Metagenomics workflow analysis of endophytic bacteria from oil palm fruits

Z A Tanjung*, R Aditama, W M Sudania, C Utomo, and T Liwang
Plant Production and Biotechnology Division, PT SMART Tbk., Sinar Mas Land Plaza, 2nd Tower, 10th Fl., Jl. M.H. Thamrin No. 51, Jakarta 10350, Indonesia
E-mail: biotechnology@sinarmas-agri.com

Abstract. Next-Generation Sequencing (NGS) has become a powerful sequencing tool for microbial study especially to lead the establishment of the field area of metagenomics. This study described a workflow to analyze metagenomics data of a Sequence Read Archive (SRA) file under accession ERP004286 deposited by University of Sao Paulo. It was a direct sequencing data generated by 454 pyrosequencing platform originated from oil palm fruits endophytic bacteria which were cultured using oil-palm enriched medium. This workflow used SortMeRNA to split ribosomal reads sequence, Newbler (GS Assembler and GS Mapper) to assemble and map reads into genome reference, BLAST package to identify and annotate contigs sequence, and QualiMap for statistical analysis. Eight bacterial species were identified in this study. Enterobacter cloacae was the most abundant species followed by Citrobacter koseri, Seratia marcescens, Lactococcus lactis subsp. lactis, Klebsiella pneumoniae, Citrobacter amalonaticus, Achromobacter xylosoxidans, and Pseudomonas sp. respectively. All of these species have been reported as endophyte bacteria in various plant species and each has potential as plant growth promoting bacteria or another application in agricultural industries.

1. Introduction
Next-Generation Sequencing (NGS) or high-throughput sequencing has become a powerful sequencing tool for microbial study. It has led to the establishment of the field of “meteganomics” [1]. Metagenomics (also referred to as environmental and community genomics) is the microorganisms genomic analysis of microorganisms by direct extraction and cloning of DNA from an assemblage of microorganisms [2]. Generally, it is used to investigate marines [3] and soils [4] complex microbial communities. NGS consists of different platforms, such as Roche 454 pyrosequencing and Illumina/Solexa as a leading platforms. Roche 454 pyrosequencing has increase its average read lengths to ~500 base pairs compared to Illumina which only ~150 base pairs [5].

Endophytic bacteria have wide range of effects on their hosts, such as: increasing phosphate solubilization and nitrogen fixation, repressions of soil-borne pathogens, breakdown of plant-produced ethylenes by bacterial production of 1-aminocyclopropane-1-carboxilate deaminase, induced systemic

* Corresponding author
resistances, and synthesis of phytohormones [6]. Oil palm fruits are the most efficient source of vegetable oil. An endophytic bacteria that may live in oil palm fruits certainly have produced enzyme activity which could utilize vegetable oil for industrial application. This study described a workflow to analyze metagenomics data via direct sequencing from endophytic bacteria in oil palm fruits.

2. Material and Methods

2.1 Materials
The raw data of this study was Sequence Read Archive (SRA) file under accession ERP004286. This SRA was uploaded by University of Sao Paulo. SILVAdb was used for identification of RNA sequences [7]. SWISS-PROT database [8] and NCBI-NR database were used for functional annotations.

The bioinformatics software used for this study were SRA Toolkit [9], Filter FASTQ [10], SortMeRNA [11], Newbler (GS Assembler and GS Mapper) [12], BLAST package [13], and QualiMap [14].

2.2 Methods
All methods used in this study were described in Figure 1. This workflow was designed for analyzing direct sequencing data from metagenomics analysis of oil palm fruits using 454 pyrosequencing platform.

![Figure 1. Metagenomics workflow analysis of endophytic bacteria from oil palm fruits](image-url)
3. Results and Discussion

This study used Sequence Read Archive (SRA) file under accession ERP004286 that contained raw data of metagenomics sequencings of endophytic bacteria in oil palm fruits. The bacterial endophytes within the oil palm fruits (*Elaeis guineensis*) were cultured using an oil-palm enriched medium and the bacterial cells present in the microcosm were separated by filtration through a 14 micrometer pore filter. The DNA obtained by these cells was then sequenced using 454 pyrosequencing technology on the 454 GS FLX platform. A single fragment (shotgun) library was sequenced in two separate runs. It was generated from 454 pyrosequencing platforms and has single-end reads.

The raw data have 548,757 reads. After trimming process, it was assembled by GS Assembler (Newbler). It produced 74 contigs from ribosomal sequence and 18,600 contigs from non-ribosomal sequence. The ribosomal sequence was identified using blastn against SILVAdb. This identified 8 species which live in oil palm fruits. The microbial diversity profile of endophytic bacteria in oil palm fruits was described on Figure 2. There were 7 species which belong to Phylum Proteobacteria and 1 species belong to Phylum Firmicutes. *Lactococcus lactis* subsp. *lactis* was the only species belong to Firmicutes. It was known as plant-associated lactic acid bacterium. *Enterobacter cloacae* was the most abundant in oil palm fruits. It is followed by another Enterobacteriaceae, which were *Citrobacter koseri*, *Seratia marcescens*, *Klebsiella pneumoniae*, *Citrobacter amalonaticus*, and *Pseudomonas* sp., *E. cloacae* was known as plant growth-promoting bacteria (PGPB). It was also an endophytic symbiont of corn, cacao, and pines [15, 16, 17]. *C. koseri* was also reported as endophyte in corn and cotton [18]. *S. marcescens* has been reported as cold tolerance and plant growth-promotion potential from summer squash [19]. *K. pneumoniae* was reported as nitrogen-fixing bacteria in wheat [20]. *C. amalonaticus* have phytase gene which potentially to degrade phytic acid [21]. *A. xylosoxidans* was the only bacteria member of Alcaligenaceae. It was reported as plant growth promoting bacterium in wheat [22] and as a copper-resistant plant growth promoting bacteria in mustard [23].

![Figure 2. Microbial diversity profile from endophytic bacteria of oil palm fruits](image-url)
There are limited publication of metagenomics profile of bacteria in plant fruits. It was done in strawberry and rambutan fruits by selected-colony to 16S sequencing [24,25] also in citrus by PCR-DGGE [26]. Bacterial profile in strawberry is more likely to oil palm fruits than citrus and rambutan fruits. Genus *Enterobacter, Lactobacillus,* and *Pseudomonas* also present in strawberry fruits [24].

Metagenomics in plant fruits analysis using whole genome shotgun sequencing give the real abundance of bacterial community profile because it reduces PCR bias compare to 16S-NGS method. In this study, 454 pyrosequencing platform was used. It give medium long reads around 600-700 bp compare to Illumina platform which generate 250 bp at most. This study provide a workflow to analyse whole genome shotgun metagenomics 454 pyrosequencing platform data of oil palm fruits to produce bacterial community profile.

4. Conclusion
This workflow performed well in revealing ribosomal and non-ribosomal metagenomics sequences of endophytic bacteria from oil palm fruits. According to their coverage, *E. cloacae* is the most abundant and followed by *C. koseri, S. marcescens, L. lactis* subsp. lactis, *K. pneumoniae, C. amalonaticus, A. xylosoxidans* and *Pseudomonas* sp.. Each species has their own potential in agricultural applications.

References
[1] Oulas A, Pavloudi C, Polymenakou P, Pavlopoulos GA, Papanikolaou N, Kotoulas G and Iliopoulos I 2015 *Bioinform Biol Insights* 9 75–88
[2] Handelsman J 2004 *Microbiol Mol Biol Rev.* 68 669–85
[3] Gilbert J, Field D, Huang Y, Edwards R, Li W, Gilna P and Joint I 2008 *PloS one.* 3 e3042
[4] Urich T, Lanzén A, Qi J, Huson DH, Schleper C and Schuster SC 2008 *PloS one.* 3 e2527
[5] Balzer S, Malde K, Lanzén A, Sharma A and Jonassen I 2010 *Bioinformatics* 26 i420–5
[6] De Melo Pereira GV, Magalhães KT, Lorenzetii ER, Souza TP and Schwan RF 2012 *Microb. Ecol.* 63(2) 405–17
[7] Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J and Glöckner FO 2007 *Nucleic Acids Res.* 35 7188–96
[8] Consortium TU 2014 *Nucleic Acids Res.* 43 204–12
[9] Leinonen R, Sugawara H and Shumway M 2011 *Nucleic Acids Res.* 39 D19–21
[10] Blankenberg D, Gordon, Von Kuster G, Coraor N, Taylor J and Nekrutenko 2010 *Bioinformatics* 26 1783–5
[11] Kopylova E, Noë L and Touzet H 2012 *Bioinformatics* 28 3211-7
[12] Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA and Rothberg JM 2005. *Nature* 437 376–80
[13] Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ, 1990. *J. Mol. Biol.* 215 403-10.
[14] García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, Tarazona S, and Conesa A 2012 *Bioinformatics* 28 2678–9
[15] Hinton DM, Bacon CW and Ars U 1995 *Mycopathologia* 129 117–25
[16] Zayda PMM, Elizabeth AAD, Thiago ASDO, Fernando PM, Leandro LL, and Jorge TDS 2015 *Afr. J. Microbiol. Res.* 9 1352–56
[17] Liu WY, Chung KMK, Wong CF, Jiang JW, Hui, RKH and Leung FCC 2012 *J. Bacteriol.* 194 5965
[18] McNamara JD and Klepper JW 1995 *Plant Soil* 173 337-42
[19] Selvakumar G, Mohan M, Kundu S, Gupta AD, Joshi P, Nazim S and Gupta HS 2008 *Lett. Appl. Microbiol.* 46 171–5
[20] Iniguez AL, Dong Y and Triplett EW 2004 *Mol. Plant. Microbe. Interact.* 17 1078–85
[21] Li C, Lin Y, Huang Y, Liu X and Liang S 2014 PloS One 9 e114728
[22] Jha P and Kumar A 2009 Microb. Ecol. 58 179–88
[23] Ma Y, Rajkumar M and Freitas H 2009 J. Environ. Manage. 90 831–7.
[24] Pereira GVDM, Magalhães KT, Lorenzetii ER, Souza TP and Schwan RF 2012 Microb Ecol. 63 405–17
[25] Suhandono S, Kusumawardhani MK and Aditiawati P 2016 HAYATI J. Biosci. 23 39–44
[26] Nguyen DDL, Gemrot E, Loiseau G and Montet D 2008 Fruits 63 75–84

Acknowledgement
The authors gave utmost gratitude to the management of PT SMART Tbk, for supporting the publication of this research paper.