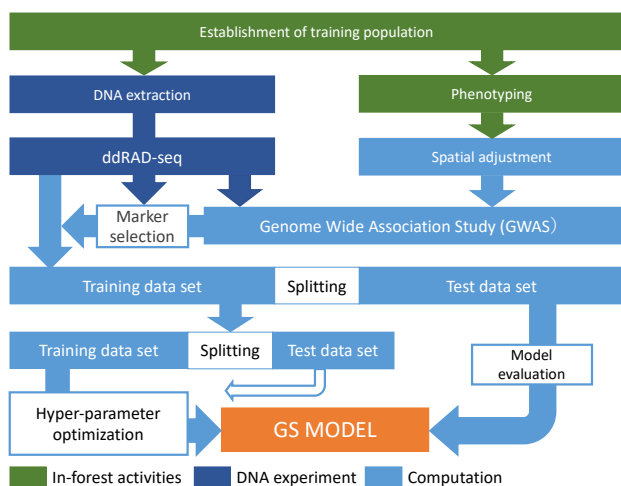


## Development of models to predict stem diameter and tree height of *Shorea platyclados* (Dipterocarpaceae) based on genomic information from seedlings

In recent years, rapid advancements in DNA sequencing technology have made it relatively inexpensive to collect genomic information even for woody plants that are close to wild species. However, for woody plants with large individual sizes and long lifespans, evaluating the performance of next generations (known as progeny trial) after breeding superior individuals requires vast land and time for their growth. This becomes a limiting factor in selecting superior individuals. Therefore, attention is now focused on methods that utilize genomic information to predict specific phenotypes. Specifically, if we can predict future phenotypes using genomic information from offspring, it becomes possible to overcome previous limiting factors. In this study, we aimed to develop a genomic selection model for *Shorea macrophylla*, a tree species in the Dipterocarpaceae, which is expected to be intensively planted in Southeast Asian tropical forests.

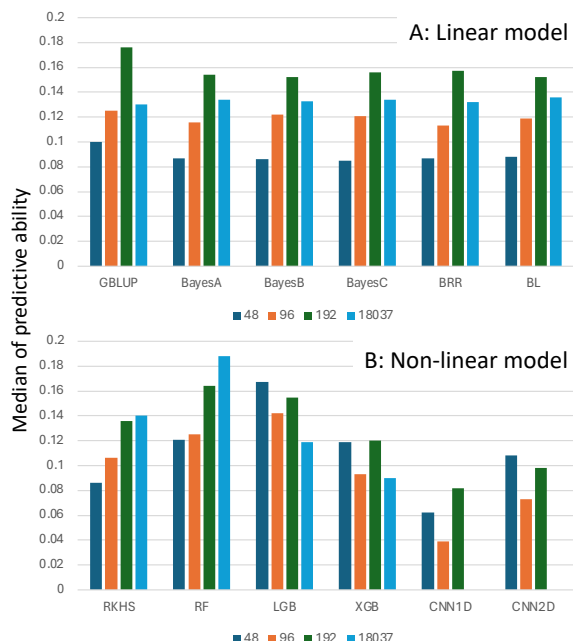
We established a workflow using a training population of a progeny trial forest with relatively less environmental heterogeneity for genomic selection (GS) model. We collected genomic information and focal phenotypes. Using these data, we corrected the spatial structure of the phenotypes due to environmental heterogeneity, detected important genetic markers by genome-wide association study (GWAS), and established a workflow for developing GS models (Fig. 1). The workflow was implemented using scripts in R and Python, which we have made publicly available. By selecting highly correlated markers based on GWAS results, we achieved usable predictive accuracy using both linear models (6 algorithms) and nonlinear models (6 algorithms). With this model, it becomes possible to select offspring based on genomic information from the training population without the subsequent progeny testing (Table 1). This approach promises to enhance breeding outcomes related to growth. Specifically, when predicting the 7-year diameter growth, we calculated the median of predictive accuracy for 100 randomly generated models using linear and nonlinear methods. Given that diameter growth results from a combination of primary and secondary growth, the use of nonlinear models, which can account for interactions among numerous genes, yielded higher predictive accuracy (Fig. 2). Conversely, tree height, being only influenced by primary growth, was associated with fewer relevant genes. Therefore, narrowing down the number of associated genes through GWAS led to a more accurate GS model. (Fig. 3)

Authors: Tani, N. [JIRCAS],  
Na'iem, M., Widiyatno, Indrioko, S., Sawitri [UGM], Akutsu, H., Tsumura, Y. [Univ. of Tsukuba]



**Fig. 1. Workflow for GS model**

This script, which includes the calculation process highlighted in light blue, has been publicly released.



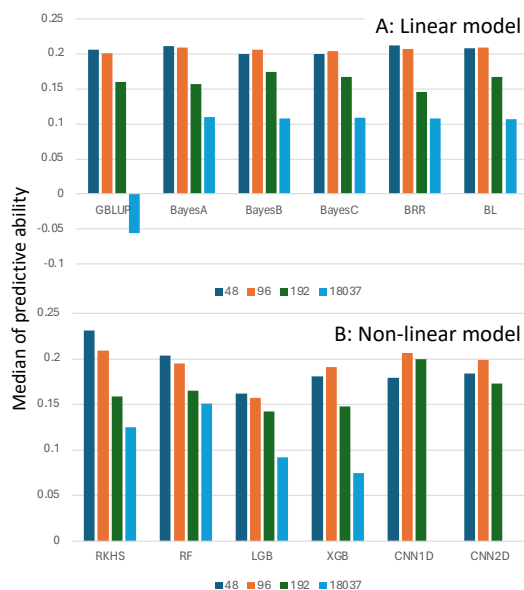
**Fig. 2. Median of predictive accuracy of GS models for diameter growth in the 7th year of planting**

By GWAS, 48, 96, and 192 DNA markers were selected from a total of 18037 DNA markers, Median was calculated from the 100 replicates for linear algorithms in A and non-linear algorithms in B. Due to the computational burden, CNN1D and CNN2d do not use all markers.

**Table 1. The maximum predictive ability for diameter growth and tree height**

Trait	Number of markers	Maximum predictive ability	Algorithm
D7	48	0.279	BRR
	96	0.342	CNN1D
	192	0.291	CNN1D
	18037	0.440	XGB
H7	48	0.343	XGB
	96	0.427	BayesB
	192	0.354	LGB
	18037	0.297	LGB

D7 and H7 represent diameter growth and tree height in the seventh year of planting, respectively. The maximum predictive accuracy was obtained from 10 trials of 10 training data (total of 100 trials) and using the algorithm when the 48, 96, and 192 markers were selected by GWAS from a total of 18037 markers.



**Fig. 3. Median of predictive accuracy of GS models for tree height in the 7th year of planting**

See footnote in Fig. 2.