Example of Metagenomic Analysis Using Metagenome@KIN

Analysis of changes in bacterial flora associated with fermented food, Kimchi

White paper

The development of next generation sequencing technology and the improvement of genetic analysis technology in recent years, have enabled thorough detection and determination of bacteria that were previously hard to observe due to the difficulties of culturing them.

Amplicon sequence analysis of 16S rRNA is the core technology for metagenomic analysis from next generation sequencing data. This method requires software to identify bacterial flora, annotate hierarchical information, and analyze bacterial frequency from the NGS data. Metagenome@KIN is a software application that was developed to enable researchers to perform such analyses without bioinformatics training. This paper explains a study using Metagenome@Kin, and details the process of correctly performing the bacterial flora analysis.

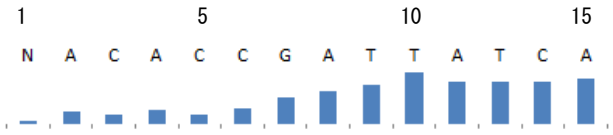
Bacterial flora analysis by Metagenome@KIN

Step 1: Trimming based on read quality

In recent years, next generation sequencers export read data in a Fastq file format.

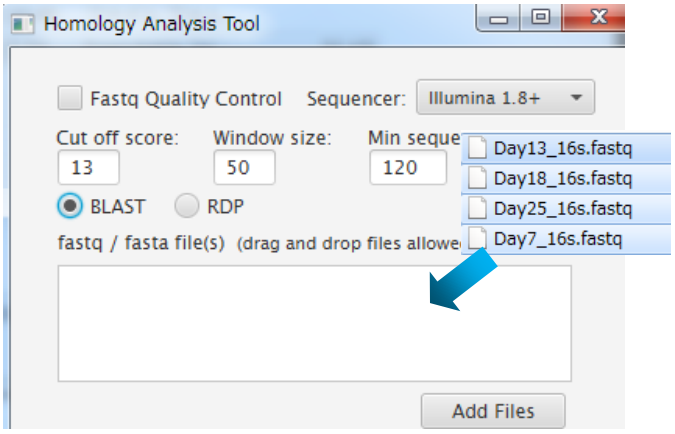
Typically, such exported read data from a sequencer tends to have low quality of base calls in the head part.

* Correlation of base length and base call quality (bar graph)



Differences of 16s rRNA sequences between species are sometimes below 10% (ref1). Thus, using short read sequences with about 100bp, the uncertain base calls would cause wrong detection of bacterial flora. To avoid this problem, it is important to remove low quality bases which cannot guarantee the correct base calls from the read prior to assigning bacterial species. Before starting analyses with Metagenome@KIN, we recommend using a free tool provided by World Fusion called, “homology_analysis_tool”, which can remove the low quality bases. This trimming step is highly recommended before the identification of bacterial species by BLAST or RDP Classifier.

* Trimming reads by homology_analysis_tool. Just drag-and-drop Fastq files.

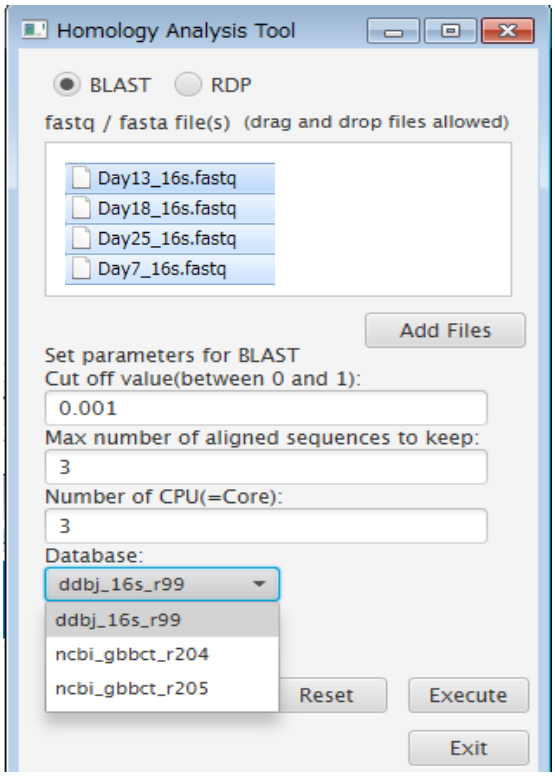


Step 2: Bacterial flora determination by BLAST

BLASTN detects which bacterial species is derived from sequence reads.

Since identification of bacterial species can be affected by the database being used, and the same database needs to be assigned for a continuous project, selecting the most appropriate database is important. The homology search tool as an accessory to Metagenome@KIN provides not only 16S rRNA databases from NCBI gbbct but also databases from DDBJ or GreenGene.

* Running Blast on homology_analysis_tool. Select database and click "Execute".



BLAST will take time to process. For example, if there are 815 16S sequence reads to BLAST, DDBJ_16s DB and ncbi_gbbct will take 10 minutes, and GreenGene takes 35 minutes. Databases for BLAST contain 16S rRNA sequences with species hierarchy and identify bacterial species against sequence reads.

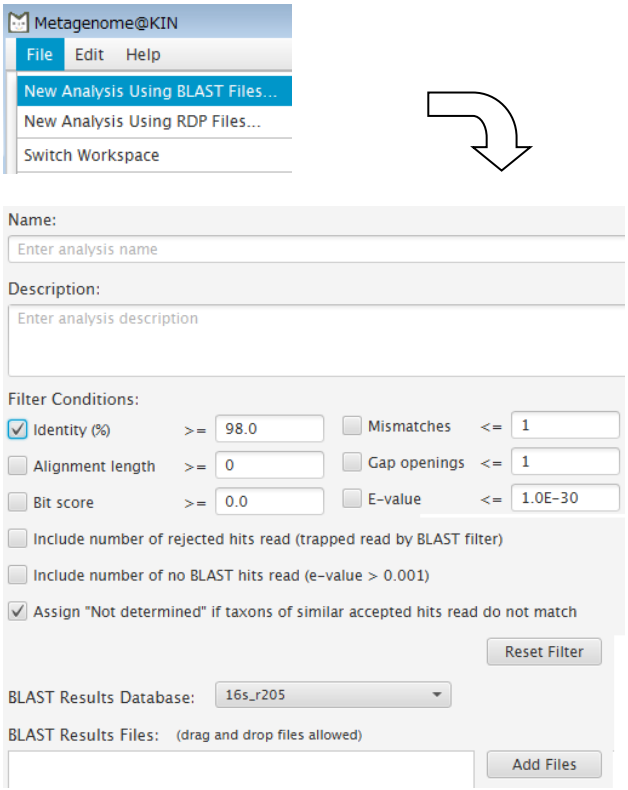
To perform fast bacterial identification, using a combination of RDP Classifier, the homology_analysis_tool, and Metagenome@KIN can be considered. When using 16S rRNA for RDP Classifier, the bacterial identification contains between domain and genus ranks, whereas using fungal ITS1/2 regional sequences, the fungal flora identification can be detected up to species rank.

Step 3: Bacterial flora visualization and clustering analysis by KIN

After the BLAST by homology_analysis_tool finishes, a folder will be created (numerically named for the date & time of execution), containing the bln extension file(s).

Drag & drop the bln file(s) into Metagenome@KIN, and setup the analysis.

From the File menu in KIN, select "New Analysis Using BLAST Files…", set filter conditions, and select the same database used for the homology_analysis_tool to execute the analysis.



After the analysis is done, pie charts, spread sheets, and sunburst graphs for each sample will be created. Following pie charts show studies of changes in bacterial flora (SRP010991) using traditional fermented food, kimchi along by fermentation days, Day1 ~ Day25. Fig1 shows the pie charts as analyzed results from 16S amplicon sequence samples at day7 using multiple 16S databases in KIN. According to the results, *mesenteroides* has the highest occurrence ratio in the samples, having the same result from each database. However, from the second database, different bacterial species appear. Even though this shows that registered bacterial names at the species rank are different between 16S reference databases, this flora difference becomes smaller at the genus rank between NCBI_gbbct and DDBJ, while GreenGene still shows the different flora. Thus, types of bacteria available in each of the databases varies, and selection should be made accordingly, as it will influence the analysis results.

Fig1. The difference of bacterial flora by different 16S rRNA database

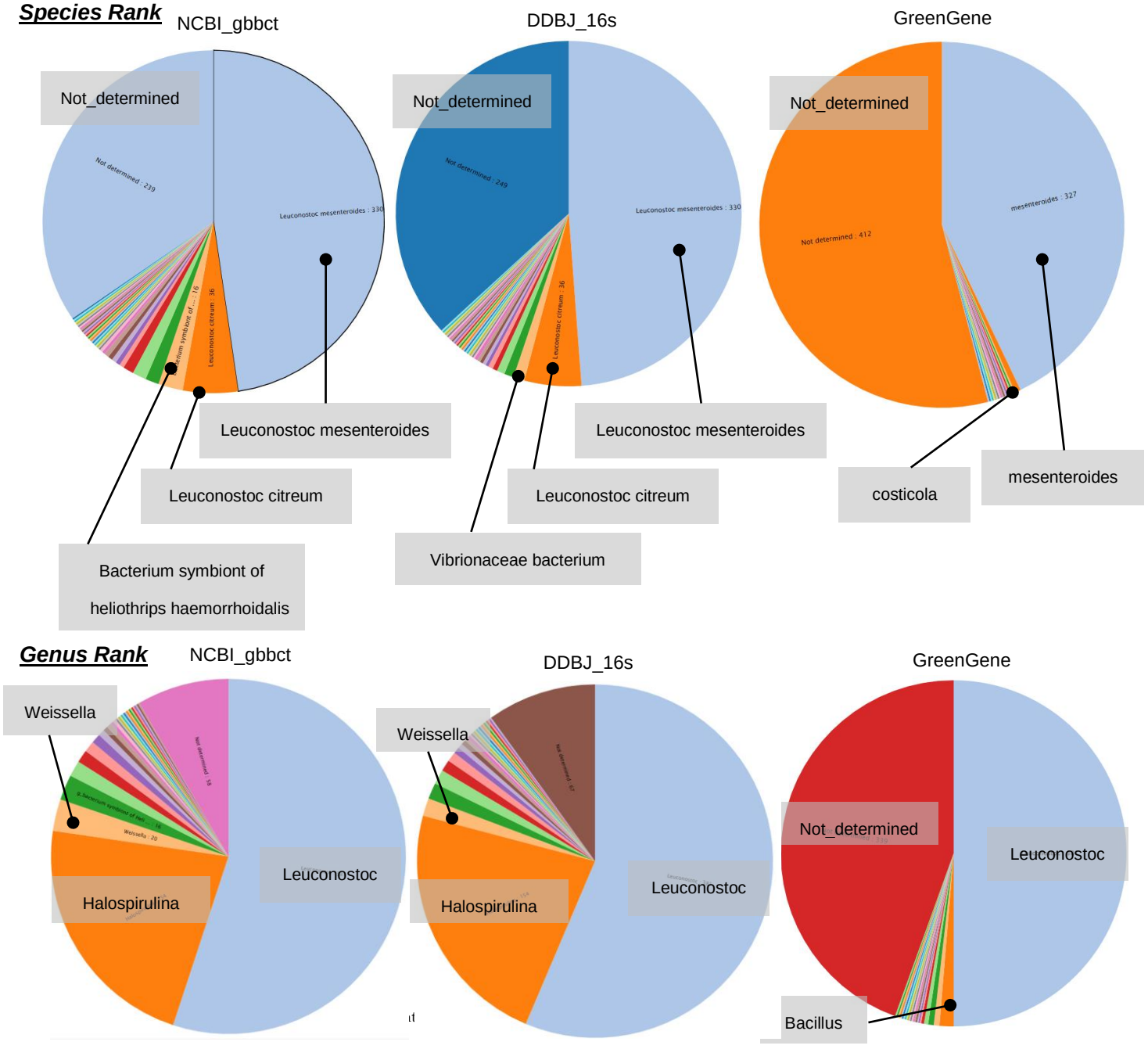
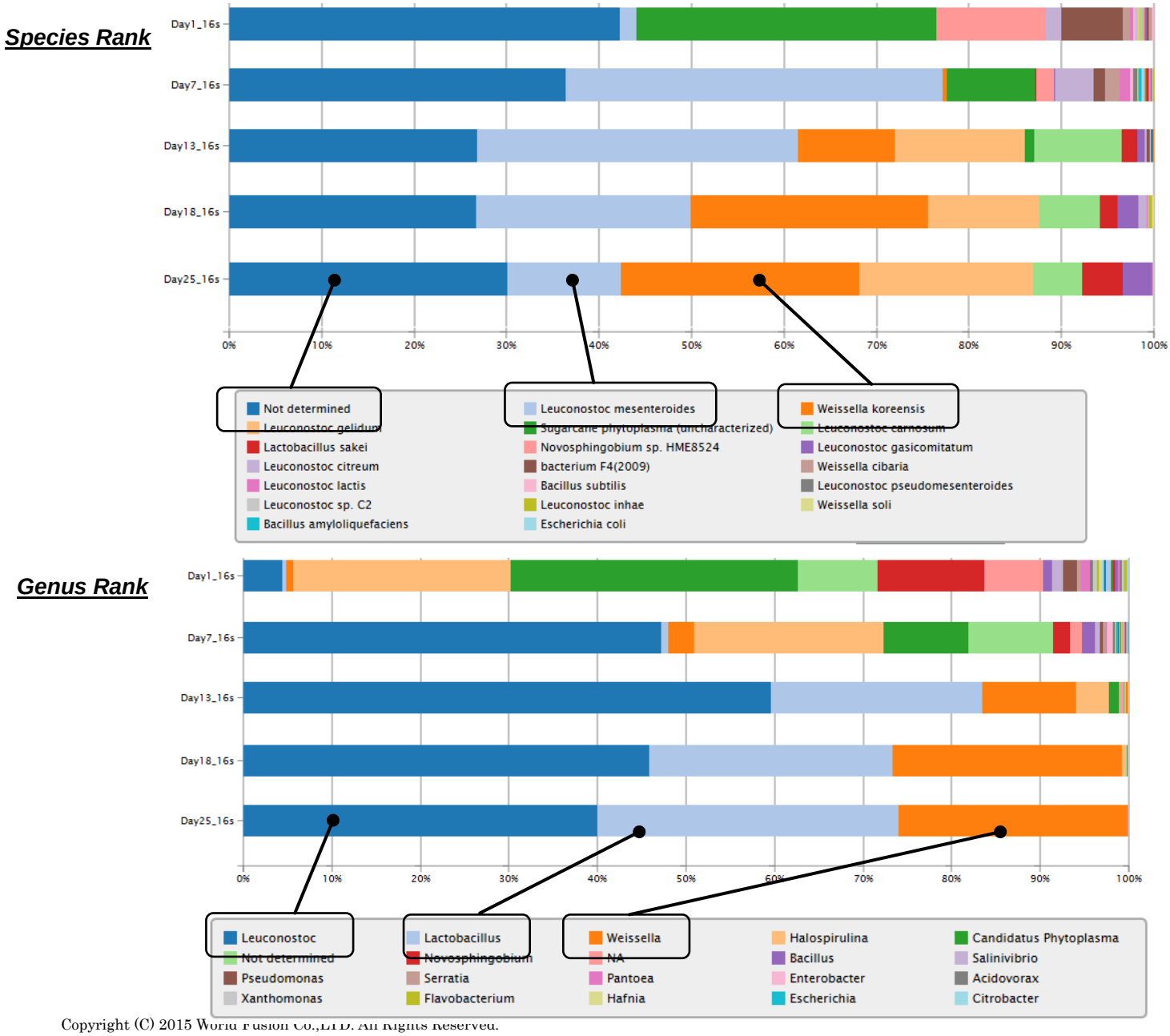


Fig 2 shows the changes in bacterial flora in kimchi throughout a 25 day process, using identification from NCBI_gbbct, and the resulting bar graph was created by Metagenome@KIN. At the species rank, a read can be assigned into multiple bacteria species with the same homology and this read becomes undeterminable (Not determined). Although this situation partially happens at the species rank, bacterial species such as *Leuconostoc mesenteroides* and *Weissella koreensis* started increasing as dominants after Day 7 of fermentation.

From the observation at the Genus rank, *Halosspirulina* disappeared at Day 13 even though it was the dominant at Day 1. On the other hand, *Leuconostoc* started increasing, and the ratio of *Lactobacillus* genus started increasing from Day 13. Few changes were observed after Day 18, so the "undeterminable" problem, assigning a read to multiple bacterial species as homology hit with the same ratio, was resolved at the species rank.

Fig 2. Changes of bacterial flora between fermentation process days in bar graphs



Step 4: Bacterial flora visualization and clustering analysis in KIN

Pie charts and bar graphs display that the drastic changes in bacterial flora throughout the kimchi fermentation process.

Using Metagenome@KIN, the bacterial flora similarity from collected samples from the observed days can be visualized through clustering analyses: the three kinds of analyses as described below.

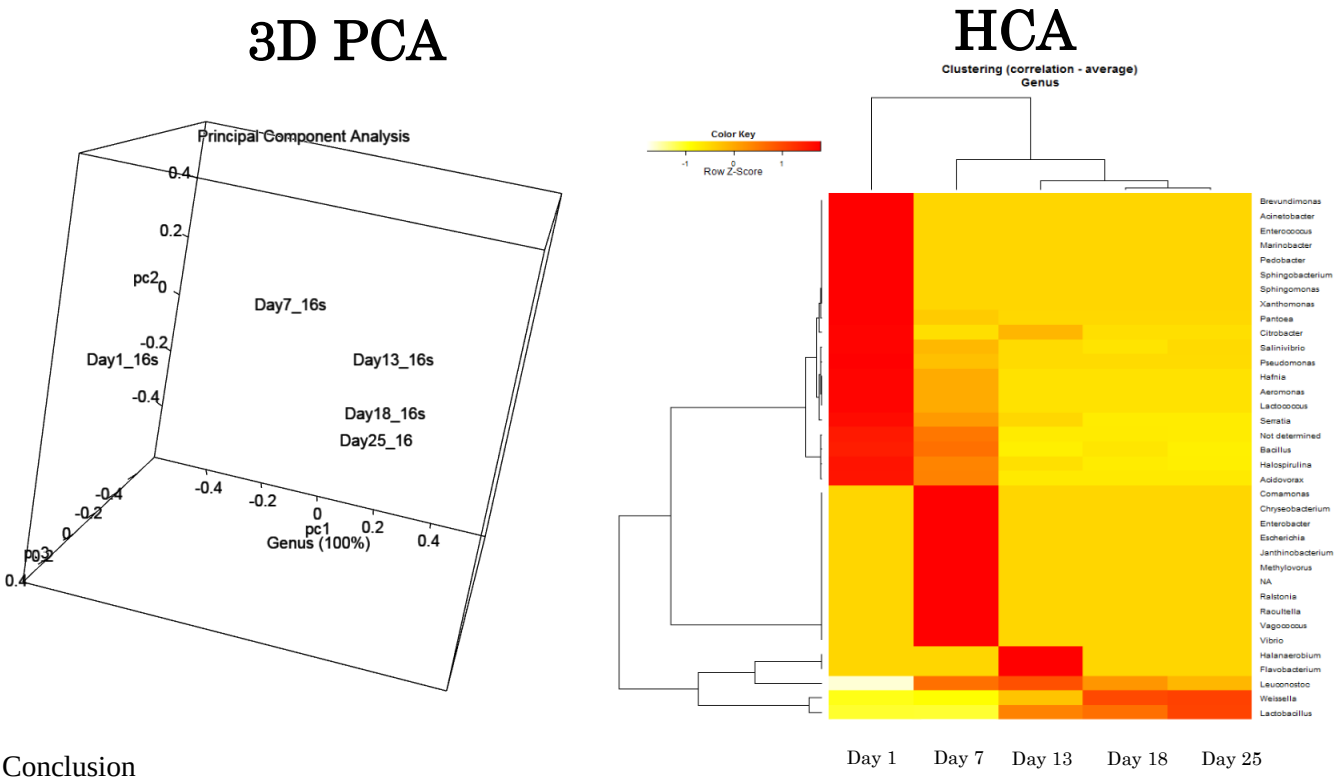
PCA (Principal Component Analysis)

HCA (Hierarchical Clustering Analysis)

SOM (self organizing map)

Fig. 3 shows clustering analysis results from the collected samples of the bacterial flora throughout the kimchi fermentation process from Day 1 through Day 25. While there is a clear difference between samples at Day 1 according to PCA and HCA, at the fermentation process of Day 13, Day 18, and Day 25, the result indicates the samples maintained similar bacterial flora.

Fig3. Clustering analysis based on the bacterial flora



Conclusion

As a result of running BLAST through the homology_analysis_tool and analysis and visualization by Metagenome@KIN, the changes of bacterial flora during the fermentation process of kimchi, a traditional Korean fermented food, were easily observed with usage of the software. According to the observation of the fermentation process from Day 7 through Day 13, bacteria which belong to *Lactobacillales* phylum became dominant. However, after Day 13, the result showed that multiple bacteria species among *Lactobacillales* phylum were coexisting at the levels of genus and species rank.

Ref1
GEORGE E. FOX, et al
INTERNATIONAL JOURNAL OF SYSTEMATIC BACTERIOLOGY, Jan. 1992, p. 166-170
How Close Is Close: 16s rRNA Sequence Identity May Not Be Sufficient To Guarantee Species Identity

More information on Metagenome@KIN:
http://www.w-fusionus.com/#!metagenomekin/c1khf

Contact:
For more information about this software, please contact
kinsupport@w-fusion.com