

# 研究報告書

## 「遺伝子情報に基づく表現型予測モデルの構築とコンピューターシミュレーション育種への応用」

研究タイプ： 通常型

研究期間： 2016 年 10 月～2020 年 3 月

研究者： 山本 英司

### 1. 研究のねらい

気候変動などの環境変化や人口増加による食料不足、あるいは社会の発展に伴う消費者ニーズの多様化など、日本を含む世界が抱える食料問題の解決には栽培技術の向上だけでなく、植物そのもののパフォーマンスを上げること(新品種開発)が重要である。実際に品種を開発する上では、収量性や品質性、栽培地域や作型に代表される環境適応性など、様々な要素を高水準で兼ね備える必要がある。これらは各々が多数の遺伝子に制御される複雑形質であるだけでなく、互いに相互作用も示すため、これらを迅速かつ効率的に同時改良することは現在の育種法では困難である。申請者らはこのような問題を解決するために、ゲノム情報に基づく表現型予測モデル(GS モデル)とコンピューターシミュレーションを利用した育種計画設計を提案し、トマトの果実糖度と収量の同時改良を例とした研究を行っている(Yamamoto et al. 2016)。しかし、このような手法をさらに多数の形質に応用するには、連鎖崩壊に伴う表現型予測精度の低下や形質間相互作用(遺伝子座の多面発現性)の説明に対する曖昧さなど、現在用いられている GS モデルの問題点を克服する必要がある。

本研究では、多数の形質の迅速かつ効率的な改良を実現する基盤として、全ゲノム配列解読によって解析集団に含まれる遺伝子アリルを検出、この情報をもとに選抜されたマーカー群を説明変数とした GS モデルを構築する。さらには、その GS モデルとコンピューターシミュレーションによって多数形質を同時改良する効率的な育種計画を設計し、その実証実験を開始する。本研究で得られる一連の解析技術および知見は他の作物種にも応用可能であり、将来的には幅広い作物種の品種開発を通して、日本を含む世界の食料問題解決に貢献することを目指す。

### 2. 研究成果

#### (1) 概要

日本国内の大玉トマト F1 品種群 (*Solanum lycopersicum*, 99 品種) の全ゲノム配列を解析し、ゲノムワイド関連解析 (GWAS) によって農業形質に関連する多数の遺伝子座を検出した。GWAS によって検出された有意シグナルから GS モデル構築に用いるマーカーを選抜し GS モデルを構築した。従来のゲノムワイドなマーカーを使った GS モデルとの間で予測精度の比較を行った結果、選抜マーカーを用いた GS モデルが実際の育種過程の集団においても高い精度を維持していることを示した。トマトのように表現型調査に多大な労力を要する作物においては、規模の小さいデータであっても精度の高い GS モデルを構築する必要がある。そこで、個体別の環境情報を取得するセンシングシステムを開発し、このシステムから得られる環境情報を用いた GS モデルの精度向上も行った。

## (2) 詳細

### ① 全ゲノム配列解析および GWAS

トマト品種群(96 品種)のゲノム配列解読には Illumina HiSeq X を用いた。リード数の平均深度は 40.6x (26.5~54.8x) であり、これはヘテロ遺伝子型のコールに推奨されている深度 (21x) を上回っている。またトマト参照配列に対するカバレッジも平均 99.4% (96.6~99.8%) と高い値を示した。トマト参照配列へのマッピングおよびバリエーションコールを行った結果、1,714,257 ヶ所の変異サイトを検出した。果実糖度などの農業関連形質に対して GWAS を行ったところ、各形質に対して多数の有意シグナルを検出した(図1)。一方、検出されたシグナルの中で既報の遺伝子に該当するものはなく、統計値と変異情報のみを用いた原因遺伝子の特定は困難であった(図2)。例えば、果実糖度に対して検出された第9染色体上の有意シグナルは、トマト果実糖度の向上に寄与することが知られている *Lycopersicon Invertase5* (*LIN5*) の近傍に座乗していたが、今回用いた品種群において *LIN5* の塩基配列に変異は認められず、有意シグナルも *Lin5* より染色体上流側に位置していた(図2)。

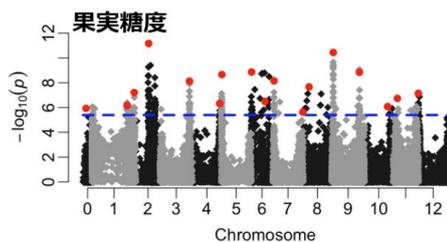


図1. 国内大玉トマトF1品種の全ゲノム配列を用いたゲノムワイド関連解析 (GWAS) . 破線は5%有意水準を示す. ●で示されたプロットは、以降のGSモデル構築に用いられたマーカーを示す.

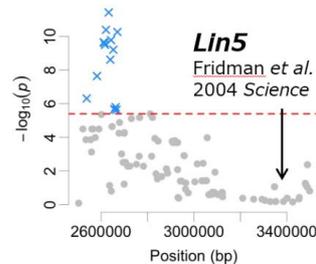


図2. 果実糖度のGWASにおいて第9染色体上に検出されたピーク. 本研究で検出されたピークは、他の研究でよく検出される *Lin5* 遺伝子とは異なる場所に位置している.

### ② 表現型予測モデルの構築とその精度評価

高い予測精度を示す GS モデルの構築には、表現型と関連するマーカーがより多く含まれること、表現型とは無関係のマーカーができるだけ排除されることが望ましい。①で実施した GWAS の結果、果実糖度に対しては 17 ヶ所、収量に対しては 12 ヶ所の遺伝子座を検出した(図1)。もしこれら GWAS の結果が妥当であるならば、検出された遺伝子座上のマーカーを用いた GS モデルは、ゲノムワイドなマーカー(本研究では全ゲノム配列)を用いた GS モデルよりも予測精度が高いはずである。そこで、①の GWAS の結果に基づいて GS モデルのためのマーカー選抜を行い(図1)、その表現型予測精度の検証を行った。予測精度の検証には、モデル構築に用いた 96 品種とは異なる 49 品種の大玉トマト F1 品種を用いた。予測モデルトレーニング用の表現型は 2011~2014 年にかけて、テスト用表現型は 2018 および 2019 年に取得した。比較を行ったほとんどの形質において、選抜マーカーに基づくモデルが従来の全ゲノムに基づくモデルの予測精度を上回った(図3)。この結果は、全ゲノム配列に基づく GWAS が診断バイアスに由来する偽陰性を回避することで適切なマーカー選抜を可能とし、これによって GS 予測精度を向上させられることを示している。なお、2018 年の収量に対する GS 予測精度は 0 となっているが、これは同年に発生した異常気象の影響と考

えられる。

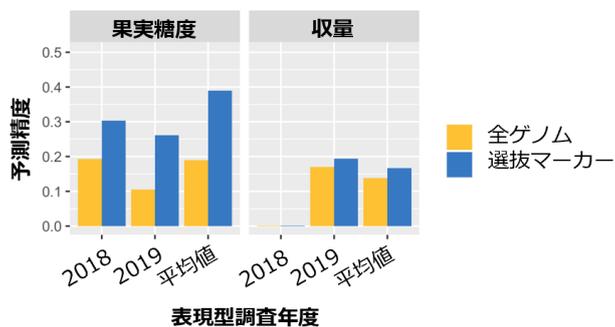


図3. 全ゲノム配列を用いたGSモデル（全ゲノム）とGWASにより選抜されたマーカーを用いたGSモデル（選抜マーカー）の予測精度比較. GSモデルは大玉トマト $F_1$ 品種96点のデータを用いて構築した. 予測精度はGSモデル構築用とは異なる大玉トマト品種49点のデータを用いて評価した. 予測精度は予測値と実測値との相関係数. 平均値は2018年度と2019年度の実測値の平均値.

### ③ 育種過程におけるGSモデルの利用

一般に、交雑回数の増加に伴ってGSモデルの予測精度は低下するが、これは複数世代先までの育種効果を予測するシミュレーション育種設計では深刻な問題となる(図4a)。この問題は、ランダムに選ばれたマーカーではなく、表現型のバリエーションの原因となる変異、あるいはそれらと強い連鎖関係にある変異をGSモデル構築に用いることで理論上回避できる(図4a)。本研究で選抜されたマーカーに基づくGSモデルの予測精度を、交雑の進んだ育種集団を用いて評価したところ、従来のランダムに選ばれたゲノムワイドマーカーを用いたGSモデルと比較して、交雑の進んだ世代でも高い予測精度を維持していた。これにより、本研究で用いた手法が複数世代の連続交雑を想定したシミュレーションベースの育種設計に適用可能であることを示した。

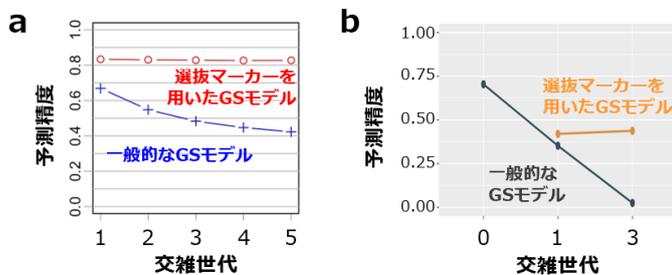


図4. GSモデルの予測精度と育種交雑世代との関係.  
 a. GSモデル予測精度と交雑世代数との関係のシミュレーション.  
 b. 果実糖度に対するGSモデル予測精度の比較. 交雑世代0はマーカー選抜そのものを実施した集団であるため、選抜マーカーを用いたGSモデルの予測精度評価は実施していない。

### ④ 個体別環境センシングによる形質評価およびGSモデル予測精度の向上

本研究でも示されたように、育種集団ごとに含まれる遺伝子資源は異なることが考えられる(図2)。そのため、GSモデル構築のための表現型調査は、集団が変わるごとに実施されることが望ましい。しかしトマトなどの園芸作物では、設備や栽培・収穫調査にかかる労力の問題から、小規模のデータであっても、より精度の高いGSモデルが構築できることが望ましい。そこで、個体ごとの日射量、気温、湿度、地温、土壌水分の環境条件を測定する環境セ

ンシングシステムを開発し、これら細かな環境バイアスが表現型値に与える影響を推定するとともに GS モデルへの反映を試みた。本試験の結果、環境が比較的安定していると期待される養液栽培ハウスにおいても、特に土壌水分レベルにおいて大きな違いが明らかとなった(図5a)。また、これら環境情報を GS モデルの説明変数に加えることで、表現型予測精度を向上させられることも示した(図5b)。

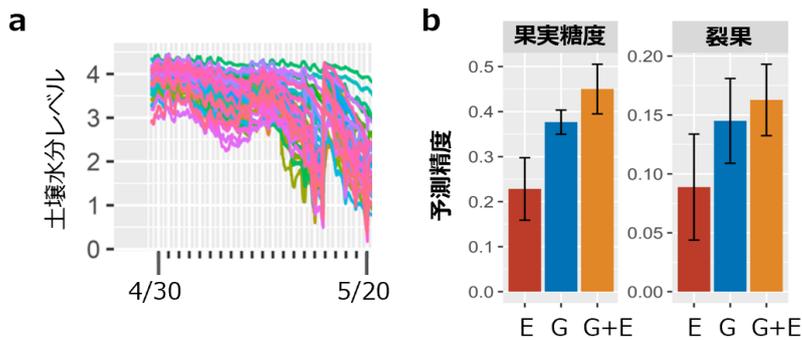


図5. 個体別環境情報の取得とGS予測モデルへの応用. a. ハウス養液栽培における個体ごとの土壌水分レベルの違い. b. 環境情報を用いた表現型予測. 予測精度は2分割交差検証における予測値と実測値との相関係数. E: 環境データのみ, G: ゲノム情報のみ, G+E: 環境情報とゲノム情報の両方を説明変数とした予測モデル.

### 3. 今後の展開

本課題で示した戦略は、民間種苗会社などと連携しながら、実際育種への展開を進めていく。また、環境センサーは生理障害など環境変化に高感受性な形質の遺伝解析に応用し、より複雑な形質に対するゲノム育種の実現を目指す。

### 4. 自己評価

#### ・研究目的の達成状況

GS モデルのためのマーカー選抜には原因遺伝子の特定が望ましかったが、既知の遺伝子と対応しない、機能喪失型など明瞭な変異が存在しないなどの理由から思うように進まなかった。今後は、トランスクリプトームなど異なる情報を統合した解析が必要である。実際の育種過程における GS モデルの挙動は概ね期待通りであった。本研究で提案した戦略が実際の育種にどれ程貢献できるか、今後の実践によって示していく必要がある。

#### ・研究の進め方

研究実施体制については、他機関との連携の強化や研究補助の増員などにより研究の範囲や展開をより進めるべきであったと省察している。研究費の執行状況は概ね順調である。

#### ・研究成果の科学技術及び社会・経済への波及効果

データサイエンスに基づく作物育種は理論的効果が示されている一方で、その実装には予算・労力などの面で障壁が存在する。本研究で示したマーカー選抜や環境センサーを用いた GS モデルの精度向上は改良の必要性はあるものの、必要な予算・労力を抑えることで、データサイエンスに基づく作物育種に貢献しうると期待している。

## 5. 主な研究成果リスト

### (1)論文(原著論文)発表

なし

### (2)特許出願

研究期間累積件数: 0 件

### (3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

1. Yamamoto E. “A high-density environmental sensing system revealed effect of point-specific environmental differences on phenotypes of greenhouse tomato.” *The 16<sup>th</sup> Solanaceae Conference*, Sep 2019, Israel
2. Yamamoto E. “Genome-wide association study, genomic selection and other technologies for tomato breeding in Japan.” *The 15<sup>th</sup> Solanaceae Conference*, Oct 2018, Thailand
3. 山本英司 情報科学を作物育種に利用する～展望と課題～ **第1回植物バイオインフォマティクス研究会**、2018年9月、神奈川
4. Yamamoto E. “Genome-wide association study, genomic selection and other technologies for tomato breeding in Japan.” *Plant & Animal Genome Asia 2018*, May 2017, Korea
5. Yamamoto E. “Application of whole-genome prediction in Japanese big-fruited tomato varieties.” *The East Asia Agricultural Genome Science Forum 2017*. Dec 2017, Korea