Research project of counterparts funded at Syiah Kuala University in 2021

Name	Counterpart	Title
Essy Harnelly	**B14**	Assembling chloroplast genome sequences of petai, jengkol, sungkai and jelutung using both short-read and long-read sequencing

Background

Since diversity is thought to be lower in oil palm plantations compared to primary and secondary forest, an enrichment planting experiment was recently conducted utilizing 6 tree species, including Durian (*Durio zibethinus* of Malvaceae), Jengkol (*Archidendron pauciflorum* of Fabaceae), Petai (*Parkia scpeciosa* of Fabaceae), Jelutung (*Dyera* spp of Apocynaceae), Sungkai (*Peronema canescens* of Verbenaceae) and Jelutung (*Dyera costulata* of Apocynaceae). Although all of these species are well known as favorite community trees and seemed to grow well throughout Indonesia, recent field observations showed that not all species were able to adapt and grow well in the experiment. The families of these tree species, especially Apocynaceae and Fabaceae, are among the best studied in the world. Those two families already have reference genomes, both nuclear and chloroplast genomes, deposited at NCBI, thanks to increasingly cheap NGS technology (Tang *et al*. 2014; Wang *et al*. 2018; Weitemier *et al*. 2019). Accurate genome assembly is the first key important step in genomic studies. It is well known that both short-read and long-read sequencing have their own advantages and limitations in the assembly process.

Objective

The objective of this research is to assemble the chloroplast genomes of four tree species commonly found among local community gardens in Indonesia, i. e., Jengkol (*Archidendron pauciflorum*), Petai (*Parkia scpeciosa*), Sungkai (*Peronema canescens*) and Jelutung (*Dyera costulata*).

Methods

Sample collection of Jengkol, Petai, Sungkai and Jelutung was conducted in *EFForTS-BEE* plots, PT Humusindo, Bugku, Jambi Province. The leaves samples were collected from three individuals for each species and extracted through a modified CTAB method. The DNA library preparation was followed the Nanopore Protocol for Native barcoding genomic DNA (with EXP-NBD104, EXP-NBD114, and SQK-LSK109), version NBE_9065_v109_revAC_14Aug2019. Sequencing was done in two rounds using two flowcells (FLO-MIN106). The sequencing run of genomic DNA samples was performed using the MinKnow v4.4.3.

Results

The data received from the sequencing process is in the form of FAST5. The high-accuracy basecalling mode was used to basecall the signal in FAST5 files and outputted FASTQ files. FASTQ files obtained from Jengkol, Petai, Sungkai, and Jelutung species were 1.2 Gb, 1.5 Gb, 1.3 Gb and 832 Mb respectively. All samples were uploaded to the usegalaxy websites for analyze (*https://usegalaxy.eu*). The reads quality and reads statistics were calculated using NanoPlot. After statistical analysis, all reads quality was filtered and assembled through Flye assembler. The draft assembly was then polished using medaka_consensus. The resulting polished assembly statistics was calculated using QUAST. Afterwards, completeness of the data would be checked by BUSCO. Jengkol species had 180,949 reads with 100 % reads quality >Q7 (nanopore default passed quality) (Fig. 1).

The contig was 492 (N50 43,210 bp, GC 37.42 %) and had 2 complete BUSCOs and 422 missing BUSCOs. For Petai species (Fig. 2), we obtained 320,291 reads with 100 % reads quality >Q7. The contig was 390 (N50 20.749, GC 35.06 %). Petai had only 1 complete BUSCOs and 423 missing BUSCOs.

Furthermore, we acquired 256,699 reads with 100% reads quality >Q7 for Sungkai species (Fig. 3). The contig was 1,201 (N50 21,952 bp, GC 36.29%) and had 13 complete BUSCOs and 407 missing BUSCOs.

Lastly, Jelutung species has 264,789 reads with 100% reads quality >Q7. The contig was 145 (N50 18,219 bp, GC 37.82%). Jelutung (Fig. 4) has no complete BUSCOs and had 425 missing BUSCOs.

The contig from medaka_consensus were then annotated by using GeSeq platform for Organellar Genomes, resulted in the GenBank annotation and their visualization.

References

Tang H, Krishnakumar V, Bidwell S, Rosen B, Chan A, Zhou S, Gentzbittel L, Childs KL, Yandell M, Gundlach H, Mayer KFX, Schwartz DC, Town CD. 2014. An improve genome release (version Mt4.0) for the model legume Medicago truncatula. BMC Genomics15:312-326. https://pubmed.ncbi.nlm.nih.gov/24767513/

Wang W, Schalamun M, Morales-Suarez A, Kainer D, Schwessinger B, Lanfera R. 2018. Assembly of chloroplast genomes with long- and short- read data: a comparison of approaches using Eucalyptus pauciflora as a test case. BMC Genomics 19:977-992. *https://doi.org/10.1186/s12864-018-5348-8*

Weitemier K, Straub SCK, Fishbein M, Bailey CD, Cronn RC, Liston A. 2019. A draft genome and transcriptome of common milkweed (Asclepias syriaca) as resources for evolutionary, ecological, and molecular studies in milkweeds and Apocynaceae. PeerJ 7:e7649. *https://doi.org/10.7717/peerj.7649*

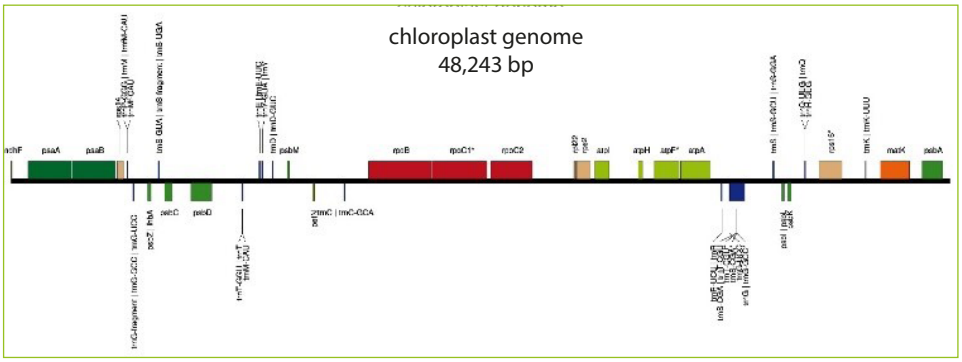

Figure 1. Gene annotation of jengkol species from long-read sequencing.

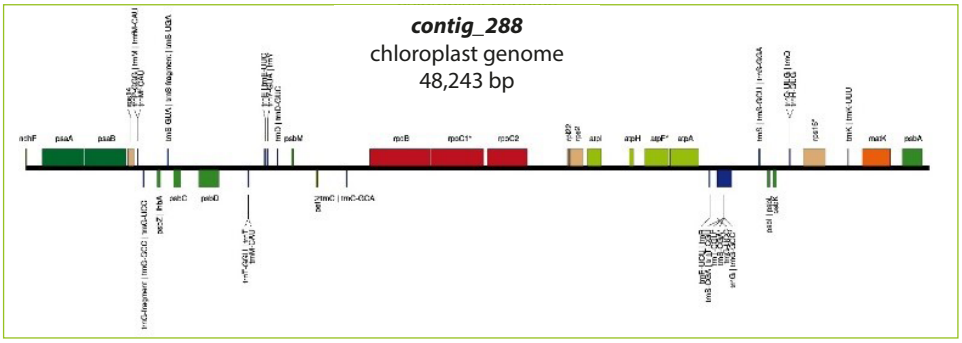

Figure 2. Gene annotation of Petai Species from long-read sequencing.

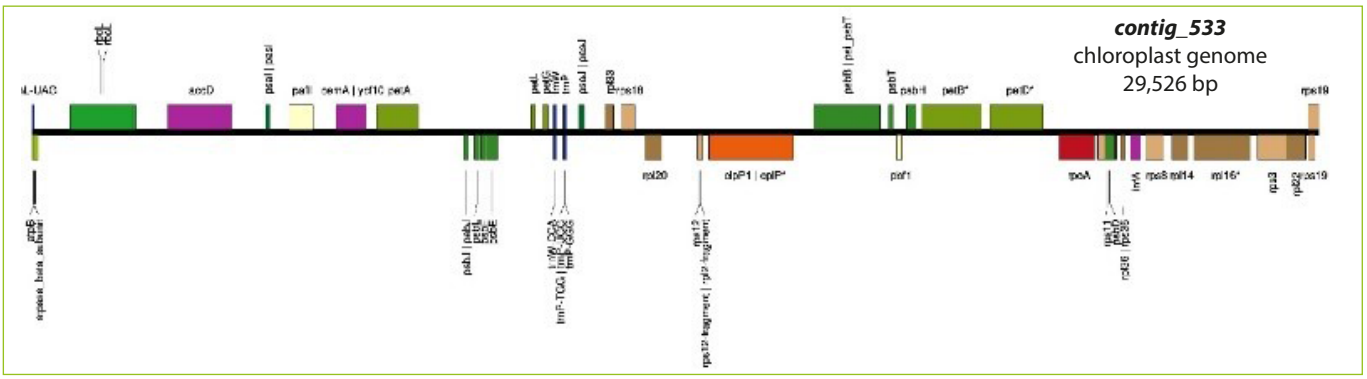

Figure 3. Gene annotation of Sungkai species from long-read sequencing.

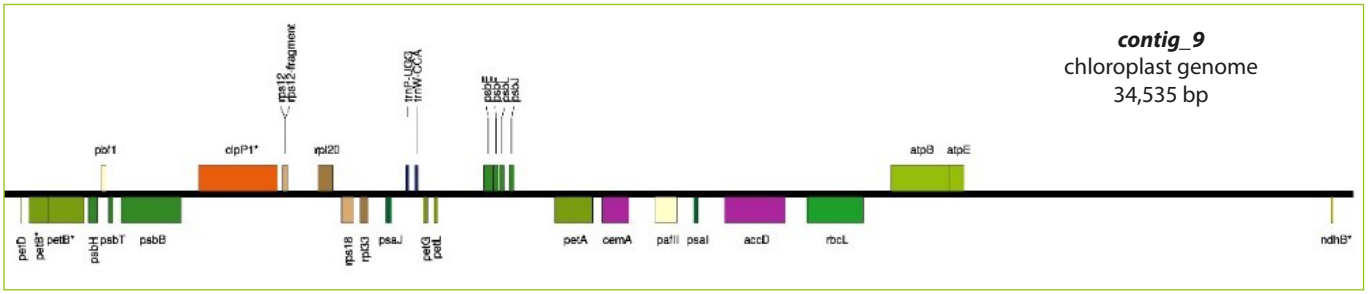

Figure 4. Gene annotation of Jelutung Species from long-read sequencing.