〔総説〕

ゲノミックセレクションの作物育種への展開

林 武司

(国研) 農業・食品産業技術総合研究機構 次世代作物開発研究センター
茨城県つくば市観音台 2-1-2

Application of genomic selection to crop breeding

Takeshi HAYASHI

Institute of Crop Science, NARO Kannondai, Tsukuba, Ibaraki

はじめに

現在、DNA 配列の解読技術の進展により、各種の作物においてゲノムに関する詳細な情報が急速かつ大規模に蓄積されており、SNP などの大量のゲノム上の変異が明らかにされている。膨大な数の SNP に対する遺伝子型判定を低コストで実現する技術や機器も開発されており、作物の品種・個体における SNP の遺伝子型情報の効率的な収集が可能となっている。このような状況の下で、ゲノムワイドなマーカー情報を活用した効率的な選抜方法であるゲノミックセレクション（以下 GS と略す）に対して、作物育種への適用を目指した研究が盛んに行われている。GS においては、まず表現型データとゲノムワイドに配置された SNP などのマーカーの遺伝子型データを持つ個体・品種群（トレーニング集団）を用いてマーカー遺伝子型から育種価を予測するためのモデル（育種価予測モデル）を構築し、次に構築されたモデルを用いて育種対象となる個体・系統のマーカー遺伝子型データから育種価を予測し、その予測された育種価（Genomic estimated breeding value、以下 GEBV と略す）にもとづいて個体選抜が行われる（Meuwissen ら 2001）。GS はもともと家畜育種を念頭に提案された選抜手法であり、乳牛においてすでに GS を用いた育種システムが運用されており（Dalton 2009）、他の家畜でも実用化が進んでいる（Knol ら 2016; Wolc ら 2016）。

GS による選抜では選抜対象となる個体・系統の表現型データを測定する必要がなく、作物育種においても幼苗から抽出された DNA による個体選抜が可能のため、育種サイクルを大幅に短縮した効率的な形質改良の実現が期待される。しかしながら、作物育種にお

いては、多くの作物種について作物種ごとに固有の育種方法が確立されており、GS の実用化のためには GS を既存の育種方法にどのように組み込むかが課題となる。また、作物育種においては、通常、栽培地域や栽培環境に適した品種の作出が目的となるため、環境応答性を考慮した育種が重要となる。そのために、遺伝的能力（育種価）だけでなく、環境応答をつかさどる遺伝子型と環境との交互作用（Genotype by Environment Interaction、以下 GxE と記す）についても評価する必要がある。したがって、GS の予測モデルにおいて育種価だけでなく GxE も取り込んで予測を行うことにより、環境に応じた品種の作出に GS を利用することも可能である。作物育種への GS の応用に関連して、GxE を予測するための予測モデルの研究もこれまでに数多く報告されている（例えば、Burgueño ら 2011, 2012; Schulz-Streeck ら 2013; Helslot ら 2014; Jarquine ら 2014; Malosetti ら 2016）。

果樹などの作物においては、異なる品種・系統間の交配により、多数の F1 個体を作り、この中から育種対象の形質に優れた個体を選抜・育成して新品種を作出することが行われている。しかしながら、果樹は一般に世代間隔が長く、重要な形質である果実形質については表現型の測定まで多大の年月が必要で、例えば、リンゴでは 5 年、カンキツでも 3 年ほどの年月を要する。したがって、選抜対象となる個体の表現型データを必要としない GS は、果樹育種の高速化を実現する有効な選抜方法となり得る可能性があり、リンゴ、

連絡先：林 武司 農研機構 次世代作物開発研究センター
〒305-8518 茨城県つくば市観音台 2-1-2
(e-mail: hayatk@affrc.go.jp)

ニホンナシ、モモ、カンキツなどでGSの有用性に関する研究が行われている (Kumar ら 2012; Muranty ら 2015; Iwata ら 2013a, 2013b; Biscarini ら 2017; Minamikawa ら 2017)。品種・系統間の交配によるF1が育種材料となる果樹育種に適したGSの興味深い応用として、GSの予測モデルを用いたF1個体の形質分離の予測 (Iwata ら 2013a) が挙げられる。連鎖地図などの組換え価の情報をもとに交配に用いる2つの親品種のマーカー遺伝子型が与えられたとき、シミュレーションを用いて交配により生じるF1個体の可能なマーカー遺伝子型を確率的に生成することができるが、GSの予測モデルを生成されたマーカー遺伝子型に適用することにより、F1個体における形質値の分布、すなわち形質分離の程度を予測することが可能となる。予測された形質分離より、交配親の組合せに対して優良なF1個体の得られる確率を事前に評価することができ、望ましい交配組合せを選択することが可能となる。果樹では樹体の大きさのために、育種に必要なスペースは広大となるが、新品種になり得る有望個体の含まれる確率が高い交配組合せを限定することにより、品種の作出効率を減らすことなく供試するF1個体数を削減することが可能となり、果樹育種の大幅な効率の向上が期待される。

このような状況を踏まえて、本レビューにおいては、まず、GSの実用化の検討が進んでいるトウモロコシやコムギなどの作物におけるトレーニング集団の設定など、既存の育種方法へのGSの実装化について概観する。次に、GxEを取り入れたGS予測手法の拡張について見ていくことにする。さらに、GSの予測モデルを応用した2品種間交配によるF1集団における形質分離予測についても解説する。

ゲノミックセレクションの作物育種への実装

GSにおいては、上述したように育種価の予測モデルを構築する必要がある。予測モデルを構築するためには、多数の個体・品種におけるゲノムワイドに配置されたSNPなどのマーカーの遺伝子型データと選抜対象となる形質の表現型データが必要であり、これらのデータを供給し、予測モデルをトレーニングするための個体・品種の集団がトレーニング集団である。構築された予測モデルは実際の育種集団において選抜対象となる各個体・系統のマーカー遺伝子型から育種価を予測するために用いられる。マーカー遺伝子型から予測された育種価にもとづいて選抜が行われるため、通常は手間と時間を要する形質の表現型の測定が不要

であり早期の選抜が可能となり、育種のサイクルを早く回すことができる。ただし、予測値の精度が重要であり、精度が低い予測では、優良な個体・系統を正確に選抜することが困難になり育種効率はかえって低下することになる。高い予測精度を実現するためには、予測モデルの構築に用いられるトレーニング集団と実際に予測が行われる育種集団の遺伝的背景が類似している必要があり、作物育種へのGSの実装においてはどのようなトレーニング集団を用意するかが重要となる。ここでは、GSの導入が盛んに試みられているコムギ (Poland ら 2012; Bassi ら 2016; He ら 2016; Huang ら 2016; Michel ら 2016; Saint Pierre ら 2016; Gaynor ら 2017; Song ら 2017) とトウモロコシ (Bernardo and Yu 2007; Beyene ら 2015; Zhang ら 2015; Vivek ら 2017) について、いかなるトレーニング集団を構築し、従来の育種プロセスにおける選抜にGSをどのようなタイミングで利用していくかを見ていくことにする。

コムギ育種におけるゲノミックセレクションの実装

コムギの育種では、既存の品種や個体・系統を交配親としていくつかの交配集団を得て、その中から望ましい個体・系統を選抜して、新しい品種を作出する、いわゆる交雑育種という育種法が実践されている (図1)。品種間交配から新品種の作出までは、10年弱の時間を要する。コムギの育種では収量性、製粉性、耐病性などが対象形質となるが、GSの適用を考える場合、改良対象となるのは量的形質であり、図1では収量性を対象形質と仮定した。

コムギのような自殖性の作物において品種は遺伝的に固定されているため、交配で得られる次の世代の集団から育種を始める場合、自殖により世代を重ねて遺伝的にホモ化を進めて、ある程度遺伝的に固定された個体・系統を得た後に形質を評価して選抜を行うことになる。遺伝的な固定を目的として世代を進めるためには多大な時間が必要で、選抜の回数も限られてしまう。しかしながら、コムギでは倍加半数体 (Doubled Haploid、以下DHと略す) 作製の技術が確立されており、ヘテロの集団であっても半数体である配偶子のゲノムを倍加させて次の世代で固定系統の集団を作出することができる。コムギ育種においてもDH作製技術が利用されており、図1にあるようにF1個体の配偶子を用いて、次の世代で固定系統 (DH系統) が作出され、その後の形質調査、選抜に用いられる。図1に示された品種育成プロセスの基礎集団は交配のために選ばれた品種や個体である。このプロセスの各ス

図 1. コムギの品種育成プロセスの例 (Gaynor et al. (2017) の図を改変)

テージは同時並行的に進んでおり、品種育成プロセスの途中で得られる優良な選抜個体についても交配親として基礎世代に取り入れるることにより、育種効率を向上させることが可能である。

コムギにおけるGSを実装した育種計画はBassiら(2016)によって論じられており、トレーニング集団の構築、GSの実行のタイミング等についての考察が与えられている。図1の基礎世代の交配に用いられる個体集団として、GSによる循環選抜を経た集団を用いることにより、形質改良の効率を一層向上させることが可能であり、そのための循環選抜の方法が提案されている。例えば、品種や個体間の交配で得られるF2世代を用いてGSを利用した循環選抜を行う場合の例(図2)を見ていくと、最初に40品種を用いた交配により、20の交配集団を得て、それぞれの集団で50個体のF2を含む全部で1000個体から成る初期F2集団(C0F2)を作る。ここで1000個体すべてについてSNPの遺伝子型判定を行い、主導遺伝子のマーカー利用選抜(MAS)やこれまでのデータで作製された予測モデルにより算出されたGEBVをもとに、20集団から10集団を選び、さらに選ばれた10集団のそれぞれから4個体を次世代の集団の親として選抜する。この段階でのGSの予測モデルは過去に蓄積されたデータから構築されており、選抜対象の集団とは

血縁の遠いトレーニング集団をモデル構築に用いているため、GEBVの予測精度は十分ではない可能性もあることに注意すべきである。この世代において、遺伝子型判定を行った1000個体から一粒ずつ種子を得て、より適当なトレーニング集団を作製するための準備をしておく。次の世代においては、10集団のそれぞれにおける4個体の交配により、各集団から2つの交配集団を作成し、全部で20集団を得て各集団が25個体から成る全体で500個体から成る集団(C1F2)を作る。前の世代の集団と同様に、MASと従来のデータによるGEBVをもとに、20集団から10集団を選抜し、さらに集団内での個体選抜を行い、次世代の親とする。その次の世代(C2F2)でも同様の操作で世代を重ねて次世代(C3F2)を得る。トレーニング集団については、C0F2集団から選ばれた10集団より得られた種子からF3個体を得て、自殖によりDH個体からなる次世代の集団(DH1)を作り、各集団50個体からなる総計500個体の遺伝子型を測定しておく。さらに2回自殖により世代を重ねてDH1個体を増殖してDH3系統(3世代目のDH個体からなる系統)の集団を作り、これをトレーニング集団として圃場で栽培して表現型を測定して、DH1世代の遺伝子型と合わせて育種価予測モデルを構築する。DH1からDH3までは世代を進めるだけなので、育種集団がC2F2の段階でDH3

図2. コムギ F2 集団における GS を用いた循環選抜の例 (Bassi et al, (2016) の図を改変)

個体のトレーニング集団をもとに予測モデルが構築され、C₃F₂の世代ではその予測モデルを用いたGEBVの予測が可能となり、GSによる個体選抜が行われる(図2)。このトレーニング集団(DH3系統群)は、選抜対象となるC₃F₂集団の個体と血縁の近いDH3個体から構成されており、高い精度でのGEBVの予測が期待できる。ここで選抜された個体とその後の通常の品種育成プロセスの基礎世代に組み込まれて、新品種の作出に用いられることになる。

この手法は品種育成プロセスの基礎世代の個体をGSによってあらかじめ改良しておくというアイデアに基づいているが、Gaynorら(2017)は、このアイデアをより明確化し、育種過程を図1の品種育成プロセスとGSを活用した集団改良プロセスの2つに分割

して、育種システムを作ること提案している(図3)。集団改良プロセスは、多数の個体(図3の例では200個体)の間の交配で得られるF₁個体をGSで選抜し、選抜された個体間の交配でさらなるF₁種子を得る過程からなり、すべて温室内で行われる。必要に応じて優性不稔を引き起こす薬剤を用いて他殖化するため、自殖性作物の交配作業に必要な労力も軽減し、1年に2回この過程を回すことが可能である。この集団改良プロセスにおいて、得られる種子のいくつかはDH個体の作出に利用され、品種育成プロセスの基礎個体として提供される。品種育成プロセスにおいては、個体・系統の収量性評価と選抜を行い、優良個体のスクリーニングを行い新品種の作出が行われる。収量調査とともにマーカーの遺伝子型判定も行い、その情報は集団

図3. 品種育成プロセスと独立したGSを活用した集団改良プロセスからなる育種システムの例 (Gaynor et al. (2017) の図を改変)

改良プロセスに受け渡され、育種価予測モデルの更新や高精度化に利用される。また、品種育成プロセスにおいては、マーカーの遺伝子型判定を行った個体については、集団改良プロセスで構築された育種価予測モデルによるGSも可能であり、品種作出の速度をさらに高めることができる。Gaynor ら (2017) はシミュレーションを用いて、育種に要するコストを同等にした条件下で、この2つのプロセスから成る育種システムは遺伝的改良の速度が長期間にわたって維持されることを示し、品種育成プロセスのみの育種システムにGSを導入した場合に比べても、品種作出の効率に優れていることを示した。

トウモロコシにおけるゲノミックセレクション実証試験

トウモロコシは他殖性作物であるが、栽培現場では主にF1品種が活用されている。そのために、他の純系品種との交配 (交雑) によって優良なF1個体を産出する純系品種の作出がトウモロコシ育種の目標となる。そこでは高いヘテロシスを発揮する組合せ能力の高い個体を見出すことが必要であり、選抜の過程では基準品種との交雑により得られたF1の形質値から組合せ能力の評価を行う。基準品種との交雑は検定交

雑と呼ばれ、組合せ能力の高さが選抜の基準の一つとなる。トウモロコシ育種へのGSの実装に関してもコムギと同様にいくつかの報告がある (Bernardo and Yu 2007; Beyene ら 2015; Vivek ら 2017)。ここでは、Vivek et al. (2017) によるトウモロコシの干ばつ耐性育種におけるGSの活用例を見ていくことにする。

熱帯アジアでのトウモロコシ栽培は、灌漑設備が完備されておらず雨水に頼っており、干ばつの被害を受けやすい。そこで、Vivek ら (2017) は干ばつ耐性を改良したトウモロコシの育種集団の作出を目指したGSの実証試験を行った。本試験の概略は図4に示されている。まず、有望な品種間の組合せによってF1、F2を作り、さらに自殖によりF3個体を作成した。本報告では2つの交配組合せにより、2つのF3集団を得た。F3個体については、基準となる品種と交雑 (検定交雑) を行って得られる個体の収量と収量に関連するその他の形質を干ばつ条件および最適条件のもとで測定した。これらの形質値を組み合わせて指数を作り、この指数をもとにF3個体を選抜し、選抜個体間の交配により次の世代の集団 (C1) を作った。F3個体については、検定交雑による形質評価 (組合せ能力評価) とともにマーカーの遺伝子型判定も行っており、F3

図4. トウモロコシにおける干ばつ耐性に対するGS検証試験の例 (Vivek et al, (2017) の図を改変)

集団をトレーニング集団として育種価の予測モデルが構築された。C1 集団の個体に対しては、自身の形質値をもとにした表現型選抜 (Phenotypic Selection, PS) および F3 集団で作成された予測モデルによる GS により次の世代の集団 (C2) を作った。表現型選抜により得られた集団は C2-PS、GS により得られた集団は C2-GS と表記する。C2-PS については、それぞれ干ばつ条件および最適条件のもとでの表現型選抜で得られた2種類の集団を含んでいる。C2-PS と C2-GS の各個体についてそれ自身の形質値を測定するとともに、基準品種との検定交雑を行い、干ばつ条件および最適条件のもとで収量等の形質を比較して GS の効率を評価した。その結果、個体自身の収量は C2-GS が C2-PS よりも 10 ~ 20% 高くなり、検定交雑による組合せ能力の比較でも、C2-GS が C2-PS よりも 4 ~ 43% 高い収量を示し、干ばつ条件および最適条件の両方で GS は表現型選抜に比べて高い遺伝的改良度を実現した。干ばつに対する耐性が育種対象の形質であるため、表現型選抜では干ばつ条件下での形質測定、個体選抜が必要となるが、GS では通常の条件で栽培した個体の DNA が利用できれば良く、育種のコストが極めて軽

減される。このように新品種の作出プロセスにおいて、その新品種の栽培・普及が予定されている環境や地域に限らず、どの場所でも選抜が可能となることは GS のメリットの一つである。我が国でも GS の実用化によって、全国に散在する農業試験場で行われている育種事業において、個体選抜の業務を特定の場所に集中することが可能となり、一層の効率化を目指した育種事業の拠点化も推進されることになるであろう。

遺伝子型と環境の交互作用を考慮したゲノミックセレクション予測モデル

イネやムギの育種においては、栽培地域や栽培環境に適した品種を作出することが目的の一つであり、普及を想定した地域に位置する試験地で品種育成が行われる。例えばイネの育種においては、図5のように日本各地の育種場所において、耐冷性や耐暑性など地域ごとに異なる育種を対象として品種育成が行われている。栽培地域に適した品種を育種するためには、育種価だけでなく環境応答性を考慮することが重要である。環境応答性は遺伝子型 (G) と環境 (E) の交互作用 (G x E) と同値であり、GxE によって環境応答

図5. イネ育種における地域拠点と地域に応じた育種対象形質.
農水省広報誌 aff(あふ)2011年11月号に掲載の図を転載
http://www.maff.go.jp/j/pr/aff/1111/spe1_01.html

性も表すことにする。作物においては、品種の環境応答性を評価するために同じ品種を複数の環境下で栽培して、各品種における表現型値の安定性や変動を調べる多環境試験 (Multi-Environmental Trials, MET) が行われてきており、GxE を解析するための統計モデルが開発されてきた (Finlay と Wilkinson 1963; Cornelius ら 1992; Gauch 2006)。作物における GS において、予測モデルに GxE の項を含めることにより、ゲノム情報から育種価だけでなく環境応答性も予測して、栽培地に適した新品種を効率的に作出することが可能となる。そのための予測モデルの開発や予測精度の評価が行われている (Burgueño ら 2011, 2012; Schulz-Streeck ら 2013; Helslot ら 2014; Jarquine ら 2014; Malosetti ら 2016)。以下に、多環境試験において従来用いられてきた統計モデルについて概観する。そして、このモデルをもとに、GxE をどのように GS の予測モデルに組

み込んでいくかを見ていくことにする。

多環境試験における従来の統計モデル

同じ品種群をいくつかの環境で栽培して形質の表現型を測定する多環境試験を行ったとする。ここでは、栽培試験を行った品種の数を n として、それぞれの品種に対して異なる m 環境で表現型を測定したものとする。 i 番目の品種の j 番目の環境における表現型を y_{ij} ($i=1,2,\dots,n; j=1,2,\dots,m$) と表す。通常の試験では、それぞれの品種について複数の個体を栽培し形質を測定することになるので、 y_{ij} は j 番目の環境で栽培された i 番目の品種の複数個体の平均値とみなされる。GxE の解析のために次のようなモデルが仮定される。

$$y_{ij} = \mu + G_i + E_j + GE_{ij} + e_{ij} \quad (i=1,2,\dots,n; j=1,2,\dots,m) \quad (1)$$

ここで μ はモデルの切片、 G_i は i 番目の品種（遺伝子型）の効果、 E_j は j 番目の環境の効果、 GE_{ij} は i 番目の品種と j 番目の環境の交互作用、すなわち GxE を表す。また、 e_{ij} はモデルの残差である。(1) のモデルより、分散分析により GxE の有意性を検定することができる。

より詳しく各品種の環境応答性を調べるために、(1) の交互作用項 GE_{ij} に対するモデル化も行われている。Finlay と Wilkinson (1963) は、 $GE_{ij} = \beta_i E_j + e_{ij}$ として、環境値 E_j に対する各品種の応答を回帰モデルで表し、回帰係数 β_i によって環境に対する安定性を評価することを提案している。また、交互作用項 GE_{ij} ($i=1,2,\dots,n$; $j=1,2,\dots,m$) を n 行 m 列の行列 $\Gamma=(GE_{ij})$ とみて特異値分解を行い、少数の特異値に対応する成分の和で Γ を近似して、各品種の環境応答性を調べるモデルも提案されている。このモデルでは、特異値分解の最初の K 個の特異値に対応する積和を用いて GE_{ij} を近似すると

$$GE_{ij} \approx \sum_{k=1}^K \lambda_k \alpha_{ik} \gamma_{jk}$$

と書くことができる。ここで、 λ_k は k 番目の特異値、 α_{ik} と γ_{jk} は k 番目の特異値に対応する列ベクトルと行ベクトルの第 i 成分と第 j 成分であり、 α_{ik} は i 番目の品種の j 番目の環境値の k 番目の成分 γ_{jk} に対する応答性を表すものと解釈される。このモデルは、AMMI (Additive Main effects and Multiplicative Interaction, 相加的な主効果と相乗的な交互作用) モデルと呼ばれ (Gollob 1968; Mandel 1969; Gauch 2006)、作物の多環境試験においては主要な解析モデルとして用いられている。

GxE を組み込んだゲノミックセレクション予測モデル

GS においては、ゲノム全体に配置されたマーカー（特に SNP マーカー）のマーカー遺伝子型から遺伝子型値である育種価、すなわち前節のモデル (1) の G_i を予測し、個体選抜に利用する。モデル (1) の G_i だけでなく、遺伝子型と環境の交互作用 GE_{ij} も予測できれば、形質が未測定 of 遺伝子型の個体がある環境で栽培した時の形質値を予測することが可能となる。すなわち、従来の GS では未知の遺伝子型の育種価すなわち平均的な形質値を予測できるが、GxE を予測に含めることにより、未知の遺伝子型を様々な環境下で

栽培した時の形質値も予測できるようになる。これにより、品種の普及を想定した地域とは異なる地域で新品種を作出することが可能となる。イネ、ムギなどの主要作物については、すでに述べたように日本各地の試験地でその地域に適応した品種の育成を行っているが、GxE の効果を考慮した GS の予測モデルの構築によって、作物育種の拠点を作り各地域に適応した品種を効率的に開発することも可能となる。

GxE の効果を取り入れることによる GS の予測モデルの拡張は、多環境試験のデータとマーカー情報を用いて、注目した品種の未試験地での環境応答性を予測することや、まだ栽培試験に供されていない品種の様々な試験地での形質値を予測することを目的として提案されてきた。以下に GxE を取り入れた GS の予測モデルの構築について見ていくことにする。

モデル (1) において、 $\mu_j = \mu + E_j$, $g_{ij} = G_i + GE_{ij}$ とおくと、 μ_j は j 番目の環境で栽培された品種群の平均値であり、 g_{ij} は GxE も含めた i 番目の品種の j 番目の環境における遺伝的能力とみなすことができる。このときモデル (1) は次のように書くことができる。

$$y_{ij} = \mu_j + g_{ij} + e_{ij} \quad (i=1,2,\dots,n; j=1,2,\dots,m) \quad (2)$$

ここで、 y_{ij} , g_{ij} , e_{ij} について m 環境における n 品種の値を次のようにベクトル $\mathbf{y}$, $\mathbf{g}$, $\mathbf{e}$ でそれぞれ表す。

$$\mathbf{y} = \begin{bmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n1} \\ y_{12} \\ \vdots \\ y_{n2} \\ \vdots \\ y_{1s} \\ \vdots \\ y_{ns} \end{bmatrix}, \quad \mathbf{g} = \begin{bmatrix} g_{11} \\ g_{21} \\ \vdots \\ g_{n1} \\ g_{12} \\ \vdots \\ g_{n2} \\ \vdots \\ g_{1s} \\ \vdots \\ g_{ns} \end{bmatrix}, \quad \mathbf{e} = \begin{bmatrix} e_{11} \\ e_{21} \\ \vdots \\ e_{n1} \\ e_{12} \\ \vdots \\ e_{n2} \\ \vdots \\ e_{1s} \\ \vdots \\ e_{ns} \end{bmatrix}.$$

また、 j 番目の環境の環境値とみなすことができる μ_j についても、 m 環境のすべての環境値をまとめて

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{bmatrix}$$

と表す。このとき、(2)は

$$y = \mu \otimes \mathbf{I}_n + \mathbf{g} + \mathbf{e} \quad (3)$$

と書くことができる。ここで、記号 $\otimes$ はクロネッカー積を表す。(3)における $\mathbf{g}$ の予測値によって品種の環境に対する適応性(GxE)も考慮したGSが可能となる。

$\mathbf{g}$ と $\mathbf{e}$ は変量効果であり、それぞれが多変量(nm 変量)の正規分布に従うと仮定する。 $\mathbf{g}$ の分散共分散行列は、 $\Sigma_E \otimes \Sigma_G$ と表すことができ、 Σ_E は形質の m 環境の間の分散共分散行列で、環境の間の類似性を反映している。また、 Σ_G は n 品種の間の遺伝的關係を表す関係行列である。 $\mathbf{e}$ は残差で環境特異的な非遺伝的効果を含み、 $\mathbf{e}$ の分散共分散行列を $\mathbf{R}$ とすると、環境特異的な残差分散を j 番目の環境について σ_j^2 とみると、 $\mathbf{R} = \text{Diag}(\sigma_j^2) \otimes \mathbf{I}_n$ と表される。ただし、 $\text{Diag}(\sigma_j^2)$ は対角成分が σ_j^2 の対角行列、 $\mathbf{I}_n$ は n 次の単位行列である。

n 品種の間の遺伝的關係行列 Σ_G については、系譜情報から n 品種の間の血縁行列 $\mathbf{A}$ によって $\Sigma_G = \mathbf{A}$ と表すことができるが、GSにおいては通常マーカー遺伝子型の情報から Σ_G が算出される。今、SNPがマーカーとして利用可能であり、 n 品種について p 個のSNPの遺伝子型のデータがあるとして、このデータをまとめて、 n 行 p 列の行列 $\mathbf{W} = (w_{ij})$ ($i=1,2,\dots,n$; $j=1,2,\dots,p$)と表す。各SNPの遺伝子型については、各SNPの2つのアレルのうち着目した一方のアレルの個数によって0、1、2と表し、 $\mathbf{W}$ の要素 w_{ij} は0、1、2のいずれかの値を取るものとする。このとき $\Sigma_G = \mathbf{W}\mathbf{W}'$ と表される。あるいは、各SNPのアレル頻度(q_j , $j=1,2,\dots,p$)を用いて w_{ij} を標準化して、

$$x_{ij} = (w_{ij} - 2q_j) / \sqrt{2q_j(1 - q_j)}$$

とおき、この x_{ij} を要素とする行列 $\mathbf{X} = (x_{ij})$ を用いて、 $\Sigma_G = \mathbf{X}\mathbf{X}'$ とする場合もある (SpeedとBalding 2014)。

形質について m 環境の間の類似性を反映した分散共分散行列 Σ_E については、いくつかの設定が考えら

れる。単純な設定として形質における異なる環境の間の共分散が0であるという仮定が考えられる。各環境内の分散がすべての環境で等しいという仮定を加えると最も単純な設定となり、各環境内の形質の遺伝分散を σ_G^2 として $\Sigma_E = \mathbf{I}_m \sigma_G^2$ と表される。しかし、この設定はGxEが存在しない場合を想定しており、環境適応性を予測するには不適當である。環境間の相関は無いが、環境ごとに形質の分散は異なるなるとすると、 j 番目の環境における形質の遺伝分散を σ_{Gj}^2 として $\Sigma_E = \text{Diag}(\sigma_{Gj}^2)$ と表される。一般的な設定として、形質における異なる環境の間に相関があり、環境ごとに形質の分散がすべて異なる場合が考えられる。このとき、 Σ_E の独立な成分は $m(m+1)/2$ 個あり、 m の値とともに推定すべきパラメータの数は急速に増大してしまう。そこで、 Σ_E についてある程度の構造化を導入することもある。例えば、因子分析モデル (factor analytic model) を仮定して、 Γ を m 行 k 列の行列 ($k < m$)、 Ψ を対角行列として $\Sigma_E = \Gamma\Gamma' + \Psi$ と構造化してパラメータの数を減らす工夫もなされている (Denisら 1997; Piepho 1997; Smithら 2001)。これらの設定においては、表現型データを用いて環境間の類似性を評価して Σ_E の要素を推定することになる。したがって、まだ評価されていない(どの品種も栽培されていない)未知の新しい環境において発揮される遺伝的能力については、予測することができない。栽培試験のされていない環境における形質値の予測のためには、日照量や温度、湿度などの気象要因を環境変数として取り込んで環境間の類似性を評価して Σ_E を構成する必要がある。例えば、 m 環境のそれぞれについて、特定の期間における日々の最高・最低気温、日長時間、降水量など l 個の環境変数の値が与えられており、 j 番目の環境の k 番目の環境変数の値を z_{kj} とすると、SNPの遺伝子型データによる品種間の関係行列と同様に、 z_{kj} を用いて環境間の関係行列を構成することができる (Malosettiら 2016)。栽培試験のなされていない環境においても環境変数を観測することにより、栽培試験の行われている環境との類縁関係がわかるため、SNP遺伝子型の与えられている品種を栽培試験の行われていない環境で栽培した時の形質を予測することが可能となる。

Malosettiら (2016) は、(3)のモデルを用いて、ゲノム全体に配置された783個のSNPの遺伝子型データを持つオオムギの148個体から成るDH集団における多環境試験の収量データについて、各DH個体の環境応答性の予測精度を調べた。多環境試験は2年にわ

たって行われ、1年目は4地域、2年目は1年目の4地域を含む6地域の計10環境でDH集団が栽培された。 Σ_E については、上で述べたような4通りの設定が考えられた。すなわち、(i) $G \times E$ を考慮しない設定 ($\Sigma_E = \mathbf{I}_m \sigma_G^2$)、(ii) 各環境で形質の分散は異なるが、環境間の相関が無い設定 ($\Sigma_E = \text{Diag}(\sigma_{Gj}^2)$)、(iii) 因子分析モデル、(iv) 環境変数を用いて算出された環境間の関係行列、である。設定(iv)における環境変数としては、オオムギの生育ステージごとに測定された日々の気温、与えられた水分量、日長時間を用いて z_{kj} が算出され、 j 番目の環境と j' 番目の環境の近縁度 (Σ_E の j 行 j' 列の成分) $\sigma_{jj'}$ は、

$$\sigma_{jj'} = 1 - \frac{\sum_{k=1}^l \frac{z_{kj} - z_{kj'}}{\max(z_k) - \min(z_k)}}{l}$$

と与えられた。ここで、環境変数は全部で l 種類あり、 $\max(\cdot)$ および $\min(\cdot)$ は全10環境における各変数の最大値、最小値を表す。全体のデータをいくつかのシナリオにもとづいて、トレーニング集団と予測集団に分割して、予測精度を評価したところ、 Σ_E に関する (iv) の設定のもとでの予測が他の設定のもとでの予測よりも全般的に高い精度を示した。作物学の分野では、これまでに生物学知見にもとづいて日長や温度などの環境要因の開花時期や収量などの形質に対する影響を記述するモデルとして、作物生育モデルが研究されてきている (Cooper ら 2009; van Eeuwijk ら 2010)。環境変数を考慮することにより作物生育モデルとGSの予測モデルの連携が可能となり、より高精度の環境適応性の予測を実現するためのモデル開発が進展しつつある。

品種の交配によるF1集団における形質分離の予測

GSにおいて構築される育種価予測モデルは、選抜対象の集団から望ましい個体を効率的に選抜するために、個体の育種価をSNPなどのマーカーの遺伝子型情報から予測するためのモデルである。育種価予測モデルにマーカーの遺伝子型をインプットすると形質の育種価がアウトプットとして出力される。したがって、作物における交雑育種において適当な育種価予測モデルが手元にあれば、交配親のマーカー遺伝子型の組合せから交配後代に生じうる個体の可能なマーカー遺伝子型に対しても育種価の予測ができ、事前に交配親の組合せの評価を行い、交配すべき品種の組合せを選別できる。

リンゴやナシ、カンキツなどの果樹の育種において

は、既存の2品種の交配により得られる多数のF1個体の中から有望な個体が選抜され、新しい品種の候補として育成される。これらの果樹は他殖性であり、自殖性作物であるイネやムギとは異なり、新しい品種を作出する過程でホモ化のために世代を進める必要がなく、また長い世代間隔のために世代を重ねることが時間的に困難なために、このような育種法が実践されている。しかしながら、交配により得られるF1個体の中に有望な個体が含まれる割合は低く、交配組合せによっては、育種目標に見合った個体が全く得られない場合も多い。実際の育種では既存の品種を用いた複数の交配組合せから複数のF1集団を作りその中から有望な個体の選抜が試みられるが、多数のF1個体を扱うため栽培管理の労力や圃場のスペースなどに多大なコストを要する。交配に用いられる品種に対して、それらの品種間の交配により生じうるF1個体における育種価の分布、すなわち形質分離の様子が交配前に予測できれば、優れた個体を生じる可能性の高い有望な交配組合せをあらかじめ決定することにより育種の効率を向上させることができる。

Iwata ら (2013a) は、果樹育種において、交配に用いられる品種・系統群の表現型データとマーカー遺伝子型データから育種価予測モデルを構築した後、2つの品種・系統の交配により得られるF1個体のマーカー遺伝子型をシミュレーションによって確率的に生成し、生成されたマーカー遺伝子型に対して育種価予測モデルを適用することにより、F1集団における形質分離を予測する手法を開発した。さらに、予測されたF1集団における育種価の分布をもとに、設定された育種目標を満たす有望な個体が得られる確率を評価し、その確率が高い交配組合せを選ぶ方法を提案した。F1集団の分離予測と交配組合せの選択方法について以下に見ていくことにする。

F1個体のマーカー遺伝子型の生成

連鎖した2つのマーカー間での組換えは、連鎖地図の位置情報を備えたマーカーであれば、組換え価より確率的に計算することができる。したがって、用いるマーカーについて連鎖地図の情報があれば、交配に用いる2品種のマーカー遺伝子型から交配によって生じるF1個体のマーカー遺伝子型を確率的に予測することができる。例えば、組換え価 r で連鎖した2つのマーカーAとBについて、交配に用いる2品種XとYの遺伝子型がハプロタイプにより、それぞれ、A1B1|A2B2、A2B2|A2B2で表されるとする。ただし、

A1、A2 はマーカー A のアリル、B1、B2 はマーカー B のアリルであり、また「|」は母方由来の配偶子と父方由来の配偶子を分ける記号である。この 2 品種 X と Y の交配では、X から 4 種類の配偶子 A1B1、A1B2、A2B1、A2B2 がそれぞれ $(1-r)/2$ 、 $r/2$ 、 $r/2$ 、 $(1-r)/2$ の確率で生じ、Y からは 1 種類の配偶子 A2B2 が生じるため、F1 個体においては 4 種類の遺伝子型 A1B1|A2B2、A1B2|A2B2、A2B1|A2B2、A2B2|A2B2 がそれぞれ $(1-r)/2$ 、 $r/2$ 、 $r/2$ 、 $(1-r)/2$ の確率で生じることになる。このような配偶子が生じるプロセスをシミュレーションで模倣することにより、交配に用いる品種の（ハプロタイプで表された）遺伝子型の情報から、生じうる F1 個体の遺伝子型を確率的に生成することができる。

F1 における育種価の分布の予測

シミュレーションで生成された F1 個体の遺伝子型

に対して、育種価予測モデルを適用することにより、遺伝子型に対応した育種価を算出することができる。したがって、2 品種の交配により得られる F1 個体について、その遺伝子型と育種価を与えて、F1 集団における育種価の分布をシミュレートして形質分離を予測することができる。

Iwata ら (2013a) は、ニホンナシの 84 品種の集団を用いて、それらの形質データとゲノム全体に配置した 333 マーカーのハプロタイプで表された遺伝子型データにもとづいて育種価予測モデルを構築し、さらに 84 品種を用いたすべての可能な交配組合のそれぞれに対して、1000 個体から成る F1 集団を想定し、それぞれの F1 個体の遺伝子型を交配親の遺伝子型データからシミュレーションにより生成し、育種価予測モデルを適用して、F1 個体における育種価の分布を予測した。この予測された F1 個体の育種価の分布をもとに、育種目標を満たす F1 個体が出現する確率を評

図 6. F1 個体における予測育種価の分布 (Iwata et al. (2013) の図を改変)。

ニホンナシの 2 品種、菊水 (図中の白抜きの丸) と早生幸蔵 (白抜きの三角) の交配を想定してシミュレーションにより生成された 1000 個体からなる F1 集団の収穫期 (横軸) と果実重 (縦軸) における予測育種価の頻度分布。菊水と早生幸蔵の交配から得られた幸水の実際の育種価 (白抜きの星) は、この F1 集団の左上にプロットされている。

価して、望ましい交配組合せの選択が可能であることを示した。

予測された F1 個体の育種価分布の例として、Iwata ら (2013a) に示された 2 品種「菊水」と「早生幸蔵」の交配におけるシミュレーションの結果が図 6 に示されている。ニホンナシにおける現在の国内の主要品種である「幸水」は、この 2 品種の交配により作出されており、収穫期が早く（早生）かつ果実も中程度に大きいという望ましい特徴を備えており、長年にわたり消費者に好まれ、我が国におけるニホンナシの生産量のトップの座を占めている。「幸水」よりも早生で果実重が大きい品種を作出することが、現在のニホンナシ育種の育種目標の一つにもなっている。そこで「菊水」と「早生幸蔵」の交配を想定して、1000 個体の F1 個体の遺伝子型をシミュレーションにより生成して、収穫期と果実重における予測された育種価の分布に「菊水」、「早生幸蔵」、「幸水」の収穫期と果実重の観測値をプロットすると図 6 のようになる。ただし、収穫期と果実重の形質値については、標準化された値が示されている。この観測値は複数年にわたり得られた形質値の平均であり、育種価を反映する値とみなしてよい。図より、生成された 1000 個体の中に収穫期と果実重において「幸水」に近接した個体は存在するが、「幸水」よりも収穫期が早く、かつ果実重が重い F1 個体は得られておらず、この交配組合せでは「幸水」よりも優れた品種を得ることは困難であることがわかる。すなわち、過去において適切な個体選抜が行われて「幸水」という品種が作出されたことを意味している。

Iwata ら (2013a) においては、「幸水」よりも早生で果実重が大きい品種を作出することを育種目標として、いくつかの品種を用いた交配組合せに対してシミュレーションによる F1 個体の育種価の分布を算出し、この育種目標を達成する F1 個体の出現確率を求め、有望な交配組合せの選定が可能となることを示している。

まとめ

家畜育種で提案された GS は、近年、作物育種においてもその実用化が盛んに議論されている。作物育種においては、自殖性作物の交雑育種において、上述したような交配後代個体のホモ化を効率的に進めるための DH の作出技術や短期間に数世代を進めるための世代促進の技術など、すでに実用化されている技術がいくつかあり、それらを GS と組み合わせることにより

さらなる育種の高速化を目指す動きがある (De La Fuente ら 2013)。また、各種作物で次世代シーケンサーを用いた大規模なゲノム情報の取得が進められているが、これと並行してドローンやセンサーネットワークなどを活用することにより、圃場における作物の大規模な表現型情報の収集技術の開発も進められている。ドローンやセンサーネットワークなど形質測定をハイスループット化するためのツールも、GS と関連した文脈で論じられることが多い。作物育種において、GS はこれらの育種技術と連携しながら、さらにビッグデータの収集・解析や AI 技術の活用と結びつくことにより、「スマート育種」というキーワードのもとでより大きなうねりとなって進展していく様相を呈している。

文献

- Bassi FM, Bentley AR, Charmet G, Ortiz R, Crossa J. 2016. Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). *Plant Science*, 242: 23-36.
- Bernard R, Yu J. 2007. Prospect for genome-wide selection for quantitative traits in maize. *Crop Science*, 47: 1082-1090.
- Beyene Y, Semagn K, Mugo S, Tarekegne A, Babu R, Meisel B, Sehabiague P, Makumbi D, Magorokosho C, Oikeh S, Gakunga J, Vargas M, Olsen M, Prasanna BM, Banziger M, Cross J. 2015. Genetic gains in grain yield through genomic selection in eight biparental maize populations under drought stress. *Crop Science*, 55: 154-163.
- Biscarini F, Nazzicari N, Bink M, Arús P, Aranzana MJ, Verde I, Micali S, Pascal T, Quilot-Turion B, Lambert P, da Silva Linge C, Pacheco I, Bassi D, Stella A, Rossini L. 2017. Genome-enabled predictions for fruit weight and quality from repeated records in European peach progenies. *BMC Genomics*, 18:432
- Burgueño J, Crossa J, Cotes JM, San Vicente F, Das B. 2011. Prediction assessment of linear mixed models for multi-environment trials. *Crop Science*, 51: 944-954.
- Burgueño J, de los Campos G, Weigel K, Crossa J. 2012. Genomic prediction of breeding values when modeling genotype x environment interaction using pedigree and dense molecular markers. *Crop Science*, 52: 707-719.
- Cooper M, van Eeuwijk FA, Hammer GL, Podlich DW, Messina C. 2009. Modeling QTL for complex traits: Dissection and context for plant breeding. *Current*

- Opinion in Plant Biology, 12: 231-240.
- Cornelius PL, Seyedsadr M, and Crossa J. 1992. Using the Shifted Multiplicative Model to Search for 'Separability' in Crop Cultivar Trials. *Theoretical and Applied Genetics*, 84: 161-172.
- Dalton R. 2009. No bull: genes for better milk. *Nature*, 457: 369.
- De La Fuente GN, Frei UK, Lübberstedt T. 2013. Accelerating plant breeding. *Trend in Plant Science*, 18: 667-672.
- Denis JB, Piepho HP, van Eeuwijk FA. 1997. Modelling expectation and variance for genotype by environment data. *Heredity*, 79: 162-171.
- Finlay K, Wilkinson G. 1963. The analysis of adaptation in a plant-breeding programme. *Crop Pasture Science*, 14: 742-754.
- Gauch HG. 2006. Statistical analysis of yield trials by AMMI and GGE. *Crop Science*, 46:1488-1500.
- Gaynor RC, Gorjanc G, Bentley AR, Ober ES, Howerll P, Jackson R, Mackay IJ, Hickey JM. 2017. A two-part strategy for using genomic selection to develop inbred lines. *Crop Science*, 57: 2372-2386.
- Gollob HF. 1968. A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika*, 33: 73-115.
- He S, Schulthess AW, Mirdita V, Zhao Y, Korzun V, Bothe R, Ebmeyer E, Reif JC, Jian Y. 2016. Genomic selection in a commercial winter wheat population. *Theoretical and Applied Genetics*, 129: 641-651.
- Heslot N, Akdemir D, Sorrells M, Jannink JL. 2014. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theoretical and Applied Genetics*, 127: 463-480.
- Huang M, Cabrera A, Hoffstetter A, Griffey C, Sanford D, Costa J, McKendry A, Chao S, Sneller C. 2016. Genomic selection for wheat traits and trait stability. *Theoretical and Applied Genetics*, 129: 1-14.
- Iwata H, Hayashi T, Terakami S, Takada N, Saito T, Yamamoto T. 2013. Genomic prediction of trait segregation in a progeny population: a case study of Japanese pear (*Pyrus pyrifolia*). *BMC Genetics*, 14: 81.
- Iwata H, Hayashi T, Terakami S, Takada N, Sawamura Y, Yamamoto T. 2013. Potential assessment of genome-wide association study and genomic selection in Japanese pear *Pyrus pyrifolia*. *Breeding Science*, 63: 125-140.
- Jarquín D, Crossa J, Lacaze X, Du Cheyron P, Daucourt J, Lorgeou J, Piraux F, Guerreiro L, Perez P, Calus M, Burgueño J, de los Campos G. 2014. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theoretical and Applied Genetics*, 127: 595-607.
- Knol E, Nielsen B, Knap PW. 2016. Genomic selection in commercial pig breeding. *Animal Frontiers*, 6: 15-22.
- Kumar K, Chagne D, Bink MCAM, Volz RK, Whitworth C, Carlisle C. 2012. Genomic Selection for Fruit Quality Traits in Apple (*Malus × domestica* Borkh.). *PLoS One*, 7: e36674.
- Malosetti M, Bustos-Korts D, Boer M.P, van Eeuwijk FA. 2016. Predicting responses in multiple environments: Issues in relation to genotype x environment interactions. *Crop Science*, 56: 2210-2222.
- Mandel J. 1969. The partitioning of interactions in analysis of variance. *Journal of Research of the National Bureau of Standards, Series B*, 73: 309-328.
- Meuwissen THE, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157: 1819-1829.
- Minamikawa M, Nonaka K, Kaminuma E, Kajiya-Kanegae H, Onogi A, Goto S, Yoshioka T, Imai A, Hamada H, Hayashi T, Matsumoto S, Katayose Y, Toyoda A, Fujiyama A, Nakamura Y, Shimizu T, Iwata H. 2017. Genome-wide association study and genomic prediction in citrus: Potential of genomics-assisted breeding for fruit quality traits. *Scientific Reports*, 7: 4721
- Michel S, Ametz C, Gungor H, Epure D, Grausgruber H, Löschenberger F, Buerstmayr H. 2016. Genomic selection across multiple breeding cycles in applied bread wheat breeding. *Theoretical and Applied Genetics*, 129:1-11.
- Muranty H, Troggio M, Sadok IB, Rifai MA, Auwerkerken A, Banchi E, Velasco R, Stevanato P, van de Weg WE, Di Guarudo M, Kumar S, Laurens F, Bink MC. 2015. Accuracy and response of genomic selection on key traits in apple breeding. *Horticulture Research*, 2:15060.
- Piepho HP. 1997. Analyzing genotype-environment data by mixed models with multiplicative terms. *Biometrics*, 53: 761-766.
- Poland J, Endelman J, Dawson J, Rutkoski J, Wu S, Manes Y, Dreisigacker S, Cross J, Sanchez-Villeda H, Sorrells M, Jannink JL. 2012. Genomic selection in wheat breeding using genotyping-by-sequencing. *The Plant Genome*, 5: 103.

- Saint Pierre C, Burgueño J, Crossa J, Fuentes Dávila G, Figueroa López P, Solís Moya E, Ireta Moreno J, Hernandez Muel VM, Zamora Villa VM, Vikram P, Mathews K, Sansaloni C, Sehgal D, Jarquin D, Wenzl P, Singh S. 2016. Genomic prediction models for grain yield of spring bread wheat in diverse agro-ecological zones. *Scientific Reports*, 6: 27312.
- Schulz-Streeck T, Ogutu J O, Gordillo A, Karaman Z, Knaak C, Piepho HP. 2013. Genomic selection allowing for marker-by-environment interaction. *Plant Breeding*, 132: 532-538.
- Smith AB, Cullis BR, Thompson R. 2001. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics*, 57: 1138-1147.
- Song J, Caver BF, Yan L, Klapste J, Ei-Kassaby YA, Chen C. 2017. Practical application of genomic selection in a doubled-haploid winter wheat breeding program. *Molecular Breeding*, 37: 117.
- Speed D, Balding, DJ. 2014. Relatedness in the post-genomic era: Is it still useful? *Nature Review Genetics*, 17: 33-44.
- van Eeuwijk FA, Bink MCAM, Chenu K, Chapman SC. 2010. Detection and use of QTL for complex traits in multiple environments. *Current Opinion in Plant Biology*, 13: 193-205.
- Vivek BS, Krishna GK, Vengadessan V, Babu R, Zaidi PH, Kha LQ, Mandal SS, Grudloyma P, Takalkar S, Krothapalli K, Singh IS, Ocampo ETM, Xingming F, Burgueño J, Azrai M, Singh RP, Cross J. 2017. Use of genomic estimated breeding values results in rapid genetic gains for drought tolerance in maize. *The Plant Genome*, 10(1). DOI: 10.3835/plantgenome2016.07.0070
- Wolc A, Kranis A, Arango J, Settar P, Fulton JE, O'Sullivan NP, Advendano A, Watson KA, Hickey JM, de los Campos G, Fernando RL, Garrick DJ, Dekkers JCM. 2016. Implementation of genomic selection in the poultry industry. *Animal Frontiers*, 6: 23-31.
- Zhang X, Pérez-Rodríguez P, Semagn K, Beyene Y, Babu R, López-Cruz MA, San Vicente F, Olsen M, Buckler E, Jannink JL, Prasanna BM, Cross J. 2015. Genomic prediction in biparental tropical maize populations in water-stressed and well-watered environments using low-density and GBS SNPs. *Heredity*, 114: 291-299.