



Structural and Functional Changes in Prokaryotic Communities in Artificial Pit Mud during Chinese Baijiu Production

 Mao-Ke Liu,^a Yu-Ming Tang,^a Xiao-Jiao Guo,^a Ke Zhao,^b Petri Penttinen,^b Xin-Hui Tian,^a Xin-Yu Zhang,^a Dao-Qun Ren,^a Xiao-Ping Zhang^b

^aInstitute of Rice and Sorghum Sciences, Sichuan Academy of Agricultural Sciences, Deyang, People's Republic of China

^bCollege of Resources, Sichuan Agricultural University, Wenjiang, People's Republic of China

ABSTRACT The Chinese alcoholic beverage strong-flavor baijiu (SFB) gets its characteristic flavor during fermentation in cellars lined with pit mud. Microbes in the pit mud produce key precursors of flavor esters. The maturation time of natural pit mud of over 20 years has promoted attempts to produce artificial pit mud (APM) with a shorter maturation time. However, knowledge about the molecular basis of APM microbial dynamics and associated functional variation during SFB brewing is limited, and the role of this variability in high-quality SFB production remains poorly understood. We studied APM maturation in new cellars until the fourth brewing batch using 16S rRNA gene amplicon sequencing, quantitative PCR, metaproteomics, and metabolomics techniques. A total of 36 prokaryotic classes and 195 genera were detected. *Bacilli* and *Clostridia* dominated consistently, and the relative abundance of *Bacilli* decreased along with the APM maturation. Even though both amplicon sequencing and quantitative PCR showed increased abundance of *Clostridia*, the levels of most of the *Clostridium* proteins were similar in both the first- and fourth-batch APM samples. Six genera correlated with eight or more major flavor compounds in SFB samples. Functional prediction suggested that the prokaryotic communities in the fourth-batch APM samples were actively engaged in organic acid metabolism, and the detected higher concentrations of proteins and metabolites in the corresponding metabolic pathways supported the prediction. This multi-omics approach captured changes in the abundances of specific microbial species, proteins, and metabolites during APM maturation, which are of great significance for the optimization of APM culture technique.

IMPORTANCE Strong-flavor baijiu (SFB) accounts for more than 70% of all Chinese liquor production. In the Chinese baijiu brewing industry, artificial pit mud (APM) has been widely used since the 1960s to construct fermentation cellars for production of high-quality SFB. To gain insights at the systems level into the mechanisms driving APM prokaryotic taxonomic and functional dynamics and into how this variation is connected with high-quality SFB production, we performed the first combined metagenomic, metaproteomic, and metabolomic analyses of this brewing microecosystem. Together, the multi-omics approach enabled us to develop a more complete picture of the changing metabolic processes occurring in APM microbial communities during high-quality SFB production, which will be helpful for further optimization of APM culture technique and improvement of SFB quality.

KEYWORDS strong-flavor baijiu, prokaryotic community, artificial pit mud, metaproteomics, metabolomics

Strong-flavor baijiu (SFB), a Chinese liquor produced by solid fermentation, is one of the oldest distillates in the world, with an annual production of 9.1 million tons (1). SFB contains over 1,300 different kinds of flavor compounds, including esters of ethyl

Citation Liu M-K, Tang Y-M, Guo X-J, Zhao K, Penttinen P, Tian X-H, Zhang X-Y, Ren D-Q, Zhang X-P. 2020. Structural and functional changes in prokaryotic communities in artificial pit mud during Chinese baijiu production. mSystems 5:e00829-19. <https://doi.org/10.1128/mSystems.00829-19>.

Editor Danilo Ercolini, University of Naples Federico II

Copyright © 2020 Liu et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Mao-Ke Liu, liumaoke0307@163.com, or Xiao-Ping Zhang, zhangxiaopingscau@126.com.

Received 4 December 2019

Accepted 7 March 2020

Published 24 March 2020

caproate, ethyl butyrate, ethyl lactate, and ethyl acetate (2, 3). Based on the ethyl caproate content, SFB quality can be assigned into different grades, with >1.20 g liter⁻¹ (4) corresponding to high quality. An increase in the ratio of ethyl caproate to ethyl lactate can enhance SFB quality (5). In the cyclical SFB production process, the anaerobic fermentation in cellars lined with pit mud lasts for 60 to 90 days (1) (see Fig. S1 in the supplemental material). The major cereal for SFB fermentation is sorghum. The fermentation involves numerous microbes originating from pit mud and a saccharifying and fermenting agent called daqu (6, 7). Microbes in the pit mud produce various small-molecule organic acids (e.g., caproic, butyric, acetic, and lactic acids) that are the key precursors of flavor esters, which largely determine the quality of SFB (1, 8, 9). The starting material for making the pit mud is fresh soil. The maturation of the pit mud over time includes the acclimation and domestication of liquor-brewing microbes during the fermentation process (10). In practice, only mature pit mud can be used to brew high-quality SFB due to a special microbial community in the mature mud (10, 11). However, the required natural maturation period of over 20 years critically restricts the production of high-quality SFB (10); the scarcity of natural mature pit mud hinders the growth and development of the baijiu industry.

Based on the knowledge on the natural pit mud microbiology, efforts to produce artificial pit mud (APM) to increase high-quality SFB production have been made since the 1960s (12). Generally, APM is prepared by inoculating a starter culture into a mixture of fresh soil, natural mature pit mud, wheat, and soybean meal, followed by incubation under anaerobic conditions for 30 to 60 days (13). This APM can be used to line the fermentation cellars, and high-quality SFB can be obtained from the first three production batches in only 1 year (14, 15), showing that a beneficial microbial community structure for SFB production can be successfully created in a short brewing period. Microbial communities in APM have been studied using a range of cultivation and molecular methods (12, 16, 17). However, our knowledge of the molecular mechanisms of APM microbial dynamics and associated functional variation and of the role of this variation in high-quality SFB production remains limited. Identifying the key microbes that drive these changes and sustain APM functionality is essential to further optimize the APM culture technique.

Recently, metagenomic analyses based on DNA sequencing have provided unprecedented profiling of microbial communities in various environments. However, this approach provides little information about the functions of the communities (18). Metaproteomics, measuring the complete protein composition expressed by a microbial community, can link proteins to specific microbes and thus provide insight into the functions and diversity of microbial communities (18). Similarly, metabolomics, characterizing the downstream products of gene expression, depicts the active metabolic pathways and provides information on their importance in biological processes (19). Together, the multi-omics approaches can provide a global perspective on microbial dynamics and their roles in driving ecosystem functions.

Our aim was to gain insights into the mechanisms driving APM prokaryotic compositional and functional dynamics during APM maturation and their connection with high-quality SFB production. The prokaryotic communities in the newly made APM and after the first four SFB brewing batches were analyzed using 16S rRNA gene amplicon sequencing, quantitative PCR (qPCR), and function prediction based on the sequencing results. To further understand the functional variation of the APM prokaryotic communities, metaproteomes and metabolomes from the first and fourth batches were analyzed using mass spectrometry (MS) techniques. Together, the multi-omics approaches enabled us to develop a more complete picture of the changing metabolic processes during APM maturation.

RESULTS

APM physicochemical properties. The humus and available P contents in APM samples increased by 41.84% and 200.00% ($P < 0.05$), respectively, during APM mat-

TABLE 1 The physicochemical and prokaryotic community properties in APM samples from SFB brewing process^a

Parameter	Value for the parameter by brewing stage				
	Initial stage	First batch	Second batch	Third batch	Fourth batch
Moisture (%)	36.32 ± 0.20	35.96 ± 0.82	35.08 ± 1.19	34.97 ± 0.71	35.89 ± 1.50
Organic matter (%)	7.95 ± 0.14	7.74 ± 0.72	8.26 ± 1.07	9.52 ± 1.60	9.11 ± 1.19
Humus (%)	4.78 ± 0.28 B	4.93 ± 0.88 B	5.29 ± 0.93 B	5.14 ± 0.74 B	6.78 ± 0.80 A
pH	5.21 ± 0.11	5.02 ± 0.42	4.88 ± 0.49	5.72 ± 0.80	5.58 ± 0.56
Total N (g/kg)	6.61 ± 0.17	6.42 ± 0.85	6.37 ± 1.07	8.22 ± 1.51	8.31 ± 1.29
Total P (g/kg)	1.62 ± 0.10	1.63 ± 0.45	1.73 ± 0.27	2.21 ± 0.42	1.94 ± 0.46
Total K (g/kg)	16.93 ± 0.40	17.33 ± 2.08	16.67 ± 2.63	16.45 ± 2.39	15.21 ± 1.96
NH ₄ ⁺ (g/kg)	0.74 ± 0.07	0.87 ± 0.16	0.96 ± 0.24	0.93 ± 0.16	0.88 ± 0.25
Available P (g/kg)	0.01 ± 0 C	0.02 ± 0.01 BC	0.04 ± 0.01 A	0.02 ± 0.01 BC	0.03 ± 0.01 AB
Available K (g/kg)	2.03 ± 0.18	2.23 ± 0.35	2.14 ± 0.40	2.62 ± 0.48	2.47 ± 0.53
No. of OTUs	546.17 ± 9.57 A	420.20 ± 53.39 B	438.73 ± 26.82 B	438.63 ± 6.64 B	508.27 ± 10.84 A
Chao1	577.43 ± 16.33 A	470.61 ± 57.47 B	478.85 ± 25.55 B	464.84 ± 6.69 B	539.08 ± 28.54 A
Shannon	6.63 ± 0.50 A	3.85 ± 2.33 B	6.22 ± 0.06 A	6.83 ± 0.12 A	7.02 ± 0.11 A
Coverage (%)	99.87 ± 0.06	99.87 ± 0.06	99.87 ± 0.06	99.90 ± 0	99.90 ± 0

^aData are presented as means ± standard deviations ($n = 3$). Values with different letters within a row are significantly different statistically ($P < 0.05$). SFB, strong-flavor baijiu; APM, artificial pit mud.

uration (Table 1). The moisture, organic matter, pH, total N, total P, total K, NH₄⁺, and available K contents remained similar over time.

SFB flavor compounds. A total of 31 flavor compounds were detected in the SFB samples: 12 esters, 8 alcohols, 6 acids, 3 aldehydes, and 2 ketones. The major flavor compounds (average content of >0.1 g liter⁻¹) included ethyl caproate, ethyl acetate, ethyl lactate, ethyl butyrate, 1-propanol, 3-methyl butanol, *n*-hexanol, 2-butanol, acetic acid, butyric acid, caproic acid, isovaleric acid, and acetaldehyde diethyl acetal. These 13 flavor compounds accounted for 95.73% of the flavor compounds present in the samples (Fig. 1a; see Table S1 in the supplemental material).

Among esters, the contents of ethyl caproate and ethyl butyrate increased during the brewing process ($P < 0.05$), reaching maximum levels in the fourth and third batches of SFB samples, respectively. The ethyl acetate content decreased significantly from the first to the later batches ($P < 0.05$) (Fig. 1b). The ratio of ethyl caproate/ethyl lactate increased steadily from 0.81 to 2.77 ($P < 0.05$), indicating an improvement in the quality of SFB (Fig. 1c).

Amplicon sequencing and qPCR of 16S rRNA genes. The 1,241,812 sequences that passed quality filtering were binned into 420.20 to 546.17 operational taxonomic units (OTUs) per APM maturation stage (Table 1). The coverage was $\geq 99.8\%$, indicating that the samples represented well the sampled prokaryotic communities. The Chao1 index was lower in samples from the first to third batch than in those of the initial and fourth batches ($P < 0.05$). The Shannon index was lower in samples from the first batch than in the other samples ($P < 0.05$) (Table 1). In the principal-component analysis (PCA), the prokaryotic communities showed batch-dependent clustering, with the exception of the first-batch APM samples, which were scattered along the PCA axis 1 (Fig. 2a).

The OTUs were classified into 36 prokaryotic classes. The nine most prevalent classes, with average abundances of $>1\%$ in at least one APM maturation stage, were *Bacilli*, *Clostridia*, *Coriobacteriia*, *Actinobacteria*, *Negativicutes*, *Gammaproteobacteria*, *Alphaproteobacteria*, *Bacteroidia*, and *Synergistia* (Fig. 2b). *Bacilli* and *Clostridia* were the only classes that consistently dominated throughout the whole brewing process, representing 31.37 to 86.60% and 11.16 to 53.33%, respectively, of the total abundances. The relative abundance of *Clostridia* was higher in the third batch than in the first and second batches, whereas the relative abundances of *Bacilli* were higher in the first and second batches than in the other stages ($P < 0.05$). The relative abundances of *Coriobacteriia* and *Alphaproteobacteria* were highest in the initial stage, and those of *Actinobacteria* and *Synergistia* were highest in the fourth batch ($P < 0.05$) (Table 2).

A total of 195 genera were detected (Table S2), out of which 30 genera dominated the APM prokaryotic community, representing 63.00 to 93.41% of the total abundances

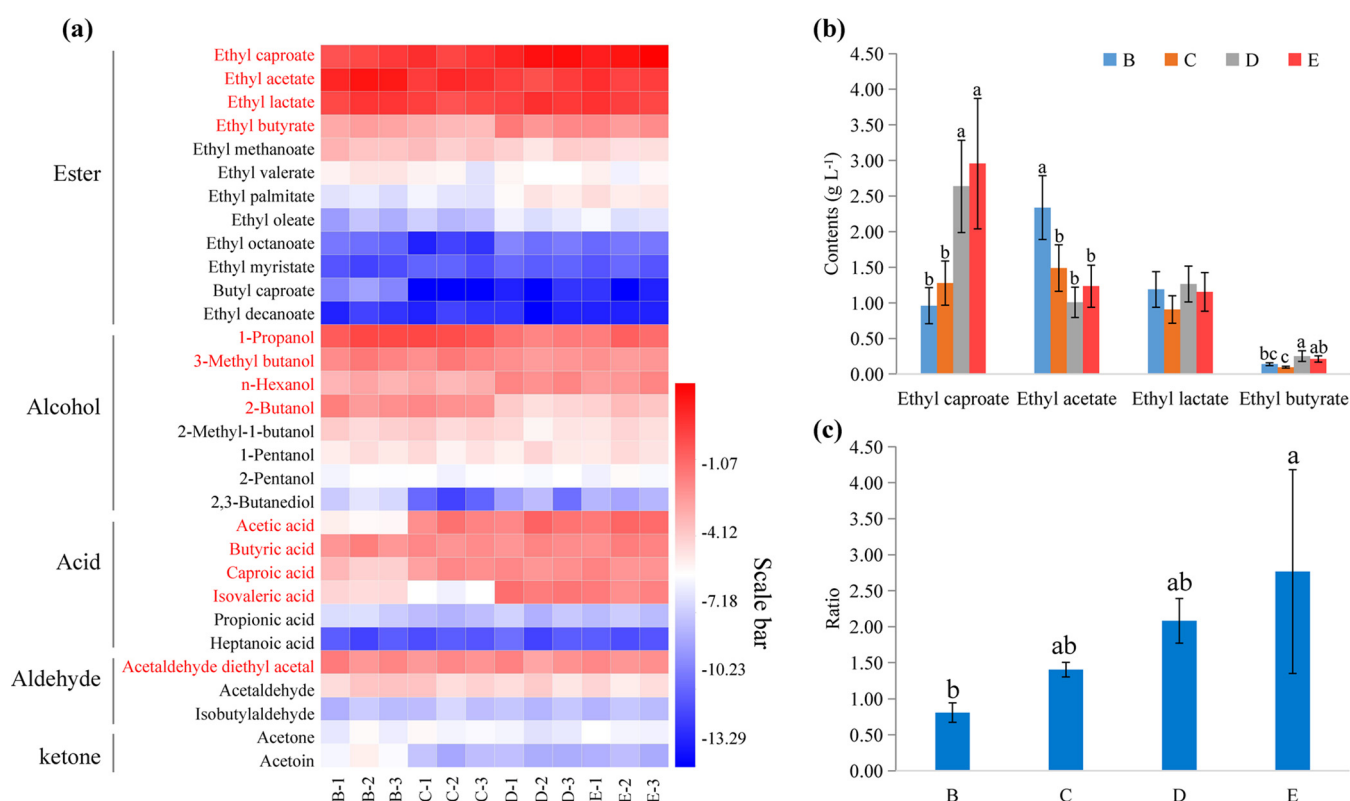


FIG 1 Dynamics of flavor compounds during the strong-flavor baijiu (SFB) brewing process. (a) Heat map of concentrations of flavor compounds (log $_2$ scale) in SFB samples. Major compounds (average content of >0.1 g liter $^{-1}$) presented are in red. (b) Concentrations of ethyl caproate, ethyl acetate, ethyl lactate, and ethyl butyrate in SFB samples. (c) Ethyl caproate/ethyl lactate ratio. B, first batch; C, second batch; D, third batch; E, fourth batch. In panels b and c, data are presented as means \pm standard deviations ($n = 3$). Different superscript letters above bars indicate statistically significant differences ($P < 0.05$).

(Table 2). In the *Clostridia*, the relative abundances of the initially rare *Desulfosporosinus*, *Fonticella*, and *Haloimpatiens* were higher in the third and fourth batches than in the other stages, and those of *Clostridium* were higher in the third and fourth batches than in the initial stage and first batch ($P < 0.05$). The relative abundance of *Lactobacillus* in the first batch was almost three times higher than that in the initial stage and over 10 times higher than that in the subsequent batches ($P < 0.05$). The relative abundances of eight other genera in the class *Bacilli* were higher in the second, third, and fourth batches than in the initial stage and first batch, and those of *Weissella* and *Lactococcus* were highest in the fourth batch, similar to abundances of *Neoscardovia* (class *Actinobacteria*) and *Aminobacterium* (class *Synergistia*) ($P < 0.05$). *Pediococcus* (*Bacilli*), *Atopobium* (*Actinobacteria*), and *Bacteroides* (class *Bacteroidia*) were abundant in the initial stage but mostly depleted in the subsequent brewing process ($P < 0.05$) (Table 2).

The most abundant taxa, *Clostridia* and *Lactobacillus*, were quantified using qPCR (Fig. 2c). The concentration of *Clostridia* decreased significantly from an initial 8.18×10^8 to 9.83×10^7 copies g $^{-1}$ measured at the first batch and then increased to 2.85×10^8 copies g $^{-1}$ at the end of the fourth batch. In line with the changes in relative abundances, *Lactobacillus* abundance increased from the initial stage to the first batch, reaching a maximum value of 4.54×10^8 copies g $^{-1}$, and was approximately 7×10^7 copies g $^{-1}$ in the subsequent batches.

Relationships of prokaryotic community with physicochemical properties and major SFB flavor compounds. The Pearson's correlation analysis between APM physicochemical properties and most abundant prokaryotic genera showed that total N, available P, and available K correlated with 15 genera (Fig. 3a). Out of the 15 genera, 6 genera correlated with eight or more major flavor compounds in SFB samples (Fig. 3b). *Clostridium*, *Rummeliibacillus*, *Desulfitobacterium*, *Fonticella*, *Haloimpatiens*, and *Desul-*

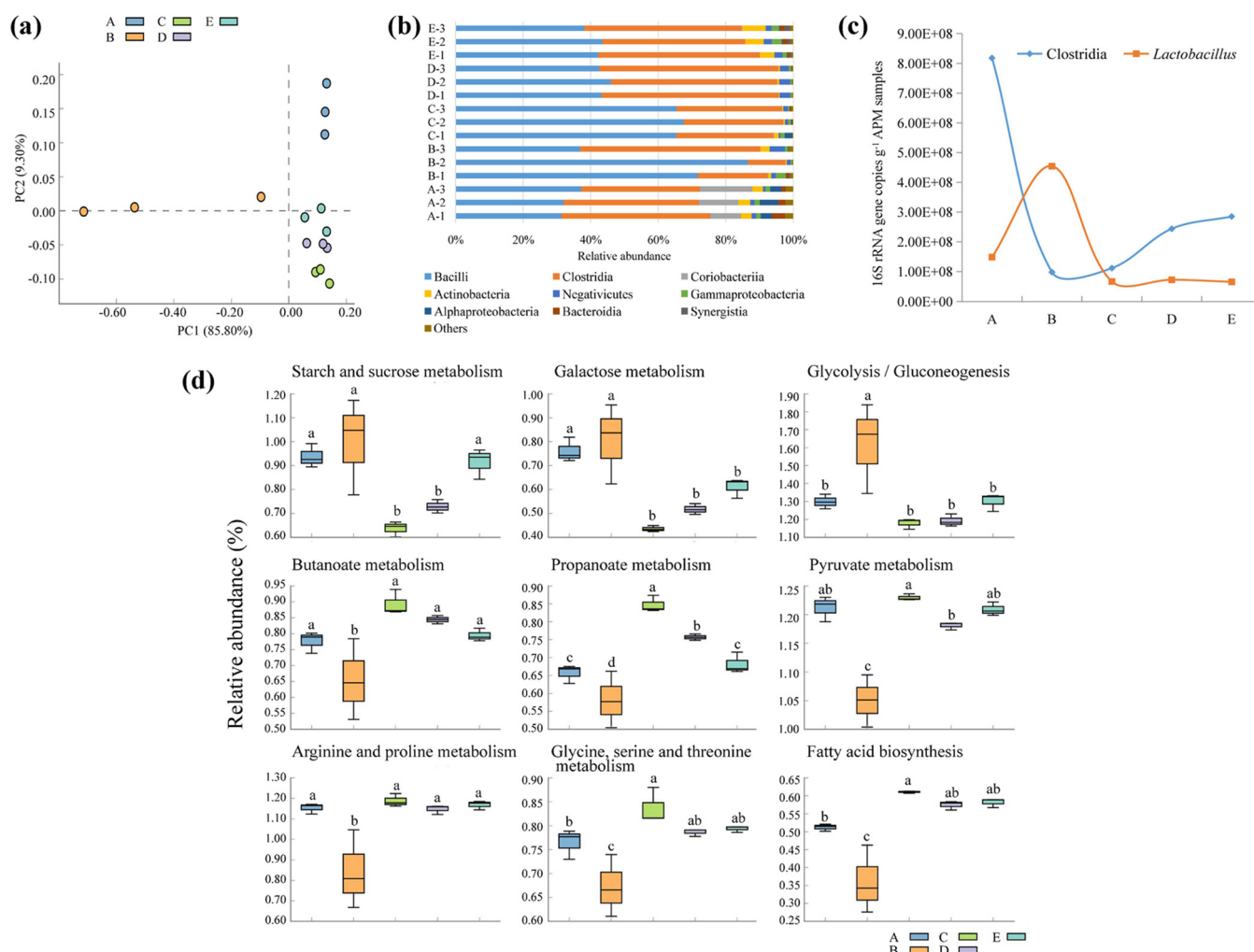


FIG 2 16S rRNA gene analysis of prokaryotic communities in artificial pit mud (APM) samples. (a) Principal-component analysis (PCA) at the operational taxonomic unit (OTU) level. (b) Proportions of the most abundant prokaryotic classes. (c) Average number ($n = 3$) of 16S rRNA gene copies of *Clostridia* and *Lactobacillus*. (d) Predicted abundances of selected KEGG orthologs (KOs) related to saccharides and organic acid metabolisms. A, initial stage; B, first batch; C, second batch; D, third batch; E, fourth batch. Different superscript letters indicate statistically significant differences ($P < 0.05$).

fosporosinus correlated positively with ethyl caproate, ethyl butyrate, isovaleric acid, and *n*-hexanol, and *Clostridium*, *Rummeliibacillus*, *Desulfotobacterium*, and *Haloimpatiens* correlated positively with caproic acid and acetic acid ($P < 0.05$). The six genera correlated negatively with ethyl acetate, 1-propanol, and 2-butanol, and all six, except *Rummeliibacillus*, correlated negatively with 3-methyl butanol ($P < 0.05$).

Predicted metabolic profile of the prokaryotic community. The PICRUSt (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States) software package identified 279 functional categories in the Kyoto Encyclopedia of Genes and Genome (KEGG) pathways. We focused on the predicted abundances of KEGG orthologs (KOs) assigned to saccharide and organic acid metabolism. Generally, the relative abundances of KOs were at the same levels in the initial stage and fourth batch, with the exception of galactose metabolism KOs (Fig. 2d). The relative abundances of KOs assigned to the starch and sucrose metabolism were highest in the initial stage and first and fourth batches, those of galactose metabolism KOs were highest in the initial stage and first batch, and those of glycolysis/gluconeogenesis KOs were highest in the first batch ($P < 0.05$). The relative abundances of KOs assigned to organic acid metabolism (butanoate metabolism, propanoate metabolism, pyruvate metabolism, arginine and proline metabolism, fatty acid biosynthesis, and glycine, serine, and threonine metabolism) were lowest in the first batch ($P < 0.05$) (Fig. 2d).

TABLE 2 Prokaryotic taxa with relative abundances of >1% in at least one stage in APM samples from the SFB brewing process^a

Class or genus	Relative abundance by brewing step				
	Initial stage	First batch	Second batch	Third batch	Fourth batch
<i>Clostridia</i>	39.77 ± 4.50 AB	28.41 ± 22.11 B	29.83 ± 1.31 B	51.65 ± 1.98 A	45.61 ± 2.94 AB
<i>Clostridium</i>	12.81 ± 2.20 CD	8.53 ± 6.07 D	14.5 ± 0.27 BC	26.41 ± 0.84 A	19.82 ± 1.72 B
<i>Caproiciproducens</i>	9.58 ± 2.77 A	0.47 ± 0.48 C	0.53 ± 0.22 C	0.94 ± 0.24 C	6.58 ± 2.29 B
<i>Hydrogenispora</i>	2.31 ± 0.46	4.81 ± 5.03	1.03 ± 0.12	2.37 ± 0.38	2.04 ± 0.14
<i>Sedimentibacter</i>	1.28 ± 0.36	1.52 ± 1.32	0.63 ± 0.12	1.10 ± 0.15	0.70 ± 0.16
<i>Tepidimicrobium</i>	0.90 ± 0.41	1.09 ± 0.91	1.09 ± 0.26	0.33 ± 0.01	0.24 ± 0.06
<i>Garciella</i>	1.34 ± 0.40 A	0.25 ± 0.22 B	1.70 ± 0.23 A	0.19 ± 0.04 B	0.10 ± 0.02 B
<i>Anaerosporebacter</i>	0.47 ± 0.28 B	0.37 ± 0.32 B	0.73 ± 0.09 AB	1.06 ± 0.18 A	0.69 ± 0.10 AB
<i>Caldicoprobacter</i>	0.93 ± 0.38	1.27 ± 1.12	0.26 ± 0.09	0.27 ± 0.05	0.23 ± 0.05
<i>Anaerosalibacter</i>	0.47 ± 0.16 B	0.62 ± 0.51 B	1.48 ± 0.13 A	0.22 ± 0.04 B	0.18 ± 0.03 B
<i>Desulfosporosinus</i>	0.18 ± 0.03 C	0.35 ± 0.27 C	0.24 ± 0.06 C	1.14 ± 0.11 A	0.83 ± 0.08 B
<i>Desulfotobacterium</i>	0.10 ± 0.02 C	0.21 ± 0.10 C	0.25 ± 0.10 C	1.28 ± 0.04 A	1.02 ± 0.15 B
<i>Fonticella</i>	0.13 ± 0.05 B	0.14 ± 0.05 B	0.13 ± 0.05 B	1.42 ± 0.53 A	1.02 ± 0.30 A
<i>Haloimpatiens</i>	0.04 ± 0.01 B	0.03 ± 0.02 B	0.06 ± 0.03 B	1.17 ± 0.06 A	0.98 ± 0.33 A
<i>Bacilli</i>	33.52 ± 3.21 B	65.07 ± 25.51 A	66.15 ± 1.22 A	43.93 ± 1.77 B	41.22 ± 2.78 B
<i>Lactobacillus</i>	19.87 ± 5.09 B	57.67 ± 31.45 A	3.13 ± 1.74 B	2.80 ± 3.74 B	4.23 ± 4.09 B
<i>Bacillus</i>	4.84 ± 1.72 B	3.12 ± 2.54 B	10.84 ± 1.19 A	8.52 ± 0.34 A	7.99 ± 1.12 A
<i>Hazenella</i>	0.26 ± 0.11 C	0.25 ± 0.14 C	9.94 ± 1.16 A	9.50 ± 0.70 A	5.96 ± 1.61 B
<i>Weissella</i>	0.68 ± 0.35 B	0.01 ± 0.01 B	0.03 ± 0.03 B	0 ± 0 B	3.71 ± 1.30 A
<i>Paenibacillus</i>	0.83 ± 0.58 D	0.27 ± 0.23 D	5.90 ± 0.16 A	3.19 ± 0.48 B	2.45 ± 0.27 C
<i>Kroppenstedtia</i>	0.45 ± 0.10 D	0.63 ± 0.42 D	6.61 ± 0.46 A	1.92 ± 0.30 B	1.24 ± 0.20 C
<i>Lactococcus</i>	0.67 ± 0.20 B	0 ± 0 C	0 ± 0 C	0 ± 0 C	1.27 ± 0.44 A
<i>Baia</i>	0.11 ± 0.04 C	0.17 ± 0.10 C	4.15 ± 1.03 A	3.50 ± 0.23 A	2.17 ± 0.59 B
<i>Rummeliibacillus</i>	0.41 ± 0.16 C	0.25 ± 0.20 C	1.15 ± 0.17 B	1.80 ± 0.22 A	1.43 ± 0.28 AB
<i>Brevibacillus</i>	0.23 ± 0.08 C	0.16 ± 0.10 C	2.06 ± 0.63 A	1.22 ± 0.20 B	1.10 ± 0.24 B
<i>Lysinibacillus</i>	0.24 ± 0.02	0.11 ± 0.06	1.15 ± 1.31	1.11 ± 0.08	0.74 ± 0.11
<i>Pediococcus</i>	1.86 ± 0.66 A	0.11 ± 0.17 B	0 ± 0 B	0 ± 0 B	0.08 ± 0.02 B
<i>Novibacillus</i>	0.02 ± 0.01 C	0.04 ± 0.03 C	1.26 ± 0.12 A	0.22 ± 0.02 B	0.17 ± 0.02 B
<i>Actinobacteria</i>	3.26 ± 0.18 B	1.29 ± 1.26 C	0.74 ± 0.59 C	0.37 ± 0.14 C	5.60 ± 1.38 A
<i>Neoscardovia</i>	0.07 ± 0.02 B	0 ± 0 B	0 ± 0 B	0 ± 0 B	4.60 ± 1.55 A
<i>Atopobium</i>	11.57 ± 3.57 A	0.10 ± 0.10 B	0.07 ± 0.07 B	0 ± 0 B	0.01 ± 0.01 B
<i>Synergistia</i>	0.05 ± 0.02 B	0.12 ± 0.18 B	0.003 ± 0.006 B	0.003 ± 0.006 B	1.05 ± 0.50 A
<i>Aminobacterium</i>	0.04 ± 0.03 B	0.11 ± 0.17 B	0.003 ± 0.006 B	0.003 ± 0.006 B	1.04 ± 0.49 A
<i>Bacteroidia</i>	2.40 ± 1.36 A	0.33 ± 0.50 B	0.24 ± 0.20 B	0.09 ± 0.11 B	1.21 ± 0.46 AB
<i>Bacteroides</i>	2.02 ± 1.18 A	0.007 ± 0.006 B	0.003 ± 0.006 B	0.03 ± 0.04 B	0.02 ± 0.02 B
<i>Coriobacteriia</i>	12.11 ± 3.28 A	0.11 ± 0.11 B	0.08 ± 0.07 B	0 ± 0 B	0.11 ± 0.05 B
<i>Negativicutes</i>	1.16 ± 0.29 B	2.34 ± 1.90 AB	0.94 ± 0.38 B	3.08 ± 0.38 A	2.14 ± 0.42 AB
<i>Gamma-proteobacteria</i>	1.42 ± 0.19 AB	1.43 ± 1.37 AB	0.86 ± 0.35 AB	0.55 ± 0.16 B	2.17 ± 0.90 A
<i>Alphaproteobacteria</i>	4.03 ± 1.18 A	0.08 ± 0.10 B	0.81 ± 1.19 B	0.03 ± 0.02 B	0.16 ± 0.06 B

^aData are presented as means ± standard deviations ($n = 3$). Values with different letters within a row are significantly different statistically ($P < 0.05$). SFB, strong-flavor baijiu; APM, artificial pit mud.

Metaproteomic analysis of APM prokaryotic communities. Based on the differences in the ratios of ethyl caproate to ethyl lactate, an SFB quality measure, and PICRUSt predictions, we next performed metaproteomics and metabolomics analyses of the first- and fourth-batch APM samples to obtain more insight into the functional changes and their relation to the SFB quality.

The peptide spectra identified by a tandem mass tags (TMT)-MS proteomics approach were searched against a prokaryotic database created by collating all UniProt entries (a total of 610,836 entries) from organisms with an abundance of >1% in the 16S rRNA gene amplicon sequences identified in the first- and fourth-batch APM samples, including *Bacilli*, *Clostridia*, *Actinobacteria*, *Synergistia*, *Negativicutes*, *Gamma-proteobacteria*, and *Bacteroidia* (Table 2). A total of 213 prokaryotic proteins were identified. The three most abundant classes collectively accounted for 78.39% of prokaryotic proteins (Fig. 4a). *Negativicutes* (32.39%) was the most abundant class, followed by *Clostridia* (23.00%) and *Bacilli* (23.00%). The proteins originated from 70 genera; 18.31% were from *Lactobacillus*, 9.86% were from *Clostridium*, and 5.16% were from *Sporomusa* (Fig. 4b). A total of 47 differential proteins with a fold change of >2.0 or <0.50 ($P < 0.05$) were found; 16 and 31 proteins were enriched and depleted, respectively, in the fourth-batch APM samples (Table S3). The differential proteins

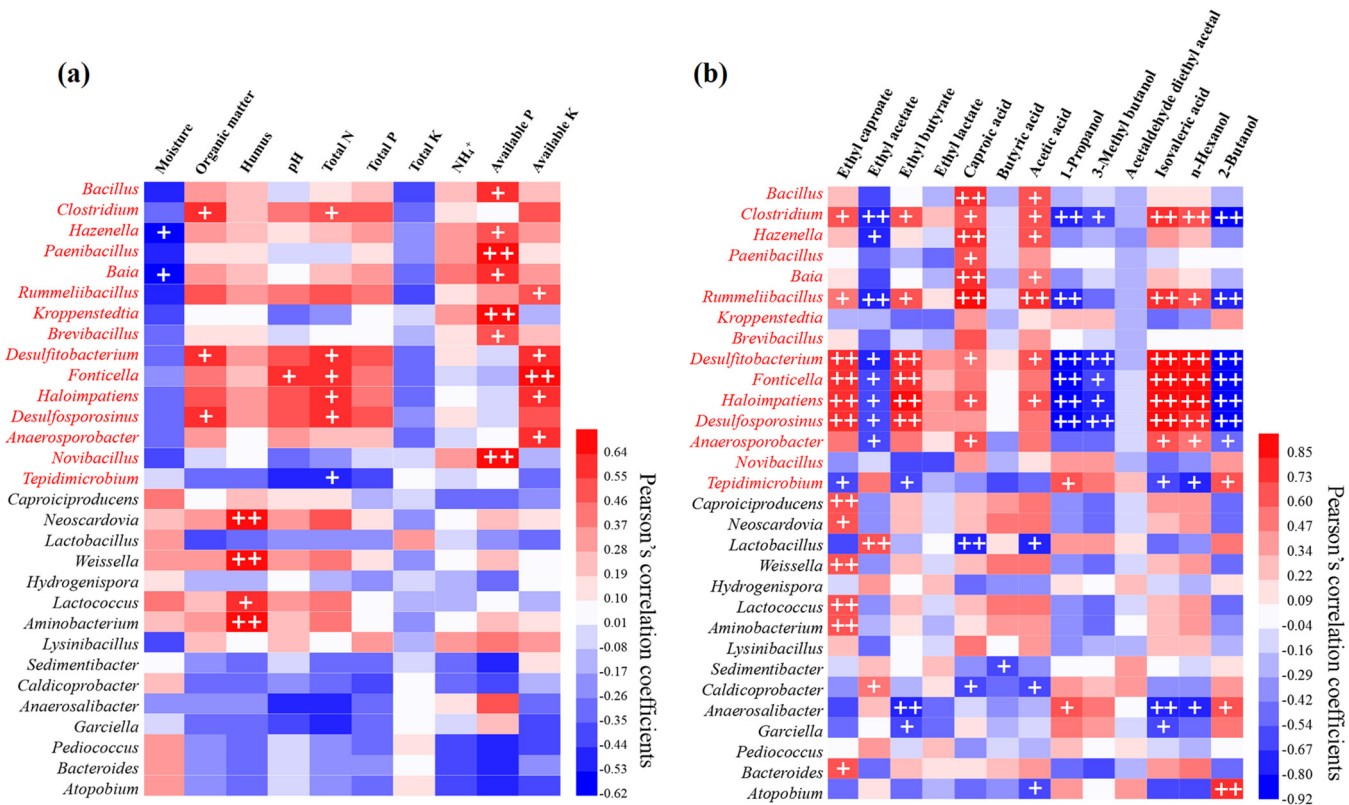


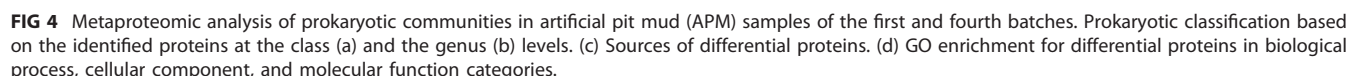
FIG 3 Pearson's correlation coefficients between dominant prokaryotic genera and physicochemical properties of APM samples (a) and flavor components of SFB samples (b). +, correlation significant at a *P* value of <0.05; ++, correlation significant at a *P* value of <0.01. The genera significantly correlated with total N, available K, and available P are indicated with a red font.

originated from 24 genera and mainly belonged to *Lactobacillus* (Fig. 4c). In line with the changes in 16S rRNA gene amplicon relative abundances and qPCR results, the 14 detected *Lactobacillus* proteins were depleted in the fourth-batch APM samples (Table S3).

To validate our metaproteomic data, 15 differentially expressed proteins, based on the Sequest HT score (Table S3), were selected, and their protein levels were measured by a parallel reaction monitoring (PRM)-MS proteomics approach. The fold changes of the proteins between the first- and fourth-batch APM samples were consistent with those determined by TMT-MS analysis (Table 3).

The differentially expressed proteins were submitted for Gene Ontology (GO) enrichment analysis (Fig. 4d). In biological process (BP) analysis, the greatest changes occurred in metabolic processes, including ADP metabolic process, purine ribonucleoside diphosphate metabolic process, and purine nucleoside diphosphate metabolic process. Cellular component (CC) analysis showed that most of the proteins acted in the cytosol. The molecular function (MF) analysis revealed that most proteins were involved in small-molecule binding, nucleoside phosphate binding, and nucleotide binding.

Metabolomic analysis of APM. In the untargeted metabolomics analysis, we identified 7,478 metabolites, out of which 246 were identified by gas chromatography (GC)-MS, and 7,232 were identified by liquid chromatography (LC)-MS from the first- and fourth-batch APM samples. Both GC-MS and LC-MS approaches led to a clear separation of the batches in the orthogonal partial least-squares discriminant analysis (OPLS-DA) scatter plots, and based on cross-validation, the predictive ability and goodness of fit of the OPLS-DA models were credible (Fig. 5a to d). In line with PCA of the prokaryotic communities (Fig. 2a), the metabolic profiles from the first batch were more dispersed than those from the fourth batch. A total of 1,014 metabolites (117 detected by GC-MS and 897 detected by LC-MS) were identified as differential (variable



Within the saccharide metabolism pathway, except for the putative aldehyde dehydrogenase AldA (gene symbol *aldA*) and ethanolamine utilization protein EutM (*SAMN04488500_10762*), most proteins were depleted in the fourth-batch APM samples. Except for pyruvate, the remaining 12 metabolites were also depleted. In contrast, except for D-lactate dehydrogenase (*ldhD*) and L-lactate dehydrogenase (*ldh*), most proteins involved in the metabolism of organic acids were enriched. Consistently, except for fumarate and L-allothreonine, the remaining 17 metabolites were also enriched in the fourth-batch APM samples. The changes were in agreement with the results of PICRUST predictive metagenome analysis (Fig. 2d).

TABLE 3 Comparison of the fourth-batch and first-batch APM samples from the strong-flavor baijiu brewing process by TMT- and PRM-based quantitative proteomics approaches^a

Accession no. ^c	Protein name	Gene symbol	Fold change by: ^b		Protein source
			TMT	PRM	
A0A1M4T8Q5	Cell division protein FtsZ	<i>ftsZ</i>	0.46	0.48*	<i>Schwartzia succinivorans</i>
A5N5B3	Formate-tetrahydrofolate ligase	<i>fhs</i>	0.48	0.20*	<i>Clostridium kluyveri</i>
A7Z207	60-kDa chaperonin	<i>groL</i>	0.39	0.26*	<i>Bacillus velezensis</i>
A8YU15	Glucose-6-phosphate isomerase	<i>pgi</i>	0.35	0.42*	<i>Lactobacillus helveticus</i>
B1MW06	50S ribosomal protein L29	<i>rpmC</i>	0.34	0.15*	<i>Leuconostoc citreum</i>
C9LN83	Methylmalonyl-CoA mutase domain protein	<i>GCWU000321_01033</i>	2.10	8.06*	<i>Dialister invisus</i>
D5EGB9	S-layer domain protein	<i>Amico_1484</i>	3.85	82.97*	<i>Aminobacterium colombiense</i>
O32755	Glyceraldehyde-3-phosphate dehydrogenase	<i>gap</i>	0.34	0.29*	<i>Lactobacillus delbrueckii</i> subsp. <i>bulgaricus</i>
O32765	L-Lactate dehydrogenase	<i>ldh</i>	0.40	0.25*	<i>Lactobacillus helveticus</i>
P80583	Flagellin (fragment)	<i>fla</i>	3.71	9.53*	<i>Clostridium tyrobutyricum</i>
Q0AVM3	Acetyl-CoA acetyltransferase	<i>Swol_1934</i>	2.16	8.61*	<i>Syntrophomonas wolfei</i> subsp. <i>wolfei</i>
Q0AZ33	Electron transfer flavoprotein subunit alpha	<i>etfA</i>	2.02	10.65*	<i>Syntrophomonas wolfei</i> subsp. <i>wolfei</i>
Q74IT6	Chaperone protein DnaK	<i>dnaK</i>	0.37	0.21*	<i>Lactobacillus johnsonii</i>
Q7MVY0	Hydroxylamine reductase	<i>hcp</i>	0.43	0.12*	<i>Porphyromonas gingivalis</i>
R7KAP8	Acetate kinase	<i>ackA</i>	2.41	573.92*	<i>Acidaminococcus</i> sp. CAG:917

^aAPM, artificial pit mud; TMT, tandem mass tag; PRM, parallel reaction monitoring.^bThe asterisks indicate statistically significant differences between values of the fourth batch and those of the first-batch APM samples ($P < 0.05$).^cNCBI or UniProt.

Among the multiple pathways, butanoate metabolism and pyruvate metabolism pathways are directly involved in the production of caproic acid, butyric acid, and lactic acid. In pyruvate metabolism, the levels of three proteins, acetyl-coenzyme A (CoA) acetyltransferase (*Swol_1934*), acetate kinase (*ackA*), and *aldA*, and two metabolites (pyruvate and succinate) were enriched, while two proteins (*ldhD* and *ldh*) and one metabolite (fumarate) were depleted in the fourth-batch APM samples. In butanoate metabolism, except for fumarate, the remaining metabolites (pyruvate, succinate, succinate semialdehyde, and *R*-3-hydroxybutanoate) and the protein *Swol_1934* were enriched in the fourth-batch APM samples.

DISCUSSION

Strong-flavor baijiu (SFB) is a fermented and distilled alcoholic beverage that is fermented in production cellars lined with pit mud, a fresh soil derivative with characteristic microbial communities. Pit mud affects both the quality and safety of the end product. The maturation time of natural pit mud of over 20 years and the stochastic nature of maturation have promoted attempts to produce artificial pit mud (APM) with shorter and more predictable maturation (13). We studied APM maturation and its connection with high-quality SFB production by assessing the composition and functions of the APM prokaryotic community using combined metagenomics, metaproteomics, and metabolomics.

The major flavor compounds of SFB are ethyl caproate balanced with ethyl lactate, ethyl acetate, and ethyl butyrate (3). Ethyl caproate, produced from the esterification of caproic acid and ethanol, is a key compound affecting SFB quality (4). The ethyl caproate contents of SFB obtained from a new APM-lined cellar increased from 2.24 g liter⁻¹ in the first batch to 5.24 g liter⁻¹ in the third batch (14). Although the ethyl caproate contents of our SFB samples were lower, the contents increased during the APM maturation, and both the ethyl caproate content and ethyl caproate/ethyl lactate ratio reached levels required for high-quality SFB (4, 5). Hence, the acquired amplicon sequencing, metaproteomics, and metabolomics data are relevant in studying the role of APM microbial dynamics in high-quality SFB production.

Amplicon sequencing targeting the 16S rRNA gene revealed that after the first batch of fermentation, the prokaryotic community was less diverse and more dispersed than that after the subsequent batches, possibly due to niche selection (10). In the early stage of brewing, the species that could not adapt to the brewing environment were eliminated, resulting in the differentiation of microbial community structure. The

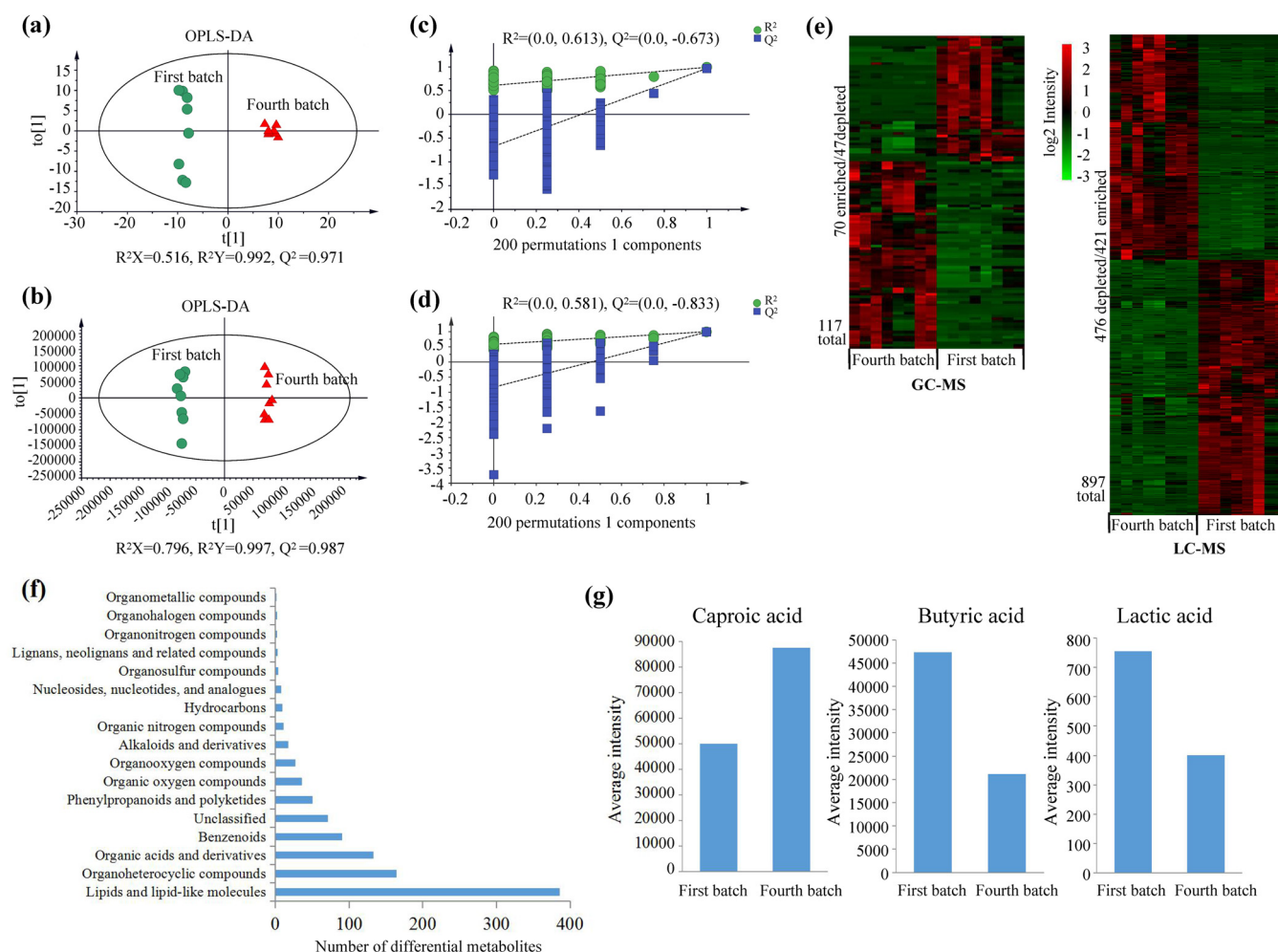


FIG 5 Metabolomic analysis of the artificial pit mud (APM) samples of the first and fourth batches ($n = 8$). Orthogonal partial least-squares discriminant analysis (OPLS-DA) scatter plots (a and b) and the corresponding cross-validation plots (c and d) derived from gas chromatography-mass spectrometry (GC-MS) (a and c) and liquid chromatography-mass spectrometry (LC-MS) (b and d) data sets. Intensity levels (e) and functional groups (f) of differential metabolites are shown. (g) Average levels of caproic, butyric, and lactic acids. For panels a to d, the R^2X and R^2Y values show the total amount of variation explained by the model in the X and Y matrix, respectively. The Q^2 value represents the predictive ability of the models.

subsequent adaptation of brewing microbes to the environment may partially explain the increased species diversity and stability in the community. Interestingly, the changes were similar to those observed in 1-, 10-, 25-, and 50-year-old natural pit muds (10), suggesting that the maturation processes in both natural and artificial pit muds proceed similarly yet at a different pace. APM is degraded relatively frequently, leading to poor SFB quality (17), and this may be an unavoidable consequence of the quicker maturation process.

In earlier studies, *Clostridium* and *Hydrogenispora* were predominant microbes in natural mature pit mud (20, 21), *Bacillus* was predominant in immature pit mud (1), and the *Clostridia/Bacilli* ratio correlated positively with pit mud quality (22). Similarly, in our study, the relative abundances of the classes *Bacilli* and *Clostridia* and the genera *Clostridium*, *Hydrogenispora*, *Lactobacillus*, and *Bacillus* changed with APM age. Again, similar to characteristics of the natural pit mud (10), the relative abundances of *Clostridia* increased, and those of *Bacilli* decreased along with the APM maturation. Lactic acid bacteria have been reported to be predominant in both new and degraded pit mud (17, 22), and the quantification of *Lactobacillus* 16S rRNA genes has been applied to evaluate the state of pit mud (23). In agreement, both the relative abundance and gene copy numbers of *Lactobacillus* were highest in the first-batch APM samples. The dominance of *Lactobacillus* was the likely reason for the low diversity at

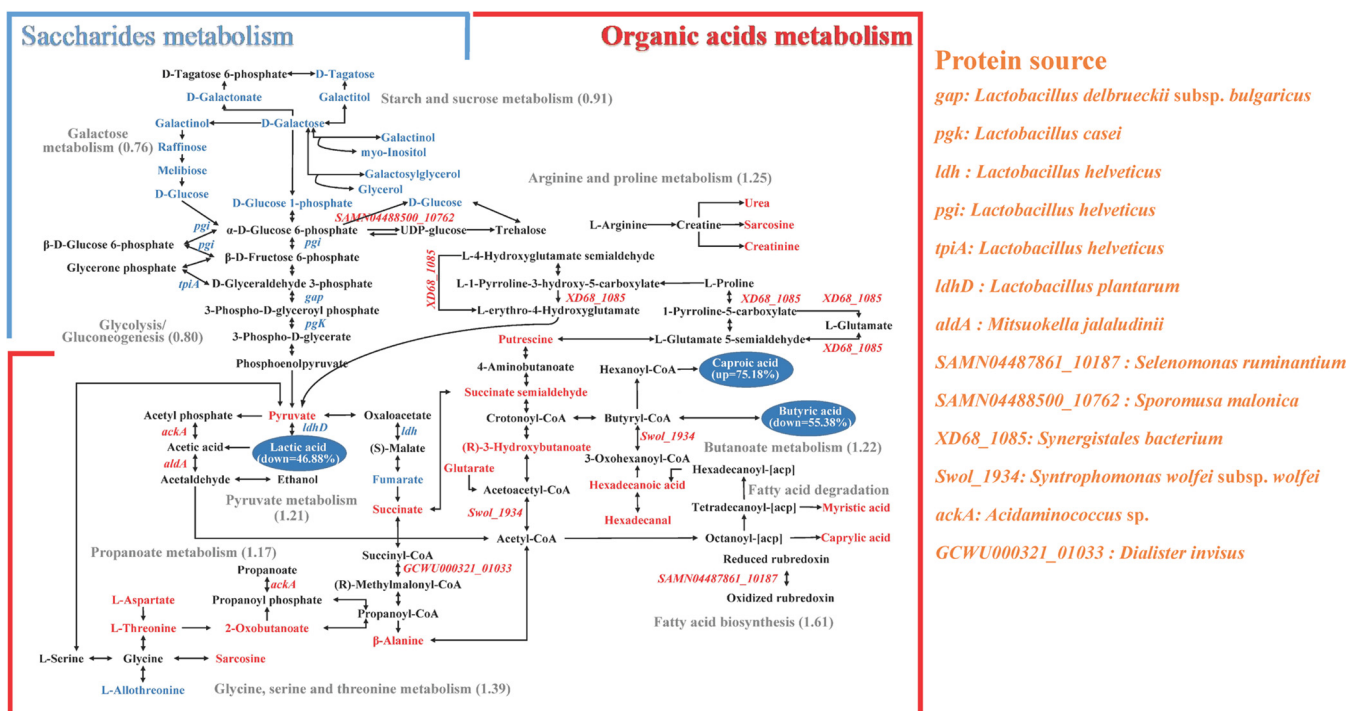


FIG 6 Schematic presentation of pathways, proteins, metabolites, and species related to the synthesis of caproic, butyric, and lactic acids in artificial pit mud (APM) samples. The pathways were organized based on the *Escherichia coli* K-12 MG 1655 genome with KEGG (Kyoto Encyclopedia of Genes and Genomes) organism identifier *eco* (<http://www.kegg.jp/>). Metabolites and proteins that are enriched or depleted in the fourth-batch APM samples compared to levels in the first-batch APM samples are indicated in red and blue, respectively. The average fold changes of the predicted abundances of KEGG orthologs (KOs) based on the 16S rRNA gene amplicon sequencing (Fig. 2d) are indicated in parentheses.

that stage; possibly the early-stage brewing environment was favorable for *Lactobacillus* (10). The subsequent changes in the environment may at least partially explain the observed increase in species diversity of APM. Among the genera that increased during the APM maturation, the increase of *Fonticella*, *Haloimpatiens*, and *Desulfosporosinus* organisms might be related to their slow growth as strictly anaerobic bacteria (24–26). However, the genera that increased included also facultatively anaerobic bacteria, for example *Weissella* (27). To our knowledge, *Fonticella* and *Hazenella* were detected for the first time in SFB production, suggesting that the APM microbial communities and their succession are more complex than previously known (12, 16, 17).

Interestingly, compared with communities identified in natural mature pit mud (10, 28), the abundances of methanogens were very low in our study. For example, the average abundance of *Methanobrevibacter* was 1.94 to 14.34% in natural mature pit mud samples (10, 28) but less than 0.05% in our APM samples. *Methanobrevibacter* is a hydrogenotrophic methanogen that produces CH_4 from H_2 and CO_2 (29). Together with fermentative bacteria, methanogens can enhance organic acid production through syntrophic interactions (30). Thus, the high abundances of *Methanobrevibacter* in natural mature pit mud may indicate efficient organic acid production, which can increase flavor ester formation during SFB production. Further studies are required to determine the contribution of the differences in microbial community structure between artificial and natural pit mud to SFB production.

Physicochemical properties play an important role in shaping microbial community structure and richness in the environment, including brewing ecosystems (10, 31). During long-time, natural maturation processes, the changes in pit mud microbial community composition were related to changes in pH and NH_4^+ content (10). In our study, the genera *Clostridium*, *Rummeliibacillus*, *Desulfotobacterium*, *Fonticella*, *Haloimpatiens*, and *Desulfosporosinus* correlated positively with total N, available P, and available K contents. These genera correlated positively with the flavor compounds

ethyl caproate, ethyl butyrate, and caproic acid and negatively with 3-methyl butanol and 2-butanol. The accumulation of fusel alcohols such as 3-methyl butanol and 2-butanol in SFB may have toxic effects, such as nervous hyperemia, dizziness, and headaches (32). This may suggest that an appropriate increase in the contents of total N, available K, and available P in APM can improve overall SFB flavor and the degree of comfort and satisfaction after drinking. Interestingly, except for *Clostridium*, which can produce caproic acid, the major metabolites of the other dominant genera are not directly related to the flavor compounds. For example, *Rummeliibacillus* produces indole (33), *Fonticella* and *Haloimpatiens* ferment saccharides to formate, acetate, lactate, ethanol, and CO₂ (24, 25), *Desulfitobacterium* ferments pyruvate to lactate and acetate (34), and *Desulfosporosinus* includes species able to produce acetate (26). However, the significant correlation between these genera and the SFB flavor compounds may indicate that the APM microbes have indirect roles and synergistic effects in the flavor compound formation.

Compared with the proteins previously detected from soil (35), the number of prokaryotic proteins (a total of 213 proteins) detected from the APM in this study was low. In addition to the different nature of the samples, the selection of the database for protein identification may affect the number of identified proteins (35, 36). For meta-proteomic analysis, one particular challenge is the choice of an appropriate database for peptide matching (37). For example, we initially identified the proteins through the UniProt database covering bacteria, archaea, and fungi and obtained a total of 324 microbial proteins (data not shown). However, to identify proteins more accurately, based on the sequencing data, a database was generated by collating entries in the UniProt database covering the dominant bacterial classes (>1%) in the first- and fourth-batch APM samples. This enabled us to limit the size of the database, which can reduce the search space and decrease the potential for false positives, and to allow the peptides to be matched against genomic data that are specific to the same taxa (37–39). In general, the results show good consistency between the amplicon sequencing and the metaproteomic data with respect to taxonomic structure. For example, *Clostridium* and *Lactobacillus* were abundant in both analyses. Consistent with the observed decrease in the relative abundance and gene copy number of *Lactobacillus*, the levels of *Lactobacillus* proteins were lower in the fourth-batch APM than in the first-batch APM. However, there were outliers. For example, *Syntrophomonas* contributed less than 0.01% to the 16S rRNA gene community, and *Hazenella* and *Weissella* contributed almost 6% and 4%, respectively; yet in the metaproteome more than 4% of all proteins were assigned to *Syntrophomonas*, and none were assigned to *Hazenella* and *Weissella*. Furthermore, even though both amplicon sequencing and qPCR showed increased abundance of *Clostridia*, the levels of most of the *Clostridium* proteins were similar in both the first- and fourth-batch APM samples. The reasons for these conflicting results may include some of the following: (i) the presence of DNA from dead organisms (40); (ii) bias in the DNA extraction, PCR, sequencing, and bioinformatics steps in amplicon sequencing (41); (iii) expression of the proteins below the threshold of detection at the time of sampling (39); or (iv) underrepresentation of the proteins in the protein sequence database (35). The last two points may explain the small number of proteins identified in this study. Together, these conflicting results add further evidence to previous findings that phylogenetic composition alone cannot reflect changes in the functions expressed by bacterial communities (42), highlighting the importance of multi-omics study designs to reveal the effects of brewing microbial communities and their contribution to SFB production.

To illustrate the main effects of the microbial dynamics in APM on SFB production, the multi-omics data were summarized in an overview model. Functional prediction suggested that the prokaryotic communities in the fourth-batch APM samples were actively engaged in organic acid metabolism, and the detected higher concentrations of proteins and metabolites in the corresponding metabolic pathways supported the prediction. Theoretically, the predicted enrichment of butanoate metabolism may be favorable for butyric acid production, providing more precursors for caproic acid

synthesis (21, 43). We speculate that the decreased levels of butyric acid might result from efficient caproic acid synthesis in the fourth-batch APM since a higher level of caproic acid was detected. Additionally, we found succinate semialdehyde and acetyl-CoA as two important links in multiple metabolic pathways that drain into the butanoate metabolism. The detected enrichment of acetyl-CoA acetyltransferase (*Swol_1934*) in the butanoate metabolism from *Syntrophomonas* suggests that rare bacterial groups may play important roles in SFB production. One challenge in SFB production is how to reduce lactic acid accumulation in pit mud since an excess of lactic acid is related to the increase in ethyl lactate that deteriorates SFB quality (21). In our study, the detected decrease in the abundance of D-lactate dehydrogenase (*ldhD*) in pyruvate metabolism is desirable in SFB production as this enzyme can generate lactic acid from pyruvate (44). The detected higher pyruvate and lower lactic acid levels in the fourth-batch APM samples agreed with the prediction. Taken together, the model provides a first systems-level overview of the effects of microbial succession on SFB quality.

Conclusion. This multi-omics study allowed the identification of taxon-specific functions and metabolic pathways that significantly differ during artificial pit mud maturation. The results revealed key linkages between several members of the prokaryotic community and metabolites and proteins in the artificial pit mud. These candidate microbial species, proteins, and metabolites may provide targets for predicting the quality of APM and its fermentation state, which will be helpful for further optimization of APM culture technique and improve the quality of strong-flavor baijiu. The results also emphasize the importance of longitudinal, multi-omics study designs to unravel the effects of brewing microbial communities and their contribution to strong-flavor baijiu production.

MATERIALS AND METHODS

Materials. APM samples were obtained from new fermentation cellars from an SFB-producing enterprise in Luzhou, Sichuan, China (105°27'13.80" E, 28°53'6.90" N). APM samples from eight cellars were selected for metabolomic analysis, and physicochemical properties, flavor compounds, 16S rRNA gene amplicons, and metaproteomes were analyzed from three of these. APMs were sampled at the initial state of SFB brewing and at the end of the first four brewing batches from September 2017 to December 2018. The brewing period of each batch was 90 days. APM samples (50 g) were collected from all corners and the center of the cellar walls, combined, and stored at −80°C. For each brewing batch, approximately 100 ml of SFB was collected into glass bottles, sealed, and stored at 4°C.

The APM for the cellars was prepared by mixing fresh soil with natural mature pit mud that had been in use for 50 years, wheat bran, and soybean meal (13:1:1:0.5, wt/wt/wt/wt) to form a solid matrix. The matrix was incubated with a starter culture (10:1, wt/wt) and appropriate amounts of SFB, tap water, and yellow water (a by-product formed during SFB brewing). Finally, the mixture was covered with polyester film and incubated at ambient temperature (20 to 34°C, with an average of 24°C) for 34 days (from 28 July 2017 to 30 August 2017). The starter culture preparation is described in detail in our previous study (45).

APM physicochemical properties. The APM moisture was measured using a dry/wet weight measurement method after drying the APM at 105°C for 3 h. The pH was determined in a 1:2.5 APM/water (wt/vol) slurry using a pH meter (PB10; Sartorius, Gottingen, Germany). The concentration of humus was determined as described by Mehlich (46). The concentrations of organic matter, total N, total P, total K, NH_4^+ , available P, and available K were measured as described previously (47).

SFB flavor compounds. Flavor compounds in SFB samples were extracted as described previously (48) and quantified using a Clarus 500 gas chromatograph (PerkinElmer, Waltham, MA) equipped with flame ionization detectors (FIDs) and a 50-m by 0.25-mm by 0.2- μm CP-Wax 57 CB column (Agilent, Santa Clara, CA) as described previously (28). The compounds were quantitated by comparison with the internal standards 2-methyl-2-butanol, pentyl acetate, and 2-ethylbutyric acid (Sigma-Aldrich, Shanghai, China).

16S rRNA gene amplicon sequencing and qPCR. DNA was extracted from the APM using a FastDNA Spin Kit for Soil (MP Biomedicals, Solon, OH) according to the manufacturer's instructions. For amplicon sequencing, the V3-V4 regions of the 16S rRNA genes were amplified using the primers 338F and 806R (49). PCR amplification, library preparation, and Illumina MiSeq sequencing were done as described previously (28).

The raw sequences were processed using the QIIME platform (version 1.9.1). Briefly, the forward and reverse reads were joined and assigned to samples based on the barcode sequences and then truncated by removing the barcode and primer sequences. Quality filtering of the joined sequences was performed, and sequences with lengths of <200 bp, ambiguous bases, and mean quality scores of ≤ 20 were discarded. The quality-filtered sequences were binned into OTUs at a 97% similarity level using the clustering program VSEARCH (version 1.9.6) against the Silva 132 database (<https://www.arb-silva.de/>) (50). The OTUs were assigned to taxa using the Ribosomal Database Project (RDP) Classifier at an 80%

confidence level (51). The Chao1 richness index and Shannon diversity index were calculated using Mothur (version 1.30.1 [<http://www.mothur.org/>]) (52). Comparative community analysis at the OTU level was done using PCA with STAMP (version 2.1.3 [<http://kiwi.cs.dal.ca/Software/STAMP>]) (53). Metabolic functions of the microbial communities were predicted using PICRUSt (version 1.0.0 [<http://picrust.github.com/picrust/>]) with KO classification (54).

Clostridia and *Lactobacillus* were quantified with qPCR using a 7900HT real-time PCR system (Applied Biosystems, Foster City, CA) and primers SJ-F/SJ-R (55) and Lac1/Lac2 (56) for *Clostridia* and *Lactobacillus*, respectively. Amplification was done in 20- μ l reaction mixtures containing SYBR Green Realtime PCR Master Mix (Transgen, Beijing, China), 2 pmol of each primer, template DNA, and distilled water, as described previously (22).

Protein preparation. For protein extraction, 1 g of APM sample was ground in liquid nitrogen with 10% (wt/wt) polyvinylpyrrolidone (PVP). The proteins in the samples were then extracted using the citrate/SDS-phenol (C/S-P) method (57). Briefly, the proteins were recovered by citrate (0.25 M, pH 8) and SDS buffer (1.25% [wt/vol] SDS, 0.1 M Tris-HCl, pH 6.8, 20 mM dithiothreitol [DTT]), extracted with phenol (pH 8), and precipitated by methanol and cold acetone. Subsequently, 100 μ g of protein was trypsin digested as described previously (58). For the PRM experiments, the digested peptides were desalted on 100-mg C₁₈ reverse-phase SPE columns according to the manufacturer's instructions (Waters, Milford, MA).

TMT-MS proteomics. TMT labeling was carried out using a TMT six-plex kit (Thermo Fisher Scientific, San Jose, CA) according to the manufacturer's instructions. The first-batch APM samples were labeled with 126, 127, and 128 mass tags, the fourth-batch APM samples were labeled with 129, 130, and 131 tags, and the samples were mixed at equal amounts. Labeled peptides were fractionated using an Agilent 1100 HPLC (high performance liquid chromatograph) and lyophilized. The lyophilized peptide fractions were dissolved into 1 μ g μ l⁻¹ using 0.1% formic acid (vol/vol), and 1 μ l of the peptide solution was analyzed using liquid chromatography-tandem MS (LC-MS/MS) with an EASY-nLC 1200 system and a Q-Exactive hybrid quadrupole-orbitrap mass spectrometer (Thermo Fisher Scientific). The detailed methods for peptide fractionation and LC-MS/MS analysis are provided in Text S1 in the supplemental material.

Peptides and proteins were identified with the Sequest HT search engine using Proteome Discoverer (version 2.2; Thermo Scientific) software, and the raw MS/MS data were compared against a database created by collating all UniProt entries corresponding to organisms (class level) with an abundance of >1% in the 16S rRNA gene amplicon sequencing data from the first- and fourth-batch APM samples. We used the setting "enable protein group" in Proteome Discoverer to group redundant proteins that share sets of identified peptides. The proteins within a group are scored according to the number of unique peptides, the number of peptide spectrum matches (PSMs), and the sequence coverage. The highest-scoring protein of a group was selected as the representative protein (master protein). Database searches were performed using the following parameters: sample type, TMT six-plex (peptide labeled); Cys alkylation, iodoacetamide; digestion, trypsin; instrument, Q-Exactive. The global false-discovery rate (FDR) was ≤ 0.01 , and protein groups considered for quantification required at least one unique peptide. Proteins with fold change values of ≥ 2 or ≤ 0.5 ($P < 0.05$) were considered differential. GO enrichment and KEGG pathway analysis were performed using OmicsBean software (<http://www.omicsbean.cn>), using *Escherichia coli* K-12 MG1655 as a reference organism.

Target analysis by PRM-MS proteomics. The protein expression levels obtained using TMT-MS analysis were confirmed by quantifying the expression levels of 15 selected proteins by PRM-MS analysis with an EASY-nLC 1200 system and a Q-Exactive mass spectrometer (Thermo Fisher Scientific). The detailed LC-MS/MS analysis methods are described in Text S1. MS raw data from data dependent analysis (DDA) were compared against the database used for TMT-MS proteomics using Proteome Discoverer (version 2.2; Thermo Scientific). The FDR was set to 0.01 for proteins and peptides. Quantitative data processing was done with Skyline software (version 4.1.0) (59) using the following parameters: precursor charges, 2; ion charges, 1 and 2; ion type, y and b; product ion selection, 3 ions; special ions, N-terminal to proline; ion match tolerance, 0.05 m/z; precursor mass analyzer, Orbitrap; isotope peaks included for MS filtering, count; MS resolving power at 200 m/z, 60,000; acquisition method selected for MS/MS filtering, targeted; MS/MS resolving power at 200 m/z, 30,000.

Metabolomic analysis of APM. Untargeted metabolomic analysis to identify the differential metabolites between the first- and fourth-batch APM samples was done using GC-MS and LC-MS. GC-MS analysis was performed using a 7890B gas chromatograph (Agilent) coupled to a 5977A mass selective detector (Agilent). LC-MS analysis was performed on an Acquity UPLC system (Waters, Milford, MA) coupled with a Xevo G2-XS quadrupole time-of flight (QToF) mass spectrometer (Waters). The detailed sample preparation and MS analysis methods are described in Text S1.

GC-MS raw data were converted to computable document format (CDF) using ChemStation (version E.02.02.1431; Agilent). Data processing was done using Chroma TOF (version 4.34; Leco, St. Joseph, MI). Metabolites were annotated by searching the databases of Kind et al. (60). LC-MS raw data were processed using Progenesis Q1 (Waters). Metabolites were annotated using the human metabolome database (HMDB) (61) and LIPID MAPS (62). Statistical tests were done using SIMCA (version 14.0; Umetrics, Umeå, Sweden). Differential metabolites were identified using an OPLS-DA model; variables with a VIP value of >1 and a Student's *t* test *P* value of <0.05 were considered differential. The quality of the OPLS-DA models was verified by the typical cross-validation procedure of leaving one-seventh of the samples out of each round and then assessing the model. The pathways of the metabolites were identified using MBROLE, version 2.0 (63), based on the KEGG database using *Escherichia coli* K-12 MG1655 as a reference organism.

Statistical analysis. Statistical analysis was performed with SPSS, version 20.0, software for Windows (SPSS, Inc., Chicago, IL). Pearson's correlation coefficient analysis was applied to investigate the associations between APM prokaryotic community, physicochemical properties, and SFB flavor compounds. Differences in the APM physicochemical properties, SFB flavor compounds, prokaryotic diversity indices, and the relative abundance of the taxa and KOs among groups were tested by one-way analysis of variance (ANOVA) with Duncan's multiple-range tests.

Data availability. The sequencing data were submitted to the Sequence Read Archive of the NCBI database under BioProject [PRJNA577744](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA577744) (BioSample accession numbers [SAMN13037336](https://www.ncbi.nlm.nih.gov/biosample/SAMN13037336) to [SAMN13037350](https://www.ncbi.nlm.nih.gov/biosample/SAMN13037350)). The mass spectrometry proteomics data were submitted to the iProX database (64) (accession number IPX0001834000, <https://www.iprox.org/page/SCV017.html?query=IPX0001834000>) and ProteomeXchange (accession number PXD016140, <http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PX016140>).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

TEXT S1, DOCX file, 0.02 MB.

FIG S1, TIF file, 1 MB.

TABLE S1, XLSX file, 0.01 MB.

TABLE S2, XLSX file, 0.03 MB.

TABLE S3, XLSX file, 0.02 MB.

TABLE S4, XLSX file, 0.1 MB.

TABLE S5, XLSX file, 0.02 MB.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their constructive suggestions that improved this paper.

This study was supported by the National Natural Science Foundation of China (grant number 31801534), the Innovation Ability Promotion Project of Sichuan Academy of Agricultural Sciences (grant number 2019KXJJ003), and the Frontier Science Foundation of Sichuan Academy of Agricultural Sciences (grant number 2019QYXK012).

We declare that we have no competing financial interests.

REFERENCES

1. Zou W, Zhao C, Luo H. 2018. Diversity and function of microbial community in Chinese strong-flavor baijiu ecosystem: a review. *Front Microbiol* 9:671. <https://doi.org/10.3389/fmicb.2018.00671>.
2. Yao F, Yi B, Shen C, Tao F, Liu Y, Lin Z, Xu P. 2015. Chemical analysis of the Chinese liquor Luzhoulaojiao by comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry. *Sci Rep* 5:9553. <https://doi.org/10.1038/srep09553>.
3. Zheng XF, Han BZ. 2016. Baijiu (白酒), Chinese liquor: history, classification and manufacture. *J Ethn Foods* 3:19–25. <https://doi.org/10.1016/j.jef.2016.03.001>.
4. Li DH. 2015. Discussion on product quality standards of Chinese liquor. *Liquor Mak* 42:7–9. (In Chinese.)
5. Li J, Liu Y, Sui L, Wang Y. 2019. Research progress on ethyl hexanoate-increasing and ethyl lactate-decreasing of strong-flavor Baijiu. *China Brew* 38:1–4. (In Chinese.)
6. Wang X, Du H, Xu Y. 2017. Source tracking of prokaryotic communities in fermented grain of Chinese strong-flavor liquor. *Int J Food Microbiol* 244:27–35. <https://doi.org/10.1016/j.jfoodmicro.2016.12.018>.
7. Zheng XW, Tabrizi MR, Nout MJR, Han BZ. 2011. *Daqu*—a traditional Chinese liquor fermentation starter. *J I Brewing* 117:82–90. <https://doi.org/10.1002/j.2050-0416.2011.tb00447.x>.
8. Wang CD, Chen Q, Wang Q, Li CH, Leng YY, Li SG, Zhou XW, Han WJ, Li JG, Zhang XH, Li YZ. 2014. Long-term batch brewing accumulates adaptive microbes, which comprehensively produce more flavorful Chinese liquors. *Food Res Int* 62:894–901. <https://doi.org/10.1016/j.foodres.2014.05.017>.
9. Hu XL, Du H, Xu Y. 2015. Identification and quantification of the caproic acid-producing bacterium *Clostridium kluyveri* in the fermentation of pit mud used for Chinese strong-aroma type liquor production. *Int J Food Microbiol* 214:116–122. <https://doi.org/10.1016/j.jfoodmicro.2015.07.032>.
10. Tao Y, Li J, Rui J, Xu Z, Zhou Y, Hu X, Wang X, Liu M, Li D, Li X. 2014. Prokaryotic communities in pit mud from different-aged cellars used for the production of Chinese strong-flavored liquor. *Appl Environ Microbiol* 80:2254–2260. <https://doi.org/10.1128/AEM.04070-13>.
11. Xiong ZS. 2001. Origin and development of Chinese first fermentation pit-record of Luzhou laojiao liquor. *Liquor Mak Sci Technol* 104:17–22. (In Chinese.)
12. Ding X, Wu C, Huang J, Li H, Zhou R. 2014. Eubacterial and archaeal community characteristics in the man-made pit mud revealed by combined PCR-DGGE and FISH analyses. *Food Res Int* 62:1047–1053. <https://doi.org/10.1016/j.foodres.2014.05.045>.
13. Liu MK, Tang YM, Zhao K, Gu YF, Ren DQ, Yao WC, Tian XH, Liu Y, Zhang XP. 2017. Recent advances in research on the community, isolation, and application of microbes in the pit mud used in manufacture of Chinese strong-flavor baijiu. *Microbiol China* 44:1222–1229. (In Chinese.)
14. Wu S, Lai D, Li J, Fu J. 2014. Pit construction and pit mud culture in Gansu. *Liquor Mak Sci Technol* 10:68–71. (In Chinese.)
15. Zheng Y, Zhou Y, Jiang P, Xu S, Ma X, Ke T. 2017. Culture of quality manmade aged pit mud in Shedian Laojiu distillery. *Liquor Mak Sci Technol* 9:57–60. (In Chinese.)
16. Zhang L, Zhou R, Niu M, Zheng J, Wu C. 2015. Difference of microbial community stressed in artificial pit muds for Luzhou-flavour liquor brewing revealed by multiphase culture-independent technology. *J Appl Microbiol* 119:1345–1356. <https://doi.org/10.1111/jam.12943>.
17. Sun Z, Chen C, Hou X, Zhang J, Tian F, Li C. 2017. Prokaryotic diversity and biochemical properties in aging artificial pit mud used for the production of Chinese strong flavor liquor. *3 Biotech* 7:335. <https://doi.org/10.1007/s13205-017-0978-0>.
18. Xiao M, Yang J, Feng Y, Zhu Y, Chai X, Wang Y. 2017. Metaproteomic strategies and applications for gut microbial research. *Appl Microbiol Biotechnol* 101:3077–3088. <https://doi.org/10.1007/s00253-017-8215-7>.

19. Bouchez T, Blieux AL, Dequiedt S, Domaizon I, Dufresne A, Ferreira S, Godon JJ, Hellal J, Joulian C, Quaiser A, Martin-Laurent F, Mauffret A, Monier JM, Peyret P, Schmitt-Koplin P, Sibourg O, D'oirion E, Bispo A, Deportes I, Grand G, Cuny P, Maron PA, Ranjard L. 2016. Molecular microbiology methods for environmental diagnosis. *Environ Chem Lett* 14:423–441. <https://doi.org/10.1007/s10311-016-0581-3>.
20. 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>.
21. Tao Y, Wang X, Li X, Wei N, Jin H, Xu Z, Tang Q, Zhu X. 2017. The functional potential and active populations of the pit mud microbiome for the production of Chinese strong-flavour liquor. *Microb Biotechnol* 10:1603–1615. <https://doi.org/10.1111/1751-7915.12729>.
22. Hu X, Du H, Ren C, Xu Y. 2016. Illuminating anaerobic microbial community and cooccurrence 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>.
23. Zhang QY, Yuan YJ, Liao ZM, Zhang WX. 2017. Use of microbial indicators combined with environmental factors coupled to metrology tools for discrimination and classification of Luzhou-flavored pit muds. *J Appl Microbiol* 123:933–943. <https://doi.org/10.1111/jam.13544>.
24. Fraj B, Ben HW, Postec A, Hamdi M, Ollivier B, Fardeau ML. 2013. *Fonticella tunisiensis* gen. nov., sp. nov., isolated from a hot spring. *Int J Syst Evol Microbiol* 63:1947–1950. <https://doi.org/10.1099/ijso.0.041947-0>.
25. Wu D, Zhang NF, Sun C, Zhang WW, Han SB, Pan J, Wu M, Th D, Zhu XF. 2016. *Haloimpatiens lingqiaoensis* gen. nov., sp. nov., an anaerobic bacterium isolated from paper-mill wastewater. *Int J Syst Evol Microbiol* 66:628–632. <https://doi.org/10.1099/ijsem.0.000765>.
26. Stackebrandt E, Sproer C, Rainey FA, Burghardt J, Pauker O, Hippe H. 1997. Phylogenetic analysis of the genus *Desulfotomaculum*: evidence for the misclassification of *Desulfotomaculum guttoideum* and description of *Desulfotomaculum orientis* as *Desulfosporosinus orientis* gen. nov., comb. nov. *Int J Syst Bacteriol* 47:1134–1139. <https://doi.org/10.1099/00207713-47-4-1134>.
27. Fusco V, Quero GM, Cho GS, Kabisch J, Meske D, Neve H, Bockelmann W, Franz CM. 2015. The genus *Weissella*: taxonomy, ecology and biotechnological potential. *Front Microbiol* 6:155. <https://doi.org/10.3389/fmicb.2015.00155>.
28. 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>.
29. Rea S, Bowman JP, Popovski S, Pimm C, Wright AD. 2007. *Methanobrevibacter millerae* sp. nov. and *Methanobrevibacter olleyae* sp. nov., methanogens from the ovine and bovine rumen that can utilize formate for growth. *Int J Syst Evol Microbiol* 57:450–456. <https://doi.org/10.1099/ijso.0.063984-0>.
30. Felchner-Zwirello M, Winter J, Gallert C. 2013. Interspecies distances between propionic acid degraders and methanogens in syntrophic consortia for optimal hydrogen transfer. *Appl Microbiol Biotechnol* 97:9193–9205. <https://doi.org/10.1007/s00253-012-4616-9>.
31. Li P, Lin W, Liu X, Wang X, Luo L. 2016. Environmental factors affecting microbiota dynamics during traditional solid-state fermentation of Chinese daqu starter. *Front Microbiol* 7:1237. <https://doi.org/10.3389/fmicb.2016.01237>.
32. Kim HM, Yang G, Kim JY, Yoon SJ, Shin BK, Lee J, Park JH, Kwon SW. 2017. Simultaneous determination of volatile organic compounds in commercial alcoholic beverages by gas chromatography with flame ionization detection. *J AOAC Int* 100:1492–1499. <https://doi.org/10.5740/jaoacint.17.0006>.
33. Her J, Kim J. 2013. *Rummeliibacillus suwonensis* sp. nov., isolated from soil collected in a mountain area of South Korea. *J Microbiol* 51:268–272. <https://doi.org/10.1007/s12275-013-3126-5>.
34. Utkin I, Woese C, Wiegand J. 1994. Isolation and characterization of *Desulfotomaculum dehalogenans* gen. nov., sp. nov., an anaerobic bacterium which reductively dechlorinates chlorophenolic compounds. *Int J Syst Bacteriol* 44:612–619. <https://doi.org/10.1099/00207713-44-4-612>.
35. Zampieri E, Chiappello M, Daghighi S, Bonfante P, Mello A. 2016. Soil metaproteomics reveals an inter-kingdom stress response to the presence of black truffles. *Sci Rep* 6:25773. <https://doi.org/10.1038/srep25773>.
36. Heyer R, Schallert K, Zoun R, Becher B, Saake G, Benndorf D. 2017. Challenges and perspectives of metaproteomic data analysis. *J Biotechnol* 261:24–36. <https://doi.org/10.1016/j.jbiotec.2017.06.1201>.
37. Jagtap P, Goslinga J, Kooren JA, McGowan T, Wroblewski MS, Seymour SL, Griffin TJ. 2013. A two-step database search method improves sensitivity in peptide sequence matches for metaproteomics and proteogenomics studies. *Proteomics* 13:1352–1357. <https://doi.org/10.1002/pmic.201200352>.
38. Russo DA, Couto N, Beckerman AP, Pandhal J. 2016. A metaproteomic analysis of the response of a freshwater microbial community under nutrient enrichment. *Front Microbiol* 7:1172. <https://doi.org/10.3389/fmicb.2016.01172>.
39. Verberkmoes NC, Russell AL, Shah M, Godzik A, Rosenquist M, Halfvarson J, Lefsrud MG, Apajalahti J, Tysk C, Hettich RL, Jansson JK. 2009. Shotgun metaproteomics of the human distal gut microbiota. *ISME J* 3:179–189. <https://doi.org/10.1038/ismej.2008.108>.
40. Nocker A, Richter-Heitmann T, Montijn R, Schuren F, Kort R. 2010. Discrimination between live and dead cells in bacterial communities from environmental water samples analyzed by 454 pyrosequencing. *Int Microbiol* 13:59–65. <https://doi.org/10.2436/20.1501.01.111>.
41. McLaren MR, Willis AD, Callahan BJ. 2019. Consistent and correctable bias in metagenomic sequencing measurements. *Elife* 8:e46923. <https://doi.org/10.7554/eLife.46923>.
42. Zuniga C, Zaramela L, Zengler K. 2017. Elucidation of complexity and prediction of interactions in microbial communities. *Microb Biotechnol* 10:1500–1522. <https://doi.org/10.1111/1751-7915.12855>.
43. Barker HA, Kamen MD, Bornstein BT. 1945. The synthesis of butyric and caproic acids from ethanol and acetic acid by *Clostridium Kluyveri*. *Proc Natl Acad Sci U S A* 31:373–381. <https://doi.org/10.1073/pnas.31.12.373>.
44. Zhang J, Gong G, Wang X, Zhang H, Tian W. 2015. Positive selection on D-lactate dehydrogenases of *Lactobacillus delbrueckii* subspecies *bulgaricus*. *IET Syst Biol* 9:172–179. <https://doi.org/10.1049/iet-syb.2014.0056>.
45. Liu MK, Tang YM, Zhao K, Liu Y, Guo XJ, Tian XH, Ren DQ, Yao WC. 2019. Contrasting bacterial community structure in artificial pit mud starter cultures of different qualities: a complex biological mixture for Chinese strong-flavor Baijiu production. *3 Biotech* 9:89. <https://doi.org/10.1007/s13205-019-1622-y>.
46. Mehlich A. 1984. Photometric determination of humic matter in soils, a proposed method. *Commun Soil Sci Plant Anal* 15:1417–1422. <https://doi.org/10.1080/00103628409367569>.
47. Zhang GL, Gong ZT. 2012. Soil survey laboratory methods. Science Press, Beijing, China. (In Chinese.)
48. Zheng J, Liang R, Wu C, Zhou R, Liao X. 2014. Discrimination of different kinds of Luzhou-flavor raw liquors based on their volatile features. *Food Res Int* 56:77–84. <https://doi.org/10.1016/j.foodres.2013.12.011>.
49. Wei H, Wang L, Hassan M, Xie B. 2018. Succession of the functional microbial communities and the metabolic functions in maize straw composting process. *Bioresour Technol* 256:333–341. <https://doi.org/10.1016/j.biortech.2018.02.050>.
50. Quast C, Priesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41:D590–D596. <https://doi.org/10.1093/nar/gks1219>.
51. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73:5261–5267. <https://doi.org/10.1128/AEM.00062-07>.
52. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75:7537–7541. <https://doi.org/10.1128/AEM.01541-09>.
53. Parks DH, Tyson GW, Philip H, Beiko RG. 2014. STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* 30:3123–3124. <https://doi.org/10.1093/bioinformatics/btu494>.
54. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Vega Thurber RL, Knight R, Beiko RG, Huttenhower C. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* 31:814–821. <https://doi.org/10.1038/nbt.2676>.
55. Hu XL, Wang HY, Wu Q, Xu Y. 2014. Development, validation and application of specific primers for analyzing the clostridial diversity in dark fermentation pit mud by PCR-DGGE. *Bioresour Technol* 163:40–47. <https://doi.org/10.1016/j.biortech.2014.04.008>.

56. Xu W, Huang Z, Zhang X, Li Q, Lu Z, Shi J, Xu Z, Ma Y. 2011. Monitoring the microbial community during solid-state acetic acid fermentation of Zhenjiang aromatic vinegar. *Food Microbiol* 28:1175–1181. <https://doi.org/10.1016/j.fm.2011.03.011>.
57. Wang HB, Zhang ZX, Li H, He HB, Fang CX, Zhang AJ, Li QS, Chen RS, Guo XK, Lin HF, Wu LK, Lin S, Chen T, Lin RY, Peng XX, Lin WX. 2011. Characterization of metaproteomics in crop rhizospheric soil. *J Proteome Res* 10:932–940. <https://doi.org/10.1021/pr100981r>.
58. Fang D, Yang W, Deng Z, An X, Zhao L, Hu Q. 2017. Proteomic investigation of metabolic changes of mushroom (*Flammulina velutipes*) packaged with nanocomposite material during cold storage. *J Agric Food Chem* 65:10368–10381. <https://doi.org/10.1021/acs.jafc.7b04393>.
59. Brendan ML, Tomazela DM, Nicholas S, Matthew C, Finney GL, Barbara F, Randall K, Tabb DL, Liebler DC, Maccoss MJ. 2010. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26:966–968. <https://doi.org/10.1093/bioinformatics/btq054>.
60. Kind T, Wohlgemuth G, Lee DY, Lu Y, Palazoglu M, Shahbaz S, Fiehn O. 2009. FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal Chem* 81:10038–10048. <https://doi.org/10.1021/ac9019522>.
61. Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vázquez-Fresno R, Sajed T, Johnson D, Li C, Karu N, Sayeeda Z, Lo E, Assempour N, Berjanskii M, Singhal S, Arndt D, Liang Y, Badran H, Grant J, Serra-Cayuela A, Liu Y, Mandal R, Neveu V, Pon A, Knox C, Wilson M, Manach C, Scalbert A. 2018. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res* 46:D608–D617. <https://doi.org/10.1093/nar/gkx1089>.
62. Eoin F, Manish S, Dawn C, Shankar S. 2007. LIPID MAPS online tools for lipid research. *Nucleic Acids Res* 35:W606–W612. <https://doi.org/10.1093/nar/gkm324>.
63. López-Ibáñez J, Pazos F, Chagoyen M. 2016. MBROLE 2.0-functional enrichment of chemical compounds. *Nucleic Acids Res* 44:W201–W204. <https://doi.org/10.1093/nar/gkw253>.
64. Ma J, Chen T, Wu S, Yang C, Bai M, Shu K, Li K, Zhang G, Jin Z, He F, Hermjakob H, Zhu Y. 2019. iProX: an integrated proteome resource. *Nucleic Acids Res* 47:D1211–D1217. <https://doi.org/10.1093/nar/gky869>.