Name	Counterpart	Title
Ulfah Juniarti Siregar	**B14**	Assembling chloroplast genome sequences of petai, jengkol, sungkai and leprosula using both short-read and long-read sequencing: WP1: Generating raw sequences of four species from short read and long read sequencing

Background and Objectives

Jengkol (*Archidendron pauciflorum* of Fabaceae), Petai (Parkia scpeciosa of Fabaceae), Jelutung (*Dyera* spp of Apocynaceae), Sungkai (Peronema canescens of Apocynaceae) and Leprosula (*Shorea leprosula* of Dipterocarpaceae) are well known and widely distributed in Indonesia due to their economic value to the local population, however, very little basic genetic information is available on these species. The lack of such information has resulted in some planted species being less successful in surviving and adapting to the new planting environment. The families of those tree species, especially Apocynaceae and Fabaceae, are among the best studied in the world because of their importance for human beings. This research aimed to assemble the chloroplast genomes of the four tree species mentioned above. The chloroplast genome has been subject of evolutionary studies in plants due to its small size and conservative nature. The chloroplast genome also contains important genes related to physiological mechanisms for adaptation and stress.

Methods

DNA was extracted from leaf samples of four species, i.e. Jengkol, Petai, Jelutung and Sungkai, then subjected to shotgun sequencing using Illumina NovaSeq 6000, through Genetika Science Co. (representative of BGI in Indonesia). Short-read sequences data analysis was performed in Maser platform (*https://cell-innovation.nig.ac.jp*). Quality control was conducted using FASTQC and Filter FASTQ. Data with quality score (Q) > 30 were further assembled de novo due to lack of reference genome for each species using Platanus (Kajitani *et al.* 2019), SOAPdenovo (Li *et al.* 2010), and Ray (Boisvert *et al.* 2010). The draft genome from the assembly was run in BUSCO v3.0.2 (Genome assembly [nucleotide]) (Seppey *et al.* 2019). Results from Platanus assembler was annotated for chloroplast genes using GeSeq and visualized by OGDraw which is available in MPI-MP CHLOROBOX (*https://chlorobox.mpimp-golm.mpg.de/geseq.html*).

Major Results and Conclusion

From the three assembler software used, only SOAPdenovo and Ray provided good results when run in BUSCO (Table 1). Only partially annotated chloroplast genes could be assembled, and further attempts are needed to find as many genes as possible to assemble the entire chloroplast genomes of the four species.

Table 1. Results of BUSCO analysis from SOAPdenovo and Ray assemblers

Assembler	Species	Complete and Single Copy BUSCOs (S)		Complete and duplicated BUSCOs (D)		Fragmented BUSCOs (F)		Missing BUSCOs (M)	
		n	%	n	%	n	%	n	%
SOAP denovo	Jengkol	27	8.9	0	0	11	3.6	265	87.5
	Sungkai	36	11.9	0	0	15	5	252	83.2
	Jelutung rawa	11	3.6	0	0	4	1.3	288	95
	Petai	32	10.6	0	0	17	5.6	254	83.8
Ray	Jengkol	33	10.9	0	0	7	2.3	263	86.8
	Sungkai	52	17.2	1	0.3	4	1.3	246	81.2
	Jelutung rawa	11	3.6	0	0	5	1.7	287	94.7
	Petai	42	13.9	2	0.7	12	4	247	81.5

References

Kajitani R, Yoshimura D, Okuno M, Minakuchi Y, Kagoshima H, Fujiyama A, Kubokawa K, Kohara Y, Toyoda A, Itoh T. 2019. Platanus-alle is a de novo haplotype assembler enabling a comprehensive acces to divergent heterozygous regions. Nature Communications 10(1702) : 1–15. *https://doi.org/10.1038/s41467-019-09575-2*

Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, *et al*. 2010. De novo assembly of human genomes with massively parallel short read sequencing. Genome Research 20 : 265-272. *https://doi.org/10.1101/gr.097261.109*

Boisvert S, Laviolette F, Corbeil J. 2010. Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. J. Comput. Biol. 17(11):1519–1533. *https://doi.org/10.1089/cmb.2009.0238*

Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: Assessing genome assembly and annotation completeness. Di dalam: Kollmar M, editor. Gene Prediction. Methods in Molecular Biology. Volume 1962. New York: Humana Press. *https://doi.org/10.1007/978-1-4939-9173-0_14*