional annotation of core genome and accessory genome were performed against the specialized database Clusters of Orthologous Groups (COG). As shown in Figure 4, the abundances of metabolism-related genes assigned to COG categories (C) (energy production and conversion), (E) (amino acid transport and metabolism), (G) (carbohydrate transport and metabolism), (F) (nucleotide transport and metabolism), (H) (coenzyme transport and metabolism) and (I) (lipid transport and metabolism) in the core genome were greater in these *A. thiooxidans* strains compared to those in the accessory genome. These findings were reasonable given that these shared genes were involved in microbial basic activities, which might support the view that core genome was essential for basic lifestyle of species [12,13]. Additionally, the core genome was highly enriched in COG category (J) (translation, ribosomal structure and biogenesis) relative to accessory genome. These features were similar to what other researchers have been found in their respective pan-genome analyses [21,22]. Especially, the core genome in all strains was commonly enriched in the COG category (M) (cell wall/membrane/envelope biogenesis). We interpreted this as an indication that *A. thiooxidans* strains inhabiting the acidic environments shared distinctive structural and functional properties to maintain a stable pH gradient, as specialized cellular structures were regarded to be important for acidophile pH homeostasis [23].

In contrast, the accessory genome among *A. thiooxidans* strains consisted of putative 5143 CDSs, and COG class assignment revealed the abundant CDSs were involved in replication, recombination and repair (COG category (L); Figure 4). Considering that the high concentration of toxic substances such as heavy metals in these acidic environments [7,24], and the high level of heavy metals concentration might cause a high rate of DNA injury [25], it appears to be reasonable that there were the abundant genes in the accessory genome assigned to COG category (L), which might be related to niche adaptation. This finding was in line with previous comparative genomics showing that the accessory genome of 48 strains of sinorhizobia *Sinorhizobium* comprising five genospecies might be related to the different strategies to interact with diverse host plant and soil environments [21]. Also, comparative analysis of *Klebsiella pneumoniae* Kp13 showed that genomic plasticity occurring at multiple hierarchical levels might play a role of the lifestyle [26]. Besides, both core genome and accessory genome had high proportion of genes in COG categories (R) (general prediction only) and (S) (function unknown). The amino acid sequences associated with these CDSs which lacked a functional assignment were then chosen for functional annotation against the NCBI-NR database ($E\text{-value} \leq 10^{-5}$). Results indicated that various proportions of CDSs (20.55%~74.03%) were hypothetical proteins, and the others were annotated as proteins with known function (Table S2). For these functional proteins, we found most CDS showing hits with *A. thiooxidans*. We emphasized the reasonability of these findings that CDS in COG categories (R) and (S) could be re-annotated as proteins with known function in our study, due to the continuously updated database.

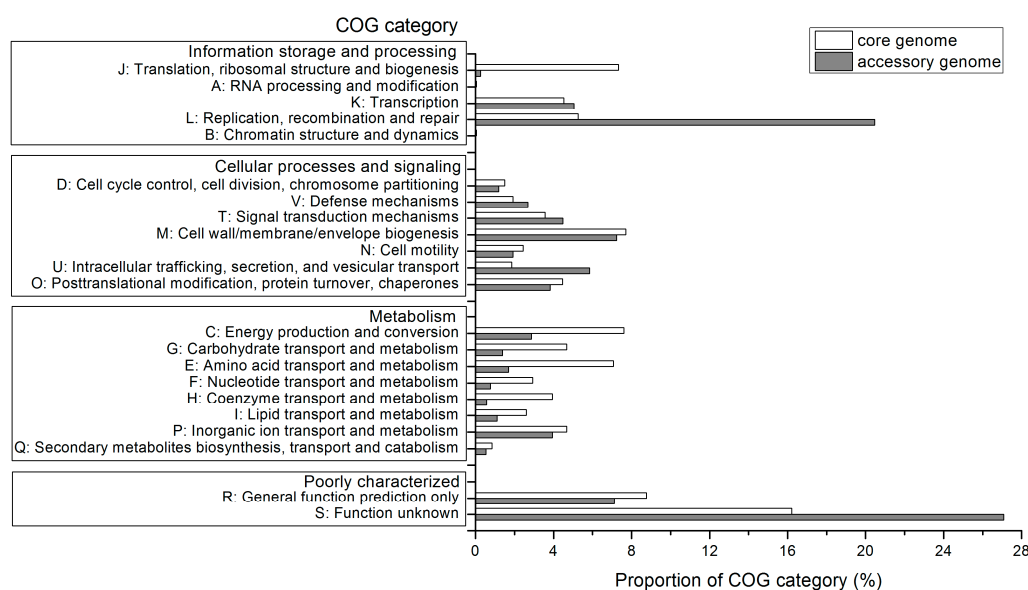


Figure 4. Distribution of core and flexible genes based on Clusters of Orthologous Groups (COG) category in *A. thiooxidans* strains. Only orthologous genes assigned to COG category were used for analysis.

We also performed a Blast search against the NCBI-nr database using the protein sequences related to strain-specific genes. Similar to several other comparative genomic analyses [27–29], a large number of CDS were annotated as hypothetical proteins. Furthermore, most of them were not assigned to the COG category (Figure S1A). Especially, we further inspected the non-shared CDS from *A. thiooxidans* Licanantay and ATCC 19377 considering that the numbers were larger compared to those in other strains. The most abundant strain-specific CDS were assigned into COG category (L) and (M) (Figure S1B), which was consistent with previous study [17]. These strain-specific genes might confer them some advantages to adapt to the environmental conditions.

family transporter, would enter into the biosynthesis of glutamate and other vital compounds. Results showed that all strains were found to harbor a full suit of genes involved in nitrogen metabolism except for strain ATCC 19377. For example, several genes associated with assimilatory and assimilatory nitrate reduction were not identified in its genome (Table S3). Also, genes encoding transporters for nitrate or nitrite were absent. One possible explanation was that strain ATCC 19377 could take up environmental ammonium as its sole pathway for the acquisition of nitrogen source.

Referred to the well-studied models for sulfur oxidation in *A. thiooxidans* species [10,35], a hypothetical model for sulfur oxidation system and electron transportation was reconstructed in our study (Figure 5). In this model, a series of complicated enzymatic reactions were performed by various enzymes with distinguished features in specific cellular position. The *sor* gene encoding sulfur oxygenase reductase, an enzyme directing the disproportionation reaction of sulfur to generate sulfide, thiosulfate, as well as sulfate [36], was absent in strain ATCC 19377. It seems that horizontal gene transfer might occur in this strain. However, it remains to be further validated.

Our results showed that the vast majority of homologous genes were identified by aligning against the public database as well as reported sequences in other literatures, although there were few exceptions such as genes involved in nitrate reduction in strain ATCC 19377. Taken together, we proposed that there were few intraspecific differences with respect to metabolic pathways, at least central metabolism.

2.5.2. Predicted Stress Tolerance Mechanisms

The comparison of metabolic profiles of *A. thiooxidans* strains was extended to potential mechanisms to respond to environmental stress, including acidic pH, high concentration of toxic substrates such as heavy metal ion and organic compounds (not shown in Figure 5). Also, cell mobility was taken into account. In this study, all strains shared a core set of genes potentially related to stress management. Like most other acidophiles, *A. thiooxidans* species highly relied on intracellular pH homeostasis, allowing it to grow at distinct ranges of pH values [37]. Numerous genes assigned to COG category (N) (cell mobility) were predicted to be involved in flagella formation (Table S3). The presence of these putative genes indicated that all *A. thiooxidans* strains had the capacity to actively move in aquatic habitats. Additionally, the identification of several gene clusters related to the resistance of heavy metals including arsenic, mercury, cadmium, and copper suggested that all strains might employ various systems to cope with high metal ion concentrations. As for organic solvents such as Lix984n, an organic extractant used for metal extraction in industrial operation [38], a putative gene cluster *acrAB-tolC* encoding a member of resistance-nodulation-cell division (RND) family protein was predicted to potentially transfer this substrate [39]. Additionally, *A. thiooxidans* strains were found to harbor a complete six-gene cluster for ATP binding cassette (ABC) transport system involved in toluene resistance (Table S3). In conclusion, all strains observed in this study appear to exhibit similar strategies to cope with the chemical constraints of their natural habitats.

2.6. Gene Turnover Analysis

Similar to earlier study in the closely related *Acidithiobacillus caldus* [40], prediction and classification of transposases using ISFinder showed that 3.41%~4.54% of the predicted CDSs of *A. thiooxidans* strains encoded transposases belonging to 25 different insertion sequence (IS) families (Table 2). The most abundant IS families in all strains were ISL3. High similarity regarding IS type were observed in all *A. thiooxidans* strains, however, closer inspection demonstrated several differences. For instance, *A. thiooxidans* strains GD1-3, DXS-W, and Licanantay with larger genome size had more transposases than the others. This finding might be reasonable considering that horizontal transfer was regarded as an evolutionary force to increase microbiological DNA content [41]. On the other hand, a frequent gene flow and genetic drift might endow species with adaptive capacity to the extreme econiche.

Table 2. The prediction of mobile elements including insertion sequences (IS) and genomic island (GI) in all *A. thiooxidans* strains observed in this study.

A. The Putative Insertion Sequences									
IS Family	DXS-W	Licanantay	A01	ATCC 19377	GD1-3	DMC	A02	BY-02	TYC-17
IS110	9	11	9	4	8	9	9	9	10
IS1182	2	0	0	0	2	0	0	0	0
IS1380	3	2	0	0	3	0	0	0	0
IS1595	5	8	8	4	5	9	7	9	8
IS1634	2	1	3	9	1	3	3	3	3
IS200/IS605	6	15	4	3	5	4	2	4	4
IS21	6	8	11	2	6	8	9	9	9
IS256	1	8	2	15	1	3	3	2	2
IS3	11	12	4	1	10	6	5	5	5
IS30	1	2	0	3	1	2	2	2	2
IS4	7	6	3	12	7	9	7	7	7
IS481	8	1	2	1	8	2	2	1	2
IS5	4	4	5	5	4	6	5	5	6
IS51	0	2	0	2	0	0	0	0	0
IS605	1	0	1	0	1	1	1	1	1
IS607	4	4	1	0	4	4	1	4	1
IS630	11	18	6	9	12	14	12	11	9
IS66	11	0	1	0	9	2	1	1	1
IS91	20	16	16	8	19	15	16	16	15
ISAs1	1	2	1	0	1	1	0	1	0
ISAZo13	1	1	0	0	1	0	0	0	0
ISKra4	7	2	4	2	7	1	4	1	4
ISL3	43	44	39	30	47	41	41	40	42
ISNCY	2	1	1	0	2	1	2	2	2
Tn3	21	16	17	16	20	19	19	20	21
Total	187	184	138	126	184	160	151	153	154
B. The Predicted Genomic Islands									
Strain	DXS-W	Licanantay	A01	ATCC 19377	GD1-3	DMC	A02	BY-02	TYC-17
GI Number	65	54	39	36	56	53	51	50	44

The four most abundant IS families were highlighted in bold.

Besides, finding presented in this study revealed that many genomic island (GI) occurred in the *A. thiooxidans* population (Table 2). Further analyses suggested that several integrases or mobile genetic elements were presented in the predicted GI (not shown in Table 2), thereby indicating that these putative GI were likely acquired by horizontal gene transfer. As stated by Wu et al. [42], GI were highly related to the niche-specific adaptation. Thus, it was inferred that these GI might play a key role in adapting to specific lifestyles and environmental niches. In short, great frequency of genetic exchange might provide *A. thiooxidans* species with adaptive advantage in extremely acidic environments.

3. Materials and Methods

3.1. Bacterial Strains Used in This Study

We included a total of six strains of *A. thiooxidans* in this study, which were preserved in the acidophile culture collection maintained at Central South University, Changsha, China (Table 1). Each strain was cultivated in liquid 9K basic medium, including the following ingredients (grams per liter): $(\text{NH}_4)_2\text{SO}_4$, 3.0; $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.5; KCl, 0.1; $\text{Ca}(\text{NO}_3)_2$, 0.01; and K_2HPO_4 , 0.5. Of this solution, elemental sulfur (autoclave-sterilized, 10g/L) was added as energy source. The culture temperature and shaking speed of rotary platform were 30 °C and 170 rpm, respectively. High-concentration cells ($1\sim 2 \times 10^8$ cells/mL) were collected at the exponential growth phase (generally 3th day), and their

genomic DNA were extracted using TIANamp Bacteria DNA Kit (TIANGEN, Beijing, China) following the manufacturer's introductions.

3.2. Genome Sequencing and Bioinformatics Analyses

The genomes of six strains of *A. thiooxidans* were sequenced using the Illumina MiSeq platform for 2×150 bp paired-end sequencing (Illumina, Inc., San Diego, CA, USA). The raw reads were filtered using NGS QC Toolkit [43] with Phred 20 as a cutoff, and then paired-end read sequences with high-quality were assembled using velvet [44]. For each strain, several kmers were run and the best assembled result was chosen for subsequent analyses. Following the cleanup step, the contigs were further clustered and assembled de novo to obtain unigene sequences using iAssembler tool [45]. The genome completeness of each strain was estimated using the program CheckM [46]. Coding sequences (CDS) prediction was performed by MetaGeneAnnotator [47], and predicted genes were then functionally annotated via homology searching against the generalist databank (NCBI-NR) and specialised databases (COG and KEGG). The hidden Markov models for the protein domains were predicted via searching against the Pfam database [48]. The online platform tRNAscan-SE was used for the identifications of tRNA [49]. Additionally, insertion sequences and transposases were identified by Blastp against the ISFinder database [50] with manual inspection of search hits (E -value $\leq 10^{-5}$). The putative genomic islands were also predicted using the online platform IslandViewer 3 [51].

3.3. Pan-Genome Analysis

Gene annotations of assembled contigs of the six strains were performed through the RAST annotation server [52]. The GenBank files containing the genome information of *A. thiooxidans* strains were then downloaded. Additionally, three other available genome sequences from homologous strains including *A. thiooxidans* strains A01, ATCC 19377, and Licanantay were obtained from NCBI. For these nine strains, the corresponding protein sequences with Fasta format were extracted using in-house Perl scripts, and were then aligned using an all-versus-all BLASTP. Output files with the m8 BLAST format were used for the identification of single-copy orthologs using PanOCT program (50% identity cut-off; E -value $\leq 10^{-5}$) [53]. Finally, the shared and dispensable genes in all strains were functionally annotated as described above.

3.4. Phylogenomic and Phylogenetic Analyses

A core of ortholog genes from nine draft genomes of *A. thiooxidans* strains were extracted using in-house Perl script. These shared genes were then used to construct a genome-based and alignment-free phylogeny using the online platform CVTree3 with K-tuple length 6 [54]. Additionally, the genome of *Acidithiobacillus caldus* SM-1 was included as an out-group. Visualization for phylogenetic tree was performed using the software MEGA v5.05 [55].

3.5. Data Deposition

These Whole Genome Shotgun projects of six *A. thiooxidans* strains have been deposited at the DDBJ/ENA/GenBank under the accession LWSC00000000 (GD1-3), LWRY00000000 (DXS-W), LWSA00000000 (A02), LWRZ00000000 (BY-02), LWSB00000000 (DMC), and LWSD00000000 (TYC-17). Additionally, the versions described in this paper are version LWSC01000000, LWRY01000000, LWSA01000000, LWRZ01000000, LWSB01000000, and LWSD01000000, respectively.

4. Conclusions

Comparative genomics provided useful information on the genetic and functional features of *A. thiooxidans* strains. Phylogeny based on their core genome revealed that genetic diversity was potentially related to the geographic distribution and geochemical conditions of their habitats. Functional assignment of common genes uncovered that abundant genes were involved in metabolism,

such as COG categories (C), (E), and (G), compared to the accessory genome, indicating that these genes were necessary for the microbial basic activities. Comprehensive analysis further showed little intraspecific diversification with respect to their predicted metabolic profiles. Additionally, most genes belonging to the dispensable genome were assigned to the COG category (L) and (M), suggesting a correlation between accessory genome and niche adaptation. Also, a considerable diverse repertoire of mobile genetic elements including insertion sequences and genomic islands were widespread in the draft genomes of *A. thiooxidans* strains obtained from various geographic origins, indicating that gain and/or loss of these elements by transferring horizontally might greatly contribute to intraspecific divergence and adaptation to acidic econiches.

Supplementary Materials: Supplementary materials can be found at www.mdpi.com/1422-0067/17/8/1355/s1.

Acknowledgments: This work was supported by the National Natural Science Foundation of China (No. 31570113 and No. 41573072) and the Fundamental Research Funds for the Central Universities of Central South University (No. 2016zzts102). We thank Wei Lin in Chinese Academy of Sciences and Guanyun Wei in Nanjing Normal University for analyzing the data. Also, we thank NCBI for providing the genome sequences of *A. thiooxidans* strains ATCC 19377, A01, and Licanantay.

Author Contributions: Conceived and designed the experiments: Xian Zhang, Xueduan Liu and Huaqun Yin; Performed the experiments: Xue Feng and Jiemeng Tao; Analyzed the data: Xian Zhang; Contributed to sample collection: Liyuan Ma and Yunhua Xiao; Wrote the paper: Xian Zhang; Revised the paper: Yili Liang.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Kennedy, S.P.; Ng, W.V.; Salzberg, S.L.; Hood, L.; DasSarma, S. Understanding the adaptation of *Halobacterium* species NRC-1 to its extreme environment through computational analysis of its genome sequence. *Genome Res.* **2001**, *11*, 1641–1650. [[CrossRef](#)] [[PubMed](#)]
- Takami, H.; Takaki, Y.; Uchiyama, I. Genome sequence of *Oceanobacillus iheyensis* isolated from the Iheya Ridge and its unexpected adaptive capabilities to extreme environments. *Nucleic Acids Res.* **2002**, *30*, 3927–3935. [[CrossRef](#)] [[PubMed](#)]
- Falb, M.; Pfeiffer, F.; Palm, P.; Rodewald, K.; Hickmann, V.; Tittor, J.; Oesterhelt, D. Living with two extremes: Conclusions from the genome sequence of *Natronomonas pharaonis*. *Genome Res.* **2005**, *15*, 1336–1343. [[CrossRef](#)] [[PubMed](#)]
- Rothschild, L.J.; Mancinelli, R.L. Life in extreme environments. *Nature* **2001**, *409*, 1092–1101. [[CrossRef](#)] [[PubMed](#)]
- Guo, X.; Yin, H.; Liang, Y.; Hu, Q.; Zhou, X.; Xiao, Y.; Ma, L.; Zhang, X.; Qiu, G.; Liu, X. Comparative genome analysis reveals metabolic versatility and environmental adaptations of *Sulfobacillus thermosulfidooxidans* strain ST. *PLoS ONE* **2014**, *9*, e99417. [[CrossRef](#)] [[PubMed](#)]
- Zhang, X.; Niu, J.; Liang, Y.; Liu, X.; Yin, H. Metagenome-scale analysis yields insights into the structure and function of microbial communities in a copper bioleaching heap. *BMC Genet.* **2016**, *17*, 21. [[CrossRef](#)] [[PubMed](#)]
- Johnson, D.B. Biodiversity and ecology of acidophilic microorganisms. *FEMS Microbiol. Ecol.* **1998**, *27*, 307–317. [[CrossRef](#)]
- Jacobsen, A.; Hendriksen, R.S.; Aarestrup, F.M.; Ussery, D.W.; Friis, C. The *Salmonella enterica* pan-genome. *Microb. Ecol.* **2011**, *62*, 487–504. [[CrossRef](#)] [[PubMed](#)]
- Tyson, G.W.; Chapman, J.; Hugenholtz, P.; Allen, E.; Ram, R.; Richardson, P.; Solovyev, V.; Rubin, E.; Rokhsar, D.; Banfield, J. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **2004**, *428*, 37–43. [[CrossRef](#)] [[PubMed](#)]
- Yin, H.; Zhang, X.; Li, X.; He, Z.; Liang, Y.; Guo, X.; Hu, Q.; Xiao, Y.; Cong, J.; Ma, L.; et al. Whole-genome sequencing reveals novel insights into sulfur oxidation in the extremophile *Acidithiobacillus thiooxidans*. *BMC Microbiol.* **2014**, *14*, 179. [[CrossRef](#)] [[PubMed](#)]
- Pallen, M.J.; Wren, B.W. Bacterial pathogenomics. *Nature* **2007**, *449*, 835–842. [[CrossRef](#)] [[PubMed](#)]
- Tettelin, H.; Riley, D.; Cattuto, C.; Medini, D. Comparative genomics: The bacterial pan-genome. *Curr. Opin. Microbiol.* **2008**, *11*, 472–477. [[CrossRef](#)] [[PubMed](#)]

13. Medini, D.; Donati, C.; Tettelin, H.; Massignani, V.; Rappuoli, R. The microbial pan-genome. *Curr. Opin. Genet. Dev.* **2005**, *15*, 589–594. [[CrossRef](#)] [[PubMed](#)]
14. Zhang, X.; Liu, X.; Liang, Y.; Fan, F.; Zhang, X.; Yin, H. Metabolic diversity and adaptive mechanisms of iron- and/or sulfur-oxidizing autotrophic acidophiles in extremely acidic environments. *Environ. Microbiol. Rep.* **2016**. [[CrossRef](#)] [[PubMed](#)]
15. Valdes, J.; Ossandon, F.; Quatrini, R.; Dopson, M.; Holmes, D.S. Draft genome sequence of the extremely acidophilic biomining bacterium *Acidithiobacillus thiooxidans* ATCC 19377 provides insights into the evolution of the *Acidithiobacillus* genus. *J. Bacteriol.* **2011**, *193*, 7003–7004. [[CrossRef](#)] [[PubMed](#)]
16. Yin, H.; Zhang, X.; Liang, Y.; Xiao, Y.; Niu, J.; Liu, X. Draft genome sequence of the extremophile *Acidithiobacillus thiooxidans* A01, isolated from the wastewater of a coal dump. *Genome Announc.* **2014**, *2*, e00222-14. [[CrossRef](#)] [[PubMed](#)]
17. Travisan, D.; Cortés, M.P.; Latorre, M.; Di Genova, A.; Budinich, M.; Bobadilla-Fazzini, R.; Parada, P.; González, M.; Maass, A. A new genome of *Acidithiobacillus thiooxidans* provides insights into adaptation to a bioleaching environment. *Res. Microbiol.* **2014**, *165*, 743–752. [[CrossRef](#)] [[PubMed](#)]
18. Nuñez, P.A.; Romero, H.; Farber, M.D.; Rocha, E.P.C. Natural selection for operons depends on genome size. *Genome Biol. Evol.* **2013**, *5*, 2242–2254. [[CrossRef](#)] [[PubMed](#)]
19. Levicán, G.; Ugalde, J.A.; Ehrenfeld, N.; Maass, A.; Parada, P. Comparative genomic analysis of carbon and nitrogen assimilation mechanisms in three indigenous bioleaching bacteria: Predictions and validations. *BMC Genom.* **2008**, *9*, 581. [[CrossRef](#)] [[PubMed](#)]
20. Douillard, F.P.; Ribbera, A.; Kant, R.; Pietilä, T.; Järvinen, H.; Messing, M.; Randazzo, C.; Paulin, L.; Laine, P.; Ritari, J.; et al. Comparative genomic and functional analysis of 100 *Lactobacillus rhamnosus* strains and their comparison with strain GG. *PLoS Genet.* **2013**, *9*, e1003683. [[CrossRef](#)] [[PubMed](#)]
21. Sugawara, M.; Epstein, B.; Badgley, B.D.; Unno, T.; Xu, L.; Reese, J.; Gyaneshwar, P.; Denny, R.; Mudge, J.; Bharti, A.; et al. Comparative genomics of the core and accessory genomes of 48 *Sinorhizobium* strains comprising five genospecies. *Genome Biol.* **2013**, *14*, R17. [[CrossRef](#)] [[PubMed](#)]
22. Kittichotirat, W.; Bumgarner, R.E.; Asikainen, S.; Chen, C. Identification of the pangenome and its components in 14 distinct *Aggregatibacter actinomycetemcomitans* strains by comparative genomic analysis. *PLoS ONE* **2011**, *6*, e22420. [[CrossRef](#)] [[PubMed](#)]
23. Baker-Austin, C.; Dopson, M. Life in acid: pH homeostasis in acidophiles. *Trends Microbiol.* **2007**, *15*, 165–171. [[CrossRef](#)] [[PubMed](#)]
24. Chen, L.X.; Li, J.T.; Chen, Y.T.; Huang, L.N.; Hua, Z.S.; Hu, M.; Shu, W.S. Shifts in microbial community composition and function in the acidification of a lead/zinc mine tailings. *Environ. Microbiol.* **2013**, *15*, 2431–2444. [[CrossRef](#)] [[PubMed](#)]
25. Silver, S.; Phung, L.T. Bacterial heavy metal resistance: New surprises. *Annu. Rev. Microbiol.* **1996**, *50*, 753–789. [[CrossRef](#)] [[PubMed](#)]
26. Ramos, P.I.P.; Picão, R.C.; de Almeida, L.G.P.; Lima, N.C.B.; Girardello, R.; Vivan, A.C.P.; Xavier, D.E.; Barcellos, F.G.; Pelisson, M.; Vespero, E.C.; et al. Comparative analysis of the complete genome of KPC-2-producing *Klebsiella pneumoniae* Kp13 reveals remarkable genome plasticity and a wide repertoire of virulence and resistance mechanisms. *BMC Genom.* **2014**, *15*, 54. [[CrossRef](#)] [[PubMed](#)]
27. Lapierre, P.; Gogarten, J.P. Estimating the size of the bacterial pan-genome. *Trends Genet.* **2009**, *25*, 107–110. [[CrossRef](#)] [[PubMed](#)]
28. Fondi, M.; Rizzi, E.; Emiliani, G.; Orlandini, V.; Berna, L.; Papaleo, M.C.; Perrin, E.; Maida, I.; Corti, G.; de Bellis, G.; et al. The genome sequence of the hydrocarbon-degrading *Acinetobacter venetianus* VE-C3. *Res. Microbiol.* **2013**, *164*, 439–449. [[CrossRef](#)] [[PubMed](#)]
29. Liu, W.; Wong, C.; Chung, K.M.; Jiang, J.; Leung, F.C. Comparative genome analysis of *Enterobacter cloacae*. *PLoS ONE* **2013**, *8*, e74487. [[CrossRef](#)] [[PubMed](#)]
30. Cannon, G.C.; Bradburne, C.E.; Aldrich, H.C.; Baker, S.H.; Heinhorst, S.; Shively, J.M. Microcompartments in prokaryotes: Carboxysomes and related polyhedra. *Appl. Environ. Microbiol.* **2001**, *67*, 5351–5361. [[CrossRef](#)] [[PubMed](#)]
31. Barreto, M.; Jedlicki, E.; Holmes, D.S. Identification of a gene cluster for the formation of extracellular polysaccharide precursors in the chemolithoautotroph *Acidithiobacillus ferrooxidans*. *Appl. Environ. Microbiol.* **2005**, *71*, 2902–2909. [[CrossRef](#)] [[PubMed](#)]

32. González, A.; Bellenberg, S.; Mamani, S.; Ruiz, L.; Echeverría, A.; Soulère, L.; Doutheau, A.; Demergasso, C.; Sand, W.; Queneau, Y.; et al. AHL signaling molecules with a large acyl chain enhance biofilm formation on sulfur and metal sulfides by the bioleaching bacterium *Acidithiobacillus ferrooxidans*. *Appl. Microbiol. Biotechnol.* **2013**, *97*, 3729–3737. [[CrossRef](#)] [[PubMed](#)]
33. Watling, H.R. The bioleaching of sulphide minerals with emphasis on copper sulphides—A review. *Hydrometallurgy* **2006**, *84*, 81–108. [[CrossRef](#)]
34. Valdés, J.; Pedroso, I.; Quatrini, R.; Dodson, R.J.; Tettelin, H.; Blake, R.; Eisen, J.A.; Holmes, D.S. *Acidithiobacillus ferrooxidans* metabolism: From genome sequence to industrial applications. *BMC Genom.* **2008**, *9*, 597. [[CrossRef](#)] [[PubMed](#)]
35. Bobadilla Fazzini, R.A.; Cortés, M.Z.; Padilla, L.; Maturana, D.; Budinich, M.; Maass, A.; Parada, P. Stoichiometric modeling of oxidation of reduced inorganic sulfur compounds (Riscs) in *Acidithiobacillus thiooxidans*. *Biotechnol. Bioeng.* **2013**, *110*, 2242–2251. [[CrossRef](#)] [[PubMed](#)]
36. Zhang, X.; Huaqun, Y.; Yili, L.; Qiu, G.; Liu, X. Theoretical model of the structure and the reaction mechanisms of sulfur oxygenase reductase in *Acidithiobacillus thiooxidans*. *Adv. Mater. Res.* **2015**, *1130*, 67–70. [[CrossRef](#)]
37. Krulwich, T.A.; Sachs, G.; Padan, E. Molecular aspects of bacterial pH sensing and homeostasis. *Nat. Rev. Microbiol.* **2011**, *9*, 330–343. [[CrossRef](#)] [[PubMed](#)]
38. Zhou, Z.; Fang, Y.; Li, Q.; Yin, H.; Qin, W.; Liang, Y.; Li, Q.; Li, N.; Liu, X.; Qiu, G.; et al. Global transcriptional analysis of stress-response strategies in *Acidithiobacillus ferrooxidans* ATCC 23270 exposed to organic extractant-Lix984n. *World J. Microbiol. Biotechnol.* **2012**, *28*, 1045–1055. [[CrossRef](#)] [[PubMed](#)]
39. Liu, J.; Qiu, G.; Ge, Y.; Xu, J. Extraction of copper from bacterial leach solution using Lix984. *Trans. Nonferrous Met. Soc. China* **2002**, *12*, 313–316.
40. Acuña, L.G.; Cárdenas, J.P.; Covarrubias, P.C.; Haristoy, J.J.; Flores, R.; Nuñez, H.; Riadi, G.; Shmaryahu, A.; Valdés, J.; Dopson, M.; et al. Architecture and gene repertoire of the flexible genome of the extreme acidophile *Acidithiobacillus caldus*. *PLoS ONE* **2013**, *8*, e78237.
41. Mira, A.; Ochman, H.; Moran, N.A. Deletional bias and the evolution of bacterial genomes. *Trends Genet.* **2001**, *17*, 589–596. [[CrossRef](#)]
42. Wu, X.; Monchy, S.; Taghavi, S.; Zhu, W.; Ramos, J.; van der Lelie, D. Comparative genomics and functional analysis of niche-specific adaptation in *Pseudomonas putida*. *FEMS Microbiol. Rev.* **2011**, *35*, 299–323. [[CrossRef](#)] [[PubMed](#)]
43. Patel, R.K.; Jain, M. NGS QC Toolkit: A toolkit for quality control of next generation sequencing data. *PLoS ONE* **2012**, *7*, e30619. [[CrossRef](#)] [[PubMed](#)]
44. Zerbino, D.R.; Birney, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **2008**, *18*, 821–829. [[CrossRef](#)] [[PubMed](#)]
45. Zheng, Y.; Zhao, L.; Gao, J.; Fei, Z. iAssembler: A package for de novo assembly of Roche-454/Sanger transcriptome sequences. *BMC Bioinform.* **2011**, *12*, 453. [[CrossRef](#)] [[PubMed](#)]
46. Parks, D.H.; Imelfort, M.; Skennerton, C.T.; Hugenholtz, P.; Tyson, G.W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **2015**, *25*, 1043–1055. [[CrossRef](#)] [[PubMed](#)]
47. Noguchi, H.; Taniguchi, T.; Itoh, T. MetaGeneAnnotator: Detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res.* **2008**, *15*, 387–396. [[CrossRef](#)] [[PubMed](#)]
48. Punta, M.; Coghill, P.C.; Lestrade, L.; Eberhardt, R.Y.; Mistry, J.; Tate, J.; Boursnell, C.; Pang, N.; Forslund, K.; Ceric, G.; et al. The Pfam protein families database. *Nucleic Acids Res.* **2012**, *40*, D290–D301. [[CrossRef](#)] [[PubMed](#)]
49. Lowe, T.M.; Eddy, S.R. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **1997**, *25*, 955–964. [[CrossRef](#)] [[PubMed](#)]
50. Siguier, P.; Perochon, J.; Lestrade, L.; Mahillon, J.; Chandler, M. ISfinder: The reference centre for bacterial insertion sequences. *Nucleic Acids Res.* **2006**, *34*, D32–D36. [[CrossRef](#)] [[PubMed](#)]
51. Dhillon, B.K.; Laird, M.R.; Shay, J.A.; Winsor, G.L.; Lo, R.; Nizam, F.; Pereira, S.K.; Waglechner, N.; McArthur, A.G.; Langille, M.G.; et al. IslandViewer 3: More flexible, interactive genomic island discovery, visualization and analysis. *Nucleic Acids Res.* **2015**, *43*, W104–W108. [[CrossRef](#)] [[PubMed](#)]

52. Aziz, R.K.; Bartels, D.; Best, A.A.; DeJongh, M.; Disz, T.; Edwards, R.A.; Formsma, K.; Gerdes, S.; Glass, E.M.; Kubal, M.; et al. The RAST Server: Rapid annotations using subsystems technology. *BMC Genom.* **2008**, *9*, 75. [[CrossRef](#)] [[PubMed](#)]
53. Fouts, D.E.; Brinkac, L.; Beck, E.; Inman, J.; Sutton, G. PanOCT: Automated clustering of orthologs using conserved gene neighborhood for pan-genomic analysis of bacterial strains and closely related species. *Nucleic Acids Res.* **2012**, *40*, e172. [[CrossRef](#)] [[PubMed](#)]
54. Zuo, G.; Hao, B. CVTree3 web server for whole-genome-based and alignment-free prokaryotic phylogeny and taxonomy. *Genom. Proteom. Bioinform.* **2015**, *13*, 321–331. [[CrossRef](#)] [[PubMed](#)]
55. Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **2011**, *28*, 2731–2739. [[CrossRef](#)] [[PubMed](#)]



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).