

# 育種と統計とデータさがけ

小野木 章雄

育種とは、ウシやブタ、コムギやイネなど人の食料となる生物を、その生産性を高めるように遺伝的に改良していくことである。今日スーパーマーケットで見かける穀物、野菜、肉のほとんどは長年にわたる育種改良で生み出されたものであり、その過程なくしては今の豊かな食生活は実現できなかった。一般の人はあまり馴染みがないと思われる育種であるが、歴史的に統計学との関わりが深く、データが重要な役割を果たしてきた応用分野である。本稿では育種における統計学とデータについて概説し、その流れの中で行われている筆者のさがけ研究「膨大なレガシー栽培データを蘇生する（データさがけ）」について紹介したい。

キーワード：農業、遺伝、ゲノム

## 1. はじめに

育種とは人間にとって有用な動植物、たとえばウシやブタ、コムギやイネを、より人間にとって好ましいように遺伝的に改良することを指す。たとえば乳牛であればより多くの乳を生産するように、またコムギであればより高い収量が得られるように改良する行為である。どのように育種するかは繁殖方法の違いに従い生物種によって多様であるが、共通しているのは優れた親同士を交配し、得られた子の中からまた優れたものを選び交配するという点を繰り返す点である。この「優れたものを人間が選ぶ」という点を「環境に適応したものを自然が選ぶ」と置き替えると、育種の過程は自然選択による生物進化のそれと同じであることがわかる。私たちが普段食するほとんどの穀物や野菜、肉は育種によりその生産性が改良されてきた。もし育種をしてこなかったら、つまり野生の生物をそのまま栽培あるいは飼育してきただけでは、今日の安定した豊かな食生活は実現不可能である。

育種は統計学と関わりの深い農学分野であり、またゲノムを始めとする膨大な生命データの登場によりデータサイエンスとも今後つながりが深くなっていくことが予想される分野でもある。そこで本稿では育種と統計学との関わりや育種で扱われるデータについてより多くの人に知ってもらうことを目的とした。統計学はこれまで特に動物の育種において多大な貢献をしてきたため、まず動物育種で広く利用されている統計学的

予測手法の紹介と、その遺伝学的背景について概説する。次に近年急速に収集が進んでいるゲノムデータにより、育種がどのように進展しているかを紹介する。さらにそのゲノムデータの登場によってこれまで統計学的予測と比較的遅かった植物（特に作物）の育種に起こりつつある変革と、その流れの中で進行中である筆者のさがけ研究「膨大なレガシー栽培データを蘇生する（データさがけ）」について紹介したい。

## 2. 統計学を必要とした動物育種

育種の効率を左右する重要な点はいかに正確に優れたものを多数の候補から選べるかにある。どの個体が優れているかは実際に飼育すれば簡単にわかるように思えるが、実はそれほど単純ではない。たとえばおいしい肉をもつウシを選ぶことを考えると、味を評価するために屠畜してしまうとその個体は子孫を残すことができなくなる。また乳牛では優れた雄牛を選ぶにも、雄は乳を出さないのどれが優れているのかわからない。また体重がよく増加してもそれは飼養管理がよかったせいで、その個体の子供を作ってもその性質は受け継がれないかもしれない。またよりよいものを見つけるには候補は多いほうがよいが、候補を作出し飼育するためにかかる多くの労力や費用を考えると、候補は少なくせざるを得ない。こうした問題の中で優れたものを選ぶためには、体重や品質などを実際に評価する前に予測する、あるいは環境による変動を除いて遺伝的な優劣を推定するという操作が必要であった。これらの必要性から動物育種では統計学的手法が活発に研究・開発され、さらにその学習に必要なデータが収集されてきた。

おのぎ あきお  
 東京大学大学院農学生命科学研究科  
 〒 113-8657 東京都文京区弥生 1-1-1  
 独立行政法人科学技術振興機構、さがけ  
 〒 332-0012 埼玉県川口市本町 4-1-8  
 onogiakio@gmail.com

### 3. 遺伝学的背景

どのような統計学的手法が用いられるかを述べる前に、まず育種の対象となる体重や品質などの生物の性質についての遺伝学的背景を概説したい。遺伝学では生物の性質や特徴のことを形質 (trait) と呼ぶ。形質は毛色や花の色など非連続的な分布 (白, 赤など) をする質的形質と、体重や収量など連続的な分布 (40 kg から 60 kg など) をする量的形質に分けられる。一般に質的形質は少数の遺伝子によって支配されることが知られている。たとえばメンデルによる遺伝法則発見の材料となったエンドウマメのしわの有無は Starch branching enzyme I という酵素を生成する遺伝子により決定される。父母から受け継いだ遺伝子双方に変異があるとしわが生じ、いずれか一方でも正常であるとしわは生じない [1]。またキンギョソウの花の色も単一の遺伝子により決定される。父母双方から受け継いだ遺伝子に変異があると白、いずれか一方ならピンク、双方正常だと赤を示す [1]。

一方で量的形質は多数の遺伝子により支配される。遺伝子が多数関わるために 1 遺伝子当たりの影響は比較的小さい。育種の対象となる形質の多く、たとえば肉質や収量などはこの量的形質に含まれる。関わる遺伝子やその数は形質および生物種に依存するが、これまですべての遺伝子が特定された形質はほとんどない。多数の遺伝子を考えると形質が連続的に分布するのは自然であることがわかる。仮に 100 個の遺伝子がある形質に関わるとして、個々の遺伝子に正常型と変異型があると、それらの形質への影響度をそれぞれ 0 と 1 で表すとすると、もし父母双方から 100 個の遺伝子すべてで変異型を受け継いだとすると、その個体の形質は (環境による変動を考えないと)  $100 \times 2$  (父母から一つずつ受け継ぐため) = 200 となり、逆に父母双方からすべての遺伝子で正常型を受け継ぐと形質の値は 0 になる。ただしいずれの場合も起こりうる頻度は低く、多くの場合は正常型と変異型を異なる割合で受け継ぐため形質は連続的な分布となる。つまり量的形質の遺伝は質的形質におけるメンデル遺伝、先のしわの有無などを多数の遺伝子へ拡張することで説明できる。

親子や兄弟でさまざまな量的形質、身長や体重、声や顔つき、あるいは疾患へのかかりやすさなどが似るのは多くの人が経験することである。この血縁者が似るという現象も多数の遺伝子が関わる量的形質の重要な特徴である。なぜ似るのかを全兄弟 (父母が同じ兄弟) で説明しよう。

ある一つの遺伝子に着目し、父親は A 型と B 型 (A/B と表記する)、母親は C 型と D 型 (C/D) をもつとする (便宜上ここでは正常型、変異型は使わない)。この父母から生まれる子の遺伝子組み合わせは A/C, A/D, B/C, B/D の四つであり、それぞれ確率は  $1/4$  である。これら組み合わせと確率は兄弟で同じであるため、たとえば兄が A/C で弟も A/C である確率は  $1/16$  となる。この場合兄弟は父母から受け継ぐ二つの型を共有していることになる。では兄弟で二つの型を共有する確率の合計はいくらになるだろうか。これはほかの組み合わせ (A/D, B/C, B/D) についても同様の計算を行えばよく、 $1/16 \times 4$  で  $1/4$  となる。

次に兄弟で一つの型を共有する確率を考える。これはたとえば兄が A/C で弟が B/C (C のみ共有) の場合などである。この確率は  $1/16$  となり、これをほかの組み合わせも同様に計算し和を取ると  $1/2$  となる。全く共有しない場合は兄が A/C、弟が B/D などの場合で、確率の合計は  $1/4$  である。そのため共有する型の数の期待値は  $1/4 \times 2 + 1/2 \times 1 + 1/4 \times 0$  で 1 となる。つまり一つの遺伝子考えた場合、全兄弟がもつそれぞれ二つずつの型のうち平均一つは共有していることになる。100 個の遺伝子考えた場合は全兄弟で合計 200 ずつの型のうち、平均 100 ずつ共有している計算となる。共有している型が正常型であれ変異型であれ、平均半分を共有しているために形質が似るのである。またあくまで平均であるために、兄弟によっては 100 以上共有していること (そっくりな場合) もあれば、100 以下 (全く似ていない場合) もありうる。ほかの血縁関係を考えると、たとえばいとこの場合は共有する数は 1 遺伝子当たり 0.25 となり、似かよいはずっと低くなる。個々の遺伝子の形質への影響力が等しいとすると、遺伝子の共有割合は血縁者間における形質の相関を決める。

まとめると形質が似るとは遺伝子が共有されていることであり、共有割合は血縁により期待値が決まる。育種における重要な論文の一つに 1918 年の Ronald Fisher による論文がある [2]。この中で Fisher は、多数の遺伝子を想定することで血縁者間に見られる形質の相関を説明できることを示した。Fisher が導入した多数 (正確には無数) の遺伝子による量的形質の支配という考えは、極微小モデル (infinitesimal model) と呼ばれ今日でも育種理論を支えている。

### 4. 統計学的手法とデータ

血縁者間における形質の相関は、前述した育種にお

ける課題、つまり品質や体重などの予測や環境による変動の除去、遺伝的な優劣の推定を行うにあたり重要な役割を果たしている。端的に言えば個体間における形質の相関を遺伝子の共有割合で定義した回帰モデルを用いる。この回帰モデルは線形混合効果モデル (linear mixed effect model) と呼ばれ、動物育種で広く用いられている。このモデルでは形質の観察値 ( $\mathbf{Y}$ ) を考慮できる環境要因を表す固定効果 (fixed effect)  $\mathbf{B}$  と遺伝要因を表す変量効果 (random effect)  $\mathbf{U}$  に回帰する。

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{Z}\mathbf{U} + \mathbf{E}$$

ここで  $\mathbf{X}$  と  $\mathbf{Z}$  はそれぞれの効果と  $\mathbf{Y}$  を結ぶ計画行列で  $\mathbf{E}$  は残差を表す。考慮できる環境要因とはたとえば飼育された年や場所の影響などを指す。考慮できない環境要因、たとえば個体ごとに微細に異なるような栄養状態などは  $\mathbf{E}$  に吸収される。育種に必要なことは「遺伝的に優れたもの」を選ぶことであるため  $\mathbf{U}$  の推定に主眼が置かれる。変量効果は確率分布に従うと想定される効果であるが (一方固定効果は確率分布を想定しない)、前述のとおり形質の相関は遺伝子の共有割合で決まるため、 $\mathbf{U}$  は以下のような多変量正規分布 (multivariate normal distribution, MVN) に従うと考える。

$$\mathbf{U} \sim \text{MVN}(\mathbf{0}, \mathbf{A}\sigma_{\mathbf{U}}^2)$$

ここで  $\mathbf{0}$  は要素がすべて 0 のベクトルで分布の平均を表す。 $\mathbf{A}$  は全個体ペアについての遺伝子の共有割合を含む行列 ( $\mathbf{A}$  行列) であり、 $\sigma_{\mathbf{U}}^2$  は分散を表す。 $\mathbf{A}\sigma_{\mathbf{U}}^2$  は分布の分散共分散行列となる。 $\mathbf{E}$  は以下のように個体間で独立と想定する。

$$\mathbf{E} \sim \text{MVN}(\mathbf{0}, \mathbf{I}\sigma_{\mathbf{E}}^2)$$

ここで  $\mathbf{I}$  は単位行列を表す。このモデルにおいて  $\mathbf{A}$  は  $\mathbf{U}$  を  $\mathbf{E}$  から分離し推定するための重要な役割を担っている。推定された  $\mathbf{U}$  は育種価 (breeding value) と呼ばれその名前のおり育種、つまり優れたものの選択と交配を行うための判断材料になる。飼育することなく  $\mathbf{U}$  を知るためには、 $\mathbf{Y}$  が既知の個体群との血縁関係 (遺伝子の共有割合) がわかっているならば、多変量正規分布の性質から以下のようにして予測できる。

$$\mathbf{U}_{\text{new}} = \mathbf{A}_{\text{new}}\mathbf{A}^{-1}\hat{\mathbf{U}}$$

ここで  $\mathbf{A}_{\text{new}}$  は予測したい個体群と  $\mathbf{Y}$  が既知の個体群との  $\mathbf{A}$  行列であり、 $\mathbf{A}^{-1}$  は  $\mathbf{Y}$  が既知の個体群の  $\mathbf{A}$  の逆行列、 $\hat{\mathbf{U}}$  は  $\mathbf{Y}$  が既知の個体群について推定され

た  $\mathbf{U}$  を表す。

線形混合効果モデルの利用を考案したのは Charles Henderson であり、彼は分散 ( $\sigma_{\mathbf{U}}^2$  と  $\sigma_{\mathbf{E}}^2$ ) が既知の場合、線形混合効果モデルの解を以下のように解くことを提案した。

$$\begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + \mathbf{A}^{-1}\sigma_{\mathbf{E}}^2/\sigma_{\mathbf{U}}^2 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{B}} \\ \hat{\mathbf{U}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{Y} \\ \mathbf{Z}^T\mathbf{Y} \end{bmatrix}$$

これは混合モデル方程式 (mixed model equation, MME) と呼ばれ、この解 ( $\hat{\mathbf{U}}$ ) は残差を最小にするという点で最良で、期待値が 0 となる不偏性をもつために最良線形不偏予測量 (Best linear unbiased predictor, BLUP) と呼ばれる [3]。MME を解くためには分散の推定が必要であるが、最尤推定法では固定効果を推定する分だけ自由度が減少し偏りが生じる。そのため固定効果の推定を回避できるようにモデル式を変換したうえで最尤推定を行う制限付き最尤推定法 (restricted maximum likelihood estimation, REML [4]) がよく用いられる。

MME と REML の組み合わせは動物育種を支える重要な統計手法である。この一連の手法は予測量の名前から BLUP 法あるいは単に BLUP と呼ばれることが多いため、本稿でも以降 BLUP と呼ぶ。なお線形混合効果モデルは、ベイズ統計の立場から見ると  $\mathbf{B}$  に無情報な事前分布を、 $\mathbf{U}$  の事前分布に多変量正規分布を想定した回帰モデルであり、MME は  $\mathbf{B}$  と  $\mathbf{U}$  の最大事後確率 (maximum a posteriori) 推定に相当する。

BLUP を実用するためには形質の観察値 ( $\mathbf{Y}$ ) と血縁 ( $\mathbf{A}$ ) のデータが必須であり、それらを収集するシステムの構築が不可欠である。たとえば日本で行われている乳牛育種では、1974 年から全国の農場における乳生産データ ( $\mathbf{Y}$ ) を収集する牛群検定が開始され、また遺伝子の共有割合を計算するために必要な血統登録事業はより古く 1911 年から行われている。2011 年時点ではおよそ 300 万頭から得られた 6400 万件以上の乳生産データが  $\mathbf{U}$  の計算に用いられた [5]。乳牛は動物育種の中でパイロット的な存在であり、世界的に最も多くのデータが収集されている家畜であるが、ほかの家畜 (肉牛、豚、鶏など) においてもデータ量はここまで多くはないものの、さまざまな育種事業体により日々データ収集が行われている。

## 5. ゲノムデータの登場

ゲノムとは生物を形作るのに必要な遺伝情報のセツ

トであり、核酸の一種である DNA から構成される。DNA はアデニン、グアニン、シトシン、チミンの 4 種類の塩基からなり、この塩基の並び方 (配列) でタンパク質の構造を始め生物に必要な情報が決定される。ゲノムの総塩基数は生物によって異なり、たとえば人は約 30 億、ウシでは約 26 億、イネは約 4 億である。

この長い塩基配列のほとんどは同じ種の中で共有されているが、一部個体によって異なる部位がある。たとえばある人ではグアニンであるが別の人はシトシンの部位がある。このような部位を一塩基多型 (single nucleotide polymorphism, SNP) と言い、人であれば約 2100 万部位存在するという報告もある [6]。このうちの一部で、たとえばグアニンかシトシンかが正常型か変異型かのように形質を左右することにより、量的形質の連続的な分布や血縁者間の相関が生じる。

近年の測定技術の進展により全塩基配列が比較的安価かつ迅速に決定できるようになってきたが、それでも多数の個体について測定することは予算上困難なため、SNP の一部を測定することによりゲノムデータを得ることが多い。主要な動植物についてはすでにコマース的な SNP の測定キットが販売されており、新たな測定技術の開発も盛んである。農学分野であれば 1 個体当たり数千から数十万部位の SNP を測定することが多い。

こうしたゲノムデータ、つまり多数の SNP の情報は疾患原因遺伝子の特定などで医療に大きな貢献をしているが、育種においても進展をもたらしている。先に線形混合効果モデルにおける  $U$  の推定で、血縁から求めた遺伝子の共有割合 ( $A$ ) を用いることに触れた。しかし血縁から求める共有割合は平均値であるという欠点がある。よく似た全兄弟は似ていない全兄弟より多くの遺伝子を共有していると予想されるが、血縁から求めると同じ共有割合になる。しかし SNP を用いれば実際に共有している割合を推定することができる。つまりゲノム全体に存在する SNP の型 (グアニンかシトシンかなど) を比較し、実際に共有している割合を計算するのである。

SNP から得られたより詳細な共有割合を用いて  $U$  を BLUP により推定する手法はゲノミック BLUP と呼ばれるが、これは機械学習で広く用いられるガウス過程 (Gaussian process) にほぼ等しい。ゲノムデータによる形質の予測はゲノミックプレディクション (Genomic prediction)、その予測に基づく育種はゲノミックセレクション (genomic selection, GS) と呼ばれ、現在その利用が盛んに研究されている。

GS の概念は 2001 年に発表されたが [7]、最初に運用されたのは北米の乳牛育種で 2009 年のことである。SNP を用いればより正確に  $U$  を予測できるため、次世代の個体に対する優劣の決定がより正確となり、育種速度の劇的な向上に寄与している [8]。乳牛への利用は現在日本でも行われ、さらに肉牛やブタ、魚類などほかの動物種でも導入・検討が進んでいる [9–11]。

## 6. ゲノムデータが変える作物育種

植物、特にイネやコムギ、ダイズなどの作物の育種において、BLUP のような統計学的な予測手法はあまり適用されてこなかった。理由は多々あるが、一つには動物育種の肉や乳生産で直面する、選択したい個体の形質を計測することができない (肉にすると子を残せない、雄は乳を出さない) という制限がなく、統計的予測の必要が動物育種ほど高くなかったことが挙げられる。作物は収量や粒数などを計測し優れたものを選んだ後、その種子を翌年に残すことができる。もっとも個体の形質を予測する利点は作物育種にもある。たとえば温室で栽培し種子を採取することを考える。基本的に作物の栽培は 1 年に 1 回であるが、温室を使えば複数回栽培を行うことができ、優れた個体の選択と交配をその分繰り返すことができる。しかし温室栽培で優れたものが屋外の農場でも優れているとは限らないため、屋外で得られたデータをもとに、温室内で得られた種子から屋外で優れているものを予測できれば利点がある。

しかしながらこれに加えて作物育種が BLUP と相性の悪い別の理由もあった。作物では一つの個体から多数の種子 (つまり子供) が得られ、その中から優れたものを選ぶ必要があるが、一つの個体から得られる種子は全兄弟であるため血縁からでは優劣が区別できない。これは雌の産子数が限られるため全兄弟が少ない動物育種とは対照的である。しかしゲノムデータの浸透により、血縁のもつこのような限界を打ち破り、統計的予測を作物育種においても有用なツールにできる可能性が出てきた。この可能性は多くの研究者や育種家を惹きつけ、作物育種における GS 研究を活発にしている [12–15]。

## 7. レガシーデータの収集とデータ収集システムの構築へ

ダイズは食用のほかに植物油の原料や家畜への飼料としても使用されるため、近年特に中国を中心に需要が高まっている。また肉に代わる高タンパク食材と

して機能性の面でも注目が集まる。一方でダイズの収量はイネやトウモロコシなどほかの作物より低いことが知られており、育種による改良の余地が未だ大きい。日本におけるダイズ育種はたとえば茨城県などで戦前から行われていたが、国主導のもとで体制が整ったのは戦後のことである。現在は国立研究開発法人農業・食品産業技術総合研究機構（農研機構）の四つのセンター（東北農業研究センター、次世代作物開発研究センター、西日本農業研究センター、九州沖縄農業研究センター）と、北海道や長野県などの地方自治体において育種が行われている。

ダイズ育種もほかの作物同様、前述のように統計的予測によらない、実際に栽培して評価する方法で行われている。ゲノムデータの利用は統計的予測を可能にし、ダイズ育種を大きく効率化させる可能性をもっているが、実はここに大きな問題がある。統計的予測である以上、その正確さは学習に用いるデータのサイズに依存する。動物育種においては統計的予測の必要性から、データの収集とそのためのシステム構築が数十年にわたり続けられてきた。そのため先に述べた乳牛のように膨大な数のデータをもとにした予測が可能となっているのである。しかしダイズなどの作物ではそういった統計的予測を前提としたデータの収集や、そのために必要なシステム構築がなされておらず、現状十分な学習データを確保することができない。

この問題を解決するために、さきがけ研究「膨大なレガシー栽培データを蘇生する（データさきがけ）」ではダイズのデータを収集することを目的とし、平成27年10月から研究を開始している。この研究ではデータ収集先をレガシーデータに求めている。実はこれまでの育種においても作物に関する詳細なデータは毎年取られており、冊子やファイルなどに記録されてきた。しかしながらある年のデータはその年の評価が終わると通常顧みられることはなく、そのまま冊子やファイル内でレガシーとなっていた。しかしその数十年に及ぶデータは統計的予測の学習データとなる可能性があり、さらに作物において重要な気象など環境への応答機構を明らかにする手がかりともなりうる。そのためさきがけ研究では前述した農研機構のセンターや地方自治体の協力を得てレガシー化したデータを収集し、データベース化する作業を行っている。冊子である場合はデジタル化を行い、クレンジングや整理を行う。

現在までに約19万件のダイズ栽培記録（一つの系統、つまり遺伝的に均一な個体群、をある年一つの試験地で栽培した結果を1件とする）を収集し処理中であるが、この規模のデータは作物としては異例である。また合わせてゲノムデータと栽培時の気象条件についても収集を行い、国内のダイズ生産に広く貢献できるようなデータベースを作成している。将来的には動物育種で行われているような全国的かつ継続的なデータ収集システムの礎とすることを目指している。

## 8. おわりに

遺伝的な改良を目指す育種にとってゲノムデータは革新的かつ最重要データであるが、育種に関わるデータは現在、遺伝子の転写産物や代謝物などのオミックスと言われるほかの生命データや、ドローンなどで撮影された画像データ、センサーによる環境データなど数多い。こういった多様かつ膨大なデータから必要な情報を抜き出し、有機的に結び付け活用するための統計学やデータサイエンスは、今後さらに重要となると考えられる。本稿を機に特に統計学やデータサイエンスなどに興味をもつ読者に、育種という分野を少しでも知ってもらえると幸いである。

謝辞 さきがけ研究においてデータおよび種子を提供していただいている国立研究開発法人農業・食品産業技術総合研究機構、地方独立行政法人北海道立総合研究機構、長野県野菜花き試験場、および茨城県農業総合センターに厚く御礼申し上げます。

## 参考文献

- [1] D. L. Hartl, *Essential Genetics: A Genomics Perspective*, 6th edition, Chapter 2, Jones & Bartlett Learning, 2014.
- [2] R. A. Fisher, "The correlation between relatives on the supposition of Mendelian inheritance," *Transactions of the Royal Society of Edinburgh*, **52**, pp. 399–433, 1918.
- [3] C. R. Henderson, *Applications of Linear Models in Animal Breeding*, 3rd edition, L. R. Schaeffer (ed.), University of Guelph, 1984.
- [4] H. D. Patterson and R. Thompson, "Recovery of inter-block information when block sizes are unequal," *Biometrika*, **58**, pp. 545–554, 1971.
- [5] 独立行政法人家畜改良センター, 乳用牛評価報告, 第31号, 2011.
- [6] M. Nagasaki, J. Yasuda, F. Katsuoka, N. Nariai, K. Kojima, Y. Kawai, Y. Yamaguchi-Kabata, J. Yokozawa, I. Danjoh, S. Saito, Y. Sato, T. Mimori, K. Tsuda, R. Saito, X. Pan, S. Nishikawa, S. Ito, Y. Kuroki, O. Tanabe, N. Fuse, S. Kuriyama, H. Kiyomoto, A. Hozawa, N. Minegishi, J. D. Engel, K. Kinoshita, S. Kure, N. Yaegashi and M. Yamamoto, "Rare variant discovery by deep whole-genome sequencing of 1,070 Japanese individuals," *Nature Communications*, **6**, pp. 8018, 2015.

- [7] T. H. Meuwissen, B. J. Hayes and M. E. Goddard, "Prediction of total genetic value using genome-wide dense marker maps," *Genetics*, **157**, pp. 1819–1829, 2001.
- [8] A. Garcia-Ruiz, J. B. Cole, P. M. VanRaden, G. R. Wiggins, F. J. Ruiz-López and C. P. Van Tassell, "Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection," In *Proceedings of the National Academy of Sciences USA*, **113**, E3995–E4004, 2016.
- [9] A. Onogi, A. Ogino, T. Komatsu, N. Shoji, K. Simizu, K. Kurogi, T. Yasumori, K. Togashi and H. Iwata, "Genomic prediction in Japanese Black cattle: Application of a single-step approach to beef cattle," *Journal of Animal Science*, **92**, pp. 1931–1938, 2014.
- [10] H. Esfandyari, P. Bijma, M. Henryon, O. F. Christensen and A. C. Sorensen, "Genomic prediction of crossbred performance based on purebred Landrace and Yorkshire data using a dominance model," *Genetics Selection Evolution*, **48**, p. 40, 2016.
- [11] H. Y. Tsai, A. Hamilton, A. E. Tinch, D. R. Guy, J. E. Bron, J. B. Taggart, K. Gharbi, M. Stear, O. Matika, R. Pong-Wong, S. C. Bishop and R. D. Houston, "Genomic prediction of host resistance to sea lice in farmed Atlantic salmon populations," *Genetics Selection Evolution*, **48**, p. 47, 2016.
- [12] J. L. Jannink, A. J. Lorenz and H. Iwata, "Genomic selection in plant breeding: From theory to practice," *Briefings in Functional Genomics and Proteomics*, **9**, pp. 166–177, 2010.
- [13] J. Crossa, G. de los Campos, P. Perez, D. Gianola, J. Burgueno, J. L. Araus, D. Makumbi, R. P. Singh, S. Dreisigacker, J. Yan, V. Arief, M. Banziger and H.-J. Braun, "Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers," *Genetics*, **186**, pp. 713–724, 2010.
- [14] A. Onogi, O. Ideta, Y. Inoshita, K. Ebana, T. Yoshioka, M. Yamasaki and H. Iwata, "Exploring the areas of applicability of whole-genome prediction methods for Asian rice (*Oryza sativa* L.)," *Theoretical and Applied Genetics*, **128**, pp. 41–53, 2015.
- [15] D. Jarquin, J. Specht and A. Lorenz, "Prospects of genomic prediction in the USDA soybean germplasm collection: Historical data creates robust models for enhancing selection of accessions," *G3*, **6**, pp. 2329–2341, 2016.