Shotgun Metagenomic Analysis Reveals New Insights on Bacterial Community Profiles in Tempeh

Adi Yulandi[1,2], Diana Elizabeth Waturangi[2], Aris Tri Wahyudi[1], Antonius Suwanto[1*]

[1]Department of Biology, Faculty of Mathematics and Natural Science, IPB University (Bogor Agricultural University), Gedung Biologi. Jalan Agatis Kampus IPB Dramaga, 16680 Bogor, Indonesia.

[2]Faculty of Biotechnology, Atma Jaya Catholic University of Indonesia, Jalan Jenderal Sudirman 51, 12930 Jakarta, Indonesia.

[*]Corresponding Author: anton.suwanto@atmajaya.ac.id

Abstract

Objective: Amplicon sequencing targeted 16S ribosomal RNA (rRNA) has been widely used for the analysis profile of the microbial community from fermented food samples. Previous results of 16S rRNA analysis metagenome showed that *Firmicutes* was the dominant *phylum* in tempeh. However, polymerase chain reaction (PCR) steps on amplicon sequencing analysis and intragenomic heterogeneity within 16S rRNA are believed to contribute to bias in the estimation of microbial community composition. An alternative approach known as shotgun metagenomic might be able to avoid this limitation. In this study, we employed total metagenomic DNA fragments sequenced directly for taxonomic dan functional profiling analysis.

Result: Taxonomic profiling showed that *Proteobacteria*, *Firmicutes*, and *Bacteroidetes* were the dominant *phyla* from direct shotgun metagenomic analysis in all tempeh samples. In terms of composition, the shotgun metagenomic study revealed that *Proteobacteria* was the most relatively abundant phylum. Functional profiling showed that iron complex outer-membrane recepter protein (KEGG ID: K02014) was the most transcribed genes based on metagenome from tempeh samples.

Keywords: Tempeh, Shotgun metagenomics, Proteobacteria, Firmicutes

Introduction

Tempeh is a fermented food that originated in Indonesia. The biochemical changes of soybean during microbial fermentation increased not only nutritional values but also health-promoting bioactive compounds in the tempeh. Compared to other indigenous soybean-based fermented food such as *nato*, *miso* (Japan), *kinema* (Nepal), and *douchi* (China), which used *Bacillus* spp. as inoculum, tempeh used *Rhizopus* spp. in the production \cite{Astuti_2000}. The nature of tempeh production processes creates consortia of microorganisms not only from tempeh inoculum but also from production materials and environment. \cite{Tamang_2016}. Over the past decade, useful tools of next-generation sequencing (NGS) such as metagenomics has been applied to study microbial consortia from fermented food microbial ecology. These culture-independent techniques have two different approaches to study the taxonomic composition of the microbial community and their relative abundances. The marker-genes are amplified from total microbial genomic DNA through PCR, followed by DNA sequencing is the commonly applied option for study fermented food microbial ecology \cite{Defilippis_2017}. Other approaches called shotgun metagenomics. The total microbial genomic DNA presents in a sample will be untargeted sequencing in this approach. Besides for profile taxonomic composition study, shotgun metagenomics could be used for functional study of microbial communities based on more objective analysis through direct whole-genome sequencing \cite{Sharpton_2014, Quince_2017}. The microbial ecology study employing shotgun metagenomic analysis on samples could enhance our knowledge on both taxonomic and functional profiling of fermented food microbial community. Previous metagenomic studies microbial community during tempeh production were conducted employing amplicon sequencing targeted region V4 16S rRNA gene \cite{Radita_2017, Radita_2018, Pangastuti_2019}. These

studies focused on the dynamic taxonomic profile of the microbial community from tempeh metagenome samples and indicated that Firmicutes was the predominant phylum. To enhance our knowledge on the taxonomic and functional profile of the microbial community in tempeh production, we conducted this study based on the shotgun metagenomic analysis.

Material and Methods

Samples

Fresh tempeh samples were collected from two local traditional producers in Bogor designated as EMP and WJB. The samples have been used as a source for microbial community analysis on tempeh for many years \cite{Radita_2018}. The EMP and WJB producer was representative for the different method in soybean boiling on tempeh production. The EMP is employing a one-time boiling while The WJB two-time.

Total DNA Extraction

The extraction process was adapted from the previous study \cite{Seumahu_2012}. One hundred-gram of tempeh sample was homogenized in 300 mL of phosphate buffer saline (PBS) using the blender for 30 seconds. The homogenate was centrifuged at 1.000 x g for 10 minutes. Supernatants were collected and centrifuged at 10.000 x g for 3 minutes. The pellets were subjected to total microbial DNA extraction employing ZymoBIOMICS DNA/RNA Mini Kit (Zymo Research, California, USA) protocols.

Metagenome Sequencing

The whole metagenome library preparation and sequencing process were using services from NovogeneAIT Genomics Singapore Pte Ltd. The whole microbial DNA was sheared to produce library fragments by restriction enzyme with a minimum of one µg of DNA as input. Precisely

quantifying microbial DNA was used Qubit 2.0 (Thermo Fischer Scientific, United States). The purity and degradation assessment for microbial DNA was employing NanoDrop (Thermo Fischer Scientific, United States) and gel electrophoresis. The pair-end sequencing library was prepared using the TruSeq DNA PCR-Free Prep Kit (Illumina, United States). The prepared library was sequenced on the NovaSeq 6000 platform (2 x 150 bp chemistry) (Illumina, United States).

Shotgun Metagenomic Sequence Data Analysis.

The sequenced reads (raw reads) were filtered from reads; containing adapters, reads containing N (the base cannot be determined) > 10% and reads containing low quality (Qscore<= 5) base which is over 50% of the total base using NovogeneAIT Genomics Singapore Pte Ltd pipeline to produce high-quality paired-end reads (clean reads). The removal of *Rhizopus* spp reads contamination from the clean reads were employed Read QC module from MetaWrap pipeline \cite{Uritskiy_2018} using *Rhizopus microsporus* var. *microsporus* str. ATCC 52814 (GCA_002083745.1) whole-genome sequence as a reference. The SqueezeMeta pipelines \cite{Tamames_2019} were employed for assembly, taxonomic, functional and bin analysis. The pipelines used co-assembly mode option where reads from all samples are pooled before the assembly using Megahit \cite{Li_2015} step performed. The SequeezeMeta binning pipeline includes DAS Tool \cite{Sieber_2018} to merge multiple binning results from Maxbin \cite{Wu_2014} and Metabat2 \cite{Kang_2019} in just set. CheckM \cite{Parks_2015}software was used to estimate the goodness of the bins in the pipeline.

Results

Metagenome Sequencing and Assembly Statistic

The total microbial DNA extracted from tempeh samples collected from two different tempeh producers in Bogor, Indonesia, was subjected to Illumina whole metagenome sequencing pipelines. The average effective rate of clean reads from two raw reads metagenomic data after the quality trimming was 99.93%. Total 9,699,953 (12.28 %) reads from 78,942,348 EMP clean reads and 5,428,509 (5.02%) from 107,981,246 WJB clean reads metagenomic were mapped to the *Rhizopus* genome reference. The number of contigs results from the co-assembly steep data were 316,661. The longest contigs were 485167 bp and N50 value 1924.

Taxonomic and Functional Profiling

The average reads it mapped with reference database using the Diamond \cite{Buchfink_2015} search on SqueezeMeta pipelines was 92.15 %. Individual gene annotation reads are used to produce a consensus for contigs. GeneBank nr was performed as a reference on taxonomic profiling. Contigs with phylum annotation were 271, 375 (85.7% from total contigs). Abundance estimation contig in each sample, the pipelines mapped clean reads onto the contigs resulting assembly step. Among bacteria, *Proteobacteria* was the abundant phylum (Figure 1). Functional profiling used the latest publicly version of KEGG database for KEGG ID annotation. Iron complex outer-membrane recepter protein (KEGG ID: K02014) was the most transcripted expression in the metagenome from tempeh samples (Figure 2).

Binning and Bin Check

The total number of bins obtained from co-assembly metagenome EMP and WJB samples result from DAS tool were 22. According to CheckM result, nine bins categorized as good-quality bins,

which completeness more than 75% with less than 10% contamination. Among good-quality bins, six categorized as high-quality bins, which completeness more than 90% (Table 1).

Discussion

Study of the microbial community of EMP and WJB tempeh samples using 16S rRNA metagenome sequencing analysis already conducted. Concordant with the result of shotgun metagenome sequencing analysis in this study, *Proteobacteria, Firmicutes, and Bacteroidetes* were the abundant phylum, although different in composition. *Firmicutes* relatively most abundant in amplicon sequencing study while in a shotgun metagenomics study was *Proteobacteria*. The data expand the understanding of bioprocess during soybean fermentation compare to previous microbial ecology study using amplicon sequencing \cite{Kumar_2019, Sarkar_2002}.

Data Availability

The tempeh metagenomic raw reads used for this study were deposit in publicly accessible NCBI's Sequence Read Archive (SRA) under the accession number: PRJNA605305.

References

1. Astuti, M., Meliala, A., Dalais, F.S., Wahlqvist, M.L.: Tempe, a nutritious and healthy food from indonesia. Asia Paci_c Journal of Clinical Nutrition 9(4), 322{325 (2000). doi:10.1046/j.1440-6047.2000.00176.x
2. Tamang, J.P., Watanabe, K., Holzapfel, W.H.: Review: Diversity of Microorganisms in Global Fermented Foods and Beverages. Frontiers in Microbiology 7 (2016). doi:10.3389/fmicb.2016.00377. Accessed 2020-03 11

3. De Filippis, F., Parente, E., Ercolini, D.: Metagenomics insights into food fermentations. Microbial Biotechnology 10(1), 91{102 (2017). doi:10.1111/1751-7915.12421. Accessed 2020-03-11
4. Sharpton, T.J.: An introduction to the analysis of shotgun metagenomic data. Frontiers in Plant Science 5 (2014). doi:10.3389/fpls.2014.00209. Accessed 2020-03-11
5. Quince, C., Walker, A.W., Simpson, J.T., Loman, N.J., Segata, N.: Shotgun metagenomics, from sampling to analysis. Nature Biotechnology 35(9), 833{844 (2017). doi:10.1038/nbt.3935. Accessed 2020-03-11
6. Kumar, J., Sharma, N., Kaushal, G., Samurailatpam, S., Sahoo, D., Rai, A.K., Singh, S.P.: Metagenomic insights into the taxonomic and functional features of kinema, a traditional fermented soybean product of sikkim himalaya. Frontiers in Microbiology 10 (2019). doi:10.3389/fmicb.2019.01744
7. Sarkar, P.K., Hasenack, B., Nout, M.J.R.: Diversity and functionality of Bacillus and related genera isolated from spontaneously fermented soybeans (Indian Kinema) and locust beans (African Soumbala). International Journal of Food Microbiology 77(3), 175{186 (2002). doi:10.1016/S0168-1605(02)00124-1. Accessed 2020-03-11
8. Radita, R., Suwanto, A., Kurosawa, N., Wahyudi, A.T., Rusmana, I.: Metagenome analysis of tempeh production: Where did the bacterial community in tempeh come from? Technical Report 4 (2017)
9. Radita, R., Suwanto, A., Kurosawa, N., Wahyudi, Rusmana, I.: Firmicutes is the predominant bacteria in tempeh. International Food Research Journal 25(6), 2313{2320 (2018)
10. Pangastuti, A., Al_sah, R.K., Istiana, N.I., Sari, S.L.A., Setyaningsih, R., Susilowati, A., Purwoko, T.: Metagenomic analysis of microbial community in over-fermented tempeh. Biodiversitas Journal of Biological Diversity 20(4), 1106{1114 (2019). doi:10.13057/biodiv/d200423

11. Seumahu, C.A., Suwanto, A., Rusmana, I., Solihin, D.D.: Comparison of DNA extraction methods for microbial community analysis in indonesian tempe employing ampli_ed ribosomal intergenic spacer analysis. HAYATI Journal of Biosciences 19(2), 93{98 (2012). doi:10.4308/hjb.19.2.93
12. Uritskiy, G.V., DiRuggiero, J., Taylor, J.: MetaWRAP|a exible pipeline for genome-resolved metagenomic data analysis. Microbiome 6(1), 158 (2018). doi:10.1186/s40168-018-0541-1. Accessed 2020-03-11
13. Tamames, J., Puente-S_anchez, F.: SqueezeMeta, A Highly Portable, Fully Automatic Metagenomic Analysis Pipeline. Frontiers in Microbiology 9 (2019). doi:10.3389/fmicb.2018.03349. Accessed 2020-03-11
14. Li, D., Liu, C.-M., Luo, R., Sadakane, K., Lam, T.-W.: MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics (Oxford, England) 31(10), 1674{1676 (2015). doi:10.1093/bioinformatics/btv033
15. Sieber, C.M.K., Probst, A.J., Sharrar, A., Thomas, B.C., Hess, M., Tringe, S.G., Ban_eld, J.F.: Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nature Microbiology 3(7), 836{843 (2018). doi:10.1038/s41564-018-0171-1. Accessed 2020-03-11
16. Wu, Y.-W., Tang, Y.-H., Tringe, S.G., Simmons, B.A., Singer, S.W.: MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome 2(1), 26 (2014). doi:10.1186/2049-2618-2-26. Accessed 2020-03-11
17. Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., Wang, Z.: MetaBAT 2: an adaptive binning algorithm for robust and e_cient genome reconstruction from metagenome assemblies. PeerJ 7 (2019). doi:10.7717/peerj.7359. Accessed 2020-03-11

18. Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W.: CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Research, 186072{114 (2015). doi:10.1101/gr.186072.114. Accessed 2020-03-11

19. Buch_nk, B., Xie, C., Huson, D.H.: Fast and sensitive protein alignment using DIAMOND. Nature Methods 12(1), 59{60 (2015). doi:10.1038/nmeth.3176. Accessed 2020-03-1

Figure 1. The taxonomic abundance of the microbial community tempeh samples at the rank of phylum

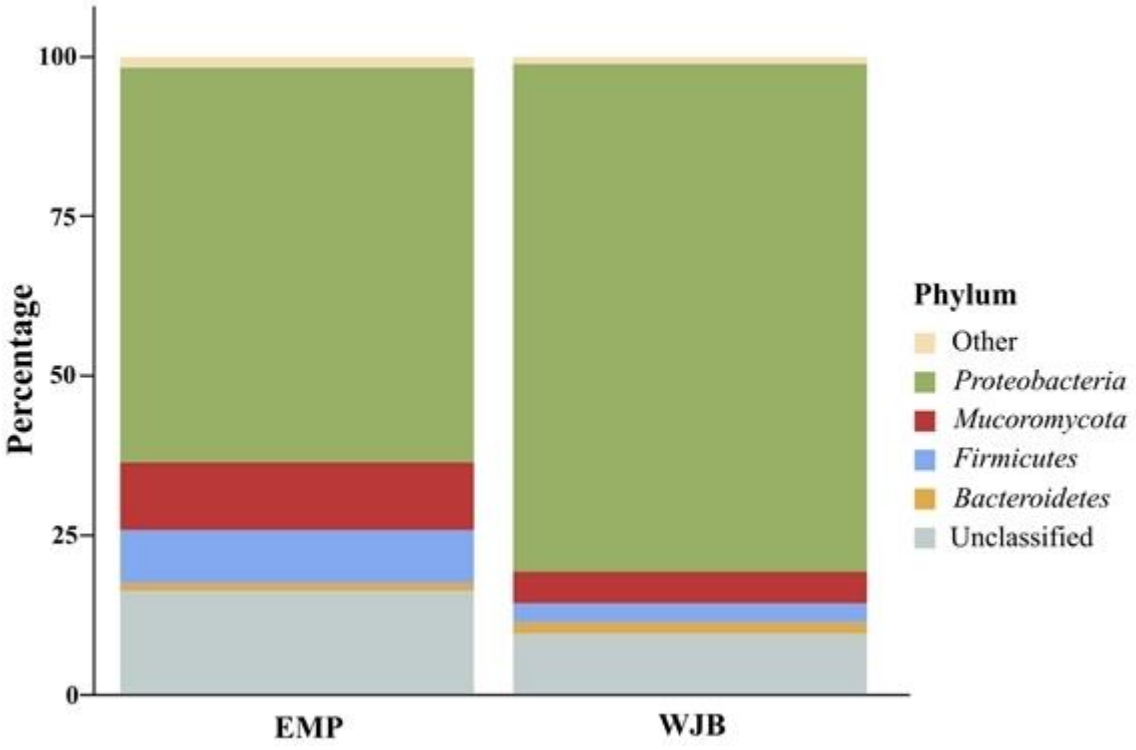
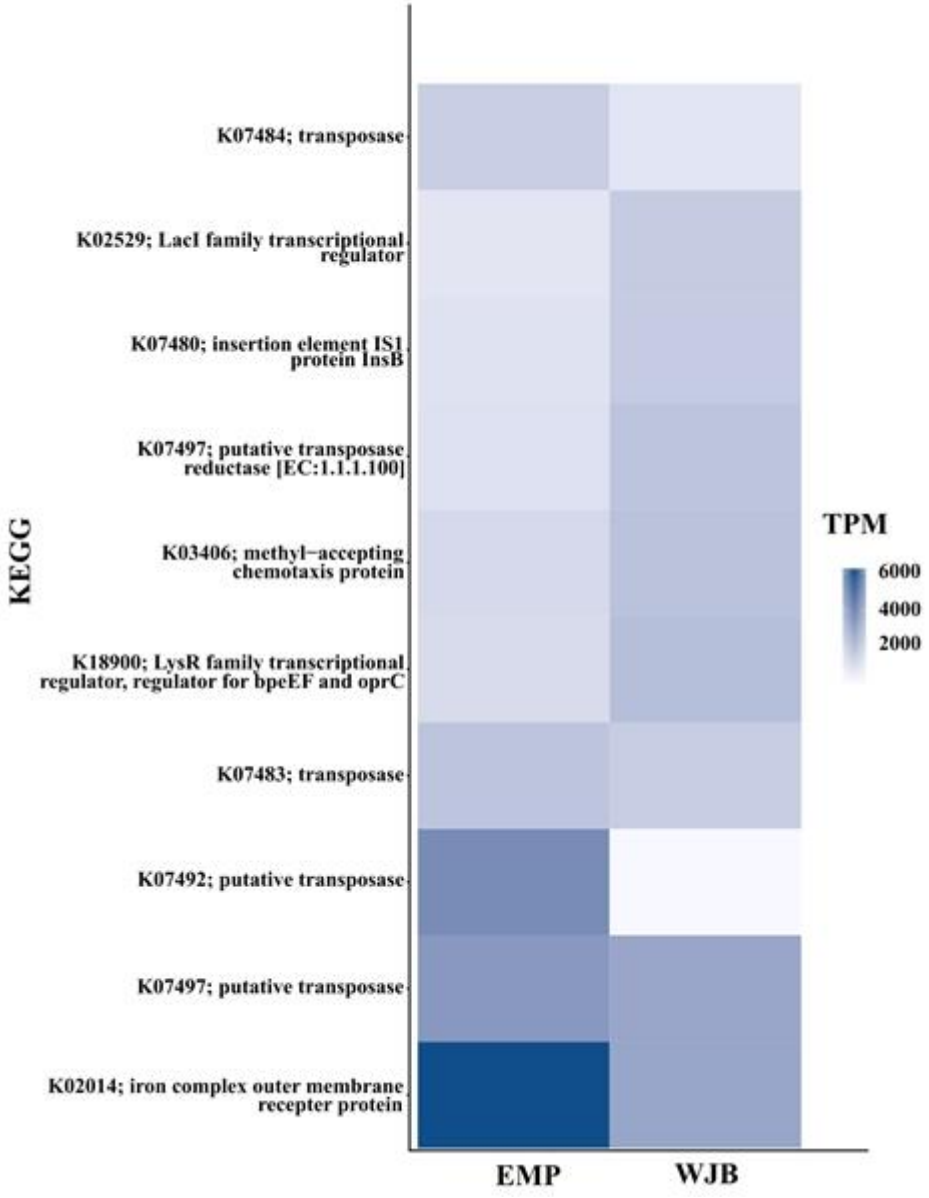


Figure 2. The functional profile of the metagenome tempeh samples using KEGG annotation in TPM (transcripts per kilobase million)



Tabel 1. High quality bins (>90% completion, <10% contamination) obtained by co-assembly mode of EMP and WJB.

Taxa	Size (bp)	Completeness	Contamination	Taxonomic rank
Bacteria	4,989,656	100%	0.16%	*Kingdom*
Lactobacillus delbrueckii	1,651,262	98.33%	0.16%	*Species*
Bacteria	1,657,902	96.55%	0.78%	*Kingdom*
Bacteria	2,925,287	95.69%	3.61%	*Kingdom*
Lactobacillus fermentum	2,113,084	95.56%	2.26%	*Species*
Bacilli	4,300,435	93.32%	6.62%	*Class*
Escherichia coli	4,496,589	85.79%	4.3%	*Species*
Bacteria	6,993,346	85.61%	4.23%	*Kingdom*
Gammaproteobacteria	3,538,428	81.82%	7.79%	*Class*