



### **Full Length Article**

## **Analysis of the Microbiota in Naturally Fermented Cabbage of Northeastern Part of China by High-throughput Sequencing of the V3-V4 Regions of the 16S rRNA Gene**

**Xinwei Li and Xiqing Yue\***

*Food Science College, ShenYang Agriculture University, ShenYang City, 110161, LiaoNing Province, China*

\*For correspondence: yxqsyau@126.com

### **Abstract**

In order to explore the microbiota in naturally fermented cabbage of northeastern part of China, the sequencing of the V3-V4 regions of the 16S rRNA gene of the microbiota in naturally fermented cabbage of northeastern part of China was made in this experiment by using the 454 FLX+ platform. A total of 302,327 valid sequences of 16 samples were obtained by 454 pyrosequencing. Evaluation of the richness of microflora of the samples by calculating Chao1 index showed that the microbiota of the samples from Daqing presented a higher richness. Beta diversity analysis showed species overlap with a very low content in the samples from Heilongjiang, Jilin and Liaoning. Heat-map showed significant differences among the dominant bacteria of different provinces. The results of sample analysis based on microbiota showed that the Lactobacilli content of the samples from Daqing, Heihe and Songyuan was more than 50%, and that from Heihe was up to 69.6%. In addition, *Wohlfahrtiimonas* was found for the first time in fermented food in this experiment, revealing the security risks of traditional methods. © 2016 Friends Science Publishers

**Keywords:** Fermented cabbage; Beta diversity analysis; Microbiota; Pyrosequencing; 16S rRNA

### **Introduction**

Fermented cabbage, as a popular pickled vegetable in northeastern part of China, has a long history. Owing to unique geographical locations, climate conditions and national cultural traditions, naturally fermented cabbage is characterized by rich sour and full body and aroma. It is superior to the fermented cabbage of other regions in taste and aroma. Therefore, research on the microbial resources in traditional naturally fermented vegetable of northeastern part of China has been a hot spot in academia in recent years.

The method of enrichment axenic isolation cultivation and identification, which only allows an access to rare information of microflora but fails to truly reflect all the information of microflora, which is generally adopted in the research on traditional fermented food (Iacumin *et al.*, 2009; Qu and Wei, 2010). With the development of the molecular biology technologies in recent years, studies have shown that the combination of 16S rDNA with the technique of denaturing gradient gel electrophoresis (DGGE) can be used to identify the dominant microflora in fermented cabbage (Wu *et al.*, 2014a). However, this method is vulnerable to the effects of the number of gene copies, which the low copy number of the samples of bacteria lacks sensitivity. Therefore, it requires complex operation, long cycles and high costs. High-throughput sequencing technology requires

no cloning construction and presents high throughput, in which the 454 sequencing technique presents the longest sequence reads, opening the door to the comprehensive and efficient understanding of the information of microflora in fermented food (Navarro-Noya *et al.*, 2013; Poulsen *et al.*, 2013; Sul *et al.*, 2013; Yu *et al.*, 2013).

For the above reasons, the objective of this study was to make the sequencing by using 454 FLX+ platform in this experiment in order to analyze the microbiota in naturally fermented cabbage. Later the sequences were grouped into many taxonomic units with their similarities, each of which is an operational taxonomic unit (OTU) (Blaxter *et al.*, 2005) that is artificially given taxonomic units in phylogenetics or population genetics. Clustering and follow-up biological analysis of OTU for sequences are usually made under the similar level of 97% to reveal the dominant microflora in naturally fermented cabbage of northeastern part of China and microflora changes in fermented cabbage of the provinces of different latitudes, laying the foundation for the exploration of the evolution of microflora diversity.

### **Materials and Methods**

#### **Sample Collection**

Sample collection began in February 2015. A total of 48 samples of naturally fermented cabbage (each ternary

sample from 16 places, including Heihe (HH), Hulun Buir (HLBE), Qiqihar (QQHE), Daqing (DQ), Jiamusi (JMS), Baicheng (BC), Songyuan (SY), Mudanjiang (MDJ), Changchun (CC), Gongzhuling (GZHL), Yanbian (YB), Fuxin (FX), Anshan (AS), Chaoyang (CY), Dandong (DD) and Dalian (DL) ) were collected in the three northeastern provinces from north to south according to their geographical positions and climate features. Due to varied pickling crafts and techniques, detailed processes were regulated to reduce accidental errors.

### Extraction, Purification and Amplification of DNA

The steps are as follows: adding 5 mL PBS buffer to the samples, collecting the supernatant after 30s in a vortex mixer and 5 min in a 350 g centrifugal clarifier and rejecting the supernatant after 5 min in a 12000 g centrifugal clarifier, adding 800  $\mu$ L TE buffer solution to the precipitate to re-dissolve it, then making the total DNA extraction with FastPrep and CTAB method. Taking the extracted total DNA as a template, PCR of 16S rDNA gene was made by using the primer mix 520F: (5-CCTACGGGNGGCWGCAG-3) and 802R: (5-GACTACHVGGGTATCTAATCC-3) having specificity to V3-V4 regions of 16S rDNA gene to make PCR amplification; the amplification reaction system (50  $\mu$ L) include: 50 ng genomic DNA template and 2  $\mu$ L dNTP (2.5 mmol/L, TaKaRa); the concentration of both primers were 0.4 mol/L; 0.5  $\mu$ L DNA Taq polymerase (1.5U, TaKaRa); 5  $\mu$ L 10 $\times$  PCR Buffer. Reaction parameter: 5 min of initial denaturation at 94 $^{\circ}$ C, 35 cycles, including 30s of denaturation at 94 $^{\circ}$ C, 45s of primer annealing at 55 $^{\circ}$ C, 40s of elongation at 72 $^{\circ}$ C and finally 10min elongation at 72 $^{\circ}$ C. The PCR reaction products were determined by 10 g/dL agarose gel electrophoresis. The detail of ingredients used is: Genomic DNA template 50 ng; dNTP (2.5 mmol/L, TaKaRa) 2  $\mu$ L; Both primers (0.4 mol/L) 2  $\mu$ L; DNA Taq polymerase (1.5U, TaKaRa) 0.5  $\mu$ L and 10  $\times$  PCR buffer 5  $\mu$ L.

### Bioinformatics Sequence Analysis

Sequencing was made by 454 FLX+ platform. The bioinformatics analysis of resulting sequences, including quality control, removal of forward and reverse primers and barcode, classification of operational taxonomic units (OTU), diversity comparison and statistical analysis (Mark *et al.*, 2013), was made by QIIME (v1.4.0). A 97% similarity classification was made for the selected representative sequences and classifying OTU by sequence clustering. The richness of microflora of the samples was evaluated through calculating Chao1 index and the microbial diversity of samples was evaluated by Shannon index and Simpson index. Microbial diversity and sequence diversity on different OTU classification level were evaluated by using Shannon Index Curve and Rarefaction Curve to assess whether the current sequencing volume

represents the diversity of primary community and whether the sample size is reasonable. Weighted and un-weighted principal coordinate analysis (PCoA) of samples were made based on the UniFrac distance and the samples were clustered by un-weighted pair-group method with arithmetic means (UPGMA).

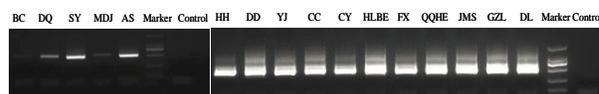
## Results

### Sample DNA Extraction Results

Sample DNA extraction results are shown in Fig. 1. The metagenomic DNA ladders of the extracted samples are clear with high brightness and without degradation.

### Richness and Diversity Analysis

A total of 302,327 valid sequences of 16 samples were obtained after 454 pyrosequencing and 158,015 high quality sequences after filtering and removing chimera, in which the most abundant sequences were obtained from the sample DQ, and the least were from the sample CY, as shown in Table 1. Statistics were done for the length of quality sequences of all samples. The results showed that the sequence length concentrates most in 430-450 bp. OTU classification was made for the obtained quality sequences with the similarity of 97%, and the numbers of OTU that were classified in phylum, class, order, family, genus and species were separately grouped.



**Fig. 1:** Agarose gel electrophoresis of metagenomic DNA

**Table 1:** The number of effective sequences and high quality sequences

Sample	Valid sequences	High quality sequences
AS	13103	9248
BC	21486	12390
CC	10912	6530
CY	9158	6157
DD	10560	5521
DL	11155	5354
DQ	49706	31708
FX	35677	11698
GZL	30737	11964
HH	18236	10100
HLBE	9298	6487
JMS	10214	5742
MDJ	9239	7602
QQHE	25716	7576
SY	11456	8729
YJ	25674	11209

### Analysis of Alpha Diversity

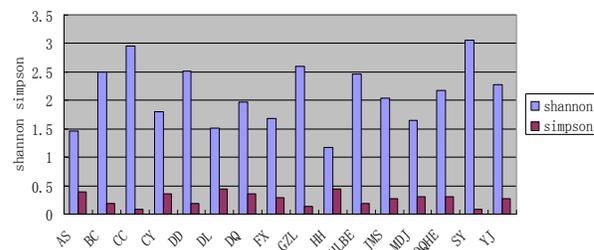
Alpha diversity refers to the diversity in a specific region or an ecosystem. Diversity index is a comprehensive indicator of richness and evenness and usually associated with the following two factors: 1, the number of species, that is, richness; 2, the evenness of the individual distribution of species. Community diversity index includes Shannon index and Simpson index. Shannon index is generally used to reflect alpha diversity index, and the greater the Shannon value, indicating that every individual belongs to different genus, the higher the species diversity is. Sample SY showed the highest Shannon value, as shown in Fig. 2, indicating that SY community diversity was the highest in sample SY. Meanwhile, the greater the Simpson index, the lower the community diversity is. Sample SY presented the smallest Simpson index, indicating that it has the highest community diversity, which is consistent with the results of Shannon index.

Chao index and ACE index represent the community richness. In Chao index, the index of OTU number in samples are estimated by Chao1 algorithm, which was first proposed by Chao and is usually used to estimate species review in ecology (Pitta *et al.*, 2010) The greater the Chao index or ACE index, the higher is the community richness. Fig. 3 shows that Chao index and ACE index had the same trends; the highest Chao index and ACE index of sample DQ was 234.2 and 230, respectively, while the lowest Chao index and ACE index of sample HH were 58.9 and 56.2 respectively. It follows that the community of the sample from Daqing had a higher richness, while that of the samples from Heihe has a lower richness. Sample SY, which is of the highest Shannon value, has higher community richness in addition to the higher community diversity, and its Chao index and ACE index is 190.8 and 184.1.

### Analysis of Beta Diversity

Beta diversity refers to the comparison of diversity among different ecological systems. It is to calculate the distance between samples using the evolutionary and richness information of sample sequences to reflect whether there are significant microflora differences among samples.

Among them, Unifrac (<http://bmf2.Colorado.edu/unifrac/>) (Lozupone and Knight, 2012) analysis was to calculate the difference among sample sequences using the differences in evolutionary information of sample sequences to reflect whether there are significant microflora differences among samples in an evolutionary tree. Unifrac analyses include un-weighted Unifrac indifference analysis (calculating the differences in species evolution of samples) and weighted Unifrac indifference analysis (calculating the differences in species evolution and richness of samples). PCoA (Principal Coordinates Analysis) is a visualization method to study data similarities or differences, which can be used to identify differences



**Fig. 2:** Comparative analysis of community diversity index in samples

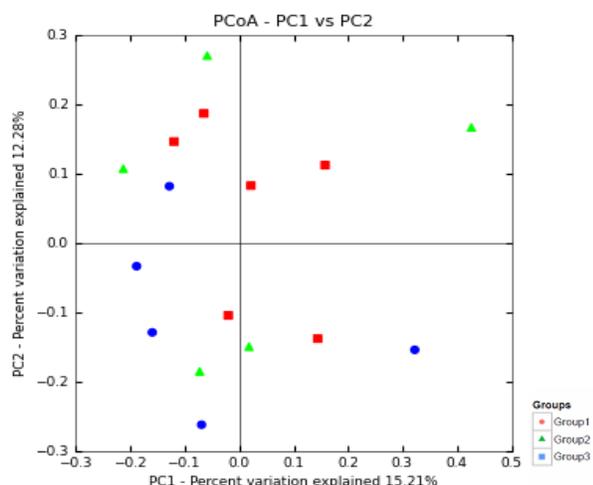


**Fig. 3:** Comparative analysis of community richness in samples

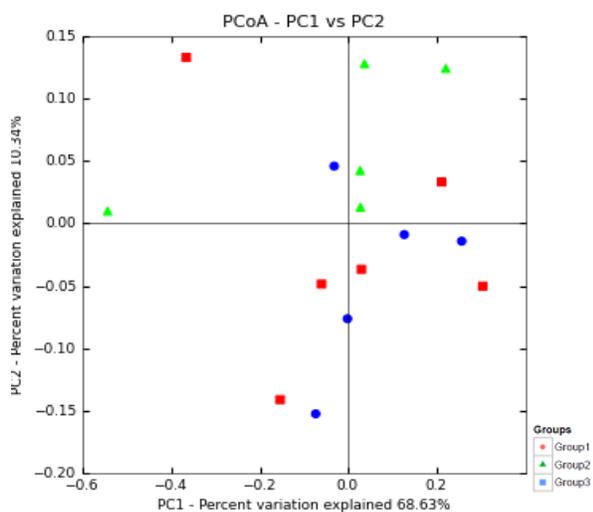
between individuals or groups. NMDS (Nonmetric Multidimensional Scaling Analysis) is a data analysis method, which can be used to position, analyze and classify the samples of multidimensional space after simplify them to low-dimensional space while retaining their original relationships. It is always used to compare the differences between samples.

By using the evolutionary information and species richness information, Unifrac analysis was made in this study to obtain the distance matrix of sample differences. PCoA analyses of distance matrix information of the samples of Group1 Heilongjiang, Group2 Jilin and Group3 Liaoning were made. Fig. 4 and Fig. 5 show that weighted and un-weighted UniFrac principal coordinate analysis based on UniFrac present the contribution rate of the first and the second principal component was 15.21, 12.2, 68.63 and 10.34. The weighted principal coordinate analysis diagram based on UniFrac shows that the distribution of the samples from Heilongjiang is absolutely isolated from the distribution of those from Jilin and Liaoning without any overlap. However, the unweighted principal coordinate analysis diagram based on UniFrac shows that the distribution of the samples from Heilongjiang presents a trend of overlap with the distribution of those from Jilin and Liaoning. This indicates that some OTUs exist in the samples from Heilongjiang, Jilin and Liaoning with a small and varied content.

The data information in two dimensional matrices or tables can be reflected in a heat-map with color



**Fig. 4:** Unweighted UniFrac analysis of Group 1 (Heilongjiang), Group 2 (Jilin) and Group 3 (Liaoning) samples



**Fig. 5:** Weighted UniFrac analysis of Group 1 (Heilongjiang), Group 2 (Jilin) and Group 3 (Liaoning) samples Cluster Analysis

changes. A heat-map can be used to intuitively reflect data values with defined light and dark colors. Richness similarity clustering of species or sample was made for the data on request, and the clustered data were expressed in the heat-map; the species of high and low richness can be clustered in blocks to reflect the similarity and difference of community composition of several samples at the classification level by color gradients and the degree of similarity.

Each of the color blocks in the heat-map represents the genus richness of a sample, and they longitudinally represent the clustering of samples, reflecting the similarity of community composition of several samples in genus. According to Alpha diversity, Shannon index and Simpson index show that the samples from Songyuan

have the highest diversity index, and in the heat-map, as shown in Fig. 6, *Lactococcus*, *Acinetobacter*, *Dysgonomonas* and *Bacteroides* are the bacteria genera of which the richness is 4 in the sample SY. In addition, the dominant bacteria genera are different in each sample. No overlap of dominant bacteria genera was found in the samples divided by provinces.

### Analysis Based on Community Structure

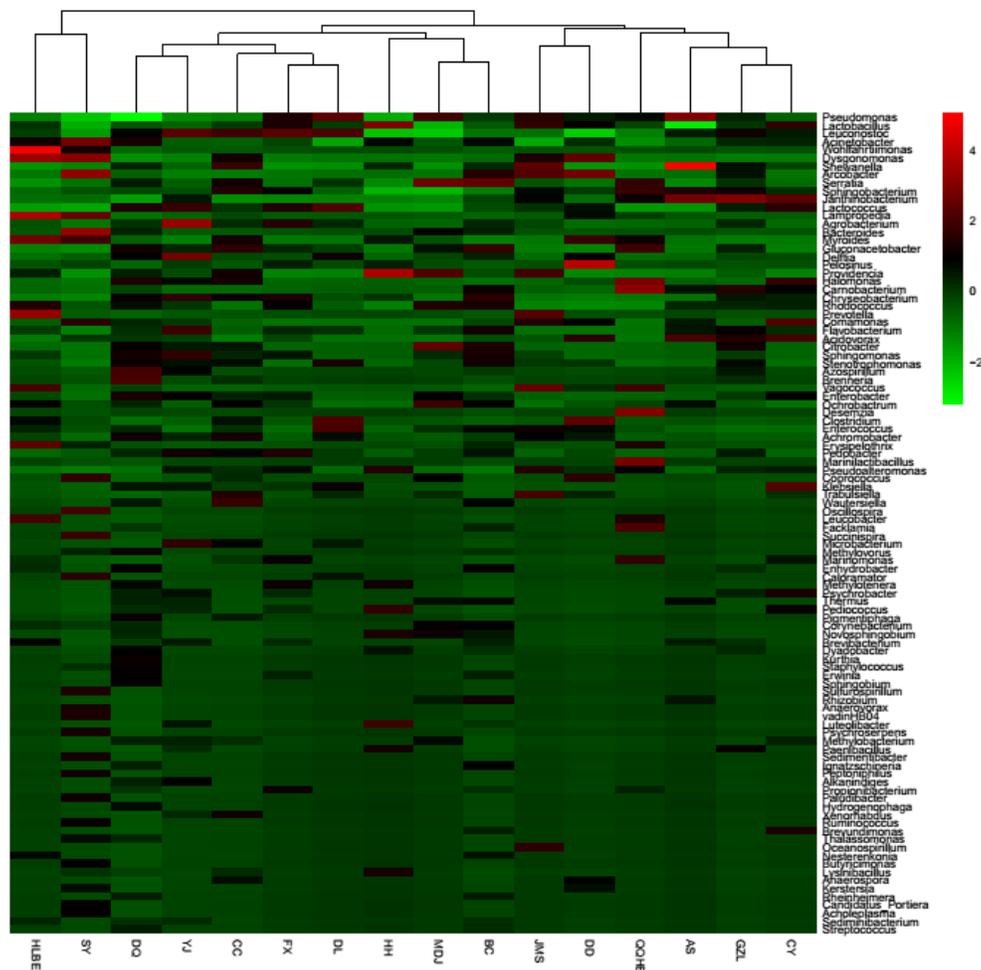
#### Classification and comparison of bacteria in phylum:

The community structures of the samples on different classification levels were observed by a statistical analysis method in this study. They are usually presented by intuitive pie charts or bar charts (Oberauer *et al.*, 2013). By combining RDP and BLAST homology sequence alignment clusterings, all the sequences of bacteria were identified as 11 strains: Actinobacteria, Bacteroidetes, Cyanobacteria, Firmicutes, Fusobacteria, Lentisphaerae, Nitrospirae, Planctomycetes, Proteobacteria, Tenericutes and Verrucomicrobia. The content of Firmicutes of the samples from Heihe, Heilongjiang (HH), is up to 95.3%, coming first among all samples; that of the samples from Anshan, Liaoning (AS) is the least (4.1%). The content of Proteobacteria of the samples from Mudanjiang, Heilongjiang (MDJ), is up to 97.5%, coming first among all samples, while that of the samples from Heihe (HH) is the least (4.5%). Bacteroidetes was only found in the samples from Songyuan (SY) and Hulunbeier (HLBE) with the content of 28.6 and 21, respectively and its content was 0.1% (HH) 7.9% (CC) in the remaining samples, while Actinobacteria was only found in the samples in Hulun Buir (HLBE) with a content of 4.5%, and that in the remaining samples is 0.1% (AS) to 0.9% (BC). So Firmicutes, Proteobacteria Bacteroidetes and Actinobacteria are assumed to be dominant bacteria phyla in all samples.

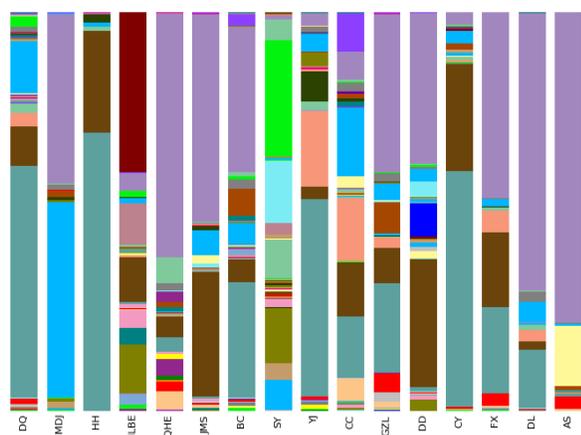
**Classification and comparison of bacteria in genus:** Fig. 7 shows that, in 16 samples from all provinces, the bacteria belonging to Firmicutes and of which the relative content is greater than 1% include 7 genera: *Pediococcus*, *Lactobacillus*, *Leuconostoc*, *Lactococcus*, *Carnobacterium*, *Clostridium* and *Erysipelothrix*. The bacteria belonging to Bacteroidetes include 5 genera: *Bacteroides*, *Dysgonomonas*, *Prevotella*, *Myroides* and *Sphingobacterium*. The bacteria belonging to Proteobacteria include 13 genera: *Agrobacterium*, *Comamonas*, *Lampropedia*, *Janthinobacterium*, *Arcobacter*, *Shewanella*, *Gluconacetobacter*, *Providencia*, *Serratia*, *Halomonas*, *Acinetobacter*, *Pseudomonas* and *Wohlfahrtiimonas*.

### Discussion

A comprehensive analysis of nine variable regions of 16S rRNA was made by (Claesson *et al.*, 2010). The comparison can indicate the proportion and accuracy in genus, and the results showed that: region V1 and V9 are the worst, followed by region V7 and V8, and then region V3, V4 and



**Fig. 6:** Heat-map of samples clustering



**Fig. 7:** Relative abundances of the bacterial in genera of the different samples

V5. Tandem primers were designed in these regions, and V1/V2, and the comparison of 6 tandem regions (V2/V3,V3/V4,V5/V6 and V7/V8) showed that V3/V4 and V4/V5 had the highest accuracy. Taking into account the

classification efficiencies (CE), region V3/V4 showed a greater significant amplification preference than other regions. The comparison of the annotation accuracy and reads of the sequencing results of variable regions in species can annotate the difference in proportions in genus, and the results showed that 454 platform was superior to the paired-end sequencing of Illumina in read length and sequencing quality of 16S rRNA (Xia *et al.*, 2014). High-throughput analysis was thereby made for the metagenome of the samples by using 454 FLX+ platform.

During natural fermentation of cabbage, the bacteria genera, including *Lactobacillus*, *Leuconostoc* and *Lactococcus* seemed to play important roles in glycolytic pathway and the process of heterolactic fermentation (Abdel-Rahman *et al.*, 2011). Through the analysis of microflora changes in fermented cabbage by PCR-DGGE technology, Wu Rina (Wu *et al.*, 2014b) found that *Lactobacillus* is the main bacterium in the fermentation broth of naturally fermented cabbage, which is consistent with the results of this study. The content of *Lactobacillus* in the samples from Daqing, Heihe and Songyuan was more than 50%; while that in the samples from Heihe was up to 69.6%. *Leuconostoc* is

the second most dominant bacterium after *Lactobacillus*. In addition, pentosaceus, *Lactococcus lactis*, *Lactobacillus hammesii* and *L. odoratitofui* were also identified, which is consistent with the research results of Wu Junrui (Wu *et al.*, 2014c). This implied that the *Lactobacillus* was in a larger proportion in naturally fermented cabbage in the provinces of higher latitudes. *Wohlfahrtiimonas* is a clinically rare facultative pathogenic bacterium that was discovered in 2011 and less reported in China (Chen *et al.*, 2013), which was discovered for the first time in the study of microorganisms in fermented food. It has been recognized as a pathogenic bacterium causing large-scale diarrhea since it has been discovered (Swain *et al.*, 2015). Out of 48 samples in this experiment, *wohlfahrtiimonas* of the samples from Hulun Buir and Songyuan presented the highest content (40.5% and 0.2% respectively), indicating that the pollution has spread to inland cities.

According to the results of Alpha diversity analysis, Shannon index and Simpson index hint that the samples from Songyuan had the highest diversity index. The results of heat-map showed that *Lactococcus*, *Acinetobacter*, *Dysgonomonas* and *Bacteroides* have the highest richness. As for *Lactobacillus*, *Leuconostoc* and *Lactococcus* that played important roles in fermentation, *Lactobacillus* in the samples HH from Heihe showed the greatest richness, while *Leuconostoc* in the samples from YJ, CC, FX and DL showed the highest richness. During the process of fermentation, the leading fermenting bodies are microorganisms, but their population varies with fermentation conditions. This is the reason for the presence of some OTUs with low and varied contents in the samples from Heilongjiang, Jilin and Liaoning according to results of Beta diversity analysis. Furthermore, it may be the differences in the natural environment of the location of samples, including soil quality and fermentation conditions (including fermenting bodies).

## Conclusion

The samples from Songyuan had the highest diversity index and *Lactococcus*, *Acinetobacter*, *Dysgonomonas* and *Bacteroides* was the dominant. In this study, most sequences concentrate in *Lactobacillus*, which is consistent with the results of previous studies. In addition, this study distinguished microflora and dominant strains in the three eastern provinces and molecularly explained the fundamental reasons for different flavors of traditional naturally fermented vegetable in different regions, providing the rationale for further studies on screening quality strains.

## Acknowledgement

Project found: National Natural Science Foundation of China: Diversity and Spatial and Temporal Distribution of Microbiota in Naturally Fermented Cabbage of Northeastern Part of China (31370502)

## References

- Abdel-Rahman, M.A., Y. Tashiro and K. Sonomoto, 2011. Lactic acid Production from lignocellulose-derived sugars using lactic acid bacteria: Overview and limits. *J. Biotechnol.*, 154: 286–301
- Blaxter, M., J. Mann, T. Chapman, F. Thomas, C. Whitton, R. Floyd and E. Abebe, 2005. Defining operational taxonomic units using DNA barcode data. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 360: 1935–1943
- Chen, T., X. Cao and J. Qi, 2013. Isolation, identification and molecular phylogeny of a new Entomophilous pathogen. *Chin. J. Chromatogr.*, 29: 1389–1391
- Claesson, M.J., Q. Wang and O. Sullivan, 2010. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucl. Acids Res.*, 38: 200–208
- Iacumin, L., F. Cecchini and M. Manzano, 2009. Description of the microflora of sourdoughs by culture-dependent and culture-independent methods. *Int. J. Food Microbiol.*, 8: 126–129
- Lozupone, C. and R. Knight, 2012. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.*, 71: 8228–8235
- Navarro-Noya, Y.E., S. Gómez-Acata, N. Montoya-Ciriaco, A. Rojas-Valdez, M.C. Suarez-Arriaga, C. Valenzuela-Encinas, N. Jimenez-Bueno, N. Verhulst, B. Govaerts and L. Dendooven, 2013. Relative impacts of tillage, residue management and crop rotation on soil bacterial communities in a semiarid agroecosystem. *Soil Biol. Biochem.*, 65: 86–95
- Oberauner, L., C. Zachow, S. Lackner, C. Hogenauer, K.H. Smolle and G. Berg, 2013. The ignored diversity: complex bacteria communities in intensive care units revealed by 16S pyrosequencing. *Sci. Rep.*, 3: 1413–1421
- Pitta, D.W., E. Pinchak, S.E. Dowd, J. Osterstock, V. Gontcharova, E. Youn, K. Dorton, I. Yoon, B.R. Min and J.D. Fulford, 2010. Rumen bacterial diversity dynamics associated with changing from bermudagrass hay to grazed winter wheat diets. *Microbial. Ecol.*, 59: 511–522
- Poulsen, P.H.B., W.A. Al-Soud, L. Bergmark, J. Magid and L.H. Hansen, 2013. Effects of fertilization with urban and agricultural organic wastes in a field trial – prokaryotic diversity investigated by pyrosequencing. *Soil Biol. Biochem.*, 57: 784–793
- Qu, Y. and L. Wei, 2010. *Principle and Application of Microbial Culture-Independent Technology*, Vol. 1, pp: 117–133. Science Press, Beijing, China
- Sul, W.J., S. Asuming-Brempong, Q. Wang, D.M. Tourlousse, R.C. Penton, Y. Deng, J.L.M. Rodrigues, S.G.K. Adiku, J.W. Jones, J. Zhou, J.R. Cole and J.M. Tiedje, 2013. Tropical agricultural land management influences on soil microbial communities through its effect on soil organic carbon. *Soil Biol. Biochem.*, 65: 33–38
- Swain, M.R., M. Anandharaj, R.C. Ray and R.P. Rani, 2015. Fermented fruits and vegetables of Asia: A potential source of probiotics. *Biotechnol. Res. Int.*, Volume 2014 (2014), Article ID 250424
- Wu, J., X. Yue and P. Shi, 2014a. PCR-DGGE Analysis of the Microbial Diversity in Naturally Fermented Suan-cai from Northeast China. *Food Sci. Biotechnol.*, 33: 127–130
- Wu, J., X. Yue, P. Shi and R. Wu, 2014b. PCR-DGGE Analysis of the Microbial Diversity in Naturally Fermented Suan-cai from Northeast China. *J. Food Sci. Biotechnol.*, 2: 127–130
- Wu, R., M. Yu and L. Meng, 2014c. Diversity of lactic acid bacteria involved in Suan-Cai using PCR-DGGE. *Mod. Food Sci. Technol.*, 30: 8–35
- Xia, Z., 2014. *Comparative Analysis and Study on 16S rRNA Sequencing Data Based on 454 and Ion Torrent Platform*, Vol. 1, pp: 157–163. Department of Biophysics, University of Electronic Science and Technology of China, Beijing, China
- Yu, L., M. Nicolaisen, J. Larsen and S. Ravnskov, 2013. Organic fertilization alters the community composition of root associated fungi in *Pisum sativum*. *Soil Biol. Biochem.*, 58: 36–41

(Received 04 January 2016; Accepted 12 August 2016)