Comparative Genomics Unveils the Habitat Adaptation and Metabolic Profiles of *Clostridium* in an Artificial Ecosystem for Liquor Production

Guan-Yu Fang,[a,b] Li-Juan Chai,[b,d] Xiao-Zhong Zhong,[b,d] Zhen-Ming Lu,[b,d] Xiao-Juan Zhang,[b,d] Lin-Huan Wu,[e] Song-Tao Wang,[f] Cai-Hong Shen,[f] Jin-Song Shi,[c] (ID) Zheng-Hong Xu[a,b,f]

[a]Key Laboratory of Industrial Biotechnology of Ministry of Education, School of Biotechnology, Jiangnan University, Wuxi, People's Republic of China
[b]National Engineering Research Center for Cereal Fermentation and Food Biomanufacturing, Jiangnan University, Wuxi, People's Republic of China
[c]School of Life Science and Health Engineering, Jiangnan University, Wuxi, People's Republic of China
[d]Jiangsu Provincial Engineering Research Center for Bioactive Product Processing, Jiangnan University, Wuxi, People's Republic of China
[e]Institute of Microbiology, Chinese Academy of Sciences, Beijing, People's Republic of China
[f]National Engineering Research Center of Solid-State Brewing, Luzhou, People's Republic of China

ABSTRACT *Clostridium* inhabiting pit mud (PM) is one of the important bacterial populations for synthesizing flavor compounds of Chinese strong-flavor baijiu. The long-term cereal fermentation with sorghum as the main raw material creates an environment rich in starch, ethanol, and organic acids (mainly lactic acid). However, the genetic factors underpinning *Clostridium*'s adaptation to PM remain poorly understood. Here, we performed comparative genomic analysis between 30 pit mud-associated (PMA) and 100 non-pit mud-associated (NPMA) *Clostridium* strains. Comparison analysis of the enrichment of KEGG pathways between PMA and NPMA *Clostridium* strains showed two-component system, flagellar assembly, and bacterial chemotaxis pathways related to environmental adaptation were enriched in PMA strains. The number of genes encoding alcohol dehydrogenase and L-lactate dehydrogenase in PMA *Clostridium* strains was significantly higher than that in NPMA, which is helpful for them to adapt to the ethanol- and lactic acid-rich environment. The analysis of carbohydrate-active enzymes demonstrated that glycoside hydrolases (GHs) was the most abundant family in all *Clostridium* strains, and genes encoding GH4 and GH13, involved in starch and sucrose metabolism, were enriched in PMA *Clostridium*. Horizontal gene transfer analysis revealed that multiple genes encoding the enzymes involved in carbohydrate and amino acid metabolism were transferred from *Bacillus* to *Clostridium* in pit mud. Most of the PMA *Clostridium* strains had good potential for butyric acid synthesis from ethanol, lactic acid, and starch. Collectively, this study furthers our understanding of the habitat adaptation and metabolic potential of PMA *Clostridium* strains.

IMPORTANCE Pit mud is a typical artificial ecosystem for Chinese liquor production. *Clostridium* inhabiting pit mud plays essential roles in the flavor formation of strong-flavor baijiu. The relative abundance of *Clostridium* increased with pit mud quality, further influencing the quality of baijiu. So far, the ecological adaptation of *Clostridium* to a pit mud-associated lifestyle is largely unknown. Here, comparative genomic analysis of pit mud-associated (PMA) and non-pit mud-associated (NPMA) *Clostridium* strains was performed. We found genes related to the metabolism of starch, ethanol, and lactic acid were enriched in PMA *Clostridium* strains, which facilitated their adaptation to the unique brewing environment. In addition, horizontal gene transfer contributed to the adaptation of *Clostridium* to pit mud. Our findings provide genetic insights on PMA *Clostridium* strains' ecological adaptation and metabolic characteristics.

KEYWORDS *Clostridium*, pit mud, CAZymes, horizontal gene transfer, short-chain fatty acids

Strong-flavor baijiu is one of the popular and traditional Chinese liquors, and it accounts for above 70% of the total baijiu sales in China (1). Raw materials (mainly sorghum) are anaerobically fermented by multiple microorganisms in underground mud cellars for 2 to 6 months and then distilled to obtain strong-flavor baijiu. Through the comparative microbiome analysis of fermented grains and pit mud, it was found that microbial assembly inhabiting pit mud played essential roles in the formation of typical flavor compounds of strong-flavor baijiu, for example, fatty acids such as caproic acid and butyric acid (2, 3). Metagenomic analysis and biomass estimation by quantitative PCR demonstrated that pit mud microbiota was dominated by bacteria and archaea (4). Clostridia dominated the pit mud bacterial community, mainly scattered in the genera including *Caproiciproducens*, *Hydrogenispora*, *Syntrophomonas*, *Clostridium*, and *Sedimentibacter* (3–5). Previous work revealed that the relative abundance of *Clostridium* increased with pit mud quality, further influencing the quality of strong-flavor baijiu (5). One of the representative characteristics and uses of *Clostridium* species was the production of short- and medium-chain fatty acids such as acetic acid, butyric acid, and caproic acid (6). Metagenomic analysis and clone library analysis based on the key genes of butyrate synthesis showed that *Clostridium* species were the main potential butyric acid producers in pit mud (7). So far, among the clostridial strains isolated from pit mud, *Clostridium* species was the most diverse population, with various fatty acid metabolic characteristics (7–9). The important role of *Clostridium* in baijiu flavor formation attracted us to further study their functions in pit mud microbial ecosystem.

The genus *Clostridium* includes 130 species according to the EzBioCloud database (data collected in December 2021; https://www.ezbiocloud.net/), and these strains were isolated from a wide range of habitats, mainly including feces, soil, and foods (10). The living environment of *Clostridium* species inhabiting pit mud is closely related to human brewing activities. During the fermentation of strong-flavor baijiu, polymers in starch-rich raw materials (mainly sorghum) are hydrolyzed and metabolized into ethanol, organic acids, and other small molecules by microorganisms affiliated with *Lactobacillaceae*, *Kazachstania*, and *Wickerhamomyces* (3). These small molecules are dissolved in the water produced by the fermentation process, formed huangshui (the liquid produced by cereal fermentation), and gradually accumulated in the cellar under the action of gravity, so as to contact with the pit mud. Ethanol content in huangshui could reach about 7% to 10% (11). The accumulation of plenty of organic acids during baijiu fermentation created an acidic environment, and pH of fermented grains and huangshui was about 3 to 4 (3, 11). Lactic acid was the most abundant organic acid, and its content in fermented grains could reach ~30 g/kg at the end of fermentation (3). Lactic acid content in huangshui was around 40 g/liter or even 80 g/liter (11, 12). Thus, it is clear that the habitat environment of pit mud microorganisms is rich in organic acids (mainly lactic acid) and ethanol due to cereal fermentation. Interestingly, study on viable bacteria in pit mud of different ages revealed that the relative abundance of *Clostridium sensu stricto* was 12 to ~18% in 20-year-old cellar, significantly higher than that in 5-year-old cellar (0.4%) (13). How these *Clostridium* species survive and resist the environmental pressure of pit mud is not yet fully understood.

Comparative genomic analysis is an effective approach to illuminate the evolution and habitat adaptation mechanism of microorganisms (14–16) and the diverse metabolic capabilities of strains within the same species isolated from different habitats (17). Five *Clostridium beijerinckii* strains were isolated from pit mud, and comparative genomics of these strains and *C. beijerinckii* from other habitats showed that hundreds of genes were shared only by strains isolated from pit mud (18). To clarify the adaptation mechanism of *Clostridium* in pit mud environment, comparative genomics of 130 *Clostridium* strains from different habitats were performed in this study, including genome sequences of 30 strains isolated from pit mud and 100 strains from other habitats. Pangenome analysis was conducted to reveal the functional differences of core genome of pit mud-associated (PMA) and non-pit mud-associated (NPMA) *Clostridium*

strains. Genes significantly enriched in PMA strains were identified to explain their adaptation to pit mud environment, and the effect of horizontal gene transfer on these enriched genes was evaluated. Considering the roles of *Clostridium* in the flavor formation of strong-flavor baijiu, we further compared the metabolic potentials of short-chain fatty acids between PMA and NPMA strains. This work provides insights into the genomic features of *Clostridium* species under the unique environment of pit mud.

RESULTS

Genome characteristics and phylogenetic composition of *Clostridium* strains. To decipher the underlying mechanism of the adaptation to pit mud environment of *Clostridium*, comparative genomic analysis was carried out between 30 pit mud-associated (PMA) and 100 non-pit mud-associated (NPMA) *Clostridium* strains (see Table S1 in the supplemental material). The 30 PMA strains were isolated from the pit mud of different fermentation cellars from strong-flavor baijiu producers in China. It should be noted that no matter what producer, the brewing process with sorghum as the main raw material creates a similar ecological environment for microorganisms inhabiting pit mud. To investigate the genomic differences between PMA and NPMA *Clostridium* strains, all of the genomes of *Clostridium* type strains isolated from other habitats and deposited in the EzBioCloud database were used for comparative genomics analysis. The genome sizes of PMA *Clostridium* strains were 4.53 ± 1.22 Mb on average, which was significantly larger than that of NPMA strains (3.92 ± 1.08 Mb) (Mann-Whitney U test, $P = 0.02$) (Fig. 1A). The mean value of GC contents of PMA *Clostridium* strains ($32\% \pm 7\%$) was significantly lower than that of NPMA strains ($33\% \pm 5\%$) (Mann-Whitney U test, $P = 0.01$) (Fig. S1).

Core genome-based phylogenetic analysis of 130 *Clostridium* strains from different habitats was conducted to examine their evolutionary relationships. The maximum-likelihood phylogenetic tree was constructed based on the amino acid sequences of 304 single-copy core genes (Fig. 1B). PMA *Clostridium* strains fell into different clades, and 63.3% of them were scattered in cluster III. A previous study found that the distribution of microorganisms in the phylogenetic tree was not necessarily related to their habitats (19). The size of pan- and core genomes will be influenced by the number of genomes. Therefore, we randomly sampled the same number of genomes from PMA and NPMA *Clostridium* strains for pangenome analysis. When more than 19 genomes were sampled, the sizes of the pangenomes of NPMA *Clostridium* strains were significantly larger than those of PMA strains (Fig. 1C). When the number of sampled genomes was greater than one, the core genome sizes of PMA *Clostridium* strains were significantly larger than those of NPMA strains (Fig. 1D).

The comparison of the core genome functions between PMA and NPMA *Clostridium* strains was performed according to the methods of a previous study (20). A random set of 30 genomes of NPMA *Clostridium* strains was selected, and the mean phylogenetic distance between PMA and NPMA strains was calculated. This step was repeated 100 times, and then 30 genomes of NPMA strains with the least deviation from the average phylogenetic distance of PMA strains were selected for the subsequent analysis. An alignment to COG and KEGG databases was performed to analyze the similarities and differences in potential functions of the core genomes of PMA and NPMA *Clostridium* strains. The results based on COG database showed that the abundant enriched function of the core genome of all *Clostridium* strains included transcription (K), translation, ribosomal structure and biogenesis (J), signal transduction mechanisms (T), cell wall/membrane/envelope biogenesis (M), and replication, recombination, and repair (L) functional categories, accounting for more than 60% of the total abundance (Fig. 2A). The relative abundance of the COG categories involved in metabolism of energy production and conversion (C, $4.1\% \pm 0.3\%$), carbohydrate transport and metabolism (G, $4.2\% \pm 0.3\%$), amino acid transport and metabolism (E, $6.2\% \pm 0.7\%$), and inorganic ion transport and metabolism (P, $5.3\% \pm 0.8\%$) was significantly higher than that of NPMA strains ($1.9\% \pm 0.5\%$, $3.1\% \pm 0.3\%$, $4.1\% \pm 1.1\%$, and $4.6\% \pm 0.8\%$), while the relative abundances of nucleotide transport and metabolism (F, $2.8\% \pm 0.4\%$) were lower than that of NPMA ($4.1\% \pm 0.4\%$). As for the functional
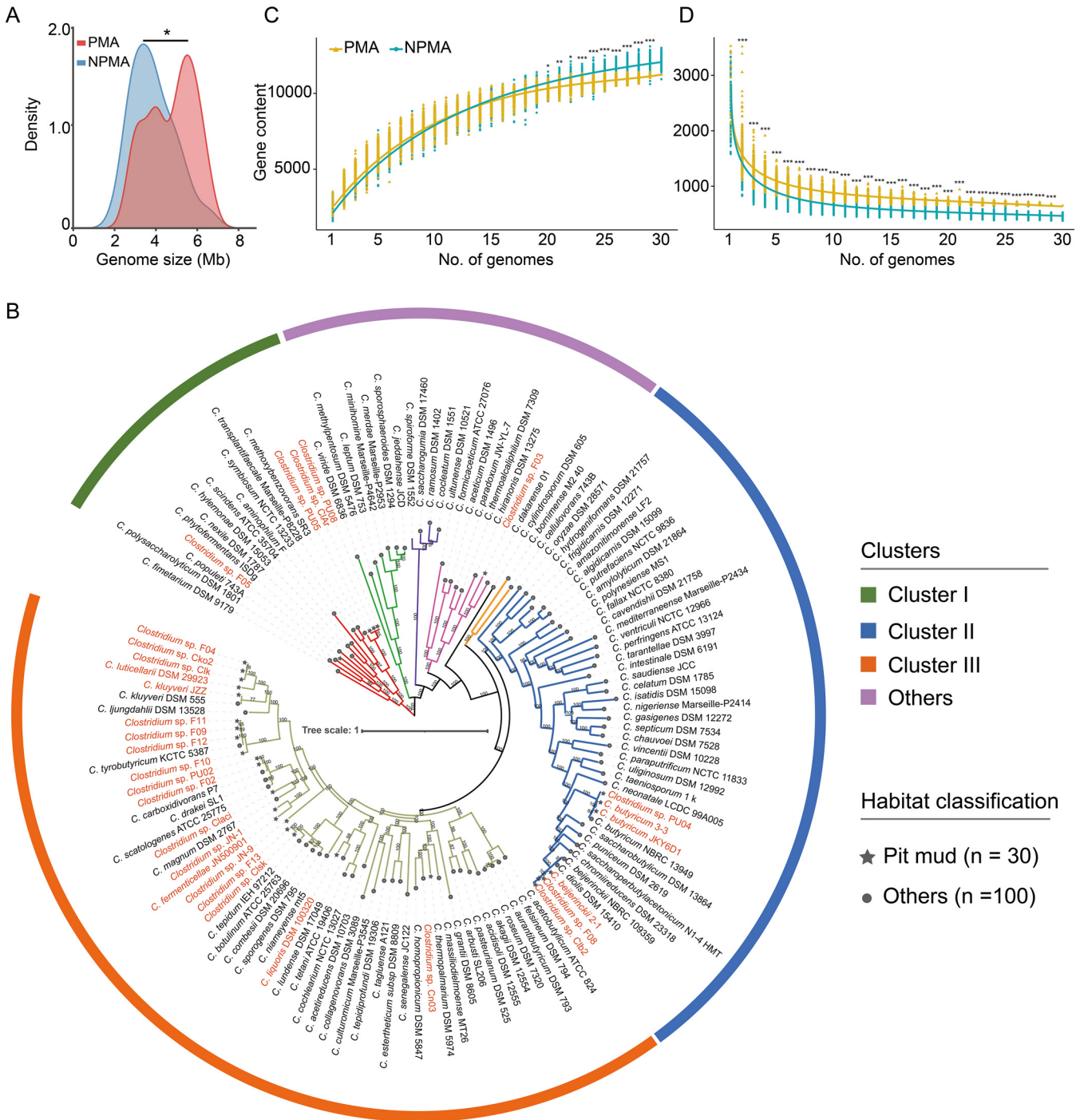
FIG 1 Comparative genomics and phylogenomics of *Clostridium* strains. (A) Genome size comparison between PMA and NPMA *Clostridium* strains. (B) The phylogenetic tree based on the protein sequences of single-copy genes of all *Clostridium* genomes using Fasttree under an approximate maximum-likelihood method. The names of *Clostridium* isolated from pit mud are highlighted in orange. The pan (C)- and core (D) genome sizes of PMA and NPMA *Clostridium* strains when randomly sampled with the same number of genomes are shown (Mann-Whitney U test, *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$).

distribution of core genes based on KEGG database, the pathways were defined as significantly enriched in PMA/NPMA *Clostridium* strains if the P value in the Mann-Whitney U test was less than 0.05 and the difference in the number of genes in the pathway between PMA and NPMA strains was more than two times. The results showed that the pathways related to environmental adaptation, including two-component system, flagellar assembly, and bacterial chemotaxis, were significantly enriched in PMA *Clostridium* strains (Fig. 2B). A large number of genes related to carbohydrate metabolism (pyruvate metabolism,
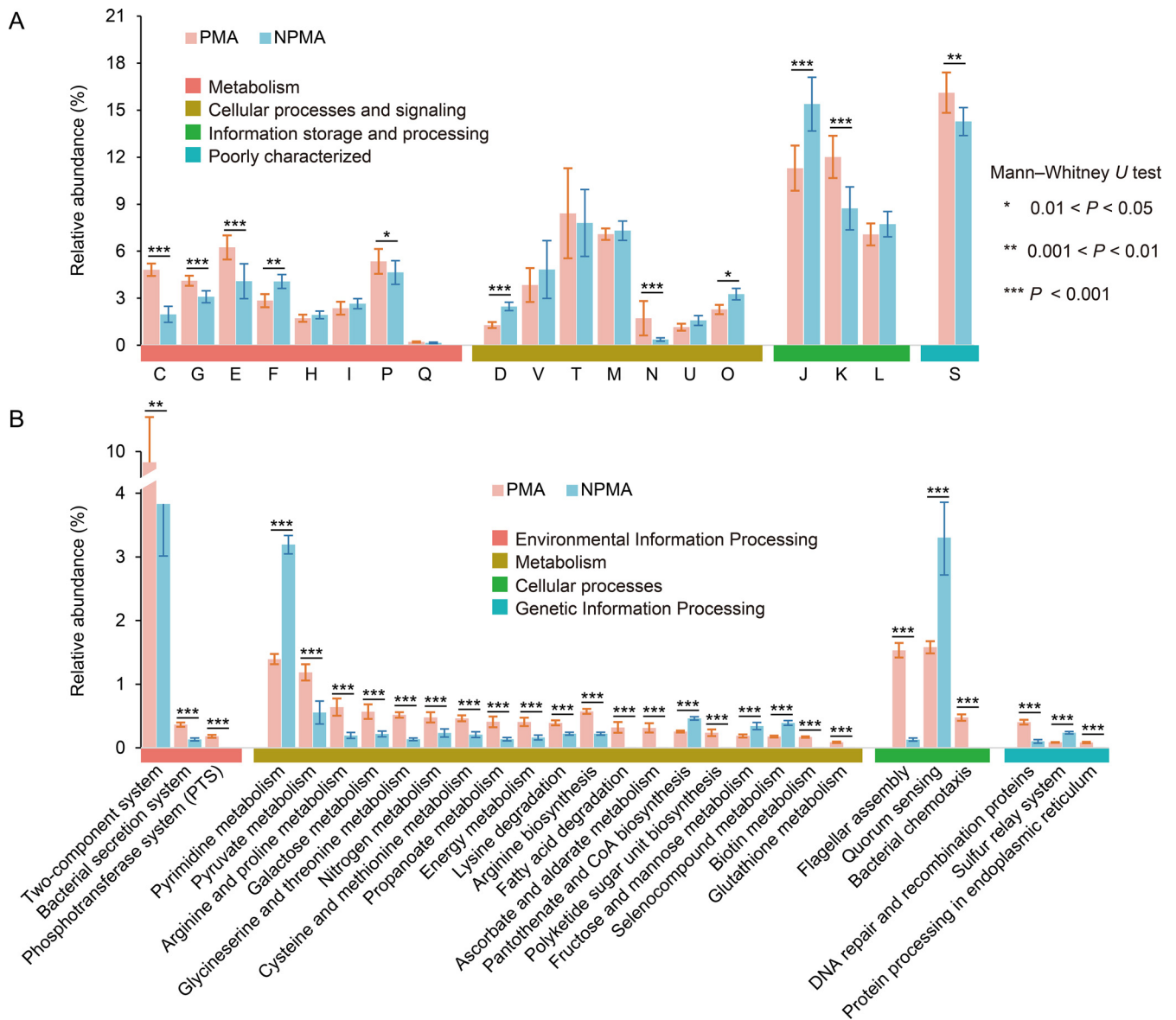
FIG 2 Functional categories of the core genome of pit mud-associated (PMA) and non-pit mud-associated (NPMA) *Clostridium* strains based on COG (A) and KEGG (B) databases (Mann-Whitney *U* test, *, *P* < 0.05; **, *P* < 0.01; ***, *P* < 0.001). Bar charts represent the number of gene categories in statistical tests, and the error bars represent the standard deviations of the number of genes in the core genomes of PMA/NPMA *Clostridium* strains.

galactose metabolism, propanoate metabolism, and ascorbate and aldarate metabolism) and amino acid metabolism (arginine and proline metabolism, glycine-serine and threonine metabolism, cysteine and methionine metabolism, lysine degradation, and arginine biosynthesis) were significantly enriched in PMA *Clostridium* strains, while the pathways of pyrimidine metabolism, pantothenate and coenzyme A (CoA) biosynthesis, fructose and mannose metabolism, selenocompound metabolism, quorum sensing, and sulfur relay system were significantly enriched in NPMA *Clostridium* strains.

Functional enrichment of *Clostridium* in different habitats. Previous studies have shown that the genomic functions of bacteria with close phylogenetic relationships are shaped by their habitats and lifestyles (21, 22). The differences of core genome functions between PMA and NPMA *Clostridium* strains were analyzed above. Here, a comparative analysis of the genome-wide functions enriched in PMA and NPMA *Clostridium* strains was conducted. The analysis based on gene copy numbers showed that more genes were significantly (Mann-Whitney *U* test, *P* < 0.05) enriched in 10 COG categories of PMA *Clostridium* strains relative to NPMA strains, especially posttranslational modification, protein turnover,

chaperones (O), energy production and conversion (C), carbohydrate transport and metabolism (G), and amino acid transport and metabolism (E) ($P < 0.001$) (Fig. 3A). Moreover, the numbers of genes enriched in these four COG categories (O, C, G, and E) were also significantly ($P < 0.001$) higher in PMA *Clostridium* strains than in NPMA strains based on presence/absence (Fig. S3A).

Comparison analysis of the KEGG pathways between PMA and NPMA *Clostridium* strains was performed by PhyloGLM test and phylogenetically informed principal component analysis (phylo-PCA) (Fig. 3B and C). The results showed that the number of genes in 45 pathways was significantly different between PMA and NPMA *Clostridium* strains (PhyloGLM test, $P < 0.05$) (Table S2), and the top 20 pathways are shown in Fig. 3B. Forty-three pathways were enriched in PMA strains, and many of these metabolic pathways were related to environmental adaptation, including two-component system (23), energy metabolism (24), and bacterial chemotaxis (25) pathways; however, only two pathways were enriched in NPMA strains. A large number of genes related to carbohydrate metabolism (glycolysis/gluconeogenesis, citrate cycle, glycosyltransferases, and propanoate metabolism) and amino acid metabolism (glycine-serine and threonine metabolism, lysine biosynthesis, alanine-aspartate and glutamate metabolism, arginine biosynthesis, tyrosine metabolism, and valine-leucine and isoleucine degradation) were enriched in PMA *Clostridium* (Fig. 3B). Phylo-PCA analysis showed that the enriched KEGG pathways in PMA and NPMA strains were divided into two clusters, which were significantly different (Mann-Whitney U test, $P < 0.01$) on both the PC1 and PC2 axes, suggesting that these pathways are related to different habitat characteristics.

Moreover, the clade-based analysis was performed to investigate the differences of gene content between PMA and NPMA *Clostridium* strains on different clades of the phylogenetic tree. The results showed that the pathways of ABC transporters, two-component system, glycolysis/gluconeogenesis, cell growth, energy metabolism, and citrate cycle (tricarboxylic acid [TCA] cycle) were significantly enriched in PMA *Clostridium* in all clusters (Fig. S3B). A large number of genes related to amino acid metabolism were enriched in PMA strains in clusters II and III. The pathways of transcription factors, purine metabolism, and arginine biosynthesis were enriched in PMA strains in clusters I and II, and the prokaryotic defense system pathway was enriched in PMA strains in clusters I and III.

Three hundred and ten KEGG orthologs (KOs) with significant differences between PMA and NPMA *Clostridium* strains were obtained by pan-genome-wide association studies (pan-GWAS) based on the presence/absence data of KOs, of which 281 KOs were enriched in PMA *Clostridium*, and the other 29 KOs were enriched in NPMA *Clostridium* (Fig. 3D). Among the enriched KOs in PMA *Clostridium* strains, it could be observed that 36, 15, 14, 8, and 8 KOs belonged to transporters, sulfur metabolism, ABC transporters, two-component system, and glycolysis/gluconeogenesis, respectively, while for NPMA *Clostridium*, eight KOs belonging to oxidative phosphorylation, two KOs belonging to ABC transporters, and two KOs belonging to two-component system were enriched (Fig. 3D). Pit mud is exposed to the environment of grain fermentation for extended periods of time, and there were a large number of yeasts and lactic acid bacteria living in this environment (3). Due to the metabolism of these microorganisms, a large amount of lactic acid and ethanol will be accumulated in pit mud (12, 26). The stress of the environment with high concentrations of lactic acid and ethanol can inhibit the growth of microorganisms. Therefore, the genes of enzymes that can directly metabolize ethanol and lactic acid were analyzed. The number of genes encoding alcohol dehydrogenase (EC 1.1.1.1, K13954), oxidizing ethanol into acetaldehyde, in PMA *Clostridium* strains was significantly higher than that in NPMA *Clostridium* strains (Fisher's exact test, Benjamini-Hochberg adjusted $P < 0.05$) (Fig. 3E). Ninety percent of PMA *Clostridium* strains had the alcohol dehydrogenase-coding genes, while that of NPMA *Clostridium* strains was only 66.0%. The proportion of strains with the genes encoding L-lactate dehydrogenase (EC 1.1.1.27, K00016) in PMA *Clostridium* (100%) was significantly higher than that in NPMA (80.0%). L-Lactate dehydrogenase plays a role in metabolizing lactic acid into pyruvate. Multiple genes encoding enzymes in the citrate cycle were enriched in PMA *Clostridium*, including
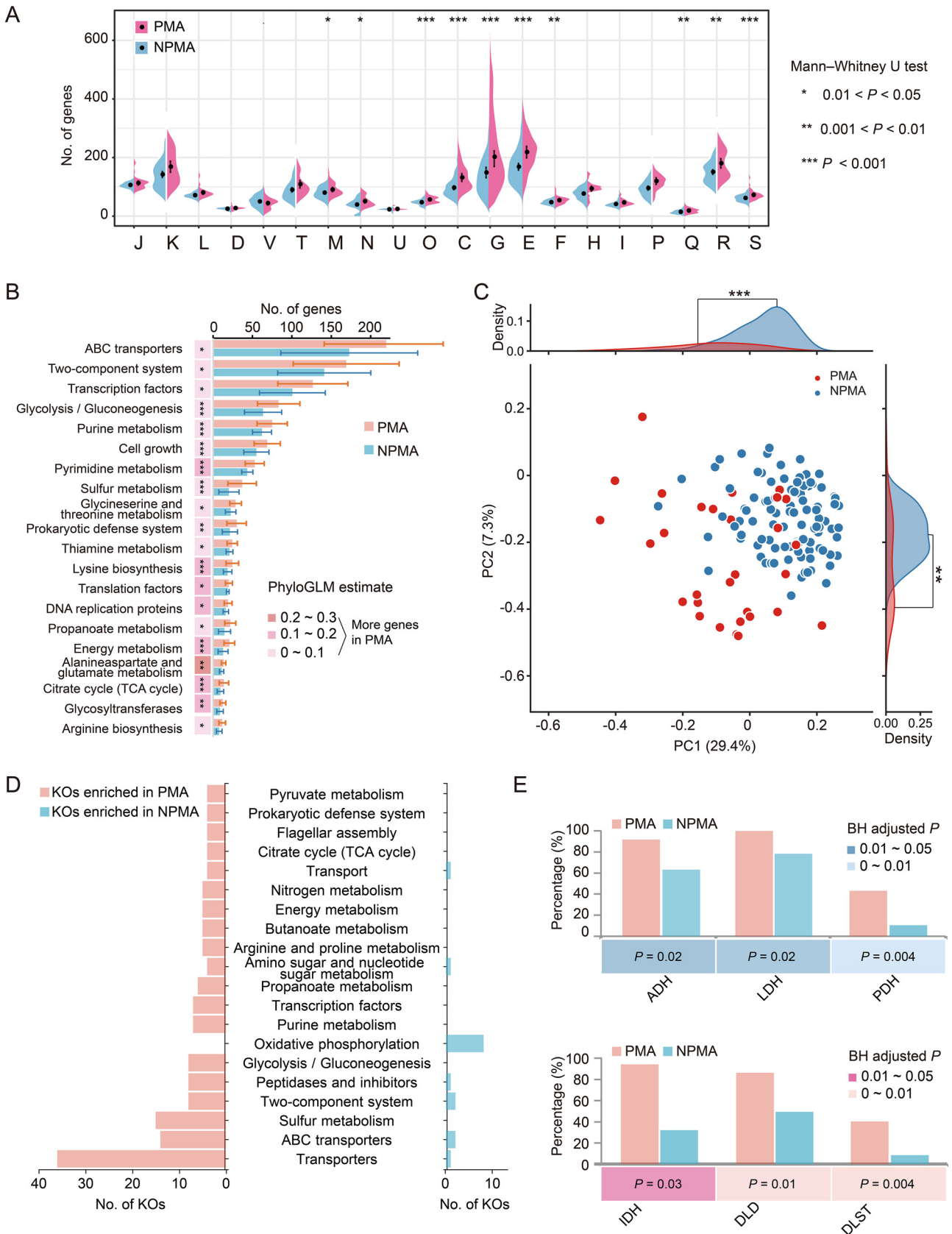
FIG 3 Enrichment analysis of functional categories in pit mud-associated (PMA) and non-pit mud-associated (NPMA) *Clostridium*. (A) Pairwise functional comparison among PMA and NPMA *Clostridium* strain genomes using COG categories. The boxes in the plot represent the gene copy

dihydrolipoamide dehydrogenase (EC 1.8.1.4, K00382), isocitrate dehydrogenase (EC 1.1.1.42, K00031), pyruvate dehydrogenase (EC 1.2.4.1, K00162), and 2-oxoglutarate dehydrogenase (EC 2.3.1.61, K00658) (Fig. 3E). These genes play critical roles in the degradation of isocitric acid, oxalosuccinic acid, succinic acid, and other organic acids (27). Therefore, the enrichment of these genes in PMA *Clostridium* strains may play an important role in adapting to the environment of pit mud rich in organic acids.

The results of phylogenetic tree clade-based analysis showed that there was no statistical difference in the content of these genes between PMA and NPMA *Clostridium* strains in cluster I (Fig. S3C). In cluster II, the number of genes encoding alcohol dehydrogenase (2.0 ± 0.0), L-lactate dehydrogenase (3.9 ± 0.2), and isocitrate dehydrogenase (1.0 ± 0.0) in PMA *Clostridium* strains was significantly higher than that of NPMA strains (1.3 ± 0.7, 1.9 ± 0.6, and 0.7 ± 0.2). In cluster III, the number of genes encoding alcohol dehydrogenase (0.9 ± 0.2), L-lactate dehydrogenase (1.0 ± 0.0), pyruvate dehydrogenase (1.0 ± 0.1), dihydrolipoamide dehydrogenase (0.7 ± 0.2), and 2-oxoglutarate dehydrogenase (0.6 ± 0.2) in PMA *Clostridium* strains was significantly higher than that of NPMA strains (0.7 ± 0.2, 0.8 ± 0.2, 0.7 ± 0.2, 0.2 ± 0.1, and 0.2 ± 0.1).

CAZyme profiling. The above-described analysis showed that a large number of carbohydrate metabolism genes encoding enzymes in glycolysis/gluconeogenesis, citrate cycle, glycosyltransferases, and propanoate metabolism were enriched in PMA *Clostridium* strains (Fig. 3B and Table S2). Pit mud is in a carbohydrate-rich environment; therefore, carbohydrate-active enzymes (CAZymes) were analyzed using dbCAN2 to investigate the genomic potential for carbohydrate utilization of all *Clostridium* strains. There were 266 CAZyme families (13,590 genes) identified in all strains, including 145 glycoside hydrolases (GHs), 15 carbohydrate esterases (CEs), 24 polysaccharide lyases (PLs), 29 glycosyl transferases (GTs), 42 carbohydrate-binding protein module (CBM) families, and seven auxiliary activities (AAs). The number of genes encoding CAZymes varied greatly between *Clostridium* strains, ranging from 20 to 284 (Fig. 4A and Fig. S4A).

GHs and GTs were the two abundant families of CAZymes in all *Clostridium* strains, with the number of genes encoding GHs and GTs accounting for 43.5% and 29.0%, respectively, of all CAZyme-coding genes (Fig. 4A and Fig. S4A). The GH superfamily is involved in glucan hydrolysis, including starch, cellulose, xylan, chitin, hemicellulose, and glycogen (28). The number of genes encoding GHs in PMA and NPMA *Clostridium* strains did not differ significantly. Accordingly, the number of genes encoding CAZymes between PMA and NPMA *Clostridium* strains was compared based on different clusters in the phylogenetic tree (Fig. 1B). The number of genes encoding GHs in PMA strains from cluster I (112.5 ± 31.0) was significantly higher than that of NPMA strains (64.1 ± 47.9), while there was no significant difference for strains on cluster II (PMA, 77.4 ± 15.6; NPMA, 58.0 ± 37.8) and III (PMA, 19.9 ± 7.1; NPMA, 27.6 ± 20.8) (Fig. S4B). GTs are essential for producing diverse and complex glycoconjugates (29). PMA *Clostridium* strains had significantly more genes encoding GTs than NPMA *Clostridium* strains (Fig. 4A). PMA strains on clusters I (31.3 ± 1.5) and II (39.1 ± 5.7) possessed significantly more genes encoding GTs than NPMA strains (cluster I, 26.1 ± 7.4; cluster II, 32.1 ± 10.6), while there was no significant difference between PMA (30.8 ± 9.8) and NPMA (29.2 ± 12.6) *Clostridium* strains on cluster III (Fig. S4C).

A PhyloGLM test was performed to detect the significantly different CAZymes between PMA and NPMA *Clostridium*. The results showed that the number of genes

FIG 3 Legend (Continued)
number of COG categories (Mann-Whitney *U* test, *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$). (B) The main KEGG pathways with significant differences in the number of genes (PhyloGLM test, $P < 0.05$, estimate > 0). Pink color (PhyloGLM estimate > 0) indicates more genes in the genomes of PMA *Clostridium* strains than NPMA strains, while blue color (PhyloGLM estimate < 0) indicates more genes are enriched in NPMA strains. PhyloGLM test, *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. Bar charts represent the number of genes in this pathway in statistical tests, and the error bars represent the standard deviations of the number of genes in PMA/NPMA *Clostridium* strains. (C) Phylogenetically informed principal-component analysis (phylo-PCA) of the pathways with significant differences in gene content between PMA and NPMA *Clostridium* strains. (D) The 310 KOs with significant differences in the number of genes between PMA and NPMA *Clostridium* strains obtained in a stringent version of Scoary, a gene presence/absence approach that combines Fisher's exact test, a phylogenetic test, and a label permutation test. (E) Comparison of the number of genes encoding alcohol dehydrogenase (ADH), lactate dehydrogenase (LDH), pyruvate dehydrogenase (PDH), isocitrate dehydrogenase (IDH), dihydrolipoamide dehydrogenase (DLD), and 2-oxoglutarate dehydrogenase (DLST).
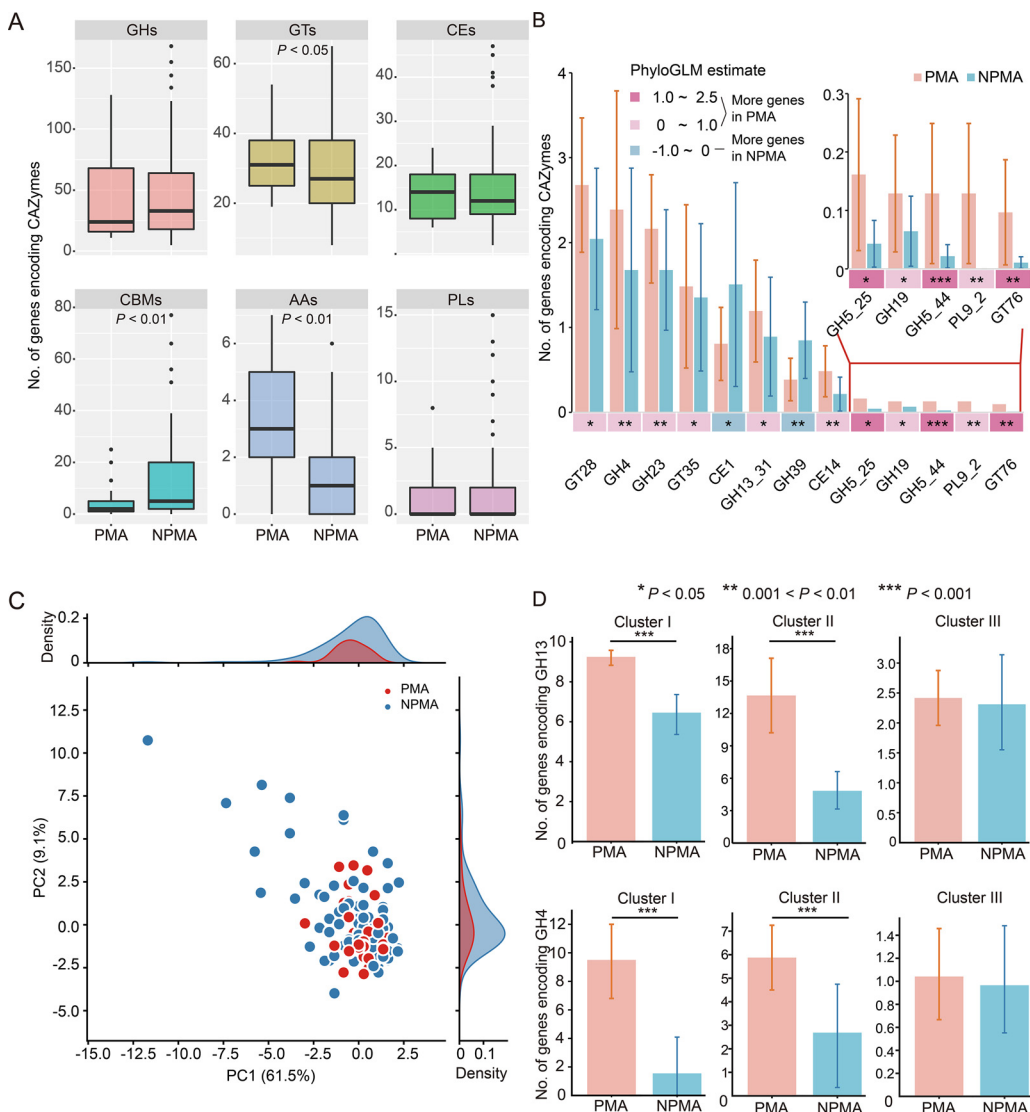
FIG 4 Carbohydrate-active enzymes (CAZymes) in all *Clostridium* strains. (A) Pairwise gene number of CAZyme comparisons among PMA and NPMA *Clostridium* strains. Box plot represents the diversity of CAZyme family content based on copy number data set (Mann-Whitney U test, *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$) of auxiliary activities (AAs), carbohydrate-binding molecules (CBMs), carbohydrate esterases (CEs), glycoside hydrolases (GHs), glycosyltransferases (GTs), and polysaccharide lyases (PLs). (B) The significantly enriched CAZymes between PMA and NPMA *Clostridium* strains (PhyloGLM test, $P < 0.05$). Pink color (PhyloGLM estimate > 0) indicates more genes in the genomes of PMA *Clostridium* strains than NPMA strains, while blue color (PhyloGLM estimate < 0) indicates more genes are enriched in NPMA strains. PhyloGLM test, *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. Bar charts represent the number of genes in this CAZyme family in statistical tests, and the error bars represent the standard deviations of the number of genes in PMA/NPMA *Clostridium* strains. (C) Phylogenetically informed principal component analysis (phylo-PCA) of the significantly enriched genes encoding CAZymes. (D) Comparison of the number of genes encoding GH4 and GH13 between PMA and NPMA *Clostridium* strains.

encoding 13 CAZymes varied significantly between PMA and NPMA *Clostridium* strains (PhyloGLM test, $P < 0.05$). Eleven CAZymes (GT28, GH4, GH23, GT35, GH13_31, CE14, GH5_25, GH19, GH5_44, PL9_2, and GT26) were enriched in PMA *Clostridium*, and only two CAZymes (CE1 and GH39) were enriched in NPMA *Clostridium* (Fig. 4B). Phylo-PCA analysis showed the NPMA *Clostridium* strains were scattered on the first two principal components while the PMA *Clostridium* strains were aggregated (Fig. 4C), suggesting that the composition of the genes encoding CAZymes in PMA *Clostridium* were similar. The differences in CAZymes among NPMA *Clostridium* strains may be due to the wide range of their habitats.

GH4 and GH13 mainly included α-glucosidase, α-amylase, maltose-6-phosphate glucosidase, and maltogenic amylase, and these enzymes were involved in starch,

oligosaccharide, and glycogen metabolic pathways (30). Sorghum is the main material for baijiu fermentation, providing a starch-rich environment for pit mud microorganisms. Therefore, the analysis of GH4- and GH13-coding genes was conducive to understanding the potential of *Clostridium* to metabolize carbohydrates. The abundant GH13 families included GH13_1, GH13_2, GH13_4, GH13_5, GH13_8, GH13_9, GH13_11 GH13_13, GH13_14, GH13_15, GH13_18, GH13_19, GH13_20, GH13_21, GH13_28, GH13_29, GH13_31, and GH13_36. On clusters I and II, the number of genes encoding GH13 in PMA *Clostridium* strains (cluster I, 9.25 ± 0.5; cluster II, 14.5 ± 4.4) was significantly higher than that in NPMA *Clostridium* strains (cluster I, 7.3 ± 3.4; cluster II, 7.1 ± 4.5), while there was no significant difference between PMA (2.4 ± 1.2) and NPMA (2.3 ± 2.3) *Clostridium* strains on cluster III (Fig. 4D). Similar to the genes encoding GH13, the number of genes encoding GH4 in PMA *Clostridium* strains (cluster I, 9.5 ± 5.0; cluster II, 5.9 ± 2.0) on clusters I and II was significantly higher than that in NPMA strains (cluster I, 1.5 ± 1.4; cluster II, 2.7 ± 2.3), and there was no significant difference between PMA and NPMA *Clostridium* strains on cluster III (Fig. 4D).

Horizontal gene transfer in PMA *Clostridium* strains. Horizontal gene transfer (HGT) plays a crucial role in the environmental adaptation of bacteria. Bacteria can share resistance and metabolic genes through HGT under the stress of the environment (31, 32). The BLAST-based method was used to infer the HGT events among PMA strains to evaluate the roles of HGTs in pit mud adaptation of PMA bacteria. Fifty-six genomes of bacteria isolated from pit mud were used for HGT analysis, and the ecological environments of pit mud were similar due to the brewing process with sorghum as the main raw ingredient. These pure strains isolated from pit mud used for HGT analysis mainly affiliated with *Clostridia* and *Bacilli* at the class level and *Clostridium* and *Bacillus* at the genus level (Table S3), which attracted researchers' attention because of their high abundance and their important functions in the flavor formation of strong-flavor baijiu (1, 3, 4). A total of 161 HGT events were detected in all PMA bacterial strains. These HGT events mainly occurred between the classes *Clostridia* and *Bacilli*, and the number of genes transferred from *Bacilli* to *Clostridia* was higher than that from *Clostridia* to *Bacilli* (Fig. S5A to D). At the genus level, there were a lot of HGT events that occurred between the genera *Clostridium* and *Bacillus* (Fig. 5A). Functional annotation of the horizontally transferred genes was performed by running similarity searches against the COG database (Fig. 5B). The results showed that the main functions of these horizontally transferred genes were translation, ribosomal structure and biogenesis (J), transcription (K), amino acid transport and metabolism (E), and energy production and conversion (C). The results of KEGG annotation showed that multiple genes encoding the enzymes involved in starch and sucrose metabolism pathway, including maltase-glucoamylase, maltose-6′-phosphate glucosidase, 6-phospho-beta-glucosidase, and sucrase-isomaltase, were transferred from *Bacillus* to *Clostridium* (Fig. S5F), and these enzymes play important roles in the metabolism of starch. The transferred genes encoding CAZymes were detected against the CAZy database (Fig. 5C). Genes encoding CAZymes were mainly transferred from *Bacilli* to *Clostridia* (Fig. S5E). Multiple genes encoding GH13 and GH135 were transferred from *Bacillus* to *Clostridium* (Fig. S5G). GH13 can be directly involved in the metabolism of glycogen and starch (30), which was enriched in PMA *Clostridium* strains (Fig. 4D). There were several genes encoding GHs transferred among *Clostridium* strains. Multiple genes encoding GTs, CEs, and AAs were transferred from other bacterial genera to *Clostridium* or among *Clostridium* strains.

Pathway analysis of the synthesis of short-chain fatty acids. The pit mud is enriched with high concentrations of ethanol, lactic acid, and starch, which is significantly different from other habitats (3, 12). Ethanol, lactic acid, and starch are important precursors for the synthesis of short-chain fatty acids, which are the key flavor substances of baijiu (33). Therefore, the potential for synthesizing short-chain fatty acids from these substances was compared between PMA and NPMA *Clostridium* strains (Fig. 6). The proportion of PMA *Clostridium* strains (93.3%) containing complete metabolic pathways from starch to glucose was higher than that of NPMA *Clostridium* strains (77.0%). To determine significance of differences in metabolic pathways between PMA *Clostridium* strains and NPMA *Clostridium* strains, we calculated the expected number of strains with the complete pathways and the
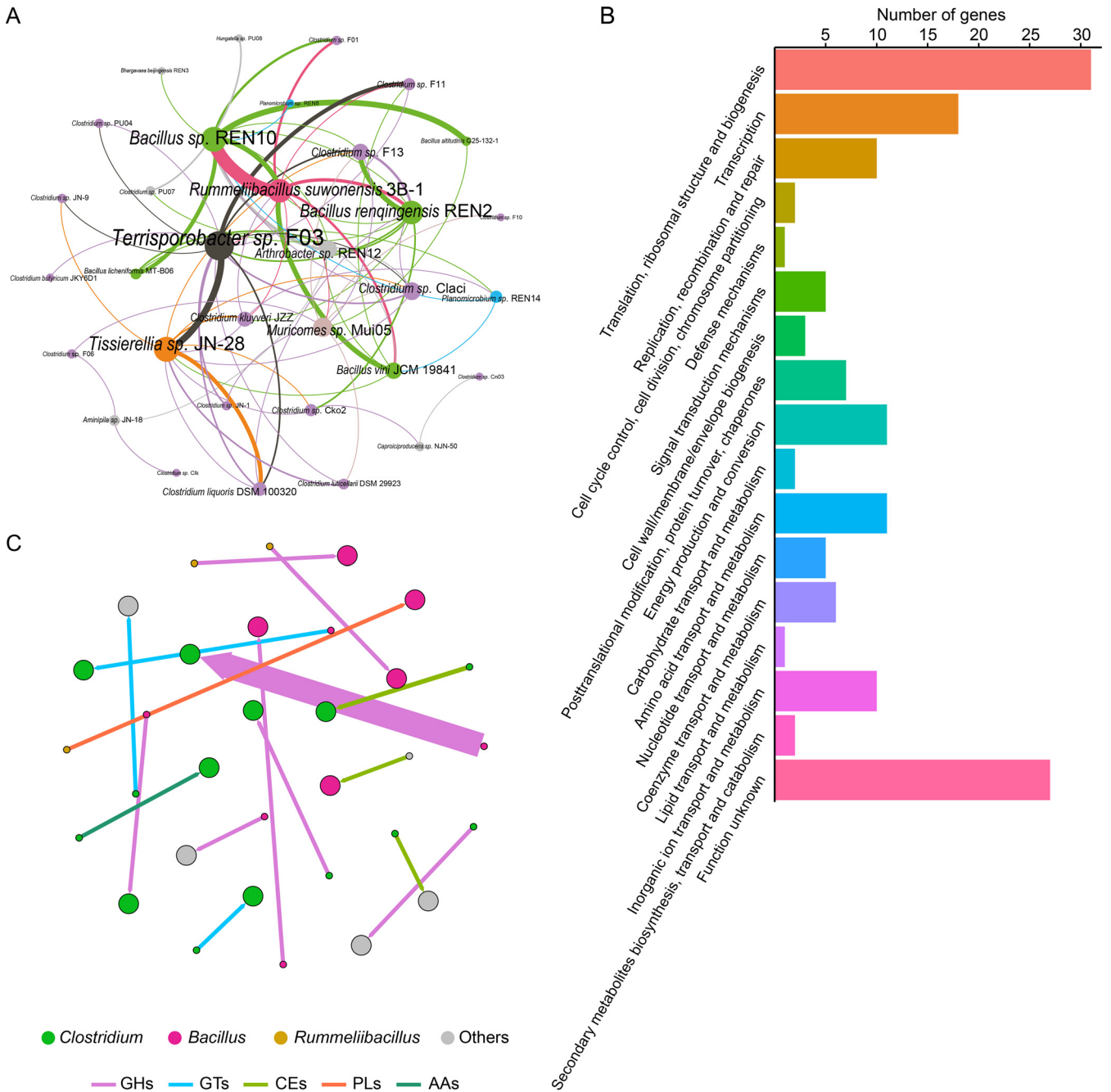
FIG 5 Horizontal gene transfer between pit mud bacteria. (A) Horizontal gene transfer among different strains of pit mud bacteria. The color of lines was the same as that for donors. (B) The functional categories of the horizontally transferred genes based on COG databases. (C) Transferred genes encoding CAZymes between pit mud bacteria.

observed number of strains and calculated a *P* value using the Poisson distribution (ppois R function). Glucose can be metabolized into pyruvate through glycolysis/gluconeogenesis, and there was no significant difference in the number of strains with this pathway between PMA and NPMA *Clostridium* strains. Lactic acid can be oxidized to pyruvate by lactate dehydrogenase (EC 1.1.1.27), and pyruvate can be metabolized to acetyl-CoA. All of the strains have genes encoding the enzymes that convert pyruvate to acetyl-CoA, while the number of strains with the gene encoding lactate dehydrogenase was significantly different between PMA and NPMA *Clostridium* strains. The proportion of the PMA *Clostridium* strains (100%) with the genes encoding lactate dehydrogenase was significantly higher than that in NPMA *Clostridium* strains (80.0%) (Poisson distribution, $P < 0.05$). There was no
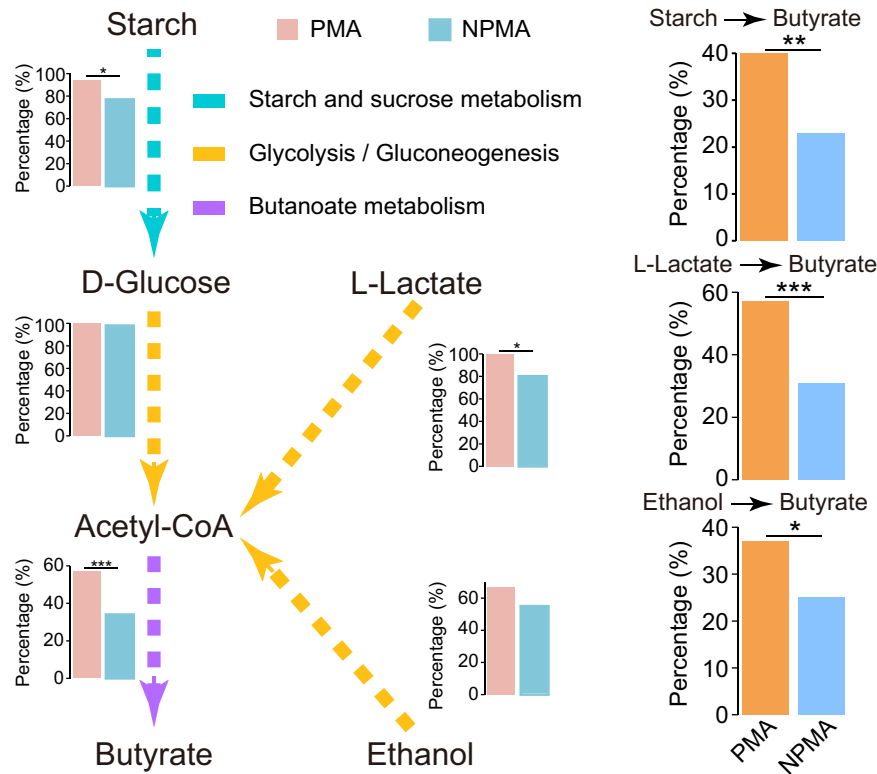
FIG 6 Metabolic pathways analysis of synthesis of butyrate from ethanol, lactic acid, and starch in *Clostridium* strains. The histogram represents the proportion of strains with these genes in this pathway, and we calculated a *P* value using the Poisson distribution (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$).

significant difference in the number of strains with complete metabolic pathway from ethanol to acetyl-CoA between PMA (66.7%) and NPMA (55.0%) *Clostridium* strains.

Butyrate can be synthesized from acetyl-CoA through butanoate metabolism, and the number of PMA *Clostridium* strains (56.7%) with this metabolic pathway was significantly higher than that of NPMA *Clostridium* strains (34.0%) (Poisson distribution, $P < 0.001$). The complete pathways of butyrate synthesis from starch, lactic acid, and ethanol of all *Clostridium* strains then were detected to estimate the potential for synthesis of butyrate. The proportion of the PMA *Clostridium* strains (40.0%, 56.7%, and 36.7%, respectively) synthesizing butyrate from starch, lactic acid, and ethanol was significantly higher than that of the NPMA *Clostridium* strains (23.0%, 31.0%, and 25.0%, respectively) (Poisson distribution, $P < 0.05$).

DISCUSSION

Microorganisms will be filtered by the specific environments of their habitats, so only a part of them can survive in the specific habitat, which often leads to significant differences in the microbial genome features among the habitats (34–36). *Clostridium* strains were isolated from a range of habitats, as summarized in the EzBioCloud database (10). *Clostridium* was the predominant bacteria in pit mud during Chinese strong-flavor baijiu production, playing key roles in the production of short-chain fatty acids in baijiu (7). Due to the long-term cereal fermentation with sorghum as the main raw material, pit mud is enriched with plenty of organic acids, ethanol, and starch (3, 12). The genomic adaptation of *Clostridium* to the habitat of pit mud remains poorly understood. Here, a comparative genomics analysis was performed on the 130 *Clostridium* genomes to provide insights into the genomic features of *Clostridium* strains isolated from pit mud.

Bacteria that live in a nutrient-enriched environment tend to have larger genomes, and gene function is more redundant in these bacteria (37). It was found that the average genome size of PMA *Clostridium* strains was significantly larger than that of NPMA *Clostridium* strains. This may be a result of pit mud's being rich in carbohydrates and amino acids, while the contents of nutrients of NPMA environments are relatively low compared with pit mud environment (10, 38). Bacteria with larger genomes accumulate secondary metabolism, and we observed that high rates of the regulatory genes in these bacteria control the large number of metabolites expressed under different growth conditions (24). Due to the great range of substrates for energy production, bacteria with a large genome size can survive in diverse habitats (24). COG and KEGG function analyses of genomes revealed that genes related to carbohydrate and amino acid metabolism were enriched in PMA *Clostridium* strains (Fig. 3A and B), suggesting that this could help them utilize carbohydrates and amino acids more effectively. Energy metabolism-related genes were important regulatory elements of the bacterial response to environmental stress (39). The pathways of ABC transporters, two-component system, flagellar assembly, and bacterial chemotaxis were also enriched in PMA *Clostridium* strains, most of which were related to environmental adaptation. The two-component system was the main form of bacterial signal transduction, which could regulate the ability of bacteria to perceive and respond to the external environment (23). Bacterial chemotaxis could help the motile bacteria respond to the chemical concentration gradient of substances, making the bacteria tend to beneficial stimuli and avoid harmful stimuli (40). These functions were enriched in PMA *Clostridium* strains, indicating these genes play key roles in the adaptation of PMA *Clostridium* strains to pit mud environment.

Large amounts of lactic acid and ethanol were produced due to the *Lactobacillaceae* and *Kazachstania* dominating the fermentation process of strong-flavored baijiu, and their relative abundances can reach up to 90% (3). High concentrations of lactic acid and ethanol became the growth inhibitors for microorganisms during the fermentation process of strong-flavor baijiu (41, 42). Previous studies showed that during the fermentation of strong-flavor baijiu, the pH of fermented grains could be lower than 4.0, and the pH of pit mud contacted with fermentation grains was 5.87 (26). It is a challenge for most microorganisms to survive in this acidic environment (43). Pan-GWAS analysis showed that the proportion of PMA *Clostridium* strains with the gene encoding alcohol dehydrogenase and L-lactate dehydrogenase was significantly higher than that of NPMA *Clostridium* strains, and the enrichment is beneficial for the PMA *Clostridium* strains to counteract the stress of lactic acid and ethanol in the fermentation environment. In previous work, good acid resistance and the potential of ethanol and lactic acid metabolism of a *Bacillus* strain and a *Clostridium* strain isolated from strong-flavor baijiu fermentation environment were detected through the analysis of genomic and phenotypic features (44, 45). Further, the enrichment of PMA *Clostridium* strains of the genes involved in citrate cycle may contribute to the metabolism of organic acids and the survival of in the organic acid-rich environment of pit mud (3).

There were significant differences in the type and amount of the genes associated with carbohydrate utilization among *Clostridium* strains in different habitats (46). GHs were the most abundant family of CAZymes in PMA *Clostridium* strains. GHs were always produced by bacteria to degrade polysaccharides to release the key resources for their growth (47, 48). The genes encoding GTs were enriched in PMA *Clostridium* strains. Exopolysaccharides could be synthesized through the family of GTs, which was conducive to the formation of biofilm by bacteria (49). Moreover, GTs were the key enzymes in the process of glycosylation, which was important for the formation of bacterial cell walls (49). The formation of biofilm and cell walls contributed to the resistance of bacteria to environmental stress (50). Therefore, GTs play key roles in the adaptation to the environment of bacteria. GH13 was observed in most bacteria that belonged to the families targeting starch and glycogen (30). GH4 is involved in starch and sucrose metabolism (51). Since starch is the most abundant carbohydrate in pit

mud, the enrichment of the genes encoding GH13 and GH4 was helpful for the utilization of resources by PMA *Clostridium* strains in pit mud.

Horizontal gene transfer plays a crucial role in the evolution of bacterial genomes (14, 31). The horizontally transferred genes among PMA strains were similar to those in recent studies, where the most frequent functions of the HGT genes belonged to COG categories C, G, and J (32, 52). Further, these functions (including C, G, E, etc.) were enriched in PMA *Clostridium* strains, indicating HGTs contributed to the functionality of PMA *Clostridium* strains. In this study, a large number of genes associated with carbohydrate and amino acid metabolism were transferred from *Bacillus* to *Clostridium*, which was consistent with a previous study that found bacteria could share nutrients in the environment by transferring genes related to metabolism (31). Multiple genes encoding GH4, GH13, and GH135 were transferred from *Bacillus* to *Clostridium*. GH13 can be directly involved in the metabolism of glycogen and starch (30). GH4 and GH13 were enriched in PMA *Clostridium* strains, indicating that HGTs increased the genes encoding the enzymes for carbohydrate utilization in PMA *Clostridium* strains, which was conducive to the utilization of resources by PMA *Clostridium* strains.

The short-chain fatty acids produced by *Clostridium* are important flavor compounds in Chinese strong-flavor baijiu (7–9). Butyric acid is synthesized via a carboxylic acid chain elongation process that employs reverse oxidation of acetic acid as an electron donor (53, 54). Our results show that the number of genes encoding the enzymes (alcohol dehydrogenase and L-lactate dehydrogenase) that metabolize ethanol and lactic acid directly in PMA *Clostridium* strains was significantly higher than that of NPMA *Clostridium* strains. Not only being easily converted to acetyl-CoA, the oxidation of ethanol and lactic acid offers a number of reducing equivalents that can support medium-chain carboxylic acid synthesis (55). The good potential of PMA *Clostridium* strains to synthesize short-chain fatty acids from ethanol and lactic acid can not only produce key flavor compounds for strong-flavor Baijiu but also alleviate the stresses of ethanol and lactic acid in pit mud (45). Short-chain fatty acids produced by *Clostridium* metabolism also can be used as stress factors for other microorganisms to improve their competitiveness in niches (56). Although there were significant differences in the proportion of *Clostridium* strains with a complete butyric acid synthesis pathway between PMA and NPMA *Clostridium* strains, in practice, we should not ignore the interaction of microorganisms because it can significantly affect the growth and metabolism of microorganisms (57).

In conclusion, the genetic basis of PMA *Clostridium* strains' habitat adaptation and metabolic profiles was investigated. Many genes related to environmental adaptation, including two-component system, flagellar assembly, and bacterial chemotaxis, were enriched in PMA *Clostridium* strains. Genes related to ethanol, organic acid, and carbohydrate metabolism were enriched in PMA *Clostridium* strains that not only contributed to their dominance in the nutrient complex environment but also ensured their great advantage in utilizing the substances enriched in pit mud. PMA *Clostridium* strains showed good potential to produce short-chain fatty acids from starch, lactic acid, and ethanol. Furthermore, it should be noted that *Clostridium* species are highly heterogeneous, and consequently, the species-level diversity of strains in the pit mud and non-pit mud groups might have a certain impact on their comparison results. On the basis of this study, expanding the species pool of *Clostridium* isolated from pit mud will help to improve the understanding of their habitat adaptability and metabolic profiles.

MATERIALS AND METHODS

Collection of data sets for comparative genomic analysis. In this study, 21 *Clostridium* strains were isolated from the pit mud of strong-flavor baijiu fermentation cellars according to our previous work (58). According to the species taxonomic information in the EzBioCloud database (10), 109 genomes of *Clostridium* strains were downloaded from the NCBI-GenBank database (https://www.ncbi.nlm.nih.gov/; data collected in December 2019), including four type and five nontype strains isolated from pit mud and 100 type strains isolated from non-pit mud habitats (e.g., soil and feces). The detailed

information of these 130 *Clostridium* strains used for comparative genomic analysis is shown in Table S1 in the supplemental material.

The genomic DNA of 21 *Clostridium* strains isolated from pit mud was extracted using a bacterial genomic DNA isolation kit (GK1072; Generay). The qualified DNA was sent to Beijing Genomics Institute (Shenzhen, China) for whole-genome sequencing using an Illumina HiSeq 4000 system (Illumina, San Diego, CA). Briefly, DNA was sheared randomly by a Bioruptor ultrasonicator (Diagenode, Denville, NJ) to construct libraries for paired-end (2 × 150 bp) sequencing. FastQC v0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to check the quality of raw reads. To obtain clean reads, raw reads were trimmed by Trimmomatic v0.39 (http://www.usadellab.org/cms/?page=trimmomatic) to filter out the low-quality reads (59). Clean reads were used for genome assembly using SOAPdenovo v1.05 software (60).

Average nucleotide identity (ANI) based on BLAST (ANIb) was calculated using PYANI v0.2.11 to compare the similarity between all of the *Clostridium* strain genomes shown in Table S1 (61). A heat map was created for the visualization of ANIb values (Fig. S2). A genome pair was considered redundant when the ANIb was greater than 99.9%, and in such cases, one of the two genomes was selected randomly and kept in the data set. Finally, 130 nonredundant *Clostridium* genomes were obtained for further analysis. OrthoFinder v2.3.7 was used to identify the orthogroups of all *Clostridium* strains (Table S1) with default parameters, and the pan-, core-, and species-specific gene families were identified (62). The amino acid sequences of 304 single-copy core genes then were extracted from the OrthoFinder output files and aligned using MAFFT v7.474 (63). A maximum-likelihood phylogenetic tree was constructed based on the single-copy core genome using Fasttree v2.1.9 (64), which was visualized via iTOL (https://itol.embl.de/) (65).

Genome function and functional enrichment analysis. Open reading frames (ORFs) of the genomes were predicted using Prokka v1.14.6 (66). The amino acid sequences of these putative ORFs were aligned against Cluster of Orthologous Groups (COG; version 2014-11-10) and Kyoto Encyclopedia of Genes and Genomes (KEGG; version 89.1) databases to obtain their corresponding annotations with an E value threshold of <1e−3, using Diamond v0.9.24 based on similarity searches (67). The pan-genome-wide association studies (pan-GWAS) analysis was performed for identification of habitat-enriched KEGG orthogroups (KOs) using Scoary v1.6.16 depending on the KO presence/absence data set (68). The metabolic pathways of short-chain fatty acids from glucose, ethanol, lactic acid, and amino acids were analyzed based on KEGG database. To identify the genes encoding carbohydrate-active enzymes (CAZymes), the putative ORFs were aligned against the CAZy database (v7) using dbCAN2 and filtered with the default parameters for bacteria (69). Mann-Whitney *U* test (cutoff, $P < 0.05$) was performed for pairwise functional comparison between PMA and NPMA strains based on the gene copy numbers of COG and CAZyme categories. PhyloGLM test was used to infer the enrichment and depletion of KEGG pathways and CAZyme-encoding genes through the R package phylolm v2.6.2 (70). Phylogenetically informed principal component analysis (phylo-PCA) was used to visualize the differences in gene compositions between PMA and NPMA *Clostridium* strains using the R package phytools v0.7-80 (71).

Identification of HGT among pit mud-associated bacteria. Horizontal gene transfer (HGT) is conducive to the rapid adaptation of bacteria to unstable environments through acquiring novel functions, and it was identified as the main factor for bacteria to adapt to the human-managed environments, including food systems (31, 72). Therefore, we collected the genomes of all pit mud-associated strains reported so far for HGT analysis, including 30 *Clostridium* strains shown in Table S1, 18 non-*Clostridium* bacterial genomes downloaded from NCBI-GenBank database by searching the keyword "pit mud," and eight non-*Clostridium* bacterial strains isolated from pit mud and sequenced in this study (Table S3). MetaChip v1.10.0 was used to analyze the HGTs among the 56 PMA strains at the phylum, class, order, family, genus, and species levels with default parameters (73). The candidate HGT genes were identified among defined taxonomic groups using a best-hit approach, and the candidate genes were assessed according to the phylogenetic analysis and reconciliation of gene and species trees. The HGTs detected at all classification levels were merged and the replicated HGTs filtered out. The results of HGTs were visualized using gephi v0.9.2 (74).

Data availability. The raw genome sequencing results of 30 strains isolated from pit mud in this study have been deposited in the Genome Sequence Archive in National Genomics Data Center (https://bigd.big.ac.cn/gsub/) (75), China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences, under the accession number CRA005913.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.4 MB.
FIG S2, PDF file, 1.1 MB.
FIG S3, PDF file, 1.1 MB.
FIG S4, PDF file, 1.5 MB.
FIG S5, PDF file, 1.3 MB.
TABLE S1, DOCX file, 0.04 MB.
TABLE S2, DOCX file, 0.02 MB.
TABLE S3, DOCX file, 0.02 MB.

REFERENCES

1. Liu MK, Tang YM, Guo XJ, Zhao K, Tian XH, Liu Y, Yao WC, Deng B, Ren DQ, Zhang XP. 2017. Deep sequencing reveals high bacterial diversity and phylogenetic novelty in pit mud from Luzhou Laojiao cellars for Chinese strong-flavor Baijiu. Food Res Int 102:68–76. https://doi.org/10.1016/j.foodres.2017.09.075.

2. Gao J, Liu G, Li A, Liang C, Ren C, Xu Y. 2021. Domination of pit mud microbes in the formation of diverse flavour compounds during Chinese strong aroma-type Baijiu fermentation. LWT 137:110442. https://doi.org/10.1016/j.lwt.2020.110442.

3. Qian W, Lu ZM, Chai LJ, Zhang XJ, Li Q, Wang ST, Shen CH, Shi JS, Xu ZH. 2021. Cooperation within the microbial consortia of fermented grains and pit mud drives organic acid synthesis in strong-flavor Baijiu production. Food Res Int 147:110449. https://doi.org/10.1016/j.foodres.2021.110449.

4. Chai LJ, Qian W, Zhong XZ, Zhang XJ, Lu ZM, Zhang SY, Wang ST, Shen CH, Shi JS, Xu ZH. 2021. Mining the factors driving the evolution of the pit mud microbiome under the impact of long-term production of strong-flavor Baijiu. Appl Environ Microbiol 87:e00885-21. https://doi.org/10.1128/AEM.00885-21.

5. Hu X, Du H, Ren C, Xu Y. 2016. Illuminating anaerobic microbial community and co-occurrence patterns across a quality gradient in Chinese liquor fermentation pit muds. Appl Environ Microbiol 82:2506–2515. https://doi.org/10.1128/AEM.03409-15.

6. Cui Y, Yang K-L, Zhou K. 2021. Using co-culture to functionalize Clostridium fermentation. Trends Biotechnol 39:914–926. https://doi.org/10.1016/j.tibtech.2020.11.016.

7. Chai LJ, Xu PX, Qian W, Zhang XJ, Ma J, Lu ZM, Wang ST, Shen CH, Shi JS, Xu ZH. 2019. Profiling the Clostridia with butyrate-producing potential in the mud of Chinese liquor fermentation cellar. Int J Food Microbiol 297:41–50. https://doi.org/10.1016/j.ijfoodmicro.2019.02.023.

8. Xu J, Sun L, Xing X, Sun Z, Gu H, Lu X, Li Z, Ren Q. 2020. Culturing bacteria from fermentation pit muds of baijiu with culturomics and amplicon-based metagenomic approaches. Front Microbiol 11:1223. https://doi.org/10.3389/fmicb.2020.01223.

9. Zou W, Ye G, Zhang K. 2018. Diversity, function, and application of Clostridium in Chinese strong flavor baijiu ecosystem: a review. J Food Sci 83:1193–1199. https://doi.org/10.1111/1750-3841.14134.

10. Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H, Chun J. 2017. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol 67:1613–1617. https://doi.org/10.1099/ijsem.0.001755.

11. Anjun L, Guoying L, Lan L, Hongkui H, Jiangjing G, Lu W. 2019. Flavor components of Chinese liquor yellow water. Food and Fermentation Industries 2019:118–123. https://doi.org/10.13995/j.cnki.11-1802/ts.018065.

12. Huimin Z, Yajing M, Yanli Y, Anjun L, Guoying L, Zhiqiang Y, Yan Z, Xinhui X. 2020. Differences between Huangshui from old and young mud pits of Chinese strong-flavor liquor and effect of static culture on it. Food Sci China 41:215–222. https://doi.org/10.7506/spkx1002-6630-20190531-382.

13. Tan G, Zhou R, Zhang W, Hu Y, Ruan Z, Li J, Zhang C, Shen D, Peng N, Liang Y, Zhao S. 2020. Detection of viable and total bacterial community in the pit mud of Chinese strong-flavor liquor using propidium monoazide combined with quantitative PCR and 16S rRNA gene sequencing. Front Microbiol 11:896. https://doi.org/10.3389/fmicb.2020.00896.

14. Chen Y, Neilson JW, Kushwaha P, Maier RM, Barberan A. 2021. Life-history strategies of soil microbial communities in an arid ecosystem. ISME J 15:649–657. https://doi.org/10.1038/s41396-020-00803-y.

15. Qi YL, Evans PN, Li YX, Rao YZ, Qu YN, Tan S, Jiao JY, Chen YT, Hedlund BP, Shu WS, Hua ZS, Li WJ. 2021. Comparative genomics reveals thermal adaptation and a high metabolic diversity in "Candidatus Bathyarchaeia." mSystems 6:e0025221. https://doi.org/10.1128/mSystems.00252-21.

16. Zhang S, Song W, Wemheuer B, Reveillaud J, Webster N, Thomas T. 2019. Comparative genomics reveals ecological and evolutionary insights into sponge-associated Thaumarchaeota. mSystems 4:e0028819. https://doi.org/10.1128/mSystems.00288-19.

17. Yu F, Cienfuegos-Gallet AV, Cunningham MH, Jin Y, Wang B, Kreiswirth BN, Chen L. 2021. Molecular evolution and adaptation of livestock-associated methicillin-resistant Staphylococcus aureus (LA-MRSA) sequence type 9. mSystems 6:e0049221. https://doi.org/10.1128/mSystems.00492-21.

18. Zou W, Ye G, Liu C, Zhang K, Li H, Yang J. 2021. Comparative genome analysis of Clostridium beijerinckii strains isolated from pit mud of Chinese strong flavor baijiu ecosystem. G3 11:2160. https://doi.org/10.1093/g3journal/jkab317.

19. de la Cuesta-Zuluaga J, Spector TD, Youngblut ND, Ley RE. 2021. Genomic insights into adaptations of trimethylamine-utilizing methanogens to diverse habitats, including the human gut. mSystems 6:e00939-20. https://doi.org/10.1128/mSystems.00939-20.

20. Levy A, Salas Gonzalez I, Mittelviefhaus M, Clingenpeel S, Herrera Paredes S, Miao J, Wang K, Devescovi G, Stillman K, Monteiro F, Rangel Alvarez B, Lundberg DS, Lu T-Y, Lebeis S, Jin Z, McDonald M, Klein AP, Feltcher ME, Rio TG, Grant SR, Doty SL, Ley RE, Zhao B, Venturi V, Pelletier DA, Vorholt JA, Tringe SG, Woyke T, Dangl JL. 2017. Genomic features of bacterial adaptation toplants. Nat Genet 50:138–150. https://doi.org/10.1038/s41588-017-0012-9.

21. Song Y, Zhao J, Liu W, Li W, Sun Z, Cui Y, Zhang H. 2021. Exploring the industrial potential of Lactobacillus delbrueckii ssp. bulgaricus by population genomics and genome-wide association study analysis. J Dairy Sci 104:4044–4055. https://doi.org/10.3168/jds.2020-19467.

22. Burke C, Steinberg P, Rusch D, Kjelleberg S, Thomas T. 2011. Bacterial community assembly based on functional genes rather than species. Proc Natl Acad Sci U S A 108:14288–14293. https://doi.org/10.1073/pnas.1101591108.

23. Ma Y, Zhang Y, Chen K, Zhang L, Zhang Y, Wang X, Xia X. 2021. The role of PhoP/PhoQ two component system in regulating stress adaptation in Cronobacter sakazakii. Food Microbiol 100:103851. https://doi.org/10.1016/j.fm.2021.103851.

24. Aujoulat F, Roger F, Bourdier A, Lotthe A, Lamy B, Marchandin H, Jumas-Bilak E. 2012. From environment to man: genome evolution and adaptation of human opportunistic bacterial pathogens. Genes (Basel) 3:191–232. https://doi.org/10.3390/genes3020191.

25. Liu L, Liu G, Zhou J, Jin R. 2021. Energy taxis toward redox-active surfaces decreases the transport of electroactive bacteria in saturated porous media. Environ Sci Technol 55:5559–5568. https://doi.org/10.1021/acs.est.0c08355.

26. Zhang H, Meng Y, Wang Y, Zhou Q, Li A, Liu G, Li J, Xing X. 2020. Prokaryotic communities in multidimensional bottom-pit-mud from old and young pits used for the production of Chinese strong-flavor baijiu. Food Chem 312:126084. https://doi.org/10.1016/j.foodchem.2019.126084.

27. Yin H, Zhang R, Chang Y, Zheng Y, Wang M. 2017. Effects of TCA cycle metabolism on the acetic acid fermentation of Acetobacter pasteurianus. Food Science 38:82–86. https://doi.org/10.7506/spkx1002-6630-201702013.

28. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–D495. https://doi.org/10.1093/nar/gkt1178.

29. Jeon JG, Rosalen PL, Falsetta ML, Koo H. 2011. Natural products in caries research: current (limited) knowledge, challenges and future perspective. Caries Res 45:243–263. https://doi.org/10.1159/000327250.

30. Berlemont R, Martiny A. 2015. Genomic potential for polysaccharide deconstruction in bacteria. Appl Environ Microbiol 81:1513–1519. https://doi.org/10.1128/AEM.03718-14.

31. Groussin M, Poyet M, Sistiaga A, Kearney SM, Moniz K, Noel M, Hooker J, Gibbons SM, Segurel L, Froment A, Mohamed RS, Fezeu A, Juimo VA, Lafosse S, Tabe FE, Girard C, Iqaluk D, Nguyen LTT, Shapiro BJ, Lehtimaki J, Ruokolainen L, Kettunen PP, Vatanen T, Sigwazi S, Mabulla A, Dominguez-Rodrigo M, Nartey YA, Agyei-Nkansah A, Duah A, Awuku YA, Valles KA, Asibey SO, Afihene MY, Roberts LR, Plymoth A, Onyekwere CA, Summons RE, Xavier RJ, Alm EJ. 2021. Elevated rates of horizontal gene transfer in the industrialized human microbiome. Cell 184:2053–2067. https://doi.org/10.1016/j.cell.2021.02.052.

32. Song W, Wemheuer B, Steinberg PD, Marzinelli EM, Thomas T. 2021. Contribution of horizontal gene transfer to the functionality of microbial biofilm on a macroalgae. ISME J 15:807–817. https://doi.org/10.1038/s41396-020-00815-8.

33. Zheng X-W, Han B-Z. 2016. Baijiu, Chinese liquor: history, classification and manufacture. J Ethnic Foods 3:19–25. https://doi.org/10.1016/j.jef.2016.03.001.

34. Chen MY, Teng WK, Zhao L, Hu CX, Zhou YK, Han BP, Song LR, Shu WS. 2021. Comparative genomics reveals insights into cyanobacterial evolution and habitat adaptation. ISME J 15:211–227. https://doi.org/10.1038/s41396-020-00775-z.

35. Reji L, Cardarelli EL, Boye K, Bargar JR, Francis CA. 2022. Diverse ecophysiological adaptations of subsurface Thaumarchaeota in floodplain sediments revealed through genome-resolved metagenomics. ISME J 16:1140–1152. https://doi.org/10.1038/s41396-021-01167-7.

36. Koeppel A, Perry EB, Sikorski J, Krizanc D, Warner A, Ward DM, Rooney AP, Brambilla E, Connor N, Ratcliff RM, Nevo E, Cohan FM. 2008. Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. Proc Natl Acad Sci U S A 105:2504–2509. https://doi.org/10.1073/pnas.0712205105.

37. Qin W, Zheng Y, Zhao F, Wang Y, Urakawa H, Martens-Habbena W, Liu H, Huang X, Zhang X, Nakagawa T, Mende DR, Bollmann A, Wang B, Zhang Y, Amin SA, Nielsen JL, Mori K, Takahashi R, Virginia Armbrust E, Winkler MH, DeLong EF, Li M, Lee PH, Zhou J, Zhang C, Zhang T, Stahl DA, Ingalls AE. 2020. Alternative strategies of nutrient acquisition and energy conservation map to the biogeography of marine ammonia-oxidizing archaea. ISME J 14:2595–2609. https://doi.org/10.1038/s41396-020-0710-7.

38. Jun X, Lili Z, Dan H, Zhiguo H, Chunhui W, Qingqing J, Huibo L. 2020. Preliminary investigation on the effect of mechanized production process on physicochemical indexes of pit mud. China Brewing 39:75–78.

39. Shoemaker WR, Jones SE, Muscarella ME, Behringer MG, Lennon JT. 2021. Microbial population dynamics and evolutionary outcomes under extreme energy-limitation. Proc Natl Acad Sci U S A 118:e2101691118. https://doi.org/10.1073/pnas.2101691118.

40. Adadevoh JST, Triolo S, Ramsburg CA, Ford RM. 2016. Chemotaxis increases the residence time of bacteria in granular media containing distributed contaminant sources. Environ Sci Technol 50:181–187. https://doi.org/10.1021/acs.est.5b03956.

41. Deng N, Du H, Xu Y. 2020. Cooperative response of Pichia kudriavzevii and Saccharomyces cerevisiae to lactic acid stress in baijiu fermentation. J Agric Food Chem 68:4903–4911. https://doi.org/10.1021/acs.jafc.9b08052.

42. Du H, Ji M, Xing M, Wang X, Xu Y. 2021. The effects of dynamic bacterial succession on the flavor metabolites during baijiu fermentation. Food Res Int 140:109860. https://doi.org/10.1016/j.foodres.2020.109860.

43. Contreras-Dávila CA, Carrión VJ, Vonk VR, Buisman CNJ, Strik D. 2020. Consecutive lactate formation and chain elongation to reduce exogenous chemicals input in repeated-batch food waste fermentation. Water Res 169:115215. https://doi.org/10.1016/j.watres.2019.115215.

44. Kun C, Rui L, Jia Z, Lei Z, Su Y. 2021. Exploration of stress resistance potential of Bacillus aquiflavi 3H-10 and analysis of ethanol tolerance mechanism. Food and Fermentation Industries. https://doi.org/10.13995/j.cnki.11-1802/ts.029692.

45. Wang H, Gu Y, Zhou W, Zhao D, Qiao Z, Zheng J, Gao J, Chen X, Ren C, Xu Y, Björkroth J. 2021. Adaptability of a caproate-producing bacterium contributes to its dominance in an anaerobic fermentation system. Appl Environ Microbiol 87:e01203-21. https://doi.org/10.1128/AEM.01203-21.

46. Palevich N, Palevich FP, Maclean PH, Altermann E, Gardner A, Burgess S, Mills J, Brightwell G. 2021. Comparative genomics of Clostridium species associated with vacuum-packed meat spoilage. Food Microbiol 95:103687. https://doi.org/10.1016/j.fm.2020.103687.

47. El Zahar Haichar F, Achouak W, Christen R, Heulin T, Marol C, Marais M-F, Mougel C, Ranjard L, Balesdent J, Berge O. 2007. Identification of cellulolytic bacteria in soil by stable isotope probing. Environ Microbiol 9:625–634. https://doi.org/10.1111/j.1462-2920.2006.01182.x.

48. Johnson DR, Goldschmidt F, Lilja EE, Ackermann M. 2012. Metabolic specialization and the assembly of microbial communities. ISME J 6:1985–1991. https://doi.org/10.1038/ismej.2012.46.

49. Zhu F, Zhang H, Wu H. 2015. Glycosyltransferase-mediated sweet modification in oral streptococci. J Dent Res 94:659–665. https://doi.org/10.1177/0022034515574865.

50. Liu L, Guo S, Chen X, Yang S, Deng X, Tu M, Tao Y, Xiang W, Rao Y. 2021. Metabolic profiles of Lactobacillus paraplantarum in biofilm and planktonic states and investigation of its intestinal modulation and immunoregulation in dogs. Food Funct 12:5317–5332. https://doi.org/10.1039/d1fo00905b.

51. Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N. 2021. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res 37:D233–D238. https://doi.org/10.1093/nar/gkab1045.

52. Caro-Quintero A, Konstantinidis K. 2015. Inter-phylum HGT has shaped the metabolism of many mesophilic and anaerobic bacteria. ISME J 9:958–967. https://doi.org/10.1038/ismej.2014.193.

53. Seedorf H, Fricke WF, Veith B, Brüggemann H, Liesegang H, Strittmatter A, Miethke M, Buckel W, Hinderberger J, Li F, Hagemeier C, Thauer R, Gottschalk G. 2008. The genome of Clostridium kluyveri, a strict anaerobe with unique metabolic features. Proc Natl Acad Sci U S A 105:2128–2133. https://doi.org/10.1073/pnas.0711093105.

54. Cavalcante WA, Leitão RC, Gehring TA, Angenent LT, Santaella ST. 2017. Anaerobic fermentation for n-caproic acid production: a review. Process Biochem 54:106–119. https://doi.org/10.1016/j.procbio.2016.12.024.

55. Ding H-B, Tan G-YA, Wang J-Y. 2010. Caproate formation in mixed-culture fermentative hydrogen production. Bioresour Technol 101:9550–9559. https://doi.org/10.1016/j.biortech.2010.07.056.

56. Krautkramer KA, Fan J, Bäckhed F. 2021. Gut microbial metabolites as multi-kingdom intermediates. Nat Rev Microbiol 19:77–94. https://doi.org/10.1038/s41579-020-0438-4.

57. Chadaideh KS, Carmody RN. 2021. Host-microbial interactions in the metabolism of different dietary fats. Cell Metab 33:857–872. https://doi.org/10.1016/j.cmet.2021.04.011.

58. Xu PX, Chai LJ, Qiu T, Zhang XJ, Lu ZM, Xiao C, Wang ST, Shen CH, Shi JS, Xu ZH. 2019. Clostridium fermenticellae sp. nov., isolated from the mud in a fermentation cellar for the production of the Chinese liquor, baijiu. Int J Syst Evol Microbiol 69:859–865. https://doi.org/10.1099/ijsem.0.003254.

59. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

60. Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674–1676. https://doi.org/10.1093/bioinformatics/btv033.

61. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. 2016. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Anal Methods 8:12–24. https://doi.org/10.1039/C5AY02550H.

62. Emms D, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 20:238. https://doi.org/10.1186/s13059-019-1832-y.

63. Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. 2019. MAFFT-DASH: integrated protein sequence and structural alignment. Nucleic Acids Res 47:W5–W10.

64. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490.

65. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296. https://doi.org/10.1093/nar/gkab301.

66. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

67. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60. https://doi.org/10.1038/nmeth.3176.

68. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biol 17:238. https://doi.org/10.1186/s13059-016-1108-8.

69. Zhang H, Tanner Y, Huang L, Sarah E, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 46:W95–W101. https://doi.org/10.1093/nar/gky418.

70. Ives AR, Garland T. 2014. Phylogenetic regression for binary dependent variables. Springer, Berlin, Germany. https://doi.org/10.1007/978-3-662-43550-2_9.

71. Revell LJ. 2012. Phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol 3:217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x.

72. Bonham KS, Wolfe BE, Dutton RJ. 2017. Extensive horizontal gene transfer in cheese-associated bacteria. Elife 6:e22144. https://doi.org/10.7554/eLife.22144.

73. Song W, Wemheuer B, Zhang S, Steensen K, Thomas T. 2019. MetaCHIP: community-level horizontal gene transfer identification through the combination of best-match and phylogenetic approaches. Microbiome 7:36. https://doi.org/10.1186/s40168-019-0649-y.

74. Bastian M, Heymann S, Jacomy M. 2009. Gephi: an open source software for exploring and manipulating networks. https://doi.org/10.13140/2.1.1341.1520.

75. Chen T, Chen X, Zhang S, Zhu J, Tang B, Wang A, Dong L, Zhang Z, Yu C, Sun Y, Chi L, Chen H, Zhai S, Sun Y, Lan L, Zhang X, Xiao J, Bao Y, Wang Y, Zhang Z, Zhao W. 2021. The genome sequence archive family: toward explosive data growth and diverse data types. Genomics Genom Proteom Bioinf https://doi.org/10.1016/j.gpb.2021.08.001.