Liu Zhenghua (Orcid ID: 0000-0002-9857-2941)
Drewniak Lukasz (Orcid ID: 0000-0002-3236-0508)
Liu Shuang-Jiang (Orcid ID: 0000-0002-7585-310X)
Wang Jianjun (Orcid ID: 0000-0001-7039-7136)
Yin Huaqun (Orcid ID: 0000-0002-1215-001X)

Mobile genetic elements mediate the mixotrophic evolution of novel *Alicyclobacillus* species for acid mine drainage adaptation

Zhenghua Liu[1,2,6,#], Zonglin Liang[3,4,#], Zhicheng Zhou[7], Liangzhi Li[1,2], Delong Meng[1,2], Xiutong Li[3,4], Jiemeng Tao[1,2], Zhen Jiang[3,4], Yabing Gu[1,2], Ye Huang[3,4], Xueduan Liu[1,2], Zhendong Yang[8], Lukasz Drewniak[8], Tianbo Liu[9], Yongjun Liu[9], Shuangjiang Liu[3,4,5], Jianjun Wang[6*], Chengying Jiang[3,4,5*], Huaqun Yin[1,2*]

[1] School of Minerals Processing and Bioengineering, Central South University, Changsha 410006, China

[2] Key Laboratory of Biometallurgy of Ministry of Education, Central South University, Changsha 410006, China

[3] State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

[4] University of Chinese Academy of Sciences, Beijing 100049, China

[5] Innovation Academy for Green Manufacture, Chinese Academy of Sciences, Beijing 100101, China

[6] State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China

[7] College of Plant Protection, Hunan Agricultural University, Changsha 410010, China

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1462-2920.15543

[8] Department of Environmental Microbiology and Biotechnology, Institute of Microbiology, Faculty of Biology, University of Warsaw, Miecznikowa 1, 02-096 Warsaw, Poland

[9] Hunan Tobacco Science Institute, Changsha 410010, China

[#] These authors contributed equally

* Corresponding authors: Huaqun Yin, yinhuaqun_cs@sina.com; Chengying Jiang, jiangcy@im.ac.cn; Jianjun Wang, jjwang@niglas.ac.cn

Running title: Adaptive evolutions of *Alicyclobacillus* species

Key findings

1. Nine *Alicyclobacillus* strains isolated from acid mine drainage (AMD) have larger genome size and more genes than the strains from other environments, four of which evolve to be mixotrophic for the utilization of iron or reduced inorganic sulfur compounds.

2. The ancestor of *Alicyclobacillus* likely originates from soil, and no early separation events occur in their evolutionary process for adaptation to diverse environments.

3. Mobile genetic elements (MGEs) are the primary drivers in adaptive evolution of *Alicyclobacillus* species. *Alicyclobacillus* genomes show variations in the numbers of four different MGE types such as genomic island (GI), insertion sequence (IS), prophage, and integrative and conjugative element (ICE). In AMD strains, furthermore, there are enrichments in the functions of MGE-associated genes such as genes encoding

for sulfide-quinone reductase, cytochrome oxidase, heat/cold shock protein, inorganic ion transport, and heavy metal resistance proteins.

4. Compared to the other strains, AMD strains have relatively greater number of GI, IS, and ICE-type MGEs and the MGE-associated genes involved in the resistance of heavy metals and energy productions.

Summary

Alicyclobacillus species inhabit diverse environments and have adapted to broad ranges of pH and temperature. However, their adaptive evolutions remain elusive, especially regarding the role of mobile genetic elements (MGEs). Here, we characterized the distributions and functions of MGEs in *Alicyclobacillus* species across five environments, including acid mine drainage (AMD), beverages, hot springs, sediments, and soils. Nine *Alicyclobacillus* strains were isolated from AMD and possessed larger genome sizes and more genes than those from other environments. Four AMD strains evolved to be mixotrophic and fell into distinctive clusters in phylogenetic tree. Four types of MGEs including genomic island (GI), insertion sequence (IS), prophage, and integrative and conjugative element (ICE) were widely distributed in *Alicyclobacillus* species. Further, AMD strains did not possess CRISPR-Cas systems, but had more GI, IS, and ICE, as well as more MGE-associated genes involved in the oxidation of iron and sulfide and the resistance of heavy metal and low temperature. These findings highlight the differences in phenotypes and genotypes between strains isolated from AMD and other environments, and the important role of MGEs in rapid environment niche expansions.

Keywords: *Alicyclobacillus*; adaptive evolution; mobile genetic element; genomic island; insertion sequence; prophage; integrative and conjugative element

Introduction

Alicyclobacillus species have adapted to a diverse range of environmental niches and colonized many environments, including acid mine drainage (AMD), various beverages (e.g., fruit juice, herbal tea), hot springs, and various sediments/soils (e.g., orchard soil, forest soil, crop soil, and solfatara soil) (da Costa et al. 2015). Most *Alicyclobacillus* are chemoorganotrophic and can grow in acidic environments (pH range of 0.5-6.5) and over a wide temperature range (4-75 ºC) (da Costa et al. 2015). Their low pH and heat resistance may be due to the ω-cyclohexane fatty acids that account for 15-91% of the total fatty acid content in their cell membranes (Hippchen et al. 1981, Jensen 1999, Smit et al. 2011, Steyn et al. 2011). Over the last couple decades, the adaptive responses to low pH and high temperature of *Alicyclobacillus* species have been widely investigated in beverages such as fruit juice and herbal tea (Bahçeci and Acar 2007, Bai et al. 2012, Jiao et al. 2015) as some species lead to food spoilage and large economic losses in the food industry. In addition, within solfataric environments, *Alicyclobacillus* species can grow mixotrophically and have developed the ability to utilize Fe^{2+} and reduced inorganic compounds including S^0 and sulfide minerals (Jiang et al. 2008). *Alicyclobacillus* sp. A4, isolated from a hot spring, is able to produce several enzymes with stable activity over broad-range pH, from 2 to 12, such as xylanase (Bai et al. 2010) and α-amylase (Bai et al. 2012). In extremely low pH environments, such as AMD, *Alicyclobacillus* species are frequently detected, but have rarely been isolated and characterized, especially at the genomic level. These findings indicate that *Alicyclobacillus* species have adapted to multiple environments, but their evolutionary trajectory and the mechanism of adaptive evolution are still unclear.

The drivers of adaptive evolution primarily originate from two aspects, phenotypic and genomic plasticity. Compared with phenotypic plasticity, genomic plasticity plays a more important role in species differentiations and niche expansions by accumulating genome variations via genetic mutations and horizontal gene transfers (Brunder and Karch 2000). Adaptive evolution, driven by the accumulation of genetic mutations, is generally accompanied by early separation events, and experiencing long-term trade-offs during life history (Mattle-Greminger et al. 2018). By contrast, the adaptive evolution driven by horizontal gene transfer with mobile genetic elements (MGE) is rapid and plays a fundamental role in environmental niche expansions (Frost et al. 2005, Shintani 2017). There are four main types of mobile genetic elements: genomic islands (GI), insertion sequences (IS), prophages, and integrative and conjugative elements (ICE), all of which are abundant in bacterial genomes and contribute to many phenotypic traits required for virulence (Busby et al. 2013, Davies et al. 2016, Melnyk et al. 2019, Rezaei Javan et al. 2019), fitness (Stoebel et al. 2009, Gaffé et al. 2011, Stuart et al. 2013, Zeng et al. 2016, Cohen et al. 2020), and metabolism (Craig et al. 2009, Penn et al. 2009, Thiaville et al. 2016, Zamarro et al. 2016). Bacteria benefit greatly from the introductions of MGEs, however, the metabolic burdens of maintaining MGEs can be high. As such, bacteria need to balance the trade-off between optimal gene acquisitions for environmental adaptation and the energy or nutrients required to maintain those adaptations. Consequently, the fates of horizontally transferred genes are generally dependent on host lifestyle and characteristics of the local environment (Jaramillo et al. 2015, Fuchsman et al. 2017, Melnyk et al. 2019). However, the role of MGEs in shaping the genomes of microorganisms across different environments is still largely unknown, particularly for widely distributed species.

To explore the adaptive evolution mechanisms of *Alicyclobacillus*, we sequenced the genomes of nine *Alicyclobacillus* strains isolated from AMD and, together with the genomes of strains from beverages, hot springs, sediments, and soils, were further characterized to assess the roles played by four types of MGEs in the evolution of their genomes. We aimed to identify the differences in distribution and function of MGE-associated genes towards the five environments and reveal the roles of MGEs play in the adaptive evolution of *Alicyclobacillus*.

Results

Physiological characteristics of *Alicyclobacillus* species in AMD

The nine *Alicyclobacillus* strains were isolated from the AMD of a copper mine in Yunnan Province, China. Their basic information and physiological characteristics are given in Table 1. We found that their optimal growth temperature (30.8 ± 2.3 ℃) were lower than that of strains isolated from beverage (50.9 ± 4.5 ℃), hot spring (56.1 ± 8.2 ℃), and soil (48.6 ± 7.8 ℃), while the optimal pH of these strains was similar. According to phylogenetic analyses based on 16S rRNA with 97% similarity threshold, strains F1 and S30H17 CGMCC 1.17123 were identified as *Alicyclobacillus tolerans*, while the remaining seven strains were potentially novel species. Notably, we found that four strains, ALEF1 CGMCC 1.17055, S30H17 CGMCC 1.17123, H1 CGMCC 1.17290, and H5 CGMCC 1.17291 had the ability to oxidize sulfur, lowering the pH of the medium from 0.66 to 1.20 (Table 1). Two strains, ALEF1 CGMCC 1.17055 and S30H17 CGMCC 1.17123, were able to utilize sulfide ore and Fe^{2+}, displaying Fe^{2+} oxidation rates of 64.53 and 183.42 $mg \cdot L^{-1} \cdot d^{-1}$, respectively (Table 1).

Inferred evolutionary history of *Alicyclobacillus* species in adapting diversified environments

Nine *Alicyclobacillus* strains from AMD, presented here, and two strains from garden and solfataric soils provided by China General Microbiological Culture Collection Center (CGMCC) were sequenced. 22 additional genomes of *Alicyclobacillus* species from beverages, hot springs, sediments, and soils were also collected from NCBI public database for comparative genome analyses. The basic information for all 33 genomes is listed in Table 1. All *Alicyclobacillus* species genomes have similar G+C content (Figure 1A). The genome sizes, ranging from 2.65 to 6.57 Mb, were significantly (Kruskal test, $P < 0.05$) different across environments, with the largest was from AMD (5.67 ± 0.69 Mb), followed by soil (3.70 ± 0.65 Mb), beverage (3.26 ± 0.42 Mb), hot spring (3.01 ± 0.13 Mb) and sediment (2.65 Mb). Similar results were found in coding DNA sequences (CDS) across the five environments (Figure 1A). For RNA genes, the number of 16S rRNA, but not tRNA or tmRNA, was significantly affected by environments ($P < 0.05$, Figure 1A). A total of 128,872 genes were clustered in 32,535 homologous genes, of which 1.3% (437) belonged to the core-genome and 43.7% of gene clusters (14,213) were singletons (Figure 1B). Among them, the genomes of AMD strains harbored more genes and gene clusters as compared to the other strains. Based on the above genomic characteristics, the 33 strains were divided into three groups with all AMD strains clustering into a single group, while the remaining strains displayed no distinct clustering based on environments (Figure 1B).

To further identify whether there are environment-specific lineages in the *Alicyclobacillus* evolutionary histories, we constructed phylogenetic trees with 16S

rRNA sequences, core-genome, and the average nucleotide identity (ANI). The phylogenetic tree of 16S rRNA sequences showed that there were no early separation events for adapting to different environments and that *Alicyclobacillus* species first evolved into multiple clades about 15 million year ago (Figure 2). For AMD strains, there were three clusters which corresponded to their abilities to oxidize iron and reduced inorganic sulfur compounds. Compared with AMD strains, strains isolated from beverage, hot spring, and soil may have experienced fast adaptive evolution in the last two million years toward different environments (Figure 2). This finding was supported by the core-genome and ANI phylogenetic trees (Figure S1). These results highlight the differences in genome characteristics between AMD strains and other strains, revealing that *Alicyclobacillus* genomes are highly plastic in diversified environments.

Genomic islands in *Alicyclobacillus* genomes

Genomic islands are common in *Alicyclobacillus* genomes (Table S1), but the numbers of GI varied among species across different environments (Kruskal test, $P = 0.04$). On average, strains isolated from AMD possess the largest number of genomic islands (29.2 ± 7.8), followed by those from soil (23.6 ± 10.9), beverage (22.2 ± 6.1), hot spring (16.7 ± 5.2), and sediment (9). In total, the 760 GIs contained 15,866 genes, and these genes could not be clustered into gene families with more than two homologous genes (that is, ≥ 50% identity), indicating that these genes have low homology. Overall, 61% of island genes (9,678 genes) were functionally assigned to COG categories (Figure 3A), of which, replication, recombination and repair (L) accounted for the largest proportion (15.24%, 2418 genes), followed by transcription

(K; 6.72%, 1066 genes), cell wall/membrane/envelope biogenesis (M; 3.59%, 569 genes), and carbohydrate transport and metabolism (G; 3.22%, 511 genes).

Further, we found that twelve COG categories varied significantly depending on the bacterial environments (Figure 3B). Among them, five COG categories, including transcription (K), replication, recombination and repair (L), signal transduction mechanisms (T), inorganic iron transport and metabolism (P), and unknown function (S), were significantly enriched in AMD strains (LSD test, $P < 0.05$), while the category of cell mobility (N) was more enriched in hot spring strains ($P < 0.05$). Specifically, five AMD strains (ALC3 CGMCC 1.16736, ALEF1 CGMCC 1.17055, S30H17 CGMCC 1.17123, F1, and H1 CGMCC 1.17290) harbored *sqr* genes encoding sulfide-quinone oxidoreductase through genomic islands (Table S2). Moreover, almost all species also improved their high or low temperature resistance via acquisition of foreign genes encoding heat and cold shock proteins in genomic islands (Table S2). For high temperature resistance, AMD strains S30H17 CGMCC 1.17123 and F1 contained the HSBP1 gene encoding heat shock factor binding protein, while strains from beverage, hot spring, and soil (including DSM17974, NBRC 100859, NBRC 103103, CGMCC 1.10793, URH17-3-68, USBA-GBX-503 and ATCC 49025) harbored the *hsp26*/*hsp42* gene encoding a molecular chaperone of small heat-shock protein. For low temperature resistance, AMD strains ALC3 CGMCC 1.16736, S30H101 CGMCC 1.17180, S30H17 CGMCC 1.17123, F1, and H5 CGMCC 1.17291, beverage strain DSM14955, and soil strains DSM 17980 and SCH, harbored various genes encoding cold-shock proteins, such as *cspA*, *cspD*, *cspG* and *cspLA*. The *cspG* gene was present only in AMD strain S30H101 CGMCC 1.17180, while *cspD* gene was present only in beverage strain DSM14955.

Additionally, there were significantly (Kruskal test, $P < 0.05$) more genes related to heavy metal resistance in the AMD strains than other environments, and many of which were present in genomic islands, such as resistance genes related to As, Co, Cu, Ni, and Zn (Figure 4A), This phenomenon corresponds to the extremely high concentration of heavy metals in AMD environments. Specifically, eight genes associated with heavy metal resistance were present only in the AMD strains (Table S3), including *arsH* for As, *chrI* and *chrB1* for Cr, *corT*/*coaT* for Co, *tcrB* and *tcrY* for Cu, *perO* for Mo, and *smtB*/*ziaR* for Zn. In addition, the resistance of AMD strains to some rare heavy metals was significantly improved by GI-associated genes, including Ag, Au and Bi (Figures 4A and 4B).

Insertion sequences in *Alicyclobacillus* genomes

There were 2,410 insertion sequences (IS) among the *Alicyclobacillus* strains studied here, belonging to 20 families, and containing 647 genes, including 235 IS transposases. Within AMD strains, the number of IS-associated genes ranged from 20 to 131, while in all other strains it ranged from 0 to 16 (Table S4). Specifically, eight strains including beverage strain NBRC 102425, hot spring strains DSM 446, Tc-4-1, URH17-3-68, and USBA-GBX-503, sediment strain RIFOXYA1_FULL_53_8, and soil strains NBRC 100866 and TC-34, had no IS-associated genes, which indicates that IS within these strains prefers to locate within non-coding sequences. After division of IS families as based on IS genetic organizations, transposases similarities, and short invert repeats (Mahillon and Chandler 1998), only family IS110 was found to be significantly (Kruskal test, $P < 0.05$) associated with environments, and the AMD strains possessed the largest number of IS110 sequences (Figure 5A). Specifically, in

AMD strains, 11 of the 30 IS110 sequences contained genes encoding potassium-transporting ATPase, while others contained genes encoding cobyric acid synthase, adaptive-response sensory-kinase SasA, copper-exporting ATPase, and cadmium-transporting ATPase (Table S5). In the beverage strain NBRC 103104, an IS110 sequence contained a fructokinase, while in the hot spring strain CGMCC 1.10793 it was D-beta-hydroxybutyrate dehydrogenase (Table S5). Finally, in soil strains CPP55 and DSM 17980, the IS110 sequences contained monofunctional biosynthetic peptidoglycan transglycosylase and multifunctional non-homologous end joining DNA repair protein LigD, respectively (Table S5).

Furthermore, some IS families were present only in a few strains or a single environment. For example, the IS6 family was present only in ALC3 CGMCC 1.16736 and DSM 14955. Interestingly, the IS6 sequences carried three tRNAs into the genome of DSM 14955, including tRNA-Gly, tRNA-Lys, and tRNA-Ser (Figure 5B). Moreover, IS91 and a new IS family only were identified in the genomes of AMD strains. In the new IS family, a majority of the insertion sequences consistently carried a pair of genes that respectively encoded tyrosine recombinase XerC and XerD (Figure 5B).

Prophages, CRISPR-Cas systems, and anti-CRISPRs in *Alicyclobacillus* genomes

Prophages were widespread in the *Alicyclobacillus* genomes (Figure 6) and the potential number of prophages ranged from 1 to 8 per each genome, accounting for 0.14% to 5.12% of the genomes, and encompassing 0.21% to 5.11% of the genes (Table S6). Among the strains, those from soil contained more prophage regions and auxiliary metabolic genes than the others. This indicates that *Alicyclobacillus* strains in soil potentially interacted more frequently with phages than strains of other environments.

In COG functional assignment, the auxiliary metabolic genes of the soil strains were associated with amino acid transport and metabolisms (E) were higher than other strains (Figure S2). For AMD strains, COG functional categories of auxiliary metabolic genes were primarily associated with host physiology improvements such as cell cycle control (D), energy and carbohydrate metabolism (C) (Figure S2). In hot spring strains, defense system (V) encoded by auxiliary metabolic genes were increased as compared to other strains (Figure S2).

In the face of bacteriophage infections, bacteria develop numerous deference systems to acquire adaptive immunity and restriction-modification systems, among these the CRISPR-Cas system is one of the most important. As expected, *Alicyclobacillus* species possessed various CRISPR-Cas systems including types I-B, I-C, I-G, V-B1, and V-F1 (Figure 6). Specifically, soil strains had the most diverse range of CRISPR-Cas systems, including more than five types, beverage strains had three types (I-B, I-C and I-G), while hot spring strains had only type I-G (Figure 6). CRISPR-Cas systems were not detected in the genomes of AMD or sediment strains (Figure 6).

As a response to CRISPR-Cas systems, phages are sometimes equipped with immunity systems, such as anti-CRISPR systems. In *Alicyclobacillus* genomes, there were primarily three types, including I-B, II-A, and III-C (Figure 6). Among them, type I-B was only detected in three AMD strains, while III-C type was identified in all strains. These results revealed that *Alicyclobacillus* species engaged in a fierce competition with bacteriophages.

Integrative and conjugative elements in *Alicyclobacillus* genomes

Integrative and conjugative elements (ICE) were merely distributed in

Alicyclobacillus species genomes from four AMD and two soil strains and their length ranged from 153.7 to 464.5 kb (median = 193.6 kb), accounting for 2.74% to 8.24% of the genome and involving 2.92% to 8.24% of genes (Table S7). A total of 1,846 genes were integrated in ICE, of which 1,116 genes could be assigned to COG functional categories. Amino transport and metabolism (E) were the most abundant, followed by energy production and conversion (C) and transcription (L) (Figure S3). Notably, there were 27 ICE-associated genes encoding cytochrome oxidase in three AMD strains (Table S8), including S30H14 CGMCC 1.17050, F1 and H5 CGMCC 1.17291. It implies that ICE play an important role in adaptive evolutions of AMD strains for energy production. Additionally, four ICE-associated genes encoding cold shock protein were also found in two AMD strains of S30H17 CGMCC 1.17123 and F1 (Table S8). These findings highlight the importance of ICE in *Alicyclobacillus* evolution for adaptation to AMD oligotrophic and low-temperature environment.

The performances of MGEs in shaping genome

To quantitatively access the performance of MGEs in shaping the genomes, we calculated the ratio of the percentage of genes introduced by or involved with MGEs to the percentage of MGE lengths in genome. We found that IS had the most inefficient performance (0.32) in shaping *Alicyclobacillus* genomes, followed by ICE (1.04), GI (1.02) and prophage (0.98; Figure 7). The performances of both of GI and IS were significantly (Kruskal test, $P < 0.05$) different across four of the five environments, with sediments being the exception (Figure S4). Further, we found that approximately 7.79% of MGE-associated genes (Figure 8A) were shared by two or more of the four MGEs types but not for prophages and ICE. Specifically, the overlaps of MGE-associated

genes between GI and prophages were the largest (4.34%), while the overlaps between IS and ICE were the lowest (0.05%). Notably, the MGE-associated genes shared by GI and IS (1.04%) in AMD (Figure 8B) were significantly higher than that of other environments (all < 0.40%; $P < 0.05$; Figures 8C-8F). The potential co-operations between various MGEs for gene transformation or modification would be a boost to genome evolution.

Discussion

The adaptive evolutions of *Alicyclobacillus* species

Our study showed that rapid evolutions occurred for *Alicyclobacillus* species to adapt to diverse environments including AMD, beverage, hot spring, sediments, and soils. In evolutionary history, *Alicyclobacillus* species evolved into clear lineages for AMD strains, while other strains from beverage, hot spring, sediment and soil did not. In general, environmental features have great influence on species evolution rate and direction, and species can differentiate into specific lineages at early stage when mediated by an environment-dependent lifestyle (Shikano et al. 2010, Sorenson et al. 2014). Compared with other environments, AMD typically is oligotrophic, with very low pH, and high concentrations of heavy metals. Consequently, the common ancestor of *Alicyclobacillus* was more likely to have an early separation event for AMD environment adaptation. Moreover, with an increasing number of environment stresses, such as pH, temperature, and heavy metal concentration, the abundance of the *Alicyclobacillus* genus decreases across natural soil (21.1%~62.3%) (Yang et al. 2014, Yang et al. 2016), hot spring (1.9%~30.3%) (Paul et al. 2016, Rincón-Molina et al. 2018), and AMD (0.61%~15%) (Brantner et al. 2014, Chen et al. 2014). These

ecological investigations revealed that *Alicyclobacillus* species were confronted with more intense survival pressures in AMD than other environments. These clues indicate that the common ancestor of *Alicyclobacillus* species have highly plasticity for extreme environment adaptations.

The roles of MGEs in *Alicyclobacillus* adaptive evolution across environment stresses

To rapidly acquire new traits or resistances for environmental adaptation, horizontal gene transfer mediated by MGEs is the most important mechanism in genome evolution (Frost et al. 2005). However, the rapid expansion of genome size through MGEs, would enhance metabolic burden and energy consumption (Bottery et al. 2017). Among the five environments of *Alicyclobacillus* in this study, AMD environment is the harshest in which to survive and is where *Alicyclobacillus* species may rely more on GI to gain novel phenotype traits than in other environments. Under this oligotrophic environment, we found that some strains evolved to be chemoautotrophic through acquisition of GI-associated genes encoding cytochromes or sulfocyanin and sulfide quinone oxidoreductase for oxidizing Fe^{2+} and sulfide (Table S2), respectively, which are necessary energy compensation mechanisms for *Alicyclobacillus* to adapt to the AMD environment. Compared with GI, the more compact elements including IS, prophage, and ICE play more flexible roles in genome evolution.

First, each IS element averagely carried 0.47 gene in AMD strains, which was far higher than in other strains (beverage: 0.12; hot spring: 0.05; sediment: 0; soil:0.09). This indicated that ISs of AMD strains preferred to be in coding sequences rather than

non-coding or regulatory sequences. This result reveals that IS played a more important role in genetic mutation for AMD strains than gene expression. Moreover, in some AMD strains, the presence of IS-associated genes encoding site-specific tyrosine recombinase XerC or XerD, that catalyze the cutting and rejoining reactions of DNA molecules, indicated that IS may further increase the rate of genome evolution. Previous studies have reported that the preferences of IS in coding or non-coding sequences varies with different environmental stressors. For example, in *Escherichia coli*, IS10 elements would rewire expressions of the systems responding to high osmotic stress by integrating into the promoter of the *otsBA* operon (Stoebel et al. 2009). However, in face of the stress from glucose and phosphate limitation, IS1 elements would insert in the *rpoS* gene which encodes the master regulator of stress conditions (Gaffé et al. 2011).

Prophages offer numerous benefits to hosts, including the improved physiology (Obeng et al. 2016), which however depends on the interplay between host and environment. Our results show that the phage regions were widely distributed in all *Alicyclobacillus* strains while the functional compositions of prophage auxiliary metabolism genes were different across AMD, soil, and hot spring. Notably, the phage-associated genes in the genomes of some AMD strains involved a greater number of genes associated with energy production and conversion (E) and carbohydrate transport and metabolism (G), such as those genes encoding ferredoxin-NADP reductase and cellulase, which are beneficial to *Alicyclobacillus* survival in oligotrophic environment of AMD. While in oligotrophic environment of hot springs, prophage-associated genes were associated with nucleotide transport and metabolism, such as nucleoside 2-deoxyribosyltransferase, adenylosuccinate lyase and phosphoribosylformylglycinamidine synthase. These genes may protect the activities

of host nucleotide metabolism from high temperature environment. Similarly, selective recruitments of prophage-associated genes across different environments have also been observed in cyanobacteria. For instance, marine strains more frequently acquire photosynthetic genes (e.g., *psbA* and *psbD*) from cyanophages (Sullivan et al. 2006, Chenard and Suttle 2008), while freshwater strains preferred genes associated with metabolic processes, such as phycobilisome degradation (Gao et al. 2012) and ribonucleotide conversion (Nakamura et al. 2014).

ICE are widespread mobile DNA that transmit both vertically and horizontally. ICE can be found in two states: (i) an integrative state where they integrate into the host genome and (ii) a conjugative state where ICE are excised from host genome and able to conjugate to a new host (Delavat et al. 2017). The enrichment of ICE in AMD strains may result from the balance between fitness cost imposed by their horizontal and vertical transmissions, as the oligotrophic condition of AMD environments would hinder high energy-consuming horizontal transmissions (Delavat et al. 2017). Notably, we found that the majority ICE in AMD strains had poor functional characterization, however, some may associate with the transmissibility of other MGEs, such as genomic islands. Just like ICESpuPO1 in marine *Pseudoalteromonas* can mediate the dissemination and loss of a biofilm-related genomic island (Wang et al. 2017).

Interactions between *Alicyclobacillus* species and phages

We found that, in contrast to the other strains, none of the AMD strains contained CRISPR-Cas systems. The CRISPR-Cas systems are bacterial immune systems for interfering with the invasion of MGEs including phages genomes, plasmids and transposons (Brouns et al. 2008, Faure et al. 2019). Lower immune barriers to phage

infection facilitate the participation of temperate phages in genome evolution, but it would consequently make the organism more vulnerable to lytic phage attack. Similarly, *Staphylococcus epidermidis* were found to stochastically lose CRISPR-Cas systems to acquire novel genetic material (Jiang et al. 2013). The group of *Bacillus cereus* is reported to selectively inactivate CRISPR-Cas systems during the evolution for adaptation to diverse environments (Zheng et al. 2020). Specifically, at the individual level, host cells could selectively inactivate the CRISPR-Cas systems to allow phage enter and take a gamble with their life (Obeng et al. 2016). Host cells make a benefit by taking up temperate phages and acquiring helpful genetic materials, while as a risk, are easily infected and killed by lytic phages. However, at the population level, heterogeneous host cell population allows novel genetic materials input from vulnerable cells via phage infections (Obeng et al. 2016), and the genetic exchange could accelerate novel gene spreads within the population (Lang et al. 2012). Such a selectively inactivation of CRISPR-Cas systems is effective to improve species evolution and avoid to be completely vulnerable to lytic phage attack. Thus, for AMD strains, this gambling behavior implies that the benefits from temperate phages outweigh the defense cost for lytic phages.

In face of the CRISPR-Cas systems in host, anti-CRISPR system in phage is one of the most important counter immunity systems. Our results showed that the type II-C anti-CRISPR system was found in all *Alicyclobacillus* strains and type II-A was also identified in the majority of strains. However, both of the corresponding subtypes of type II CRISPR-Cas systems were not found in all strains. It is curious that *Alicyclobacillus* species harbor these seemingly unnecessary subtype II anti-CRISPR systems. It has been reported that the type II CRISPR-Cas system is widely encoded by

members of the Firmicutes, including *Alicyclobacillus* (Bernheim et al. 2017). However, due to the inhibitions by anti-CRISPR and non-destruction of phages, the type II CRISPR-Cas systems in the host may be lost during long-term adaptive evolution, but the anti-CRISPR system remained. This hypothesis could be partly revealed by the type I-B anti-CRISPR in *Alicyclobacillus* evolutionary trajectory. We found that type I-B CRISPR-Cas systems were mainly present in soil strains, and the type I-B anti-CRISPR were only found in some AMD strains without corresponding CRISPR-Cas systems. This implies that target phages of the type I-B CRISPR-Cas systems interacted with *Alicyclobacillus* species in various way across different environments. Therefore, AMD strains may benefit more from temperate phages than soil strains in adaptive evolution and the type I-B CRISPR-Cas systems were removed from the host genome to avoid unnecessary conflict.

The roles of MGE on the high or low temperature resistance

Our results showed that the genes encoding heat shock proteins were present in all strains, but merely involved with one type MGE of genomic islands. Among them, nine strains from AMD, beverage, hot spring and soil, acquired the genes encoding heat shock protein from genomic islands. Similar observations were made in *Pseudomonas aeruginosa*, which is widely distributed in the natural and man-made aquatic environments and in clinical environments (Lee et al. 2015). Particularly, the thermotolerance genomic island can specifically carry a series of genes mediated heat resistance, and contribute to enhance thermal tolerance of *Cronobacter* strains (Orieskova et al. 2016). These findings imply that genomic islands are important and popular in species evolutions for adaptation to high temperature. Among the four types

of MGEs examined in this study, genes encoding cold shock proteins were only observed in the genomic islands of 8 strains isolated from AMD (5) and soil (3), and ICE of 2 AMD strains. Cold shock domain family proteins are involved in the regulation of multiple phenotypes, including virulence (Michaux et al. 2017, Liu et al. 2020), cellular aggregations (Eshwar et al. 2017), and biofilm (Ray et al. 2020). Consequently, genomic islands possessing such genes could not only enhance species low temperature resistance at the individual level, but also at the population-level through cellular aggregation or biofilm formation that provide micro-environments which would be important for *Alicyclobacillus* species in adaptation to the harsh AMD environment by creating habitable microenvironments.

To summarize, our study has revealed that mobile genetic elements play important roles in adaptive evolutions of *Alicyclobacillus* species. Mobile genetic elements can help *Alicyclobacillus* species evolved to be chemoautotrophilies and adapted harsh acid mine drainage through increasing heavy metal resistances. Although these results provide further insights into the possible mechanisms that *Alicyclobacillus* rapidly broaden environment niches, additional evidences from manipulative experiments are needed to support the inferences in this study.

Experimental Procedures

Isolations and sequencing of *Alicyclobacillus* strain from acid mine drainage

Nine *Alicyclobacillus* strains were isolated from sediment or water samples of acid mine drainage from a copper mine in Yunnan Province, China (23° 28′ 53″ N, 103° 46′ 47″ E). The pH of samples ranged from 2.3 to 2.6 and temperature ranged from 25 to 35 ℃.The samples were enriched for 7 days at 30 °C in liquid medium, which consisted

of basal salt solution (BSM), 0.5 ml·L^{-1} trace elements solution, 0.8 g·L^{-1} yeast extract and 0.05 mmol·L^{-1} $FeSO_4·7H_2O$. The compositions of BSM were L^{-1}: $(NH_4)_2SO_4$ 3.0 g, $MgSO_4·7H_2O$ 0.5 g, $Na_2SO_4·10H_2O$ 0.15 g, KH_2PO_4 0.1 g, KCl 0.1 g, $Ca(NO_3)_2·4H_2O$ 0.014 g. After, samples of the grown culture were streaked on medium solidified by supplementing with 1.0 g·L^{-1} gelrite (Sigma-Aldrich). The purified strains were incubated at 30 °C. The utilizations of Fe^{2+}, sulfur, and sulfide ore were determined by culturing the cells for 5 days in BSM supplemented with 0.2 g·L^{-1} yeast extract and 10 mM $FeSO_4$, 5 g·L^{-1} S^0 or 5 g·L^{-1} FeS_2, respectively. Oxidation of Fe^{2+} and FeS_2 was visually assessed from the color change of broth to a yellow or reddish-brown color, indicating the oxidation of ferrous-iron to ferric-iron. The oxidation was also quantified by the oxidation rate of Fe in broth (Jiang et al., 2008), and Fe^{2+} concentration was measured by the 1,10-phenanthroline spectrophotometry assay (Hallberg et al., 2010). Oxidation of sulfur was determined by detecting the decrease of pH compared to culture supplemented only with yeast extract. The genomic DNA of the *Alicyclobacillus* strains were extracted by Wizard Genomic DNA Purification Kit (Promega). After checking the quality, DNA was fragmented, end-repaired and polyadenylated, ligated to a sequencing adapter and library constructed, then sequenced by Illumina HiSeq PE150 and PacBio Sequel (PacBio) sequence platforms, according to the protocols of Guangdong MAGIGENE Biotechnology Company (http://www.magigene.com/).

Genome annotation and pan-genome analysis

Genome completeness was calculated by BUSCO v3 (Simao et al. 2015). Coding sequences (CDS) and RNAs including (e.g. 16S rRNA, tRNA and tmRNA) were identified by prokka v1.13.4 (Seemann 2014) and the function assignments of CDS

were carried out by eggNOG v5.0 (Huerta-Cepas et al. 2017) with DIAMOND method (Buchfink et al. 2015). Pan-genome analyses were performed by the function of 'anvi-pan-genome' in Anvi'o v. 5.5.0 platform (Eren et al. 2015) with BLAST 2.5.0+ (E-value < 1e-5) (Altschul et al. 1990) and Markov CLuster algorithm (MCL) (Van Dongen and Abreu-Goodger 2012), and the results were visualized by the function of 'anvi-display-pan'. The genes associated with heavy metal resistance were identified by DIAMOND against BacMet v2.0 (Pal et al. 2014). Sequence data in this study are available in Bio-Med Big Data Center (https://www.biosino.org) according to accession number in Table 1.

Phylogenetic reconstruction

The 16S rRNA sequences of *Alicyclobacillus* species were aligned using ClustalW2 (Thompson et al. 2003). The alignment files were used to construct a phylogenetic tree by Maximum Likelihood algorithms with 100 bootstraps in MEGAX software (Kumar et al. 2018). The strain of *Bacillus subtilis subsp. subtilis* NCIB 3610 was included as the outgroup for the phylogenetic tree. To assess the divergence time for *Alicyclobacillus* species, four species pairs including *A. shizuokensis-A. kakegawensis*, *A. vulcanalis-A. sendaiensis*, *A. macrosporangiidus-A. acidiphilus* and *A. pomorum-A. hesperidum*, were as constraint points in phylogenetic tree. The divergence time of these four constraint points were collected from Timetree reference (Kumar et al. 2017). The concatenated sequences of core-genome consisted of 437 core genes also were used to construct Maximum Likelihood tree in MEGAX. Average nucleotide identity (ANI) based on whole genome was calculated by pyani v0.2.8 (Pritchard et al. 2019).

Detections of mobile genetic elements

Genomic islands were identified with IslandViewer 4 (Bertelli et al. 2017) by using one of four methods including IslandPath-DIMOB v1.0.0 (Bertelli and Brinkman 2018), SIGI-HMM (Waack et al. 2006), IslandPick (Langille et al. 2008), and Islander (Hudson et al. 2015). Insertion sequences were identified by ISEScan v. 1.7.1 (Xie and Tang 2017) across the whole genomes. The prophage regions were detected by PHASTER v4.3X (Arndt et al. 2016). CRISPR-Cas systems were searched for using the web tool CRISPRFinder (Grissa et al. 2007) against CRISPRCasdb database (Pourcel et al. 2020). The anti-CRISPR system were searched by BLAST v2.10.1 with E-value cut-off of 1e-5 (Altschul et al. 1990) against antiCRISPRdb database (Dong et al. 2018). The integrative and conjugative elements in genome were identified by ICEFinder (https://db-mml.sjtu.edu.cn/ICEfinder/ICEfinder.html) against ICEberg 2.0 database (Liu et al. 2019).

Statistical analyses

Analysis of variance (ANOVA) was carried out to identify the function genes differentiated among environments. The relationships between the proportions of MGE length and the MGE-associated genes accounting for genome size and all genes were estimated by linear model (Chambers and Hastie 1992). The overlaps of MGE-associated genes among four types of MGEs were calculated by using “VennDiagram” package (Chen and Boutros 2011). The differences in the number of COG genes across environments were detected by LSD multiple comparisons and *P* values were adjusted by Bonferroni method (d Steel and Torrie 1986). Data visualizations were conducted within “ggplot2” package (Wickham 2016). Above analyses were performed in R v3.6.1 environment (Team 2000).

Author Contributions

CY Jiang and HQ Yin designed and coordinated the study. JJ Wang and XD Liu guided the genomic analysis. ZH Liu, LZ Li, DL Meng, JM Tao, and YB Gu performed the genomic analysis. ZL Liang, XT Li, Z Jiang, and Y Huang isolated the strains and carried out the physiological characteristic analysis and DNA extraction. ZH Liu wrote original draft manuscript. JJ Wang, CY jiang, HQ Yin, ZD Yang, L Drewniak, TB Liu, YJ Liu, SJ Liu and ZC Zhou reviewed and edited the manuscript.

Acknowledgements

This work was supported by National Nature Science Foundation of China (91851206, 41877345, 31670124), the Key Research Program of Chinese Academy of Sciences (ZDRW-ZS-2018-1) and the Joint Funds of Innovation Academy for Green Manufacture, Chinese Academy of Sciences (IAGM2020C24).

Conflict of Interest

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

References

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. Journal of molecular biology **215**:403-410.

Arndt, D., J. R. Grant, A. Marcu, T. Sajed, A. Pon, Y. Liang, and D. S. Wishart. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Research **44**:W16-W21.

Bahçeci, K. S., and J. Acar. 2007. Modeling the combined effects of pH, temperature and ascorbic acid concentration on the heat resistance of Alicyclobacillus acidoterrestis. International journal of food microbiology **120**:266-273.

Bai, Y., H. Huang, K. Meng, P. Shi, P. Yang, H. Luo, C. Luo, Y. Feng, W. Zhang, and B. Yao. 2012. Identification of an acidic α-amylase from Alicyclobacillus sp. A4 and assessment of its application in the starch industry. Food Chemistry **131**:1473-1478.

Bai, Y., J. Wang, Z. Zhang, P. Yang, P. Shi, H. Luo, K. Meng, H. Huang, and B. Yao. 2010. A new xylanase from thermoacidophilic Alicyclobacillus sp. A4 with broad-range pH activity and pH stability. J Ind Microbiol Biotechnol **37**:187-194.

Bernheim, A., A. Calvo-Villamanan, C. Basier, L. Cui, E. P. C. Rocha, M. Touchon, and D. Bikard. 2017. Inhibition of NHEJ repair by type II-A CRISPR-Cas systems in bacteria. Nat Commun **8**:2094.

Bertelli, C., and F. S. Brinkman. 2018. Improved genomic island predictions with IslandPath-DIMOB. Bioinformatics **34**:2161-2167.

Bertelli, C., M. R. Laird, K. P. Williams, S. F. U. R. C. Group, B. Y. Lau, G. Hoad, G. L. Winsor, and F. S. Brinkman. 2017. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Research **45**:W30-W35.

Bottery, M. J., A. J. Wood, and M. A. Brockhurst. 2017. Adaptive modulation of antibiotic resistance through intragenomic coevolution. Nature ecology & evolution **1**:1364-1369.

Brantner, J. S., Z. J. Haake, J. E. Burwick, C. M. Menge, S. T. Hotchkiss, and J. M. Senko. 2014. Depth-dependent geochemical and microbiological gradients in Fe(III) deposits resulting from coal mine-derived acid mine drainage. Front Microbiol **5**:215.

Brouns, S. J., M. M. Jore, M. Lundgren, E. R. Westra, R. J. Slijkhuis, A. P. Snijders, M. J. Dickman, K. S. Makarova, E. V. Koonin, and J. Van Der Oost. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science **321**:960-964.

Brunder, W., and H. Karch. 2000. Genome plasticity in Enterobacteriaceae. International journal of medical microbiology **290**:153-165.

Buchfink, B., C. Xie, and D. H. Huson. 2015. Fast and sensitive protein alignment using DIAMOND. Nature methods **12**:59-60.

Busby, B., D. M. Kristensen, and E. V. Koonin. 2013. Contribution of phage-derived genomic islands to the virulence of facultative bacterial pathogens. Environ Microbiol **15**:307-312.

Chambers, J., and T. Hastie. 1992. Linear models. Chapter 4 of statistical models in S. Wadsworth & Brooks/Cole.

Chen, H., and P. C. Boutros. 2011. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC bioinformatics **12**:1-

7.
Chen, Y. T., J. T. Li, L. X. Chen, Z. S. Hua, L. N. Huang, J. Liu, B. B. Xu, B. Liao, and W. S. Shu. 2014. Biogeochemical processes governing natural pyrite oxidation and release of acid metalliferous drainage. Environ Sci Technol **48**:5537-5545.
Chenard, C., and C. A. Suttle. 2008. Phylogenetic diversity of sequences of cyanophage photosynthetic gene psbA in marine and freshwaters. Appl Environ Microbiol **74**:5317-5324.
Cohen, E., M. Davidovich, A. Rokney, L. Valinsky, G. Rahav, and O. Gal-Mor. 2020. Emergence of new variants of antibiotic resistance genomic islands among multidrug-resistant Salmonella enterica in poultry. Environ Microbiol **22**:413-432.
Craig, J. P., S. Bekal, T. Niblack, L. Domier, and K. N. Lambert. 2009. Evidence for horizontally transferred genes involved in the biosynthesis of vitamin B1, B5, and B7 in Heterodera glycines. Journal of nematology **41**:281.
d Steel, R. G., and J. H. Torrie. 1986. Principles and procedures of statistics: a biometrical approach. McGraw-Hill.
da Costa, M. S., F. A. Rainey, and L. Albuquerque. 2015. Alicyclobacillus.1-18.
Davies, E. V., C. E. James, D. Williams, S. O'Brien, J. L. Fothergill, S. Haldenby, S. Paterson, C. Winstanley, and M. A. Brockhurst. 2016. Temperate phages both mediate and drive adaptive evolution in pathogen biofilms. Proceedings of the National Academy of Sciences **113**:8266-8271.
Delavat, F., R. Miyazaki, N. Carraro, N. Pradervand, and J. R. van der Meer. 2017. The hidden life of integrative and conjugative elements. FEMS Microbiol Rev **41**:512-537.
Dong, C., G. F. Hao, H. L. Hua, S. Liu, A. A. Labena, G. Chai, J. Huang, N. Rao, and F. B. Guo. 2018. Anti-CRISPRdb: a comprehensive online resource for anti-CRISPR proteins. Nucleic Acids Res **46**:D393-D398.
Eren, A. M., Ö. C. Esen, C. Quince, J. H. Vineis, H. G. Morrison, M. L. Sogin, and T. O. Delmont. 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ **3**:e1319.
Eshwar, A. K., C. Guldimann, A. Oevermann, and T. Tasara. 2017. Cold-Shock Domain Family Proteins (Csps) Are Involved in Regulation of Virulence, Cellular Aggregation, and Flagella-Based Motility in Listeria monocytogenes. Front Cell Infect Microbiol **7**:453.
Faure, G., S. A. Shmakov, W. X. Yan, D. R. Cheng, D. A. Scott, J. E. Peters, K. S. Makarova, and E. V. Koonin. 2019. CRISPR-Cas in mobile genetic elements: counter-defence and beyond. Nat Rev Microbiol **17**:513-525.
Frost, L. S., R. Leplae, A. O. Summers, and A. Toussaint. 2005. Mobile genetic elements: the agents of open source evolution. Nature Reviews Microbiology **3**:722-732.
Fuchsman, C. A., R. E. Collins, G. Rocap, and W. J. Brazelton. 2017. Effect of the environment on horizontal gene transfer between bacteria and archaea. PeerJ **5**:e3865.
Gaffé, J., C. McKenzie, R. P. Maharjan, E. Coursange, T. Ferenci, and D. Schneider. 2011. Insertion sequence-driven evolution of Escherichia coli in chemostats. Journal of molecular evolution **72**:398-412.
Gao, E. B., J. F. Gui, and Q. Y. Zhang. 2012. A novel cyanophage with a cyanobacterial nonbleaching protein A gene in the genome. J Virol **86**:236-245.
Grissa, I., G. Vergnaud, and C. Pourcel. 2007. CRISPRFinder: a web tool to identify

clustered regularly interspaced short palindromic repeats. Nucleic Acids Res **35**:W52-57.

Hippchen, B., A. Röll, and K. Poralla. 1981. Occurrence in soil of thermo-acidophilic bacilli possessing ω-cyclohexane fatty acids and hopanoids. Archives of Microbiology **129**:53-55.

Hudson, C. M., B. Y. Lau, and K. P. Williams. 2015. Islander: a database of precisely mapped genomic islands in tRNA and tmRNA genes. Nucleic Acids Research **43**:D48-D53.

Huerta-Cepas, J., K. Forslund, L. P. Coelho, D. Szklarczyk, L. J. Jensen, C. von Mering, and P. Bork. 2017. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. Molecular Biology and Evolution **34**:2115-2122.

Jaramillo, V. D. A., S. A. Sukno, and M. R. Thon. 2015. Identification of horizontally transferred genes in the genus Colletotrichum reveals a steady tempo of bacterial to fungal gene transfer. BMC Genomics **16**:2.

Jensen, N. 1999. Alicyclobacillus: a new challenge for the food industry. Food Australia **51**:33-36.

Jiang, C. Y., Y. Liu, Y. Y. Liu, X. Y. You, X. Guo, and S. J. Liu. 2008. Alicyclobacillus ferrooxydans sp. nov., a ferrous-oxidizing bacterium from solfataric soil. Int J Syst Evol Microbiol **58**:2898-2903.

Jiang, W., I. Maniv, F. Arain, Y. Wang, B. R. Levin, and L. A. Marraffini. 2013. Dealing with the evolutionary downside of CRISPR immunity: bacteria and beneficial plasmids. PLoS Genet **9**:e1003844.

Jiao, L., J. Ran, X. Xu, and J. Wang. 2015. Heat, acid and cold stresses enhance the expression of DnaK gene in Alicyclobacillus acidoterrestris. Food Research International **67**:183-192.

Kumar, S., G. Stecher, M. Li, C. Knyaz, and K. Tamura. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution **35**:1547-1549.

Kumar, S., G. Stecher, M. Suleski, and S. B. Hedges. 2017. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. Mol Biol Evol **34**:1812-1819.

Lang, A. S., O. Zhaxybayeva, and J. T. Beatty. 2012. Gene transfer agents: phage-like elements of genetic exchange. Nat Rev Microbiol **10**:472-482.

Langille, M. G., W. W. Hsiao, and F. S. Brinkman. 2008. Evaluation of genomic island predictors using a comparative genomics approach. BMC bioinformatics **9**:329-329.

Lee, C., E. Wigren, J. Trcek, V. Peters, J. Kim, M. S. Hasni, M. Nimtz, Y. Lindqvist, C. Park, U. Curth, H. Lunsdorf, and U. Romling. 2015. A novel protein quality control mechanism contributes to heat shock resistance of worldwide-distributed Pseudomonas aeruginosa clone C strains. Environ Microbiol **17**:4511-4526.

Liu, M., X. Li, Y. Xie, D. Bi, J. Sun, J. Li, C. Tai, Z. Deng, and H. Y. Ou. 2019. ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. Nucleic Acids Res **47**:D660-D665.

Liu, Y., X. Tan, H. Cheng, J. Gong, Y. Zhang, D. Wang, and W. Ding. 2020. The cold shock family gene cspD3 is involved in the pathogenicity of Ralstonia solanacearum CQPS-1 to tobacco. Microb Pathog **142**:104091.

Mahillon, J., and M. Chandler. 1998. Insertion sequences. Microbiology and molecular biology reviews **62**:725-774.

Mattle-Greminger, M. P., T. Bilgin Sonay, A. Nater, M. Pybus, T. Desai, G. de Valles,

F. Casals, A. Scally, J. Bertranpetit, T. Marques-Bonet, C. P. van Schaik, M. Anisimova, and M. Krutzen. 2018. Genomes reveal marked differences in the adaptive evolution between orangutan species. Genome Biol **19**:193.

Melnyk, R. A., S. S. Hossain, and C. H. Haney. 2019. Convergent gain and loss of genomic islands drive lifestyle changes in plant-associated Pseudomonas. ISME J **13**:1575-1588.

Michaux, C., E. Holmqvist, E. Vasicek, M. Sharan, L. Barquist, A. J. Westermann, J. S. Gunn, and J. Vogel. 2017. RNA target profiles direct the discovery of virulence functions for the cold-shock proteins CspC and CspE. Proceedings of the National Academy of Sciences **114**:6824-6829.

Nakamura, G., S. Kimura, Y. Sako, and T. Yoshida. 2014. Genetic diversity of Microcystis cyanophages in two different freshwater environments. Arch Microbiol **196**:401-409.

Obeng, N., A. A. Pratama, and J. D. V. Elsas. 2016. The Significance of Mutualistic Phages for Bacterial Ecology and Evolution. Trends Microbiol **24**:440-449.

Orieskova, M., M. Kajsik, T. Szemes, O. Holy, S. Forsythe, J. Turna, and H. Drahovska. 2016. Contribution of the thermotolerance genomic island to increased thermal tolerance in Cronobacter strains. Antonie van Leeuwenhoek **109**:405-414.

Pal, C., J. Bengtsson-Palme, C. Rensing, E. Kristiansson, and D. G. Larsson. 2014. BacMet: antibacterial biocide and metal resistance genes database. Nucleic Acids Res **42**:D737-743.

Paul, S., Y. Cortez, N. Vera, G. Villena, and M. Gutiérrez-Correa. 2016. Metagenomic Analysis of Microbial Communities in the Soil-mousse Surrounding of an Amazonian Geothermal Spring in Peru. British Biotechnology Journal **15**:1-11.

Penn, K., C. Jenkins, M. Nett, D. W. Udwary, E. A. Gontang, R. P. McGlinchey, B. Foster, A. Lapidus, S. Podell, E. E. Allen, B. S. Moore, and P. R. Jensen. 2009. Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria. ISME J **3**:1193-1203.

Pourcel, C., M. Touchon, N. Villeriot, J. P. Vernadet, D. Couvin, C. Toffano-Nioche, and G. Vergnaud. 2020. CRISPRCasdb a successor of CRISPRdb containing CRISPR arrays and cas genes from complete genome sequences, and tools to download and query lists of repeats and spacers. Nucleic Acids Res **48**:D535-D544.

Pritchard, L., P. Cock, and Ö. Esen. 2019. pyani v0. 2.8: average nucleotide identity (ANI) and related measures for whole genome comparisons.

Ray, S., R. Da Costa, S. Thakur, and D. Nandi. 2020. Salmonella Typhimurium encoded cold shock protein E is essential for motility and biofilm formation. Microbiology **166**:460-473.

Rezaei Javan, R., E. Ramos-Sevillano, A. Akter, J. Brown, and A. B. Brueggemann. 2019. Prophages and satellite prophages are widespread in Streptococcus and may play a role in pneumococcal pathogenesis. Nat Commun **10**:4852.

Rincón-Molina, C. I., J. A. Hernández-García, R. Rincón-Rosales, F. A. Gutiérrez-Miceli, D. A. Ramírez-Villanueva, E. González-Terreros, B. A. Peña-Ocaña, H. Palomeque-Domínguez, L. Dendooven, and V. M. Ruíz-Valdiviezo. 2018. Structure and Diversity of the Bacterial Communities in the Acid and Thermophilic Crater-Lake of the Volcano “El Chichón”, Mexico. Geomicrobiology Journal **36**:97-109.

Seemann, T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics

30:2068-2069.
Shikano, T., Y. Shimada, G. Herczeg, and J. Merila. 2010. History vs. habitat type: explaining the genetic structure of European nine-spined stickleback (Pungitius pungitius) populations. Mol Ecol **19**:1147-1161.
Shintani, M. 2017. The behavior of mobile genetic elements (MGEs) in different environments. Biosci Biotechnol Biochem **81**:854-862.
Simao, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics **31**:3210-3212.
Smit, Y., M. Cameron, P. Venter, and R. C. Witthuhn. 2011. Alicyclobacillus spoilage and isolation--a review. Food Microbiol **28**:331-349.
Sorenson, L., F. Santini, and M. E. Alfaro. 2014. The effect of habitat on modern shark diversification. J Evol Biol **27**:1536-1548.
Steyn, C. E., M. Cameron, and R. C. Witthuhn. 2011. Occurrence of Alicyclobacillus in the fruit processing environment--a review. Int J Food Microbiol **147**:1-11.
Stoebel, D. M., K. Hokamp, M. S. Last, and C. J. Dorman. 2009. Compensatory evolution of gene regulation in response to stress by Escherichia coli lacking RpoS. PLoS Genet **5**:e1000671.
Stuart, R. K., B. Brahamsha, K. Busby, and B. Palenik. 2013. Genomic island genes in a coastal marine Synechococcus strain confer enhanced tolerance to copper and oxidative stress. ISME J **7**:1139-1149.
Sullivan, M. B., D. Lindell, J. A. Lee, L. R. Thompson, J. P. Bielawski, and S. W. Chisholm. 2006. Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. PLoS biol **4**:e234.
Team, R. C. 2000. R language definition. Vienna, Austria: R foundation for statistical computing.
Thiaville, J. J., S. M. Kellner, Y. Yuan, G. Hutinet, P. C. Thiaville, W. Jumpathong, S. Mohapatra, C. Brochier-Armanet, A. V. Letarov, and R. Hillebrand. 2016. Novel genomic island modifies DNA with 7-deazaguanine derivatives. Proceedings of the National Academy of Sciences **113**:E1452-E1459.
Thompson, J. D., T. J. Gibson, and D. G. Higgins. 2003. Multiple sequence alignment using ClustalW and ClustalX. Current protocols in bioinformatics:2.3. 1-2.3. 22.
Van Dongen, S., and C. Abreu-Goodger. 2012. Using MCL to extract clusters from networks. Pages 281-295 Bacterial Molecular Networks. Springer.
Waack, S., O. Keller, R. Asper, T. Brodag, C. Damm, W. F. Fricke, K. Surovcik, P. Meinicke, and R. Merkl. 2006. Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. BMC bioinformatics **7**:142.
Wang, P., Z. Zeng, W. Wang, Z. Wen, J. Li, and X. Wang. 2017. Dissemination and loss of a biofilm-related genomic island in marine Pseudoalteromonas mediated by integrative and conjugative elements. Environ Microbiol **19**:4620-4637.
Wickham, H. 2016. ggplot2: elegant graphics for data analysis. Springer.
Xie, Z., and H. Tang. 2017. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. Bioinformatics **33**:3340-3347.
Yang, S., X. Wen, Y. Shi, S. Liebner, H. Jin, and A. Perfumo. 2016. Hydrocarbon degraders establish at the costs of microbial richness, abundance and keystone taxa after crude oil contamination in permafrost environments. Sci Rep **6**:37473.
Yang, S., X. Wen, L. Zhao, Y. Shi, and H. Jin. 2014. Crude oil treatment leads to shift of bacterial communities in soils from the deep active layer and upper permafrost along the China-Russia Crude Oil Pipeline route. PLoS One

9:e96552.
Zamarro, M. T., Z. Martin-Moldes, and E. Diaz. 2016. The ICEXTD of Azoarcus sp. CIB, an integrative and conjugative element with aerobic and anaerobic catabolic properties. Environ Microbiol **18**:5018-5031.
Zeng, Z., X. Liu, J. Yao, Y. Guo, B. Li, Y. Li, N. Jiao, and X. Wang. 2016. Cold adaptation regulated by cryptic prophage excision in Shewanella oneidensis. ISME J **10**:2787-2800.
Zheng, Z., Y. Zhang, Z. Liu, Z. Dong, C. Xie, A. Bravo, M. Soberon, J. Mahillon, M. Sun, and D. Peng. 2020. The CRISPR-Cas systems were selectively inactivated during evolution of Bacillus cereus group for adaptation to diverse environments. ISME J **14**:1479-1493.

Table 1. Basic information of *Alicyclobacillus* species.

Strain	Species	Accession number	Habitat category 1	Habitat category 2	Optimal growth temperature	Growth temperature range	Optimal growth pH	Growth pH range	Contigs	Bases	CDS	rRNA	tRNA	tmRNA	Completeness %	Fe^{2+} oxidation rate ($mg \cdot L^{-1} \cdot d^{-1}$)	†RISC Oxidation (Decrease of pH)
ALC3 CGMCC 1.16736	sp.	LMSG_G000001124.1	Acid mine drainge	Acid mine drainge	37	20-45	3.5	2.5-5.0	1	5073212	4993	15	52	1	94	-	-
S30H101 CGMCC 1.17180	sp.	LMSG_G000001125.1	Acid mine drainge	Acid mine drainge	30	20-40	3.5	2.5-5.0	1	4848675	4654	15	52	1	94.8	-	-
S30H31 CGMCC 1.17131	sp.	LMSG_G000001126.1	Acid mine drainge	Acid mine drainge	30	20-40	3.5	2.5-5.0	1	4807099	4654	15	53	1	94.8	-	-
ALEF1 CGMCC 1.17055	sp.	LMSG_G000001127.1	Acid mine drainge	Acid mine drainge	30	20-40	4	2.5-5.5	1	6027356	5627	21	76	3	94	+(64.53)	+(1.20)
S30H14 CGMCC 1.17050	sp.	LMSG_G000001128.1	Acid mine drainge	Acid mine drainge	30	20-40	3	2.5-5.5	1	5293537	5069	24	74	1	94.4	-	-
S30H17 CGMCC 1.17123	sp.	LMSG_G000001129.1	Acid mine drainge	Acid mine drainge	30	20-55	3.3	1.5-5.0	1	6080594	5889	34	84	1	95.3	+(188.13)	+(0.66)
F1	sp.	LMSG_G000001135.1	Acid mine drainge	Acid mine drainge	30	20-55	3.3	1.5-5.0	1	5818277	5638	31	83	1	95.3	-	-
H1 CGMCC 1.17290	sp.	LMSG_G000001137.1	Acid mine drainge	Acid mine drainge	30	20-40	3.5	2.0-5.0	1	6514169	6227	21	75	3	94	-	+(0.87)
H5 CGMCC 1.17291	sp.	LMSG_G000001139.1	Acid mine drainge	Acid mine drainge	30	20-40	3.5	2.0-5.0	1	6571288	6233	21	75	1	94	-	+(0.79)
SCH	*A. cycloheptanicus*	LMSG_G000001148.1	Soil	Garden soil	48	40-53	4.0	3.0-5.5	1	3617228	3421	18	70	1	94.8	-	-
DSM 13609	*A. herbarius*	GCA_000430585.1	Beverage	Herbal tea	57.5	35-65	4.8	3.5-6.0	57	3310509	3181	6	65	1	94.8	-	-
DSM 14955	*A. pomorum*	GCA_000472905.1	Beverage	Spoiled mixed fruit juice	48	30-60	4.3	3.0-6.0	36	3398558	3385	9	96	1	95.3	-	-
DSM 17974	*A. sacchari*	GCA_004366795.1	Beverage	Liquid sugar	48	30-55	4.3	2.0-6.0	66	2946319	2892	5	72	1	95.3	-	-
NBRC 100859	*A. acidiphilus*	GCA_001544355.1	Beverage	An ‘off’-flavoured acidic beverage	50	20-55	3.0	2.5-5.5	165	3865262	3719	3	70	1	94.4	-	-
NBRC 102425	*A. mali*	GCA_001570745.1	Beverage	Apple juice	NA	NA	4.0	3.0-5.0	85	2786970	2678	2	62	1	94	-	-
CGMCC 1.10793	*A. tengchongensis*	GCA_001447355.1	Hot spring	Hot spring	50	30-75	4.0	2.5-6.0	55	2809442	2799	5	60	1	93.1	-	-
DSM 16176	*A. vulcanalis*	GCA_900156755.1	Hot spring	Hot spring in Mojave Desert	55	NA	4.0	NA	34	2994480	2903	6	62	1	93.5	-	-
DSM 446	*A. acidocaldarius subsp. Acidocaldarius*	GCA_000024285.1	Hot spring	Hot and acidic spring in Yellowstone National Park	62.5	45-70	3.5	2.0-6.0	4	3205686	3149	18	65	1	95.7	-	-
LAA1	*A. acidocaldarius*	GCA_000173835.1	Hot spring	Acidic creek in Yellowston National Park	57	NA	3.6	NA	58	2943284	3005	8	60	1	88.8	-	-
Tc-4-1	*A. acidocaldarius subsp. Acidocaldarius*	GCA_000219875.1	Hot spring	Hot spring	70	NA	3.0	NA	1	3124048	3114	18	65	1	94	-	-
URH17-3-68	*A. hesperidum*	GCA_000294675.1	Hot spring	Hot spring sludge	53	NA	3.8	NA	78	2967160	2893	7	71	1	94	-	-
USBA-GBX-503	*A. montanus*	GCA_900142255.1	Hot spring	Hot spring, National Natural Park Los Nevados	45	25-55	3.0	1.5-4.5	74	3045912	3022	8	53	1	94	-	-
RIFOXYA1_FULL_53_8	sp.	GCA_001767765.1	Sediment	Rifle well sediment	NA	NA	NA	NA	363	2646965	2648	0	32	0	72.4	-	-
ATCC 49025	*A. acidoterrestris*	GCA_000444055.1	Soil	Garden soil	45	25-60	4.8	2.0-6.0	207	4063548	4111	2	132	1	95.7	-	-
CPP55	*A. macrosporangiidus*	GCA_000702485.1	Soil	Canopy soil	55	NA	4.2	NA	2	4091649	4042	18	56	1	94.8	-	-
DSM 12489	*A. hesperidum*	GCA_900107035.1	Soil	Solfataric soils	52	35-60	3.8	2.0-6.0	53	2859702	2807	6	71	1	93.5	-	-
DSM 17975	*A. contaminans*	GCA_000429525.1	Soil	Soil of crop fields in Fuji, Janpan	53	35-60	4.3	3.0-6.0	111	3272318	3330	7	75	1	94.8	-	-
DSM 17980	*A. macrosporangiidus*	GCA_900116805.1	Soil	Soil of crop fields in Shizuoka, Janpan	53	35-60	4.3	3.0-6.5	95	3789420	3758	8	54	1	93.5	-	-
NBRC 100866	*A. sendaiensis*	GCA_001552675.1	Soil	Soil of Aoba-yama Park, Sendai, Miyagi, Japan	55	40-65	5.5	2.5-6.5	153	2793850	2749	2	50	1	93.5	-	-
NBRC 103103	*A. shizuokensis*	GCA_001552255.1	Soil	Soil of crop fields	48	35-60	4.3	3.0-6.5	148	3628929	3541	3	57	1	91.8	-	-
NBRC 103104	*A. kakegawensis*	GCA_001552655.1	Soil	Crop field	53	40-60	4.3	3.0-6.5	71	3421583	3286	3	59	1	91.4	-	-
NBRC 106287	*A. acidoterrestris*	GCA_007991715.1	Soil	Orchard soil	45	NA	3.5	NA	168	4033884	4001	3	60	1	95.7	-	-
TC-34	*A. ferrooxydans*	LMSG_G000001149.1	Soil	Solfataric soils	28	17-40	3.0	2.0-6.0	3	5126777	4875	27	90	1	95.3	+	+

†RISC: Reduced inorganic sulfur compounds. NA: Not available. +: Positive.

Figure and legends

Figure 1. The genome characteristics and phylogenetic relationships of *Alicyclobacillus* species. **(A)** Bar plot of basic genome information. Error bars are the standard deviations. **(B)** The pan-genome analyses. Blocks represent the occurrence of gene clusters, arranged based on their distribution across genomes. The cluster analysis of 33 genomes was based on genome size (Total length), GC content, the number of genes, singleton gene clusters, and gene clusters. A gene cluster is a group of homologs identified based on the similarity of amino acid sequence.

Figure 2. Phylogenetic tree constructed from 16S rRNA sequences. The numeric labels are bootstrap confidence values greater than 60. Strain color represents the isolation environments. Purple: Acid mine drainage. Green: Beverage. Red: Hot spring. Orange: Sediment. Blue: Soil. Black: No genome data. Grey: Outgroup. RISC: Reduced inorganic sulfur compunds.

Figure 3. The gene function distributions of genomic islands (GI) in *Alicyclobacillus* species. **(A)** The COG function compositions of GI-associated genes in each strain. Bar color denotes COG function category. **(B)** Boxplots of the function categories significantly (Kruskal test, $P < 0.05$) depending on environments. Lowercase letters denote the significant differences (LSD test, $P < 0.05$).

Figure 4. The contributions of GI-associated genes to heavy metal resistance. **(A)** All genes related to heavy metal resistance in genome, the numeric labels are the number of GI-associated genes. **(B)** All genes related to heavy metal resistance except for the GI-associated genes. Heavy metal resistance genes in red indicated that it was

significantly dependent on bacterial environments (Kruskal test, $P < 0.05$).

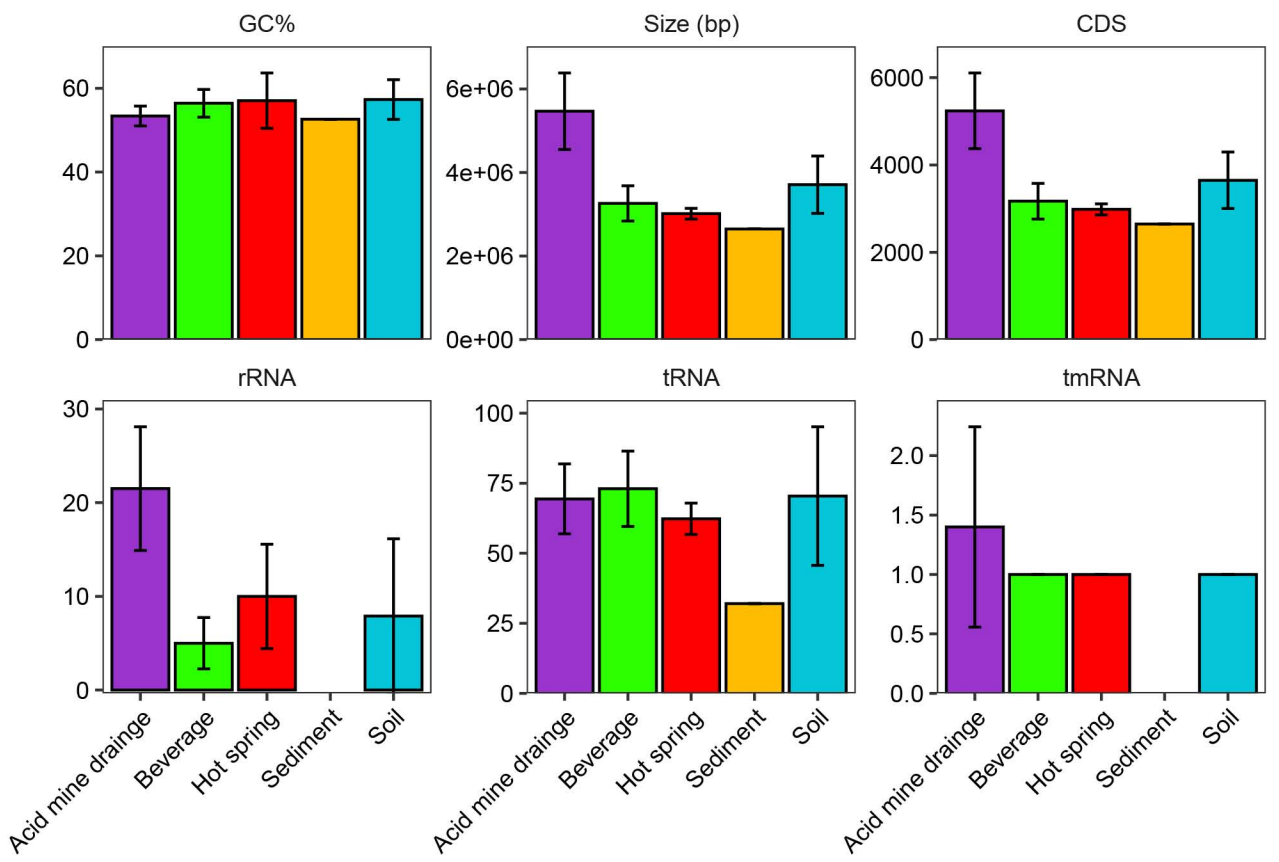
Figure 5. Insertion sequences (IS) in *Alicyclobacillus*. **(A)** The number of IS110 sequences in *Alicyclobacillus* species, which is significant dependent on environments (Kruskal test, $P < 0.05$). Lowercase letters denote the significant differences (LSD test, $P < 0.05$). **(B)** The IS6 and new families of insertion sequences in *Alicyclobacillus* genomes. The first one fragment is IS6, while the remaining fragments are new families. The arrows between two black vertical lines denote the IS genes and arrow length denotes gene size.

Figure 6. The number of prophage regions, anti-CRISPR, and CRISPR-Cas systems in *Alicyclobacillus* genomes.

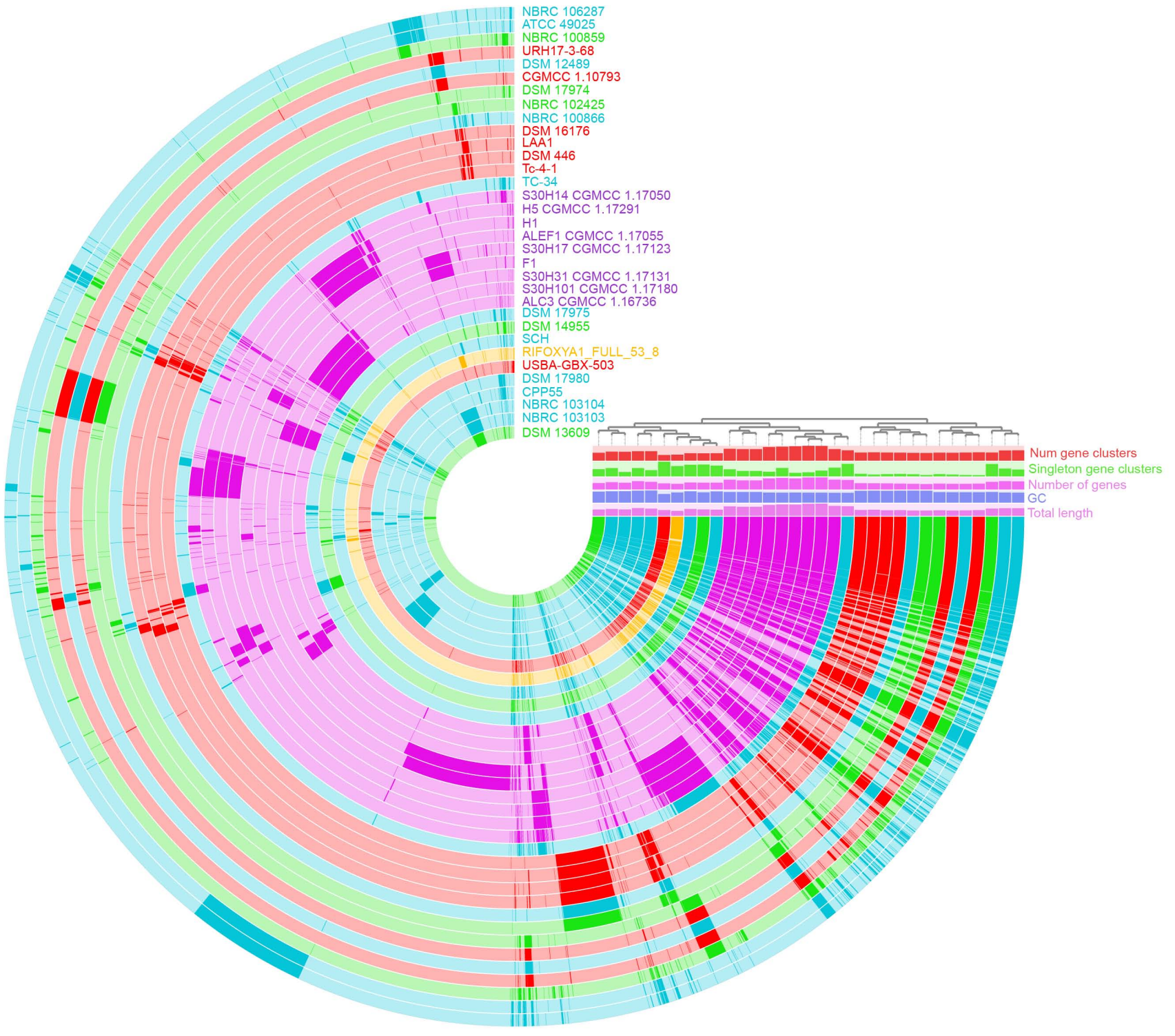
Figure 7. The performances of MGEs in shaping *Alicyclobacillus* genome. The performances of MGEs are assessed by the slope of the linear curves.

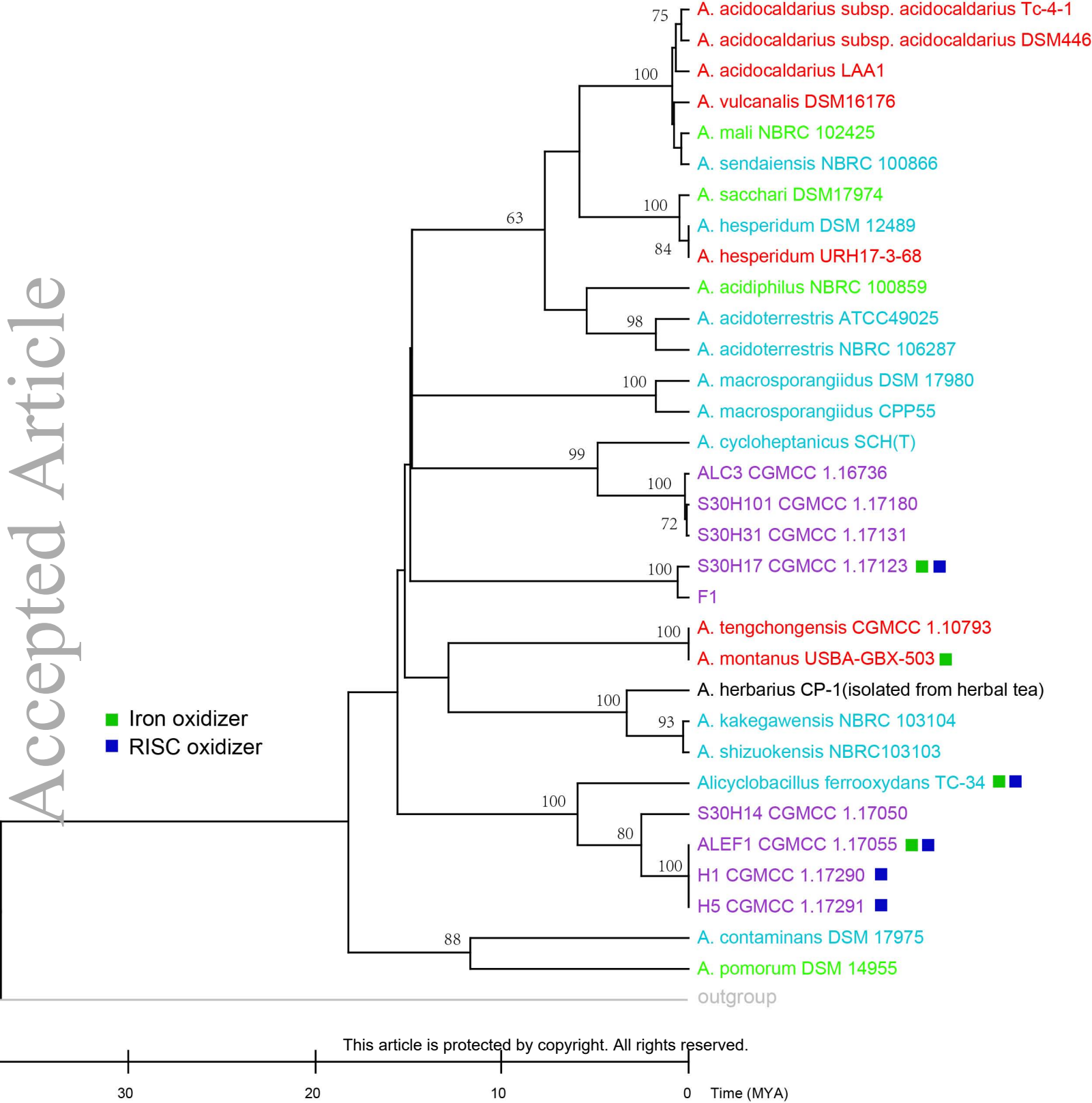
Figure 8. The MGE-associated genes shared among the four types: genomic island (GI), prophage, integrative and conjugative element (ICE) and insertion sequence (IS).

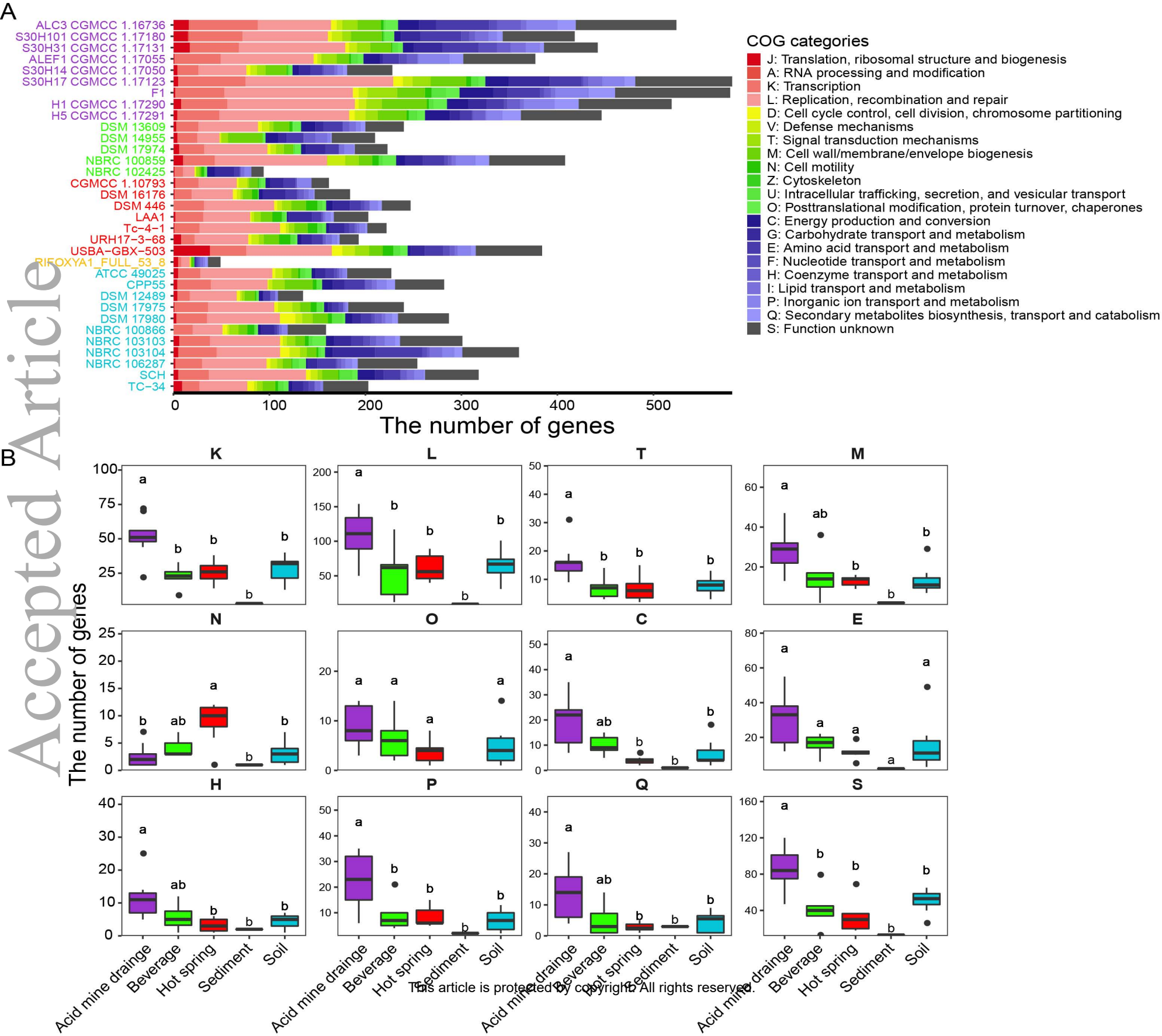
A

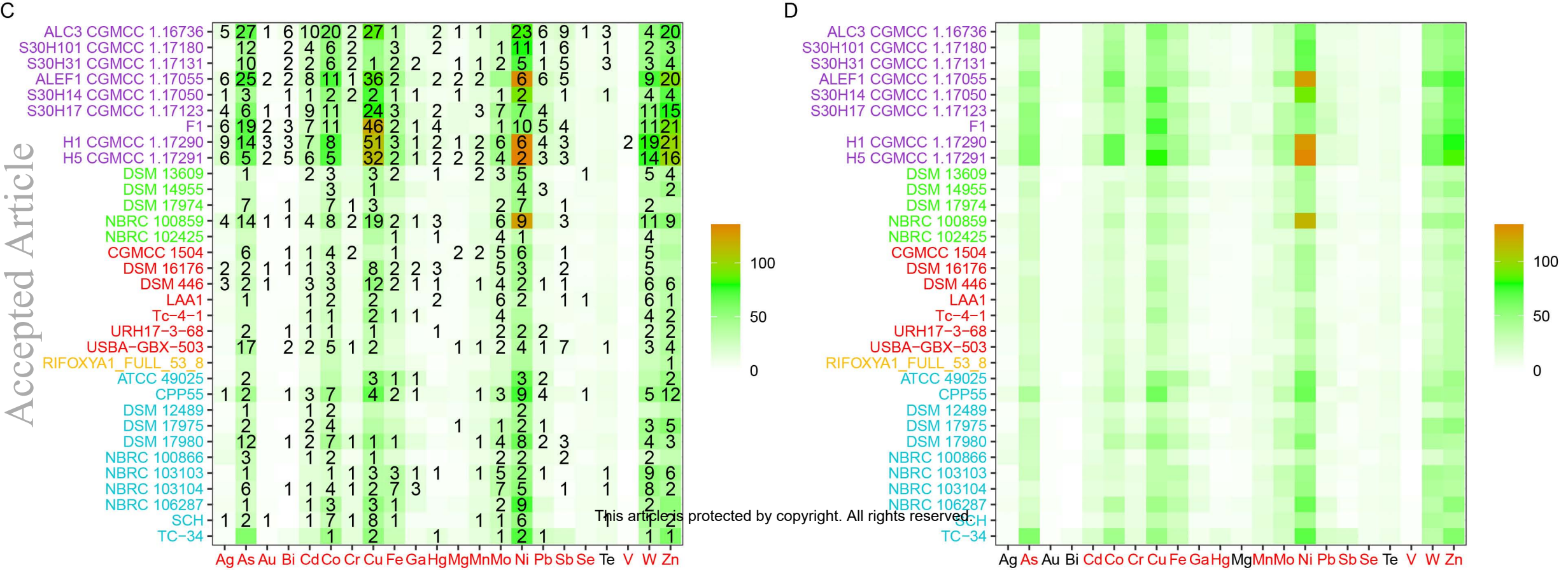


B

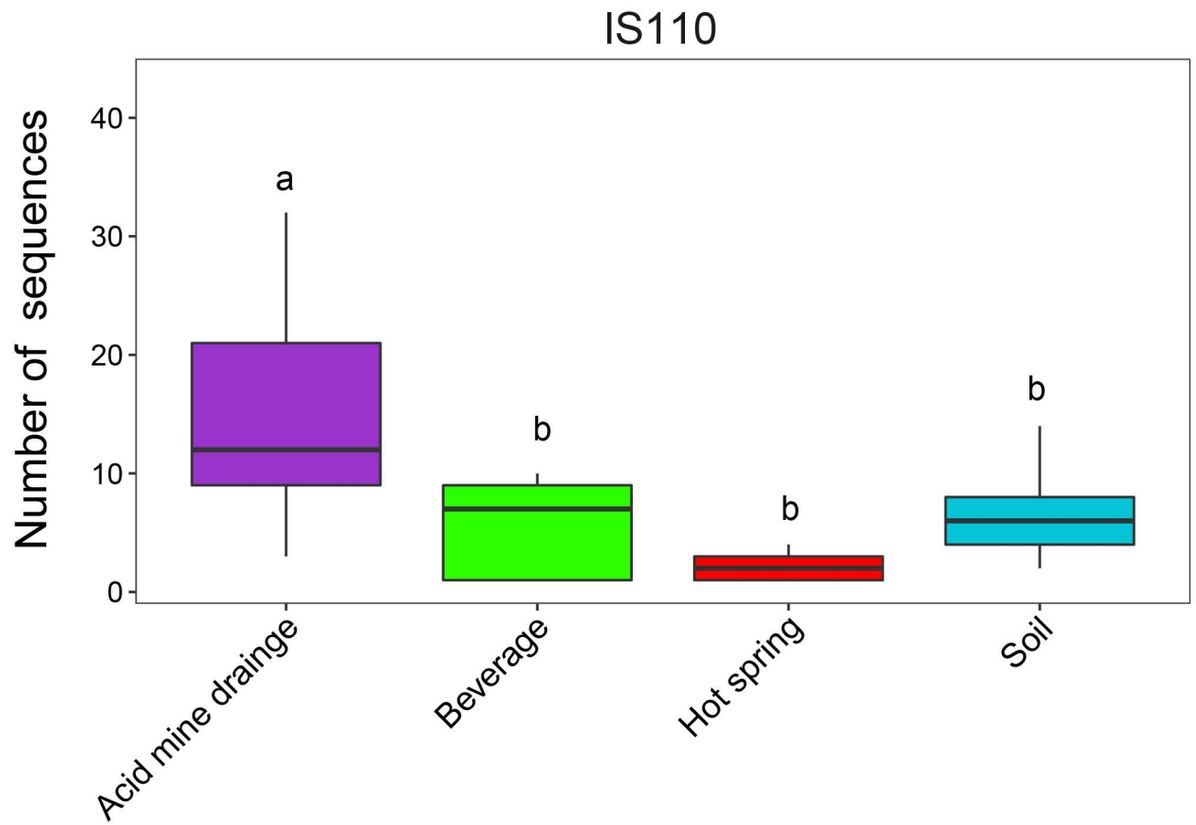




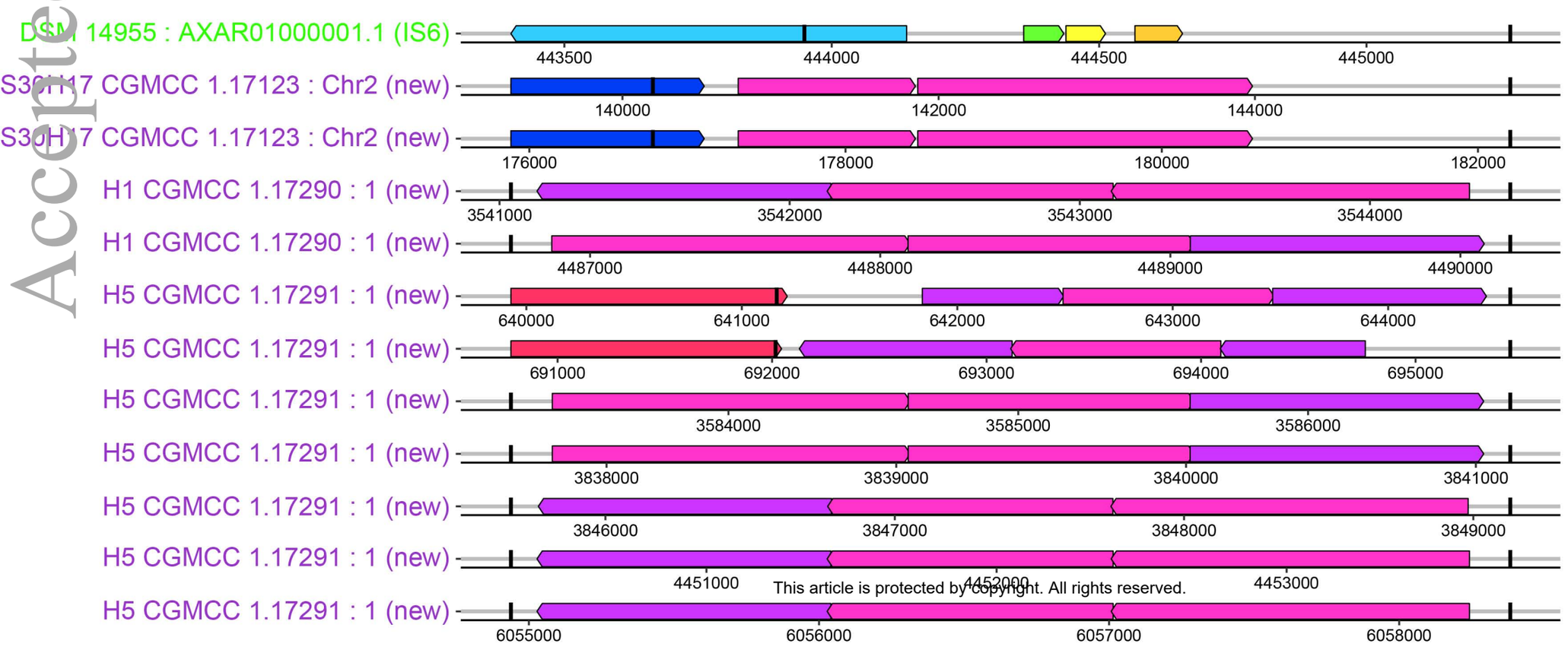


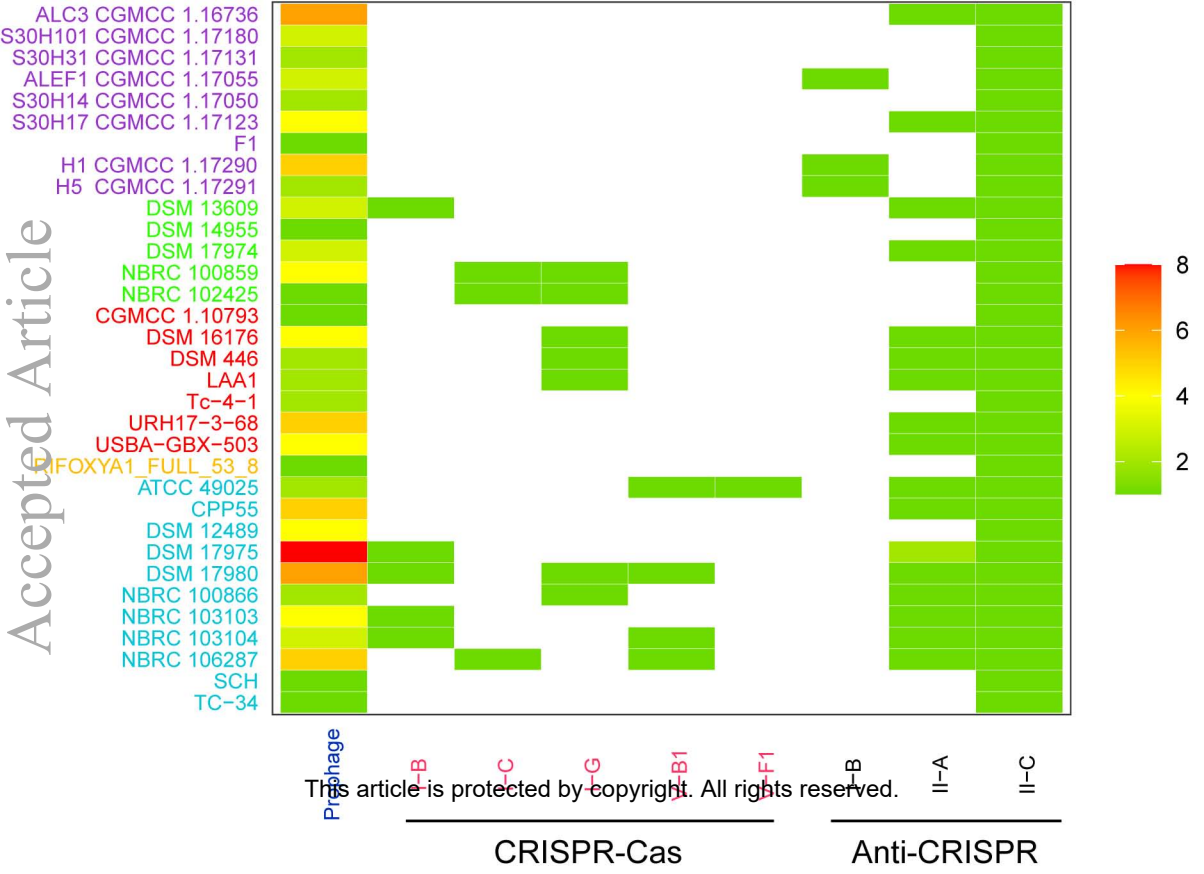

C
D
ALC3 CGMCC 1.16736
S30H101 CGMCC 1.17180
S30H31 CGMCC 1.17131
ALEF1 CGMCC 1.17055
S30H14 CGMCC 1.17050
S30H17 CGMCC 1.17123
F1
H1 CGMCC 1.17290
H5 CGMCC 1.17291
DSM 13609
DSM 14955
DSM 17974
NBRC 100859
NBRC 102425
CGMCC 1504
DSM 16176
DSM 446
LAA1
Tc-4-1
URH17-3-68
USBA-GBX-503
RIFOXYA1_FULL_53_8
ATCC 49025
CPP55
DSM 12489
DSM 17975
DSM 17980
NBRC 100866
NBRC 103103
NBRC 103104
NBRC 106287
SCH
TC-34
Ag As Au Bi Cd Co Cr Cu Fe Ga Hg Mg Mn Mo Ni Pb Sb Se Te V W Zn
100
50
0
Accepted Article
This article is protected by copyright. All rights reserved.

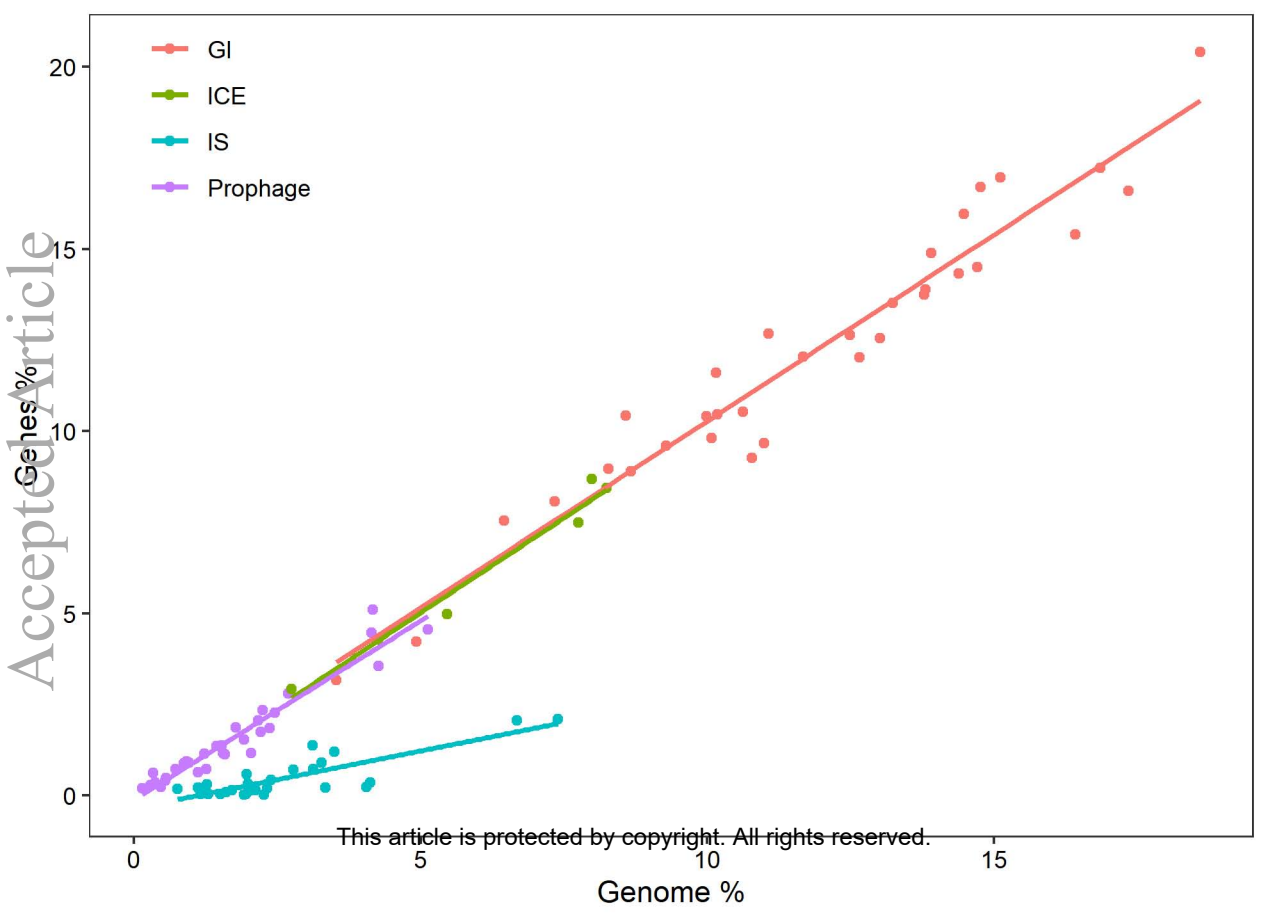
A



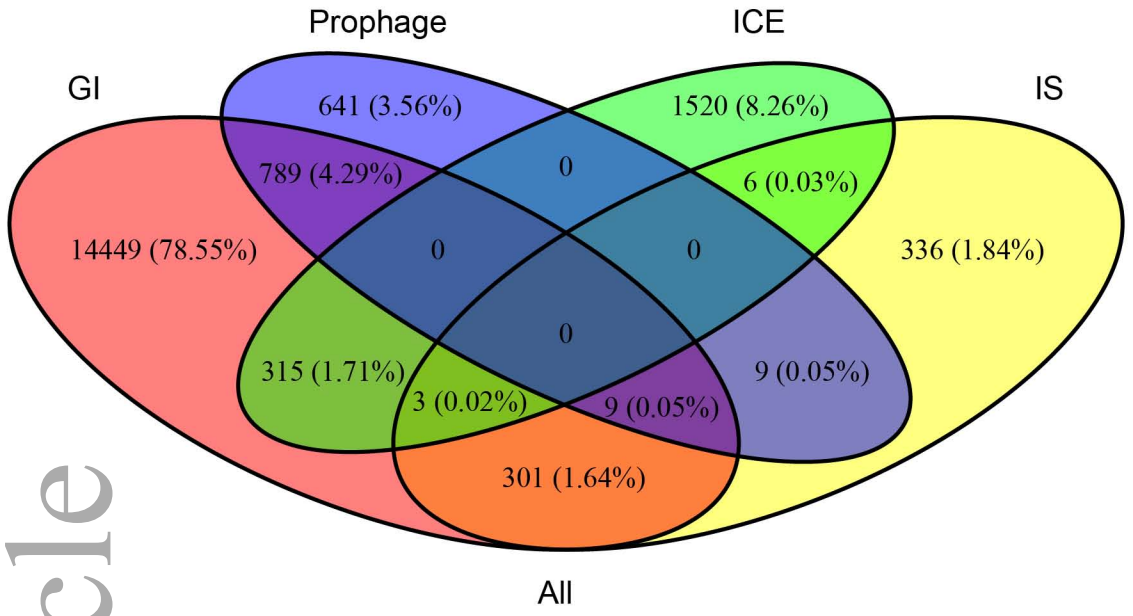
B



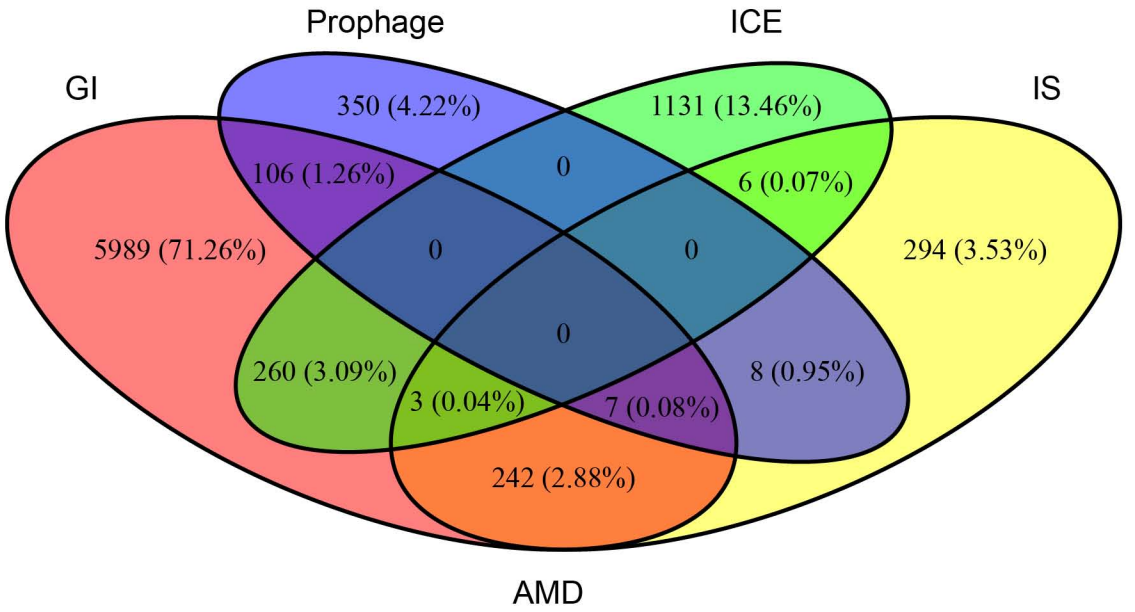




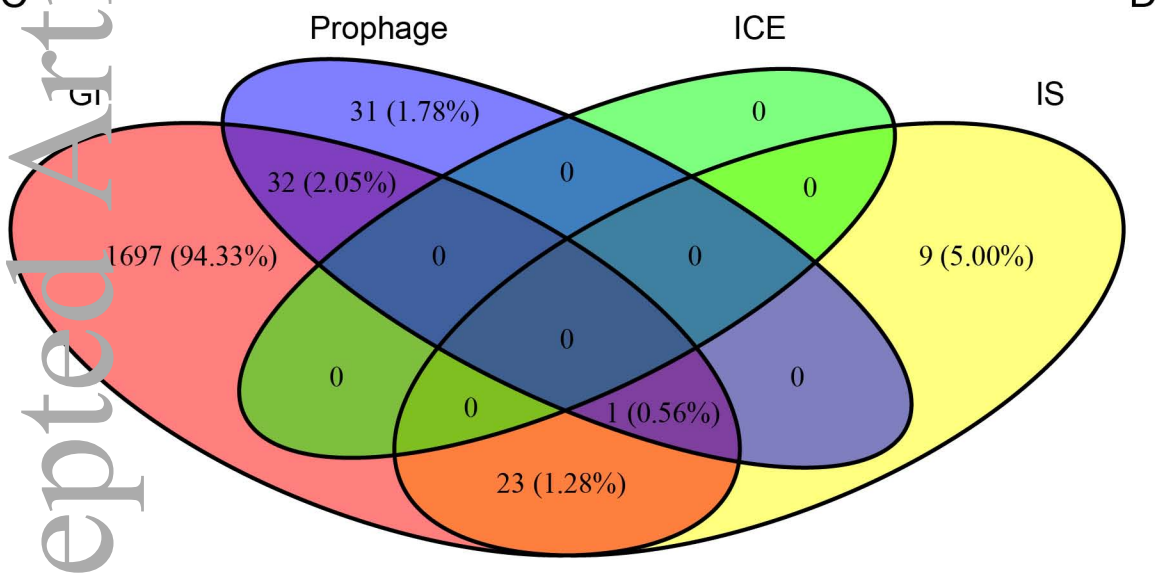
A



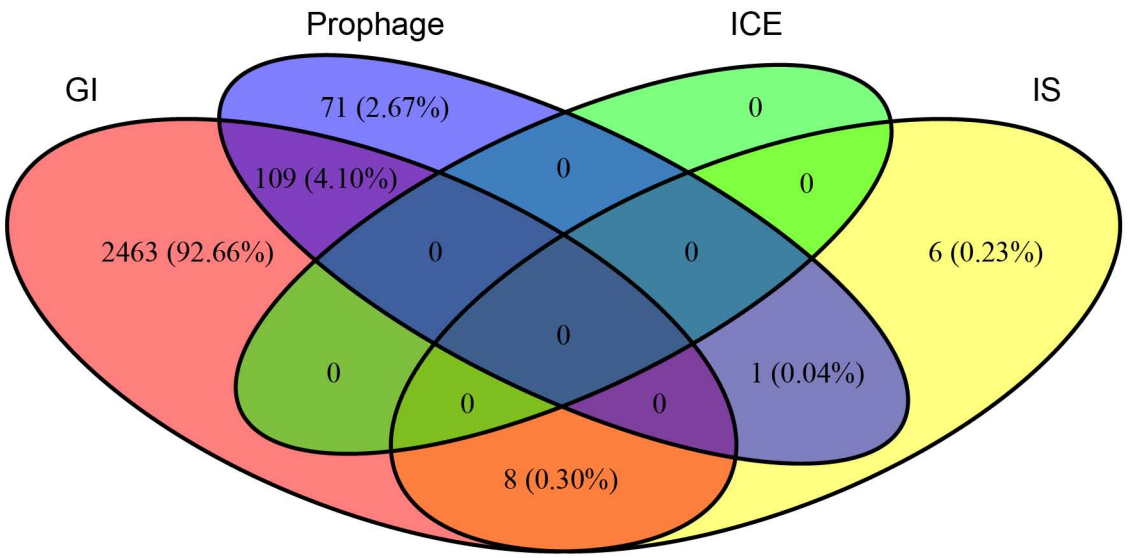
B



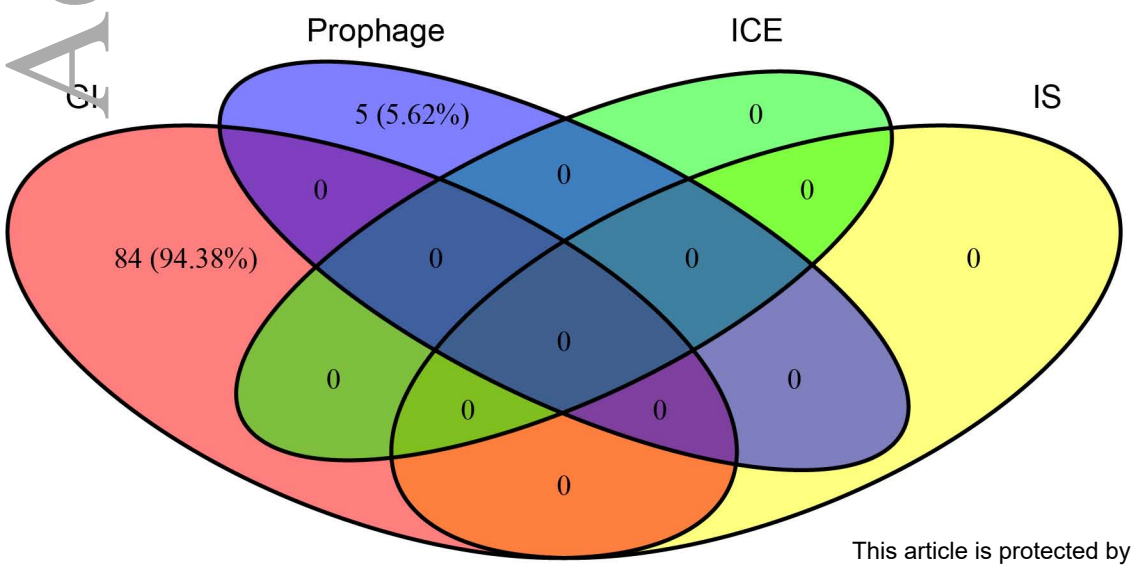
C



D



E



F