

## 実生のゲノム情報からフタバガキの幹直径と樹高を予測するモデルの開発

東南アジア熱帯雨林の主要構成樹種であるフタバガキ科の成長には多数の遺伝子が複雑に関与している。既存の次代検定林をゲノム予測モデル開発のためのトレーニング集団とし、ゲノムワイド連関解析と線型・非線形のモデル開発用アルゴリズムを組み合わせたワークフローを適用することにより、実生のゲノム情報から幹の直径成長と樹高成長における複雑な遺伝様式を説明する最良のゲノム選抜モデルが得られる。

キーワード：フタバガキ科、ゲノム選抜モデル、ゲノムワイド連関解析、成長、炭素蓄積

### 背景・ねらい

近年、ゲノム解析技術の急速な発展に伴い、野生種に近い木本植物においても比較的安価にゲノム情報の収集が可能となっている。一方で、個体サイズが大きく長寿命である木本植物においては、優良個体を交配した後、次世代のパフォーマンスを評価（次代検定）するために広大な土地と成長に係る時間が必要であり、優良個体の選抜を律速する要因である。そこで、ゲノム情報を活用し、着目する表現型を予測する手法に注目が集まっている。すなわち、交配から得られた実生のゲノム情報を活用して将来の表現型を予測できれば、これまでの律速要因を解決することが可能となる。そのためには、対象とする表現型とゲノム情報が既知の林分を用いて、ゲノム情報から表現型を予測するモデル（ゲノム選抜モデル）が必要である。本研究では東南アジア熱帯林において今後集約的林業で植栽が期待されるフタバガキ科樹種の一つである *Shorea macrophylla* の成長に関するゲノム選抜モデルを作成する。

### 成果の内容・特徴

1. 比較的環境変化の少ない検定林をゲノム選抜モデル作成のためのトレーニング集団として設定し、ゲノム情報の収集、目的とする表現型の計測、環境偏差による空間構造補正をおこなった上で、重要な遺伝マーカーの検出のためのゲノムワイド連関解析、その結果を用いたゲノム選抜モデル作成のワークフローを確立する（図1）。
2. ゲノムワイド連関解析の結果をもとに連関度の高いマーカーを選抜し、線形モデル（6 アルゴリズム）と非線形モデル（6 アルゴリズム）の中から、最適なアルゴリズムを使用すると、使用に耐える予測精度が得られる（表1）。
3. 植栽7年目の幹直径成長について線形モデルを用いた7手法と非線形モデルを用いた7手法でランダムに100モデルを作成した時の予測精度の中間値を算出する。直径成長は一次成長と二次成長の複合的な結果であるため、関連遺伝子数が多く、そのため、多くの遺伝マーカーと遺伝子間の相互作用を説明できる非線

形モデルを用いた方が高い予測精度が得られる。（図2）。

4. 樹高は一次成長の結果であるため、関連する遺伝子数が少ないと考えられる。GWAS によって関連遺伝子数を絞り込んだ方が、予測精度の高いゲノム選抜モデルが得られる。（図3）。

### 成果の活用面・留意点

1. 本研究において開発した R 言語及び Python で記述したスクリプトを公開した。これにより多くの手法を用いたゲノム選抜モデルの作成への応用が期待できる。
2. 本成果を基にすることで、次代検定を行うことなく、トレーニング集団から種子を採取し、ゲノム情報を元にその実生を選抜することで、成長に関する育種効果が期待できる。
3. 本モデルは遺伝的組成が異なる集団には適応できない。したがってトレーニング集団か、遺伝的組成の近い集団からの実生を選抜に使用する必要がある。また、同一アルゴリズムでも予測精度にばらつきがあることから、アンサンブル学習を行うなど、予測精度に留意する必要がある。

### その他

予算区分：交付金プロ [A4 環境適応型林業]、外部資金 [SATREPS]

研究実施期間：2021～2023 年度

研究担当者：谷尚樹（林業領域）、Na'iem, M., Widiyatno, Indrioko, S., Sawitri（ガジャマダ大）、阿久津春人、津村義彦（筑波大）

発表論文等：Akutsu et al. (2023) *Frontiers in Plant Science* 14:1241908. <https://doi.org/10.3389/fpls.2023.1241908>

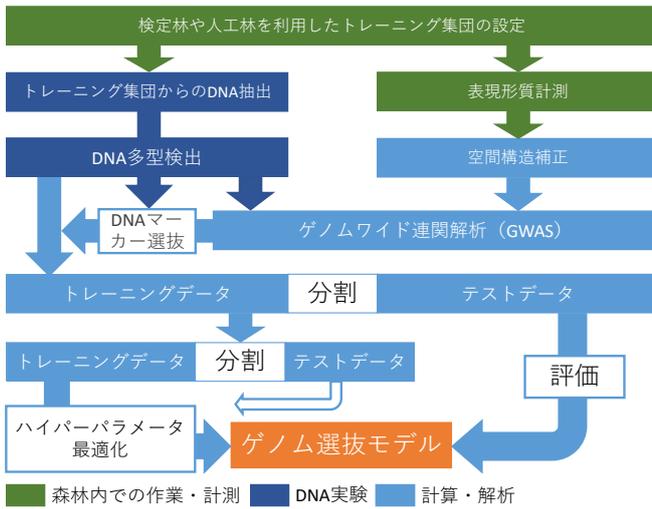


図1 ゲノム選抜モデルを作成するためのワークフロー

水色はコンピュータ上での解析のコンピュタスクリプトを示し、公開した。

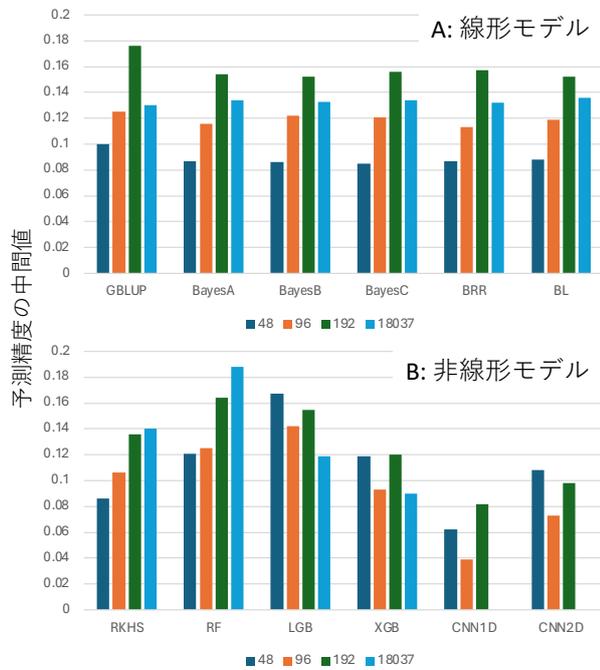


図2 植栽7年目の幹直径成長のゲノム選抜モデルの予測精度の中間値

ゲノムワイド連関解析によって、幹直径成長に関連度の高いDNAマーカーを48個、96個、192個選抜し（18037個は全DNAマーカー）、トレーニングとテストデータのペアをランダムに10個作成した後、それぞれのペアについて独立に10回計算を試行し、合計で100回の計算を行い、得られた予測精度の中間値を得た。AのGBLUP、BayesA、BayesB、BayesC、BRR、BLは線形モデルを用いたゲノム選抜モデルのアルゴリズムを表す。BのRKHS、RF、LGB、XGB、CNN1D、CNN2Dは主に機械学習を用いた非線形のゲノム選抜モデルの手法を表す。計算負荷によりCNN1DとCNN2dでは全マーカーを用いたモデル作成は行っていない。

表1 幹直径成長及び樹高で得られたゲノム選抜モデルの最大予測精度

形質	マーカー数	最大予測精度	アルゴリズム
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は植栽7年目の幹直径成長、H7は植栽7年目の樹高を表す。マーカー数はゲノムワイド連関解析で得られた結果をもとに対象形質に関連度の高い準に48、96、192個のマーカーを選抜し（18037個は全マーカー）、得られたデータセットを用いて、10トレーニングデータを10回試行した（合計100試行）際の最大の予測精度とその値を算出したアルゴリズムを示す。

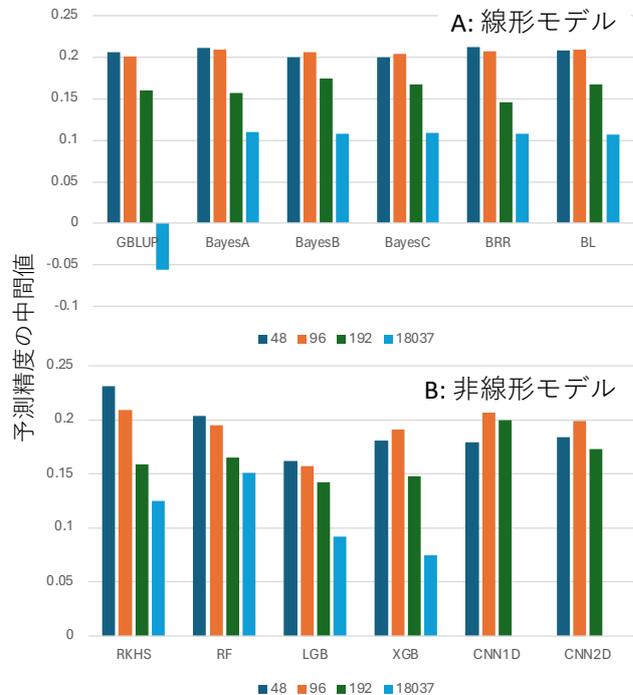


図3 植栽7年目の樹高のゲノム選抜モデルの予測精度の中間値

ゲノムワイド連関解析によって、樹高に関連度の高いDNAマーカーを48個、96個、192個選抜し（18037個は全DNAマーカー）、トレーニングとテストデータのペアをランダムに10個作成した後、それぞれのペアについて独立に10回計算を試行し、合計で100回の計算を行い、得られた予測精度の中間値を得た。AとBのゲノム選抜モデル作成手法については図2と同様である。

図表は Akutsu et al. (2023) © The Author(s) 2023 より転載/改変して作成