

Application of Advanced Genomic Technology to Improve
Timber Productivity and Wood Quality of a Commercial
Timber Species in Tropical Rainforest

July 2020

SAWITRI

Application of Advanced Genomic Technology to Improve
Timber Productivity and Wood Quality of a Commercial
Timber Species in Tropical Rainforest

A Dissertation Submitted to
the Graduate School of Life and Environmental Sciences,
the University of Tsukuba
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Agricultural Science
(Doctoral Program in Biosphere Resource Science and Technology)

SAWITRI

Abstract

Shorea platyclados (Dark Red Meranti) is a commercially important timber tree species in Southeast Asia. However, its stocks have dramatically declined due, *inter alia*, to excessive logging, insufficient natural regeneration and a slow recovery rate. Thus, population genetic analysis was conducted to evaluate the genetic diversity and population structure of this species. Moreover, there is a need to promote enrichment planting and develop effective technique to support its rehabilitation and simultaneously improve timber production and wood quality. These can be achieved in classical tree breeding improvement based on phenotype selection and indirectly through implementation of Genome-Wide Association Studies (GWAS) and Genomic Selection (GS). To assist such efforts, plant materials were collected from a half-sib progeny population in Sari Bumi Kusuma forest concession (PT. SBK), Central Kalimantan, Indonesia.

Population genetics were evaluated using 5,900 markers in sequences from 356 individuals obtained by double digest Restriction-site Associated DNA sequencing (ddRAD-seq). My study confirmed that there was a relatively high level of nucleotide diversity (π), observed (H_o) and expected heterozygosity (H_e) of 0.004811, 0.233 and 0.245, respectively. These were supported by low inbreeding coefficient (F_{IS}) of 0.047. For the population structure, I found no distinct spatial clusters with PCoA analysis explained only 4.13% and 3.40% of the total genetic variance.

Further, using number of markers obtained by ddRAD-seq method, I detected high linkage disequilibrium (LD) of *S. platyclados* extending up to >145 kb, much longer than that observed by other forest tree species. It suggests that associations between phenotypic traits and markers in LD can be more easily and feasibly detected with GWAS than with analysis of quantitative trait loci (QTLs). However, in fact, the detection power of single-locus GWAS seemed low, since fewer single nucleotide polymorphisms (SNPs) linked to any focal traits were detected with a stringent false discovery rate. It indicates that the species' phenotypic traits are mostly under polygenic quantitative control. Unlike single-locus GWAS, a multi-locus GWAS model (FASTmrMLM) showed its power by successfully identified 18 significant Quantitative Trait Nucleotides (QTNs) without any QTNs duplicated. Those QTNs were distributed in the different contigs. Therefore, identified contigs (in which significant QTNs were detected) are potentially utilized as the hotspot region for fine mapping and candidate gene analysis of phenotypic traits of *S. platyclados*.

Different with GWAS, GS is not affected by the limitation of GWAS linked to the problem of detecting markers that are significantly associated with polygenic traits, because it exploits the predictive power of large numbers of markers simultaneously across the whole genome. I explored the genomic prediction accuracy of this investigated species and found that Machine Learning method (Random Forest) performed slightly better prediction accuracies (range from 0.130 to 0.244) than all Bayesian methods (range from 0.113 to 0.210). My study showed that stem diameter, branch diameter ratio and wood density were more predictable than height, clear bole, branch angle and wood stiffness traits, therefore it proved that genomic prediction accuracy were trait dependent. This predictive model then will be used to predict the phenotypic value of selection candidates (young seedlings derived from the training population are genotyped, but not phenotyped) in the next generation. Using the accurate predictive model formed from the earlier step, the breeders can select some top ranked GEBV seedlings. In addition, these top ranked GEBV seedlings can be promoted for flower induction to conduct cross breeding. Hereafter, breeders can apply higher selection intensity among juveniles. This strategy will help forest tree breeders to reduce generation interval to gain full advantage of the early availability of GEBV in shorter cycle.

Beside statistical methods, my study also confirmed that some factors may affect the genomic prediction accuracy, such as training population set, marker density and kinship relatedness. Additionally, I found high genomic heritability, but it was not always increase high accuracy. However, high genomic heritability may help selection of candidate traits for future breeding. Generally, my study suggested that GS had potential for improving the timber productivity and wood quality of *S. platyclados*.

Keywords: *Shorea platyclados*, ddRAD-seq, genetic diversity, population structure, LD, GWAS, single-locus and multi-locus, GS, Bayesian method, Machine Learning, genomic heritability, improvement, growth, wood quality traits