Analysis of microbial Diversity in Soy Sauce fermented grains by Illumina High-throughput sequencing technique

Yongbi Liu, Qiao Lin*, Dongmei Yan, Li Cai
Xichang College, Xichang 615000, Sichuan, China

*Corresponding author e-mail: 66656736@qq.com

Abstract. This paper mainly studies the diversity of microorganisms in the sauce fermented grains during the fermentation of soy sauce, and uses Illumina high-throughput sequencing technique to sequence the 16S region and ITS1 region of 7 samples respectively, and finds that the evolution law of microbial community structure in the whole soy sauce fermentation process is from complex to simple, It also shows that soy sauce fermentation environment has the effect of inhibiting microbial growth. The results of the analysis and sequencing can be: Staphylococcus (Staphylococcus), Klebsiella, Weiss and Bacillus, representing fungal strains of Aspergillus and abnormal Wickham yeast. In this study, the microbial colony structure during the fermentation production of soy sauce was analyzed, which was helpful to control the growth and propagation of microorganisms and ensure the quality and flavor of soy sauce products.

1. Introduction
Soy sauce is a traditional seasoning in China with a long history. In the process of koji making and fermentation of soy sauce, microbes in the air would also fall into fermented grains for breeding, so they would secrete many enzymes. For example, under the action of yeast, sugars could be converted into alcohol, and lactic acid bacteria could form lactic acid. Therefore, soy sauce fermentation process is to use these enzymes in certain conditions, decomposition of raw materials, forming a special flavor of soy sauce. Therefore, it can be considered that soy sauce is formed under the systematic action of bacteria, aspergillus, yeast and other microorganisms [1-2]. In the fermentation and koji making of soy sauce, the mutual metabolism of microorganisms is used to generate flavoring substances. At the same time, during the fermentation of soy sauce, the changes in microbial reactions are complex. In general, brewing soy sauce is closely related to microbiology and biochemistry. Therefore, it is of great significance to study the microbial species and composition in soy sauce fermentation process for controlling the microbial growth and reproduction in soy sauce fermentation process and ensuring the quality of soy sauce.

2. Materials and methods

2.1. Materials and instruments

2.1.1. Samples. The test samples were collected from a fermentation tank of shengsheng brewing oil plant, and the sampling time was 7 times, respectively: fermentation 226, 154, 131, 101, 71, 24 and 2 d (samples were labeled S1, S2, S3, S4, S5, S6 and S7).
2.1.2. Main reagents and instruments. Biometra Tgradient PCR, Biometra, Germany; Biorad DCode Apparatus DGGE, Biorad, USA; GelDoc 2000 System, Biorad, USA; Soy sauce DNA extraction kit, Beijing dingguo biotechnology co., LTD., Beijing; 40% acrylamide/methylene acrylamide solution (37.5:1), Beijing puboxin biotechnology co., LTD., Beijing.

2.2. Methods

2.2.1. Macronomic preparation. Liquid nitrogen grinding and extraction with a kit were used to extract acer from mash. After mixing each group of mash samples evenly, 10 g of mash was weighed and washed twice with 20 mmoll-1 EDTA. The washed mash was transferred to Portland, where liquid nitrogen was added and the cells were fully ground at low temperature to break up. Finally using PowerHA ® Soil DNA Isolation Kit Kit and extraction and purification according to the manual [3]. Nano Drop 2000c ultra micro spectrophotometer was used to determine the concentration of the extracted macro genome of fermented mash and record relevant purity parameters. Finally, the integrity of extracted genomic DNA was determined by 0.6% agarose-gel electrophoresis.

2.2.2. Sequencing process. The bacterial sequencing region was v3-v4 variable region of 16S rRNA, and the common primers were 341F (5 'cctacgggnggcwgcag-3') and 908R (5 'ccgtcaattcmtttgagttt-3'). The sequencing region was the ITS1 variable region of 18S rRNA, and the general primers were ITS1F (5 'CTTGGTCATTtagaggaagtaa-3') and ITS4R (5 'tcctccgcttattgatatgc-3'). The amplified fragments of these primers have been proved to have good clustering effect of microbial species [4]. PCR reaction and amplification product extraction were also carried out according to the description in literature [5]. The extracted amplicons were sequenced by Illumina MiSeq platform applied by Shanghai sangon co., LTD.

2.2.3. Sequence analysis. After the original sequencing data were obtained, reads were spliced and the chimera was removed. According to Barcode, the data were returned to the samples and the sequences with a fragment length of less than 200 bp, a single base repeat of more than 8, and a fuzzy base were deleted to obtain high-quality sequence files. Then RDP software was used to compare sequences in the SILVA database under the 97% similarity threshold for species classification, and then multiple sequences were clustered according to the distance between their sequences. According to the similarity between sequences, it is divided into Operational taxonomic unit (OTU) as the threshold value, and the number of sequences contained in each OTU is counted. Finally, MOTHUR software was used to calculate the abundance index and diversity index of the microbial community [6].

3. Result analysis

3.1. Sequencing data quality assessment
The evaluation results of sequencing data of each Sample are shown in the following table: Sample ID is the name of the Sample; PE Reads is the number of double-ended Reads obtained by sequencing. Raw Tags are the number of original sequences obtained by splicing of double-ended reads; Clean Tags are the number of optimized sequences obtained after filtering the original sequence; Effective Tags are the Effective sequence Numbers after filtering chimera of Clean Tags; AvgLen (bp) is the average sequence length of the sample. GC (%) is the GC content of the sample, that is, the percentage of G and C bases in the total bases; Q 20(%) is the percentage of bases with mass value greater than or equal to 20 in the total number of bases; Q30 (%) is the percentage of bases with mass value greater than or equal to 30 in the total number of bases; Effective (%) is the percentage of Effective Tags in PE Reads.
Table 1. Statistics of data processing results of sample 16s sequencing

| Sample ID | PE Reads | Raw Tags | Clean Tags | Effective Tags | AvgLen(bp) | GC (%) | Q20 (%) | Q30 (%) | Effective (%) |
|-----------|----------|----------|------------|---------------|------------|--------|---------|---------|---------------|
| S1        | 20,885   | 19,408   | 17,260     | 17,213        | 475        | 52.47  | 96.72   | 93.6    | 82.42         |
| S2        | 9,785    | 9,166    | 8,200      | 8,105         | 474        | 51.12  | 96.66   | 93.56   | 82.83         |
| S3        | 84,338   | 78,867   | 70,368     | 67,785        | 475        | 53.55  | 96.8    | 93.7    | 80.37         |
| S4        | 82,426   | 76,761   | 68,549     | 67,657        | 475        | 53.27  | 96.81   | 93.72   | 82.08         |
| S5        | 85,802   | 79,323   | 69,888     | 67,717        | 475        | 53.11  | 96.51   | 93.21   | 78.92         |
| S6        | 87,537   | 81,356   | 71,999     | 69,962        | 475        | 52.35  | 96.56   | 93.33   | 79.92         |
| S7        | 122,669  | 113,881  | 100,361    | 100,178       | 475        | 54.34  | 96.56   | 93.24   | 81.67         |

After quality control filtration, the number of Reads sequences within the corresponding length range in each sample was counted in the length distribution diagram. The Effective Tags length distribution diagram is as follows:

Table 2. Statistics of ITS sequencing data processing results of samples

| Sample ID | PE Reads | Raw Tags | Clean Tags | Effective Tags | AvgLen(bp) | GC (%) | Q20 (%) | Q30 (%) | Effective (%) |
|-----------|----------|----------|------------|---------------|------------|--------|---------|---------|---------------|
| S1        | 52,238   | 50,458   | 50,458     | 50,377        | 305        | 53.29  | 99.27   | 98.11   | 96.44         |
| S2        | 92,897   | 54,728   | 54,728     | 54,645        | 305        | 53.39  | 99.24   | 98.03   | 58.82         |
| S3        | 95,349   | 88,958   | 88,958     | 88,705        | 305        | 54.31  | 99.28   | 98.12   | 93.03         |
| S4        | 46,863   | 44,135   | 44,135     | 44,003        | 305        | 54.32  | 99.26   | 98.06   | 93.9          |
| S5        | 110,325  | 25,470   | 25,470     | 25,408        | 305        | 54.07  | 99.24   | 98.01   | 23.03         |
| S6        | 173,724  | 163,288  | 163,288    | 163,044       | 306        | 54.56  | 99.25   | 98.06   | 93.85         |
| S7        | 100,766  | 96,974   | 96,974     | 96,800        | 306        | 53.25  | 99.25   | 98.03   | 96.06         |

After quality control filtration, the number of Reads sequences within the corresponding length range in each sample was counted in the length distribution diagram. The Effective Tags length distribution diagram is as follows:
According to the figure, in the Effective Tags length distribution diagram of 16S sequencing, the number of Reads sequences is mainly concentrated in the 430-480 length range, and the number of Reads sequences in the 470-480 length range is the most. In the Effective Tags length distribution diagram of ITS sequencing, the number of Reads sequences is mainly concentrated in the range of 250-500, and the most Reads sequences are in the range of 280-310.

3.2. OTU analysis

UCLUST[7] in QIIME[8] (version 1.8.0) software is used to cluster the Tags and obtain OTU at 97% similarity level, and make taxonomic annotation of OTU based on Silva (bacteria) and UNITE (fungi)
taxonomic database. The figure below shows the number of OTU for each sample obtained through clustering: the number on the column is the number of OTU for the corresponding sample.

![Figure 3. OTU number of 16S sequencing](image)

![Figure 4. OTU number diagram of ITS sequencing](image)

In general, if the similarity between sequences, for example, different 16SrRNA sequences is greater than 98%, it can be defined as an OTU. Each OTU corresponds to a different 16SrRNA sequence, that is, each OTU corresponds to a different bacterial (microorganism) species. Through OTU analysis, the microbial diversity and abundance of different microorganisms in the sample can be known.

From the figure OUT, it can be seen that the number of bacteria is more than that of fungi in the initial stage of soy sauce fermentation. As the fermentation progresses, the number of fungi gradually
increases, while the number of bacteria generally increases first, reaches the maximum in the middle stage of fermentation, and then decreases.

The petal diagram is another form of Venn diagram. The statistics and comparison of OTU of all samples are made to find out the common OTU in all samples and the unique OTU only in all samples. According to OTU of each sample, the petal graph is as follows: the number in the middle of the petal graph represents the number of OTU Shared by all samples, the number on the petal represents the number of OTU unique to the sample, and the OTU not unique to a single sample or Shared by all samples is not shown in the graph.

![Figure 5. 16s sequencing OTU petals figure](image)

![Figure 6. ITS sequencing OTU petals figure](image)

As can be seen from the above figure, in the OTU petal diagram of 16S sequencing, the number of OTU Shared by 7 samples is 25, and only 5 samples have a unique number of OTU, which is 2. That is, during the whole fermentation process of soy sauce, only the middle fermentation stage has unique bacterial species. In the OTU petal diagram of ITS sequencing, the number of OTU Shared by the 7 samples was 4, among which the number of OTU unique to the samples 1, 2, 3, 4 and 6 was 7, 2, 3, 2 and 1, respectively. That is, from the early fermentation stage to the late fermentation stage of soy sauce, the number of fungal species increased gradually.

### 3.3. Species annotation and taxonomic analysis

#### 3.3.1. Display of clustering results

Screening low content OTU. The original OTU clustering results may contain extremely low abundance OTU (species abundance less than 0.005%), the final OTU list will be obtained and the number of tags annotated to species in each level of samples will be counted. The results are shown in the following table: Kindom, Phylum, Class, Order, Family, Genus and Species represent seven taxonomic levels respectively. The following table is a TAB of Tags of various levels of the sample, where the value represents the total number of Tags covered in the corresponding level of the sample:

**Table 3. Tags of 16S sequencing samples**

| Sample | Kindom | Phylum | Class | Order | Family | Genus | Species |
|--------|--------|--------|-------|-------|--------|-------|---------|
| S1     | 16,050 | 16,050 | 16,050| 16,046| 16,046 | 14,678| 2,673   |
| S2     | 7,327  | 7,327  | 7,327 | 7,327 | 7,327  | 7,011 | 4,375   |
| S3     | 62,144 | 62,144 | 62,144| 62,140| 62,140 | 55,071| 27,274  |
| S4     | 63,457 | 63,457 | 63,457| 63,457| 63,457 | 57,990| 34,721  |
| S5     | 65,070 | 65,070 | 65,070| 65,070| 65,070 | 54,051 | 17,407  |
| S6     | 65,472 | 65,472 | 65,472| 65,470| 65,470 | 60,040 | 14,455  |
| S7     | 98,245 | 98,245 | 98,245| 98,089| 98,089 | 86,708 | 6,440   |

Note: Sample is listed as Sample name
Table 4. Tags statistics table of various grades of ITS sequencing samples

| Sample | Kingdom | Phylum       | Class        | Order        | Family        | Genus       | Species |
|--------|---------|--------------|--------------|--------------|---------------|-------------|---------|
| S1     | 50,118  | 49,954       | 49,781       | 49,781       | 549,781       | 49,670      | 1,510   |
| S2     | 54,507  | 54,480       | 54,480       | 54,480       | 54,480        | 54,480      | 5,4368  |
| S3     | 88,411  | 88,378       | 88,378       | 88,378       | 88,378        | 88,378      | 1,446   |
| S4     | 43,863  | 43,766       | 43,764       | 43,764       | 43,764        | 43,764      | 317     |
| S5     | 25,722  | 25,272       | 25,215       | 25,215       | 25,215        | 25,215      | 126     |
| S6     | 162,862 | 162,862      | 162,862      | 162,862      | 162,862       | 162,862     | 446     |
| S7     | 96,486  | 96,485       | 96,485       | 96,485       | 96,485        | 96,485      | 6,729   |

Note: Sample is listed as Sample name

The following table is the statistical table of species in each grade of the sample, showing the number of species types in each grade of the sample:

Table 5. Number of species types of each grade in the sample of 16S sequencing

| Sample | Kingdom | Phylum | Class | Order | Family | Genus | Species |
|--------|---------|--------|-------|-------|--------|-------|---------|
| S1     | 1       | 4      | 5     | 9     | 13     | 18    | 8       |
| S2     | 1       | 3      | 4     | 6     | 12     | 16    | 6       |
| S3     | 1       | 4      | 5     | 8     | 13     | 18    | 9       |
| S4     | 1       | 3      | 4     | 8     | 14     | 19    | 8       |
| S5     | 1       | 3      | 4     | 10    | 16     | 22    | 10      |
| S6     | 1       | 4      | 5     | 8     | 14     | 19    | 9       |
| S7     | 1       | 4      | 5     | 7     | 13     | 17    | 7       |

Note: Sample is listed as Sample name

Table 6. The number of species types of each grade in the samples of ITS sequencing

| Sample | Kingdom | Phylum | Class | Order | Family | Genus | Species |
|--------|---------|--------|-------|-------|--------|-------|---------|
| S1     | 1       | 3      | 9     | 11    | 14     | 16    | 13      |
| S2     | 1       | 2      | 5     | 7     | 10     | 12    | 10      |
| S3     | 1       | 2      | 7     | 8     | 12     | 15    | 13      |
| S4     | 1       | 3      | 5     | 5     | 8      | 9     | 6       |
| S5     | 1       | 3      | 5     | 6     | 9      | 11    | 7       |
| S6     | 1       | 1      | 3     | 3     | 4      | 5     | 4       |
| S7     | 1       | 1      | 2     | 2     | 4      | 7     | 6       |

Note: Sample is listed as Sample name

From above the level of species type number in the table, it can be seen as the fermentation of sauce fermented grains in the microbial species have increased first, after the reducing trend of anomalies, individual number of microbial species in the sample may be in sauce fermented grains fermentation environment for open state, the environment is also involved in the foreign microbes in the fermentation.

3.3.2. Histogram of species distribution. When there is only one sample or grouping, the species distribution is shown in pie chart, otherwise in bar chart. The figure below is the columnar (pie) diagram of the distribution of species at each level: from left to right are phylum, class, order, family, genus and species. A color represents a species, and the color block length (bar chart) or color block area (pie chart) represents the relative abundance ratio of species. In order to make the best view, only the species with the top ten abundance levels are shown, and the other species are merged into others. According to the figure, Unclassified represents the species with no taxonomic annotation, and the specific species information can be found in the species abundance table of the corresponding classification level.
As can be seen from the above figure, with the progress of soy sauce fermentation, the relative abundance ratio of protein bacteria in 16S sequencing gradually decreased, while that of firmicutes gradually increased. In ITS sequencing, other fungi with relative abundance ratio appeared only in the late fermentation period, and ascomycetes were the only fungi in the rest of the fermentation period.
3.3.3. *Species abundance clustering heat map*. Heatmap is a way of graphical presentation in which the color gradient represents the size of values in the data matrix and clustering is carried out according to the species or sample richness similarity. The species with high abundance and low abundance were clustered in blocks to reflect the similarity and difference of the composition of several sample communities through color gradient and similarity degree. According to the species composition and relative abundance of each sample, the species on each taxonomic level are extracted and mapped using the R language tool. The Heatmap is analyzed on the classification level of phylum, class, order, family, genus and species. In the heat map clustering results, color represents species abundance. Longitudinal clustering indicates that the abundance of different species is similar between samples. The closer the distance between two species is, the shorter the branch length is, indicating that the abundance of these two species is more similar between samples. Horizontal clustering indicates the similarity of species abundance of different samples. Similar to vertical clustering, the closer the distance between two samples is, the shorter the branch length is, indicating that the species abundance of these two samples is more similar.

The clustering heat maps of species abundance at each taxonomic level of the 7 samples are as follows: from left to right be phylum, class, order, family, genus and species; the value corresponding to the heat map is the Z value of the relative abundance of each line of species after standardized treatment. The color gradient from blue to red indicates the relative abundance from low to high. If there is sample grouping information, the first two behaviour sample grouping information in the figure (if there is only one grouping, there is only one row), and the color corresponds to the column of the figure.

Figure 9. 16 s sequencing of the species abundance clustering heat maps
According to the above species accumulation heat map, the species abundance of sample 1 and sample 4, sample 5 and sample 6 were the most similar in 16S sequencing, the species abundance of sample 2 and sample 3, and the species abundance of sample 5, 6 and 7 were the most similar in ITS sequencing. The results showed that the species abundance of bacteria changed little in the middle and late stage of fermentation. The species abundance of fungi changed little at the beginning and end of fermentation. At the same time, it can be seen that the abundance of proteobacteria and actinomycetes is the most similar among the samples, and the abundance of basidiomycetes and mortierella is the most similar among the samples.

3.4. Beta diversity analysis

3.4.1. PCoA analysis. Principal coordinates analysis [9] (Principal coordinates analysis, PCoA) is a kind of similar to PCA dimension reduction sorting method, the principle is the assumption of N samples are measured data differences or the distance between them, you can use this method to find out a right Angle coordinate system, will be represented as N sample points, and the square of the Euclidean distance between the point is exactly equal to the original difference data, qualitative data and quantitative transformation, multi-dimension data extracted from the main element and structure. Multiple samples can be classified by means of principal coordinate analysis, and the diversity of species among samples can be further displayed.

Based on the four distance matrices obtained by the Beta diversity analysis, the PCoA analysis results drawn by R language tools are as follows: the closer the sample is on the coordinate diagram, the greater the similarity will be.
Figure 11. 16 s sequencing of PCoA analysis diagram

Figure 12. ITS sequencing of PCoA analysis diagram

Note: dots represent each sample; Different colors represent different groups; Horizontal and vertical coordinates are the two characteristic values that lead to the largest difference between samples, and the main influence degree is reflected in the form of percentage.
It can be seen from the PCoA analysis diagram above that: in 16S sequencing, sample 5 and sample 6 are the most similar, and sample 4 and sample 7 are the most similar. The results showed that the number of species of bacteria changed little from the initial stage to the middle stage of fermentation. In ITS sequencing, sample 2 was most similar to sample 5, and sample 6 was most similar to sample 7. The results showed that the species of fungi changed little and tended to be stable in the early stage of fermentation.

3.4.2. UPGMA clustering tree drawing combined with histogram. UPGMA clustering tree and histogram plotting is an analytical method to display the clustering tree and the abundance histogram. The left figure is the sample cluster tree (same as UPGMA): based on the four distance matrices obtained from the analysis of Beta diversity, the samples are hierarchical clustering through the Python language tool to judge the similarity of species composition among samples. The right figure is the histogram of species abundance at the genus level of each sample to judge the similarity of species abundance among samples. The sample hierarchical cluster tree is shown as follows: sample cluster tree -- the closer the sample is, the shorter the branch length is, indicating that the species composition of the two samples is more similar; Histogram of abundance -- species diversity, abundance similarity and dominant species of each sample were compared according to the proportion of different color blocks.

Figure 13. 16 s sequencing combination of histogram clustering tree diagram

Figure 14. ITS sequencing combination of histogram clustering tree diagram
As can be seen from the figure, during the whole fermentation process, there were a large number of bacterial species, showing dynamic changes. The fungi were mainly aspergillus with little variation.

4. Alpha diversity analysis
Alpha diversity reflects the species abundance and species diversity of individual samples, with multiple measures including Chao1, Ace, Shannon and Simpson. Chao1 and Ace index measure species abundance which is the number of species. Shannon and Simpson index is used to measure species diversity, which is affected by species abundance and species evenness in sample communities. Under the condition of the same species abundance, the greater the evenness of each species in the community, the greater the diversity of the community, the higher the Shannon index value and the smaller the Simpson index value, indicating the higher the species diversity of the sample [10]. In addition, the OTU Coverage was also counted. The higher the value was, the higher the probability that the species in the sample was detected and the lower the probability that it was not detected. This index reflects whether the sequencing results represent the real situation of microorganisms in the sample.

4.1. Shannon exponential curve
The Shannon diversity index dilution curve was drawn with Mothur software and R language tool according to the Shannon index (reflecting the index of microbial diversity in the sample) at different sequencing depths based on the sequencing amount of each sample, so as to reflect the microbial diversity of each sample at different sequencing amounts. The larger the Shannon index is, the more species of OTU are, and the richer the species are, indicating that the sample has covered the majority of microbial species information. When the curve tends to be flat, it indicates that the sequencing data volume is large enough that the OTU species will not increase with the increase of sequencing volume. If the curve does not flatten, it indicates unsaturation, and more OTU can be found by increasing the amount of data. The results are shown in the figure below: each curve represents a sample and is color-coded.

![Multy samples Shannon Curves](image)

**Figure 15.** samples of 16 s sequencing Shannon Index curve
Figure 16. ITS sequencing samples Shannon Index curve

Note: the abscissa is the number of sequencing sequences randomly selected from a sample, and the ordinate is Shannon index index. As the number of sequenced species increased, the number of discovered species increased until the species became saturated, and increasing the number of samples did not lead to the discovery of new OTU.

As can be seen from the above figure, in the 16S sequencing, the species richness of the sample was from low to high: S4 in ITS sequencing, the species richness of samples was S7 the results showed that the bacterial species in fermented soy sauce grains increased and then decreased with the progress of fermentation. The changing trend of fungal species was similar to that of bacteria, and the microbial species in fermented soy sauce was the most in the later stage of fermentation.

4.2. Grade abundance curve
The hierarchical abundance curve [11] is a curve graph that sorts the OTU abundance of each sample according to size and is drawn based on its relative abundance. It is mainly used to explain the richness and evenness of species contained in the sample at the same time. The richness of species is reflected by the length of the curve on the horizontal axis. The evenness of species composition is reflected by the shape of the curve. The flatter the curve, the higher the evenness of species composition is, as shown in the figure below: each curve corresponds to a sample, marked with different colors.
Figure 17. Samples of 16 s sequencing Shannon Index curve

Figure 18. Shannon Index curve of ITS sequencing samples

Note: the abscissa is the ordinal number sorted by OTUs abundance, and the ordinate is the relative abundance of OTUs.

It can be seen from the above figure that, in 16S sequencing, the composition richness of species in each sample from low to high is S2. The results showed that the species composition of bacteria first increased and then decreased with the progress of fermentation, and was most abundant in the middle stage of fermentation. The composition of fungal species is most abundant at the end of fermentation.
5. Conclusion
According to the sample Shannon Index curve of 16S sequencing, sample 1 has the most species of OTU, that is, the most abundant species, while sample 4 has the least species of OTU and the least species. In the sample Shannon Index curve of ITS sequencing, sample 2 has the most species of OTU and the most abundant species, while sample 7 has the least species of OTU and the least species.

According to the histogram of species distribution, 16S sequencing of 7 samples showed that protein bacteria and firmicutes accounted for the largest proportion of relative abundance, and ITS sequencing showed that the species with the highest proportion of species abundance in all samples was ascomycetes.

According to the analysis of the cluster tree histogram of the samples, it can be seen that all the samples basically contain the bacterial species with the species richness ranking top 10, among which the dominant bacteria in sample 1 is bacillus, followed by candida. The dominant bacteria in sample 2 was weisiella followed by staphylococcus. The dominant bacteria of sample 3 and sample 4 were staphylococcus, followed by klebsiella. The dominant bacteria of sample 5 was klebsiella, followed by lactococcus. The dominant bacteria in sample 6 was weisiella, followed by klebsiella. The dominant bacterium of sample 7 was klebsiella SPP. Similarly, aspergillus was the dominant fungus among the fungi with the highest species abundance in the sample, followed by anomalous wickham yeast in the sample 7, picha montorchia in the sample 2, and plasmospora only existed in sample 1 and sample 7.

Acknowledgements
The authors gratefully acknowledge the financial support from: (1) Applied Basic Research Project of Sichuan Science and Technology Department (2018JY0396), Study on microbial community of fermented potato products -- a case study of Tartary buckwheat wine; (2) Key Scientific Research Project of Sichuan Education Department (17ZA351), Research on the key technology of potato liquor brewing by solid state method; (3) "Two High" Talent Research Support Project of Xichang University (LGZ201917), microbial community in potato liquor fermentation based on high-throughput sequencing.

References
[1] Li yan. Introduction to fermentation industry [M]. Beijing: China light industry press, 2000, 0.
[2] Jiang yongxin. Food production technology [M]. Beijing: China light industry press, 2001.
[3] Antony-Babu S, Murat C, Deveau A, et al. An improved method compatible with metagenomic analyses to extract genomic DNA from soils in Tuber melanosporum orchards. Journal of Applied Microbiology, 2013, 115 (1): 163 - 170.
[4] Rossard A, Ramond J B, Seely M, et al. Water regime history drives responses of soil Namib Desert microbial communities to wetting events. Scientific Reports, 2015 (5): 12263.
[5] Li J, Hu H W, Ma Y B, et al. Long-term nickel exposure altered the bacterial community composition but not diversity in two contrasting agricultural soils. Environmental Science and Pollution Research, 2015, 22(14): 10496 – 10505.
[6] Kemp P F, Aller J Y. Bacterial diversity in aquatic and other environments: what 16S rDNA libraries can tell us. Fems Microbiology Ecology, 2004, 47 (2): 161 - 177.
[7] Caporaso JG, Kuczynski J, et al. (2010). QIIME allows analysis of high-throughput community sequencing data. Nature methods, 7 (5): 335 - 336.
[8] Edgar R C. Search and clustering orders of magnitude faster than BLAST [J]. Bioinformatics, 2010, 26 (19): 2460 - 2461.
[9] Sakaki T, Takeshima T, Tominaga M, Hashimoto H, Kawaguchi S: Recurrence ofICA - PCo.
[10] Aaneurysms after neck clipping. Journal of neurosurgery 1994, 80 (1): 58 - 63.Grice EA, Kong HH, et al. (2009). Topographical and temporal diversity of the human skin microbiome. Science, 324 (5931): 1190 – 1192.
[11] Kõljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AF, Brähm M, Bates ST, Bruns TD, Bengtsson-Palme J, Callaghan TM: Towards a unified paradigm for sequence-based identification of fungi. Molecular Ecology 2013, 22 (21): 5271 - 5277.