

## 6 ゲノム予測技術を用いたサガンスギの無花粉化に関する研究

将来を見据えた新たな試験研究推進事業（開発事業）

（県単：R5（2023）～R7（2025））

江島 淳

### 概要

近年、無花粉遺伝子を有するスギ品種が極まれに存在することが分かり、それらの無花粉遺伝子を計画的な人工交配により、品種開発に利用する動きがスギ育種の分野で活発化してきている。無花粉遺伝子は DNA 上の 1 箇所の変異で決まる劣勢遺伝子で、人工交配により劣勢ホモにすれば無花粉化できることが実証されており、DNA 分析により容易に同定が可能な質的な形質であり、マーカーアシスト選抜（MAS；Marker Assisted Selection）と呼ばれる（図-1）。図-1において、無花粉遺伝子は黒色の塗り潰しの横長の長方形で表現した。

一方、サガンスギなどが有する優れた成長や材質などの形質は、多数の遺伝子に影響を受けていると考えられており、このような形質に関与する DNA 上の領域を量的形質遺伝子座（QTL；Quantitative Trait Locus）といい、DNA 情報からの予測が難しいとされてきた。しかし、2010 年代に登場した次世代シーケンサーと呼ばれる DNA の高密度な判読技術により、約 10,000 箇所の DNA 上の変異を検出し、それらの多数の箇所の DNA 上の変異とこれまでの長年の研究で蓄積した材質などの形質情報を関連付け、DNA 情報だけから形質を予測するモデル（ゲノム予測モデル）が構築可能となりつつある。現地調査から得られるクローン（品種）の形質の評価値は clonal value（遺伝子型値；genotypic value ともいう）と呼ばれ、ゲノム予測モデルから予測される形質の評価値をゲノム予測値（GEV；Genomic Estimated Value）と定義し、GEV をもとに個体選抜することをゲノム選抜（GS；Genomic Selection）と呼ぶ（図-1）。

佐賀県林業試験場では、2019 年より筑波大学の津村義彦教授と森林総合研究所（国）の協力を得ながら、第一世代精英樹とサガンスギを含む第二世代精英樹の形質情報と次世代シーケンサーにより得られた約 10,000 箇所の DNA 上の変異を関連付けたゲノム予測モデルの構築に向けた研究を行い、研究成果を英語論文（Ejima et al. 2023）および博士論文（江島 2024）としてまとめた。その有用性を育種の現場で確認する段階に来ている。

既に、人工交配により無花粉遺伝子を有する品種とサガンスギの実生は誕生しているため、MAS（マーカーアシスト選抜）により無花粉遺伝子が次世代に受け継がれた個体（1/2 の確率）を選抜し、その個体に GS（ゲノム選抜）を適用す

ることで、無花粉遺伝子をヘテロで有し成長や材質に優れた個体（無花粉ヘテロのサガンスギ）を実生段階で選抜する計画である（図-1のGS1部分）。

将来的には、本課題により選抜する「無花粉ヘテロのサガンスギ」同士を交配することで、「無花粉ホモのサガンスギ」、つまり、成長と材質はサガンスギと同等で、全く花粉を出さないスギを作出する（図-1のGS2部分）。

本報告では、博士論文（江島 2024）の摘要を掲載するとともに、2024年1月19日に筑波大学での博士論文公聴会で使用したスライドの一部を用いながらゲノム予測技術の概要について報告する。

### 博士（農学）学位論文（江島 2024） 摘要

寿命が長く、個体サイズが大きな林木の育種は挑戦的だが、DNA解析技術の発展に伴い、個体あたり数万個の一塩基多型（SNP）を迅速にジェノタイピングできるようになったことで、DNA情報から優良個体の遺伝的な能力を予測すること（以下、ゲノム予測：Genomic prediction）が可能となり、育種の効率化が進むことが期待されている（発表スライドP2）。ゲノム予測の基本原理は、SNPと樹高などの表現型を持つ個体群（トレーニング集団）を構成し、全てのSNPを含む回帰モデルをあてはめる。各SNPの表現型への寄与は回帰係数として推定され、それらを含む予測式が得られる。続いて、この式を予測群（SNP情報を持つが表現型を持たない個体群）に当てはめ、各SNPの効果の総量として各個体の遺伝的能力である遺伝子型値（genotypic value またはクローン化された個体を対象とする場合 clonal value ともいう）を予測する。そのDNAの情報のみからの予測値（GEV；Genomic Estimated Value）から優良個体を苗木の段階で選抜することができるため、1世代の更新に数十年を要していた林木育種の時間的な問題を劇的に改善する可能性がある。

本研究の調査地である佐賀県では1960年代から人工交配により育成・選抜してきたスギの第2世代精英樹101クローンと、その親世代である第1世代精英樹47クローンからなる2世代の選抜集団を構築している（発表スライドP3）。調査地の九州では、約500年前から挿し木林業が継続して続けられており、そこで活用される挿し木品種は、挿し木により増殖され複数の検定林で検定されるため、信頼度の高い樹高などの表現型データが得られる。さらに、人工交配への取り組み時期が早かったため、第2世代精英樹も既に30年生を超えており、林木育種の課題である長期性を乗り越え、世界でもゲノム予測の適用例が少ない、選抜された2世代の集団の表現型データが揃っている。本研究では、これらの表現型データを活用し、第3世代以降の育種を加速させるために、ゲノム予測の有用性について検証した。

まず、長年管理してきた育種集団の遺伝的な背景を整理するため、マイクロ

サテライトマーカー（以下、SSR マーカー；Simple Sequence Repeats）を用いて、家系構造の再構築を行った（発表スライドP4）。第2世代と親世代の第1世代の間で親子鑑定をしたところ、第2世代の約100クローンの内、25のクローンで人工交配記録上の親と実親が異なる事が判明した。しかし、親子解析により誤りがあった第2世代の19クローンの親を第1世代精英樹の中から確定することができた。長期間を要する林木育種においての、人工交配ミスについての調査例はなく、SSR マーカーの活用によりその実態を明らかにするとともに、育種集団の遺伝的な背景を整理することができた（発表スライドP5）。

次に、ゲノム予測の対象とする3つの形質について、クローン検定林での調査結果をもとに各クローンの遺伝的能力の代表値である clonal value を算出した。樹木は固着性であり植栽された環境の影響を強く受けて長期間成長するため、遺伝的な能力である clonal value を正確に測るには、表現型に含まれる環境の効果を考慮する必要がある。樹高形質では、林地の不均一性を考慮するため空間統計学を応用して、マイクロ環境である微地形の影響を取り除いた後に、マクロ環境である検定林間の違いを反映して BLUP（Best Linear Unbiased Prediction）法（線形混合モデル）により clonal value を求めた（発表スライドP6）。2つ目の木材剛性については、30年生を超える個体を伐倒し製材品の状態にしてから、生育環境の影響を強く受ける年輪幅と木材科学分野ではよく知られている未成熟材部／成熟材部の区分を考慮して、BLUP 法により各クローンの clonal value を求めた（発表スライドP7）。3つ目の雄花量は、年変動が大きいことが知られているため、複数年の調査データを元に、測定年度による違いを考慮して BLUP 法により各クローンの clonal value を求めた（発表スライドP8）。

最終目標のゲノム予測では、SSR マーカーで整理した148クローンを対象に、DNA 情報は dd-RAD-seq（Double digest restriction site-associated DNA sequencing）法により得られた8,664個の SNP を利用し、形質データは上記3形質（樹高・木材剛性・雄花量）の clonal value を用いた。形質と各 SNP の関連は、5つのベイズモデルを用いてモデル化され、その予測精度を交差検証によって評価したところ、形質ごとに最適なモデルが異なることが明らかになった（樹高は BRR モデル、木材剛性は BayesA モデル、雄花量は BayesB モデル）（発表スライドP9）。また、ゲノム予測は従来の血統に基づく方法よりも1.2～16.0倍正確であり、これはメンデルアンサンプリングをモデル化する能力に起因すると考えられた（発表スライドP10）。一般的にはトレーニング集団と予測集団の遺伝的な関係は、血縁関係がある方が予測に有利に働くとの報告が多いが、世代間の血縁関係の影響を分析した結果、選抜圧の影響を強く受ける形質では、親子関係がゲノム予測の精度に悪影響を及ぼした。まとめる

と、ゲノム予測は対象とする集団の遺伝的背景と形質の特性を考慮して、予測結果を検証し解釈しながら運用すれば、林木育種に有望なツールになることを示した (図-1)。

#### 引用文献

Ejima A, Uchiyama K, Mori H, Tsumura Y (2023) Genomic Prediction of Tree Height, Wood Stiffness, and Male Flower Quantity Traits across Two Generations in Selected Individuals of *Cryptomeria japonica* D. Don (Japanese Cedar). *Forests* 14: 2097

江島 淳 (2024) スギ交雑育種における DNA 情報を活用したゲノム予測に関する研究. 筑波大学大学院 生命環境科学研究科 生物圏資源科学専攻 博士 (農学) 学位論文

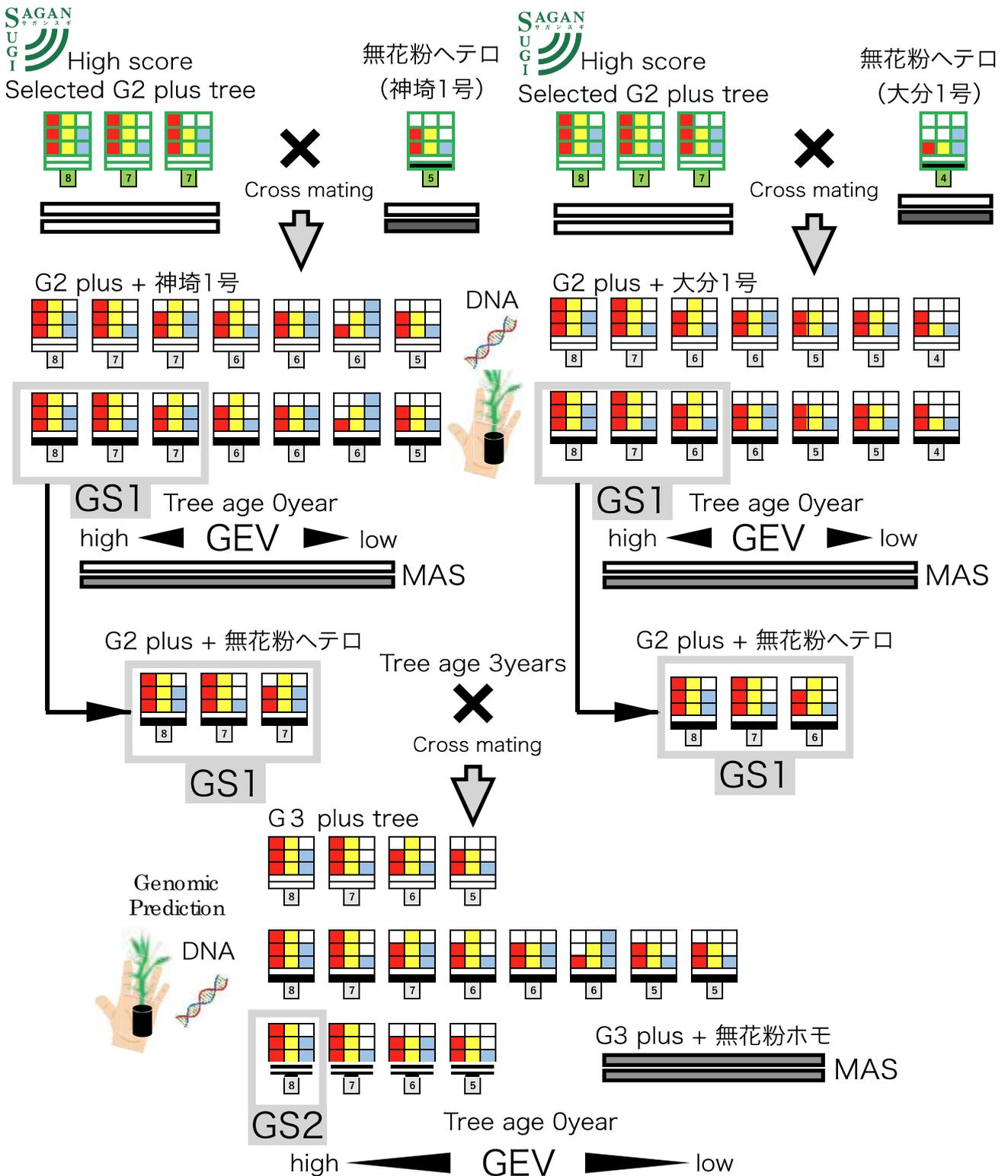
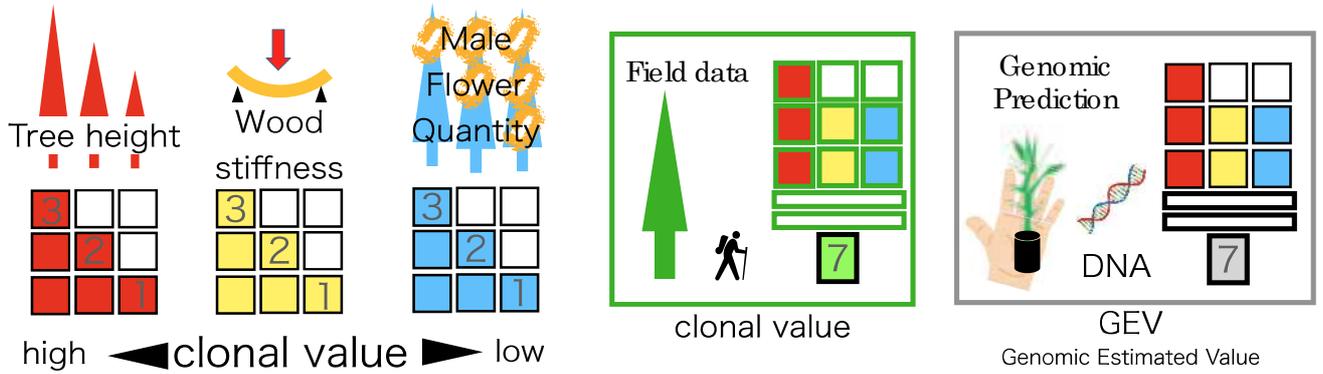


図-1. Genomic Selection (GS) と Marker Assisted Selection (MAS) による無花粉サガンスギの開発スキーム。江島 (2024) を一部改変

# スギ交雑育種におけるDNA情報を活用したゲノム予測に関する研究

## 目次

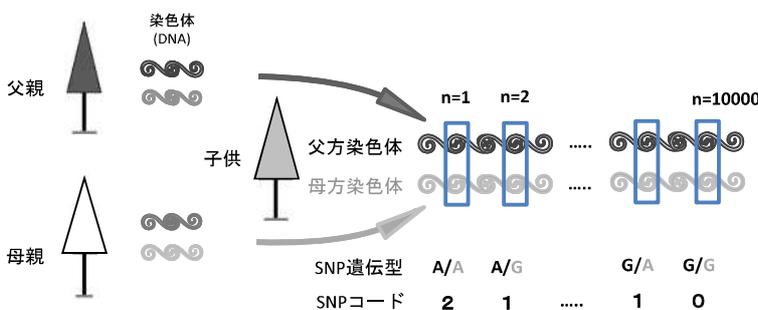
- STEP1 DNAによる育種集団の家系情報の整理**  
 佐賀県のスギ交雑育種の経緯（交配から選抜まで約60年）  
 SSR解析による育種集団の家系情報の再構築
  - 1-1 クローン鑑定
  - 2-1 親子鑑定
- STEP2 各クローン(品種)の形質情報の整理**
  - 2-1 空間情報を活用した樹高のclonal value (評価値)の算出
  - 2-2 製材品を用いた木材剛性のclonal value (評価値)の算出
  - 2-3 雄花着花量のclonal value (評価値)の算出
- STEP3 ゲノム予測(Genomic Prediction)**  
 選抜された2世代の育種集団を対象としたゲノム予測モデルの構築と精度評価
- STEP4 ゲノム予測の活用**  
 無花粉サガンスギ(早い成長・強い材質・無花粉)の開発スキーム

※STEP4は、本業務報告では、図-1(前ページ)に示す。

発表スライド 1

## ゲノム予測 Genomic prediction

多数個体の表現型（樹高など）とその個体の遺伝子型（DNA情報）を関連づけることで、DNA情報のみから表現型を予測すること。



### 理論1：極微小モデル (Fisher 1918)

多数の遺伝子による量的形質（樹高など）の支配。

### 理論2：ゲノム予測 (Meuwissen et al.2001)

大規模な遺伝情報から極微小モデル (Fisher 1918) に基づき表現型とDNA型を対応させる数理モデル。

### 技術：次世代シーケンサー (2005～)

1個体から数1,000~数10,000の遺伝的変異 (SNP)を検出できるようになった。

※SNP(一塩基多型)

多数個体を比較した際、DNAの配列の中で1塩基だけ、違いがある部分。それ以外の場所は、全個体ほぼ同じ。

### トレーニング集団

	SNPコード				表現型値 (樹高など)
	X1	X2	X3	Xn	
	0	2	2	..... 1	6.2
	1	0	2	..... 0	9.5
	2	1	1	..... 2	11.5
	1	0	1	..... 2	13.5

### 回帰モデル

$$Y = b_0 + b_1 X_1 + \dots + b_n X_n + e$$

### 回帰係数 (各SNPの微小な効果)

$$b_0 = 6.5, b_1 = 0.21, \dots, b_n = 0.02$$

### 予測式 (Genomic prediction)

$$Y = 6.5 + 0.21 X_1 + \dots + 0.02 X_n$$

### 予測

DNA	SNPコード				ゲノミック 予測値
	X1	X2	X3	Xn	
	0	2	2	..... 0	4.5
	1	2	1	..... 2	14.8

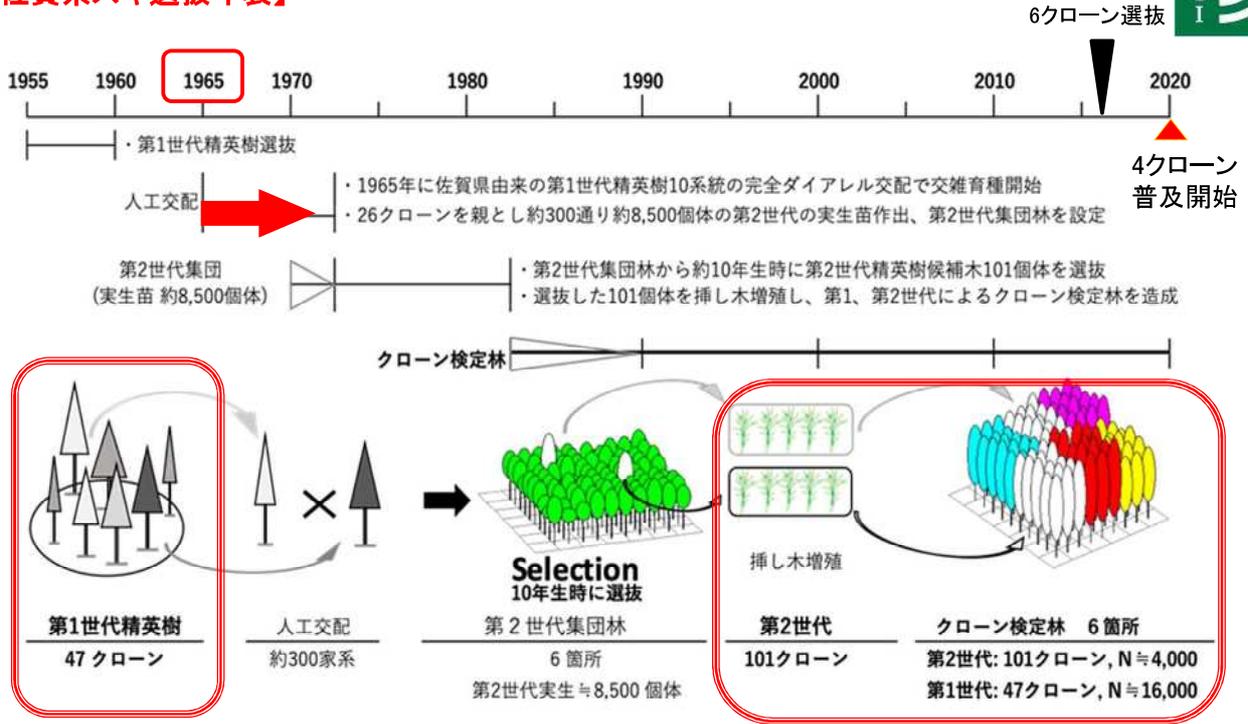
### 予測式 (Genomic prediction)

$$Y = 6.5 + 0.21 X_1 + \dots + 0.02 X_n$$

発表スライド 2

# STEP1 佐賀県の育種経緯

## 【佐賀県スギ選抜年表】



### 2016年 6クローンを選抜

①成長, ②材の強度, ③雄花量, ④挿し木発根率を評価

2021年8月 B-74を品種登録  
初期成長・材の強度を詳細評価  
2022年2月 4クローンを普及開始



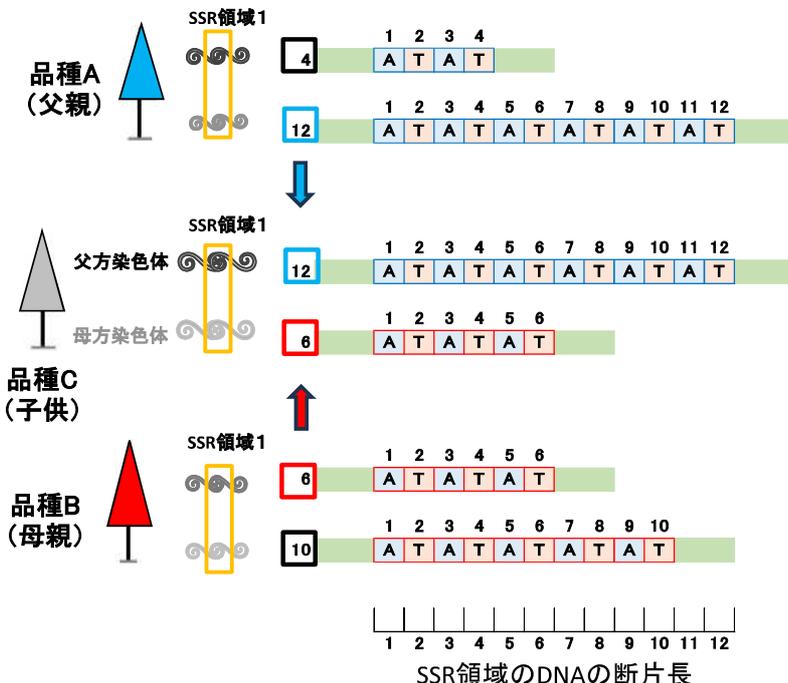
# STEP1 SSR解析

## 家系情報の整理

SSR (Simple Sequence Repeats : 単純反復配列) マーカー

### 原理

DNA上の個体間の変異が起こりやすい(多型)、単純反復箇所(SSR領域)を特定のプライマーを使用することで増幅し(PCR)、反復数の違いをDNAの断片の長さとして計測する。

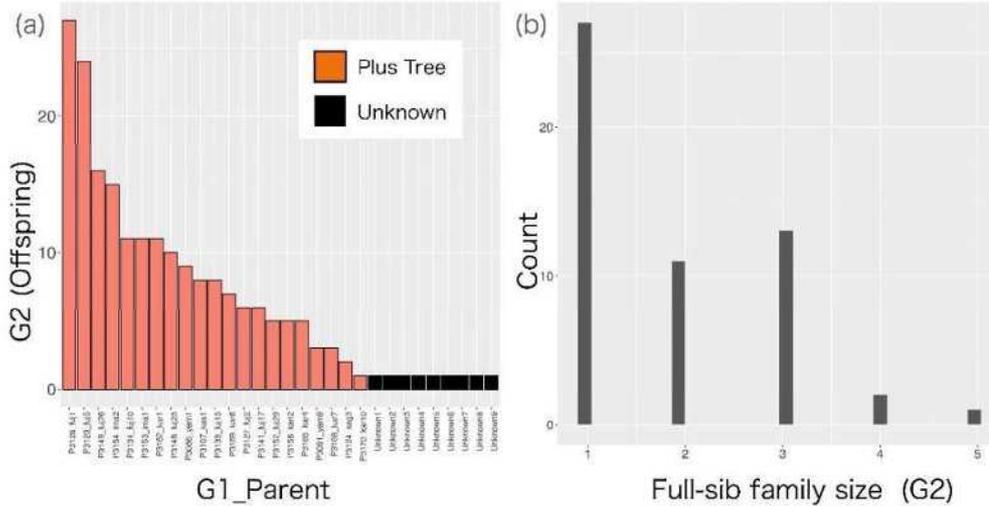


### 特徴

多型性が高く少ないマーカー数でクローン識別(挿し木品種の同定)ができる。  
左の例では、1つのマーカーで3個体から4つ(4,6,10,12)の型を識別。

共有性(父方由来と母方由来の区別)であり、親子鑑定に使用できる。

大量のSNPの取得(次世代シーケンサー)と比較し、実験と解析が簡単で安価。育種現場での必需技術。



左図 各第1世代精英樹の第2世代への寄与

右図 第2世代精英樹の全兄弟のサイズと家系数

第1世代の47クローン中、21クローンが第2世代に子孫を残した親。  
最も多くの子孫を残した第1世代クローンは28の第2世代クローンと親子関係。  
(平均=9.19クローン)。

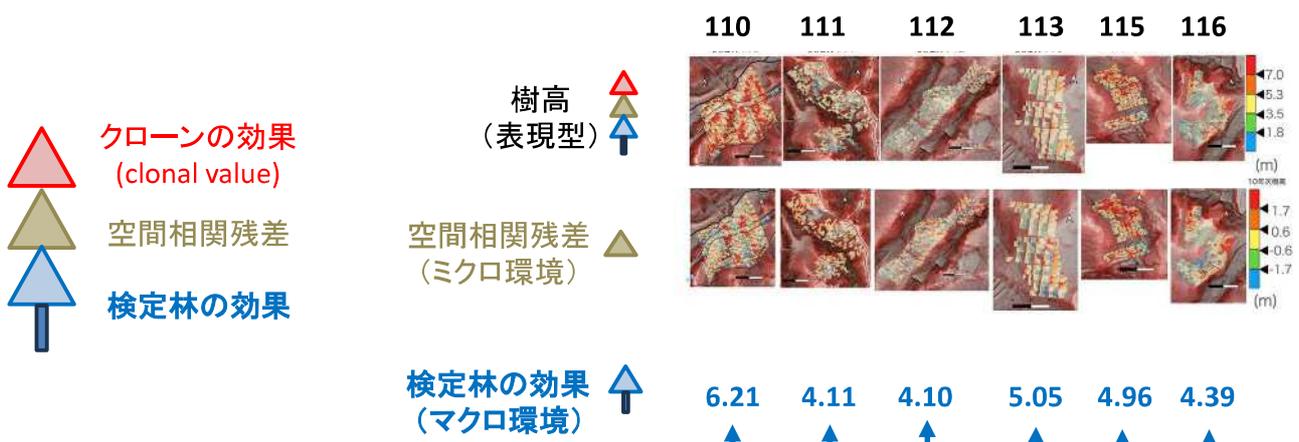
第2世代101クローンの家族構成では、最も頻度の高い全兄弟数は3クローン(13の全兄弟家系)で、全兄弟を持たないクローンは28クローンであった。

STEP2-1 全ての個体 (N = 16,444) の調整済樹高でclonal valueの算出

10年次樹高



樹高が、以下の3つの要素から構成されていると仮定する。

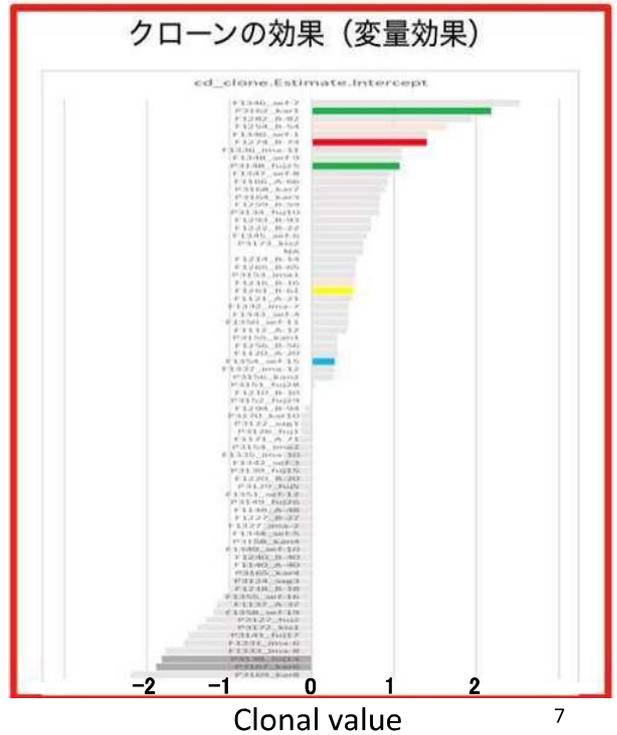
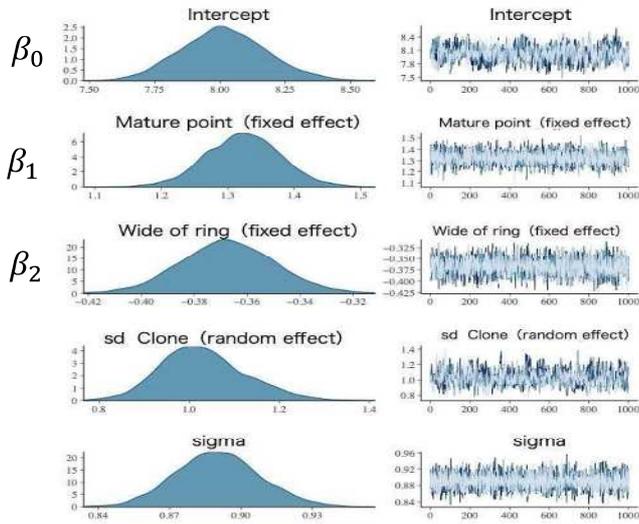


STEP2  
6検定林すべての個体の調整済樹高(樹高 $\uparrow$  - 空間相関残差 $\triangle$ )で clonal value $\triangle$ と検定林の効果 $\uparrow$ を算出。

調整済樹高  $clone_i = \beta_0 + \beta_1 \cdot test\ site + clone + e_{clone,i}$  個体  $i = 1, 2, \dots, n$

動的ヤング係数  $clone_{i} = \beta_0 + \beta_1 \cdot juvenile/mature + \beta_2 \cdot ringwide + clone + e_{clone,i}$   
 個体  $i = 1, 2, \dots, n$

【解析結果：MCMCサンプリング】



固定効果(全クローンに共通)の結果  
 未成熟材部／成熟材部の区分: 成熟材部の効果は平均値が1.32  
 95%確信区間は1.21～1.43  
 年輪幅の影響: mm単位あたり平均値が-0.37  
 95%確信区間は-0.40～-0.33

雄花着花指数  $clone_{i} = \beta_0 + \beta_1 \cdot year + clone + e_{clone,i}$  個体  $i = 1, 2, \dots, n$

【解析結果：MCMCサンプリング】

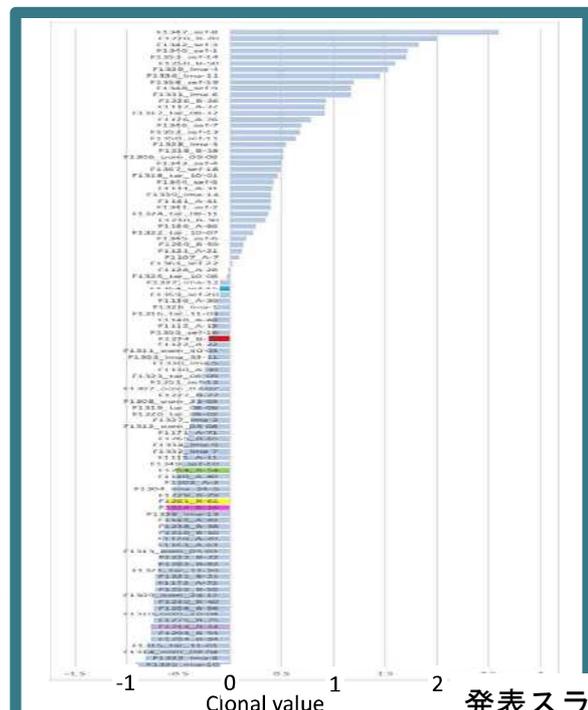
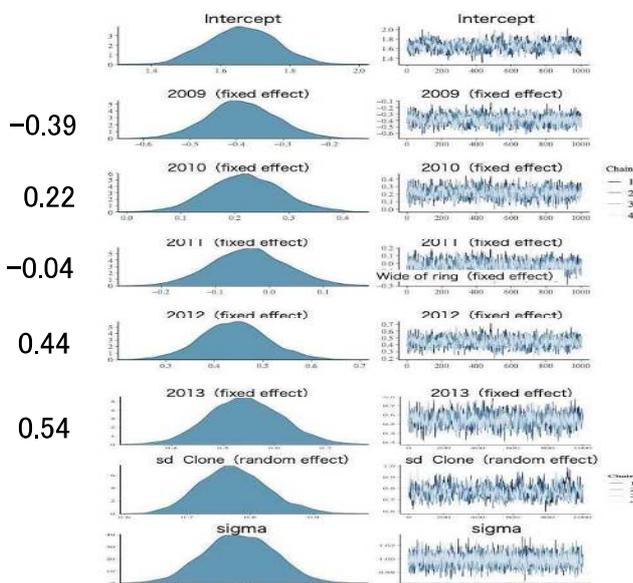
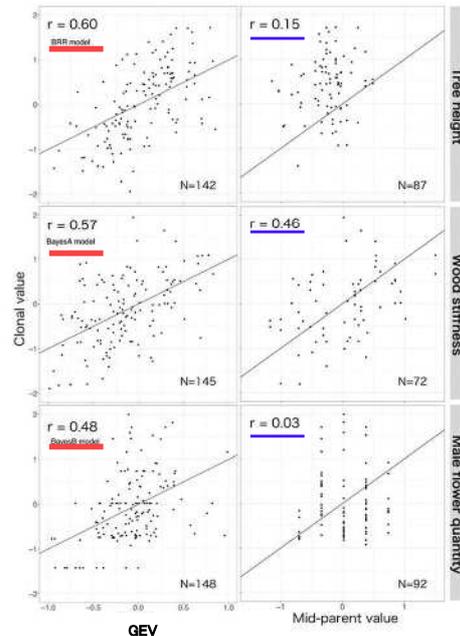


表. 5つのGPモデルによる予測精度と家系情報による予測(Mid-parent value)の精度比較。

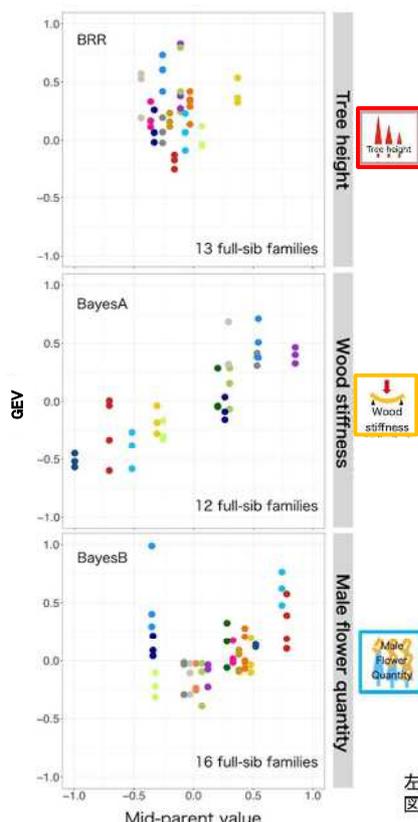
Trait	Model	N	r	(SE)
 Tree height	Genomic prediction			
	BayesA	142	0.57	(0.03)
	BayesB	142	0.59	(0.04)
	BayesC	142	0.52	(0.05)
	BL	142	0.54	(0.04)
	BRR	142	<b>0.60</b>	<b>(0.05)</b>
	Mid-parent value	87	0.15	
 Wood stiffness	Genomic prediction			
	BayesA	145	0.57	(0.01)
	BayesB	145	0.56	(0.03)
	BayesC	145	0.52	(0.07)
	BL	145	0.51	(0.05)
	BRR	145	0.50	(0.05)
	Mid-parent value	72	0.46	
 Male flower quantity	Genomic prediction			
	BayesA	148	0.46	(0.06)
	BayesB	148	0.48	(0.06)
	BayesC	148	0.46	(0.05)
	BL	148	0.45	(0.05)
	BRR	148	0.45	(0.05)
	Mid-parent value	92	0.03	



最も高い予測精度を示したモデルは各形質で異なっていた。  
 樹高はBRR( $r = 0.60$ ) : 小さい多数のQTLが関与  
 木材剛性はBayesA( $r = 0.57$ ) : 適度の数のQTLが相加的に関与  
 雄花量はBayesB( $r = 0.48$ ) : 少数のQTLが関与

図. 左図はGEVとclonal valueの関係、右図は中間親値とclonal valueの関係。各形質でGEVを算出したモデルは、左の表で最も予測精度の高いモデルを採用し、図化した。図中の黒線は完全一致である $Y = X$ 。

GEV(Genomic estimated values)はすべての形質において、中間親値を用いた血縁ベースのアプローチよりも高い予測精度を達成した。



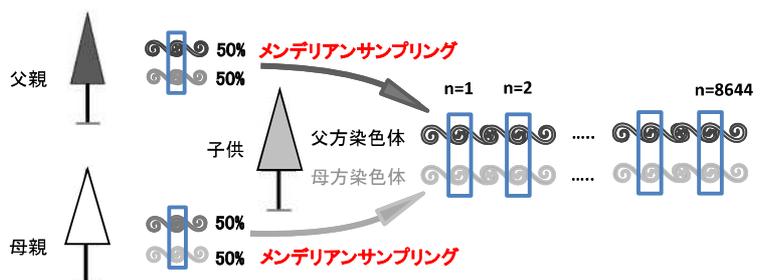
### 【結果】

従来の血縁に基づく方法では、両親のclonal valueの平均値であるため、全ての兄弟について単一の予測値が得られるのに対し、ゲノムマーカーに基づくGEVでは、全兄弟内の各兄弟においてそれぞれ固有の予測値が生成される。

### 【考察】

GPが**メンデルianサンプリング効果**をモデル化している。

GPがメンデルianサンプリング効果を正しく捉えることができれば、育種プログラムの成果を大幅に向上させ、次世代の選抜に必要な時間を短縮することができる。



左図. GEVとMid-parent value(中間親値)の関係。上段は樹高、中段は木材剛性、下段は雄花量。図中の同じ色の点は、全兄弟を示している。