

育種学情報のデータマイニングと効率的解析のためのプログラム開発

誌名	育種学研究 = Breeding research
ISSN	13447629
著者	岩田, 洋佳
巻/号	11巻4号
掲載ページ	p. 177-184
発行年月	2009年12月

農林水産省 農林水産技術会議事務局筑波産学連携支援センター
Tsukuba Business-Academia Cooperation Support Center, Agriculture, Forestry and Fisheries Research Council
Secretariat



特集記事

育種学情報のデータマイニングと効率的解析のためのプログラム開発

岩田洋佳

中央農業総合研究センター, つくば市, 〒305-8666

Development of data-mining methods and programs for plant breeding

Hiroyoshi Iwata

National Agricultural Research Center, Tsukuba 305-8666, Japan

キーワード

画像解析, 形状解析, 連鎖地図作成, アソシエーション解析, ゲノミックセレクション

はじめに

“正確な計測”は、育種にとって非常に重要な課題の一つである。作物の農業的能力を正確に計測することは、育種における選抜効率を高め、より高い能力をもつ品種を、より早く育成することにつながる。計測精度は、反復数の増加やブロックの設定などによって向上させることができるが、精度向上のための手段はそれだけではない。得られた情報を効果的・効率的に解析することで、前述した手段以上に精度を向上させることができる。

近年の分子生物学的実験技術の進歩により、DNA 多型、転写因子、タンパク質発現、代謝活性に関するデータの網羅的収集が可能になった (Ganal *et al.* 2009, Lister *et al.* 2009, Shinozaki and Sakakibara 2009)。また、形質データについても、計測技術・手法の進歩により多様なデータが収集・蓄積されている。作物の量的形質の主な変動要因のひとつである環境データも、時空間的に高密度に収集・蓄積されている。このようなデータを、作物の能力の正確な計測に積極的に活用するためには、単に多様なデータを収集・蓄積するだけでなく、得られたデータを効果的に組み合わせ、効率的に解析する手法を開発する必要がある。著者らは、これまで、そのような視点からデータ解析手法の開発と、その新手法を実用化するためのプログラム開発を進めてきた。ここでは、現在注力している研究を中心にこれまでの研究結果を紹介し、これからの研究展開について考えてみたい。

1. 画像解析により作物の形態情報を定量化する

作物の色、形、模様といった形態特徴は、嗜好、加工、流通、消費等に関連する重要形質の一つである。人間の

視覚に基づく認識力は大変優れており、形態特徴にみられる微細な違いも容易に感知できる場合が多い。しかし一方で、そのような違いを“感じる”ことはできても“計る”こと、ないし、数値化することは容易ではない。我々は、これら形態特徴を計るための“ものさし”として、画像解析に基づく計測システムを構築し、形態特徴の遺伝・育種学的研究を進めてきた。

著者らは、楕円フーリエ記述子 (elliptic Fourier descriptors: EFD; Kuhl and Giardina 1982) とよばれる数学的な形状記述法を用いて、作物の形状解析を行うシステムを構築した (図 1; 岩田 2005)。EFD は、輪郭をトレースする際の座標値の変動を波と見なし、この波の形を、様々な周波数 (調和数) の波の重ね合わせによって近似する (図 2)。波を重ね合わせる際、各周波数の波に重み付けがなされるが、この重み (フーリエ係数) に輪郭の形状情報が含まれる。そこで、この重みを直接、または、主成分分析等で縮約して解析することで、輪郭形状の定量的な評価・解析が可能となる。EFD の生物形状解析への応用の歴史は比較的長く、植物ではカバノキ属の葉形解析 (White *et al.* 1988)、動物ではカの羽形解析 (Rohlf and Archie 1984) が最初の応用例である。EFD は記述力が高く、様々な 2 次元輪郭の評価に普遍的に用いることができるため、様々な生物の形状解析に有用である。しかし、EFD を用いた形状解析には、画像解析、フーリエ変換、多変量統計解析等、様々な計算手法を必要とするため、データ解析を専門とする研究者でもなかなか容易に用いることができなかつた。そこで著者らは、ソフトウェアパッケージ SHAPE (<http://cse.naro.affrc.go.jp/iwatah/shape>; Iwata and Ukai 2002) を開発し、図 1 に示した一連の解析を誰でも容易に行えるようにした。SHAPE は、育種学だけでなく、生態学、分類学、古生物学、人類学、医学など、様々な研究分野で種々の生物形状の研究に利用され、これまでに 56 報の論文に引用されている (2009 年 7 月 5 日現在、Web of Science による)。

著者らは、EFDに基づく形状解析システムを、ダイコン根形の遺伝解析 (Iwata *et al.* 1998, 2000, 2004a, 2004b, 岩田・新倉 2005) や、カンキツ葉形の遺伝解析 (Iwata *et al.* 2002a, 2002b) に適用し、形の遺伝様式の解明や、遺

伝子型×環境交互作用の解明を試みた。ダイコン根形の遺伝解析では、EFDの主成分として抽出された形状特徴がそれぞれ異なる遺伝様式を示すこと (Iwata *et al.* 2000)、形状特徴毎に生長に伴う遺伝様式の変化 (Iwata *et al.* 2004a) や土壌の違いによる遺伝様式の変化 (Iwata *et al.* 2004b) の傾向が異なることを明らかにした (表1)。ダイコンの根径/根長比に関しては、広義・狭義ともに遺伝率が非常に高く、生長に伴う遺伝様式の変化や土壌による影響が小さいため、どの生長段階および土壌条件でも一貫した選抜が可能であることが分かった。一方、根の下部の太り (尻止まり) については、生長に伴う遺伝様式の変化が大きく、土壌による影響も大きいため、収穫適期における選抜、および、対象とする土壌条件での選抜が不可欠であることが分かった。このように、形を定量的に評価することで、これまで困難であったダイコン根形の遺伝様式について定量的・客観的に記述することが可能となり、ダイコン根形の遺伝的改良に向けて、世代の進め方や、選抜の場あるいは時期に関して具体的な指針を示すことができた (岩田・新倉 2005)。カンキツ葉形の遺伝解析では、ダイアレル分析によって葉長/葉幅比、葉身の重心の位置、翼葉の大きさの遺伝様式を明らかにし (Iwata *et al.* 2002a)、これら特徴に見られる遺

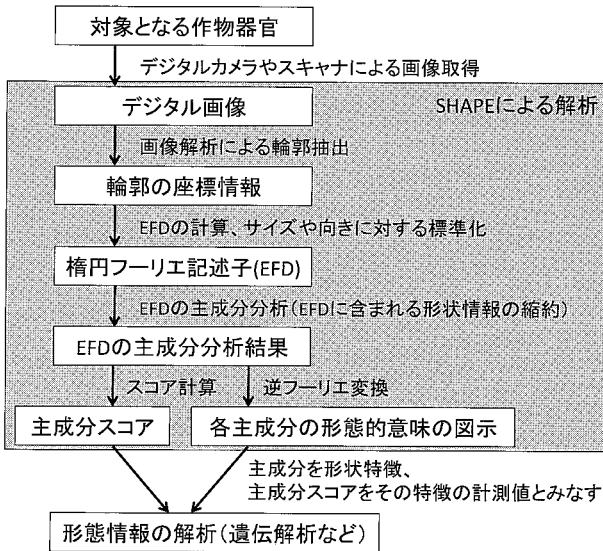


図1. 楕円フーリエ記述子(EFD)を用いた作物の形状解析システムにおける解析のながれ

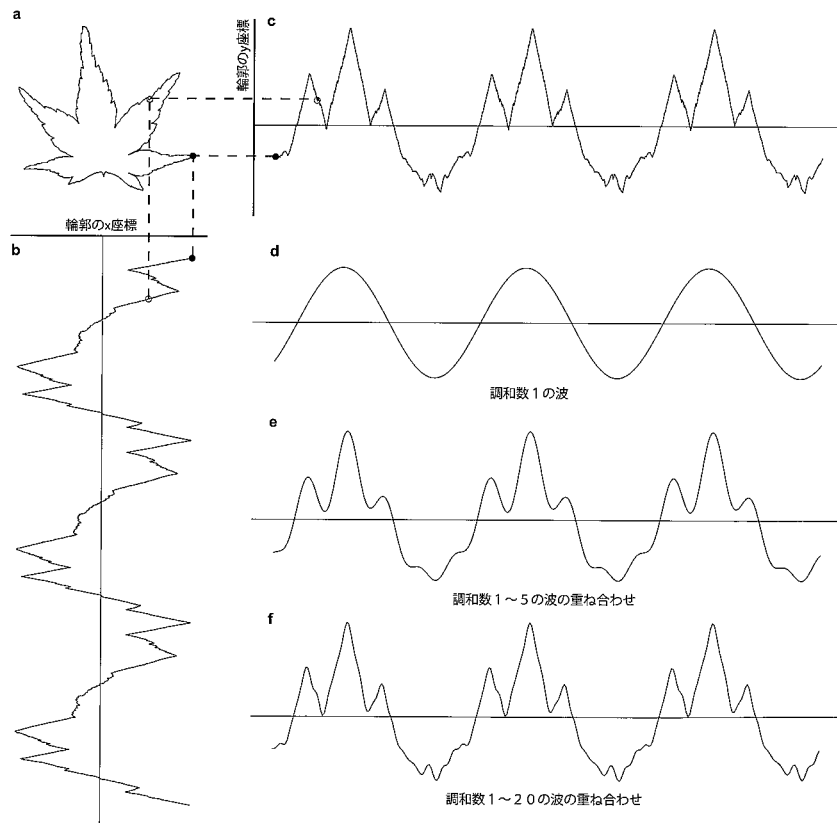


図2. 輪郭の楕円フーリエ記述子による記述

- (a) 解析対象の輪郭, (b) 輪郭を始点(●)からトレースした場合のx座標の変化, (c) 輪郭を始点からトレースした場合のy座標の変化, (d) y座標の変化を調和数1の波で近似した曲線, (e) 調和数1から5の波の重ね合わせで近似した曲線, (f) 調和数1から20の波の重ね合わせで近似した曲線

伝子型×環境交互作用の大きさとパターンを明らかにした (Iwata *et al.* 2002b). これら研究を通して、カンキツ葉形を種・品種の分類指標として利用することについて、遺伝学的な裏付けを与えることができた。

また、著者らは、EFDに基づく形状解析システムをイネの玄米形の評価に用いて、その評価値のアソシエーション解析を行った。これにより、イネ遺伝資源に見られる玄米形変異に関連する量的形質遺伝子座 (quantitative trait loci: QTL) の検出を試みた (Iwata *et al.* 2009b)。その結果、EFDの第1主成分は玄米の長幅比を評価しており、玄米形変異の90%以上を説明することが分かった。このことは、これまで玄米形のQTL解析によく用いられてきた長幅比が、玄米形変異の中で最も重要な尺度であることを示している。また、第1主成分についてアソシエーション解析を行った結果、玄米の長さや長幅比のQTLとして報告されていたQTLと一致する場所に有意なアソシエーションがいくつか検出された (図3)。その一方で、これまで評価が難しかった微細な形態特徴についても新規QTLが検出され、これら形態特徴についても遺伝的改良が可能であることが明らかとなった。また、逆フーリエ変換を用いた計算により、検出されたQTLの効果を視

覚的に表すことができることを示した (図3)。ここで用いられた方法は、閉曲線であればどのような形でも記述できるEFDの汎用性によって、イネだけでなく、多くの穀物の子実形の遺伝解析に有用である。将来は、イネ科作物における子実形変異に関連する遺伝子のシンテニー解析等も可能になるかもしれない (Iwata *et al.* 2009b)。

作物育種の現場において、育種家が視覚的に識別できても計測することが難しい形質は少なくない。画像解析に基づく計測は、このような形質の正確な計測を可能にし、選抜効率を高める可能性がある。また、画像解析に基づく形質計測は、手間や時間を要する表現型データ計測過程の効率化のためにも、今後、ますます重要になると考えられる。なお、ここでは、形状解析に関する研究例のみを紹介したが、植物の形態特性として重要なのは形だけではない。鑑賞用の園芸作物では色や模様の評価も重要である。現在では、画像解析に基づく評価により、花の色模様についても定量的解析が行えるようになっている (Yoshioka *et al.* 2004, 岩田ら 2007)。今後は、形や色模様等について定量化された情報と、その他の情報 (品質、嗜好等) 間に見られる関連を詳細に解析することにより、作物形態に関するより深い知見を得ることができよう。

表 1. ダイコン根形特徴の遺伝様式 (EFDの主成分得点のダイアレル分析の結果) (Iwata *et al.* 2000, 2004a, 2004b)

	形態的意味	遺伝率 (広義・狭義)	平均優性度	生長に伴う 遺伝様式の変化	土壌による影響
対称変異					
第1主成分	根径/根長比	0.99・0.96	0.29	小	小
第2主成分	根の下部の太り	0.86・0.71	0.60	大	大
非対称変異					
第1主成分	根の中間部の曲がり	0.67・0.54	0.57	大	小

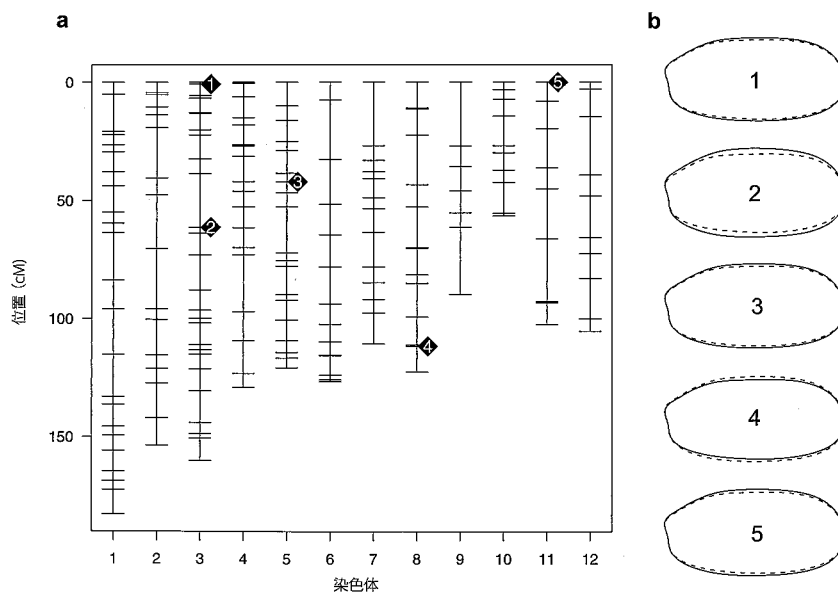


図 3. イネ玄米形のアソシエーション解析結果 (Iwata *et al.* 2009b)
(a) 有意なアソシエーションが検出されたマーカーの染色体上での位置 (◆), (b) 各 QTL が玄米形に与える遺伝効果 (実線・破線は、それぞれ、「日本晴」型・「Kasalath」型の対立遺伝子がホモ型である場合に期待される玄米形)

2. 連鎖地図を高速に作成する

近年、作物においても、多数の DNA マーカーの遺伝子型を高速・高効率にタイピングして、DNA マーカーが高密度にマップされた連鎖地図を作成できるようになってきた（例えば、Sato and Takeda 2009）。このような高密度連鎖地図は、QTL の map-based cloning や、DNA マーカー利用選抜（marker assisted selection: MAS）等を通して、望ましい特性を備えた新品種育成に大きく貢献する。しかし、多数のマーカーを用いて連鎖地図を作成する作業は簡単ではない。マーカー数が増えると、考えなければならないマーカー-遺伝子座の順序数が膨大になるからである。例えば、遺伝子座数が 5 の場合は可能な順序数は 60 だが、10 の場合で約 181 万、50 の場合は約 1.57×10^{64} となる。連鎖地図を作成するには、これら順序の中から観察データに最も適合する順序を選び出す必要があるが、遺伝子座数が多いと高速なコンピュータでも現実的な時間内に最適な順序を求めることは難しい。このような問題を解決するために、著者らは新しい方法を考案した（Iwata and Ninomiya 2006）。それは、蟻の群れが餌場までの最短距離を発見する仕組み（図 4a）にヒントを得た蟻コロニー最適化法（ant colony optimization: ACO; Dorigo *et al.* 1996）を応用した方法である。

遺伝子座の並べ替えは、巡回セールスマン問題の一種として捉えることができる（Liu 1998）。巡回セールスマン

ン問題とは、複数都市間の距離が与えられたときに全都市を巡る最短経路を求める問題である。遺伝子座の並び替えにおける巡回セールスマン問題では、“都市”が遺伝子座、“都市間の距離”が組換え価（または対数尤度の絶対値）に対応する（図 4b）。ACO は、コンピュータ上に“都市”間を移動する“蟻の群れ”をシミュレートし、“蟻の群れ”に最短経路を探索させる。蟻はその探索行動において、(1) 都市間に蓄積されたフェロモン量と都市間距離の両方を考慮した確率で、次に移動する都市を決定する（図 4c）、(2) 全都市を訪れた後、総移動距離に応じて移動してきた経路にフェロモンを放出する（量は移動距離の逆数に比例）、というルールに従う。ACO は、個々の蟻がこのような単純なルールに基づくことで、蟻が群れとして最短経路を割り出す仕組みを利用したアルゴリズムである。著者らは、ACO をもとに遺伝子座の最適な順序を高速計算するアルゴリズムを開発し、それを採用した遺伝子連鎖地図作成ソフトウェア AntMap (<http://cse.naro.affrc.go.jp/iwatah/antmap>) を開発した（図 5）。また、シミュレーション解析を行い、開発されたアルゴリズムによる順序推定の精度が高いことを確認した（Iwata and Ninomiya 2006）。AntMap が連鎖地図作製に要する時間は短く、染色体数 10、全遺伝子座数 1,200 のデータでも 10 分以内に地図が得られる（図 5, Intel Core 2 Duo 2.4 GHz で計測）。なお、高速計算が可能になったことにより、データの再抽出（ブートストラップ抽出）を繰り返

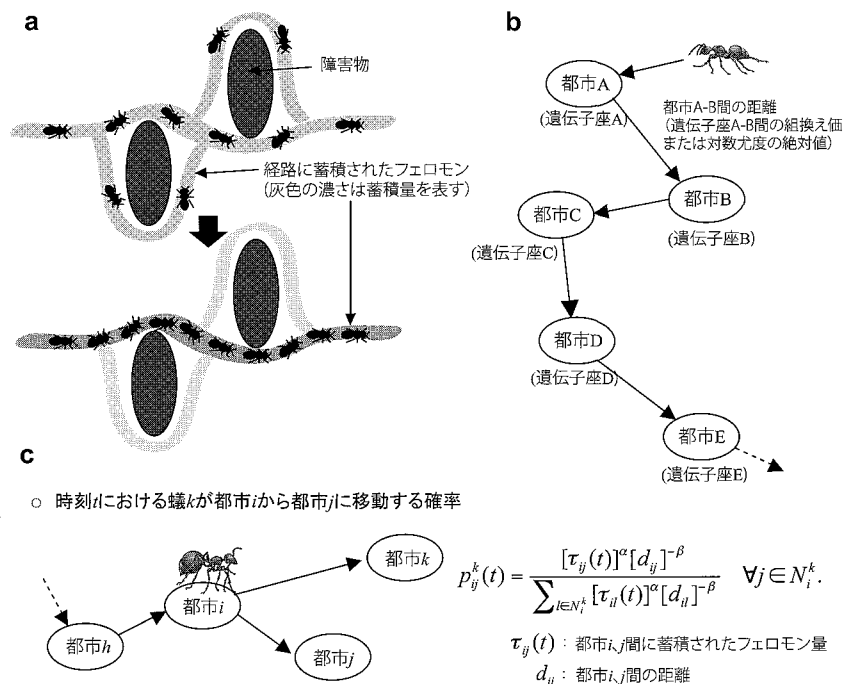


図 4. 蟻コロニー最適化 (ACO) アルゴリズムを用いた遺伝子座順序の推定

(a) 蟻は、巣から餌場までの最短距離を、群として行動する中で発見する。ACO は、この仕組みにヒントを得たアルゴリズムである。(b) 全都市（ある染色体に座乗する全遺伝子座）を最短距離（最小組換え価または最大尤度）で巡る経路を探索する。(c) 蟻は、都市間に蓄積されたフェロモン量と都市間距離に基づいて、次に移動する都市を確率的に選択する。

返して連鎖地図を作成することも可能となった。こうして繰り返し推定された順序について、繰り返し間のばらつきを求めることにより、順序が不確かな遺伝子座を検出することができる (Iwata and Ninomiya 2006)。

大量なデータが日常的に得られるようになった作物の遺伝・育種学研究分野では、今後、連鎖地図作製に限らず様々な場面で、高速・高効率な解析アルゴリズムの開発が不可欠になっていくと考えられる。

3. 作物遺伝資源を用いて形質変異を司る原因遺伝子を検出する

近年の分子生物学実験技術の向上により、さまざまな作物で、ゲノムワイドマーカーデータの収集・蓄積が始められている (Rostoks *et al.* 2005, McNally *et al.* 2009)。ゲノムワイドに収集されたマーカー情報が育種にもたらず果実のひとつとして、作物の遺伝資源を材料に用いたゲノムワイドアソシエーション解析 (Genome-wide association studies: GWAS) が挙げられる (Zhu *et al.* 2007)。アソシエーション解析 (association studies: AS) は、従来、ヒトの病気の原因遺伝子の探索に用いられてきた手法で、近年、作物においても、従来の QTL 解析法に代わる、もしくは、それを補完する手段として注目を浴びている。作物における AS は、遺伝資源等の既存の品種・系統か

らなる集団について DNA 多型と表現型多型を調査し、表現型多型と有意なアソシエーションが見られる DNA 多型を検出することで原因遺伝子の検出を試みる手法である (図 6; Balding 2006)。GWAS は、このうち、ゲノム全体に分布する DNA 多型を多数用いて解析するものをいう。AS は、従来の QTL 解析法に比べ、(1) 解像度が高い、(2) 多数の対立遺伝子を調査できる、(3) 交配実験を要しない等、様々な利点をもつ。その一方で、解析集団内にある潜在的な遺伝的分化により擬陽性・擬発見

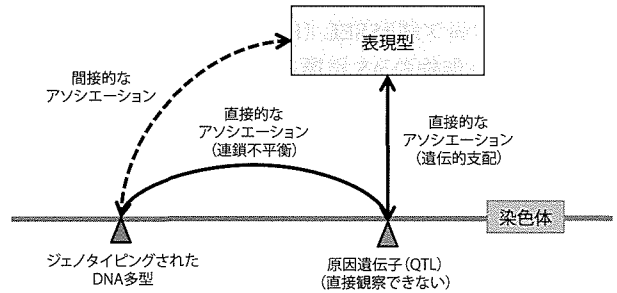


図 6. アソシエーション解析の仕組み (Balding (2006) より改変) 原因遺伝子 (QTL) の近傍にある DNA 多型が、原因遺伝子と連鎖不平衡の状態にあれば、原因遺伝子と DNA 多型間の連鎖不平衡 (直接的なアソシエーション) および原因遺伝子による表現型多型の支配 (直接的なアソシエーション) を介して、DNA 多型と表現型間に有意なアソシエーションが間接的に検出される。

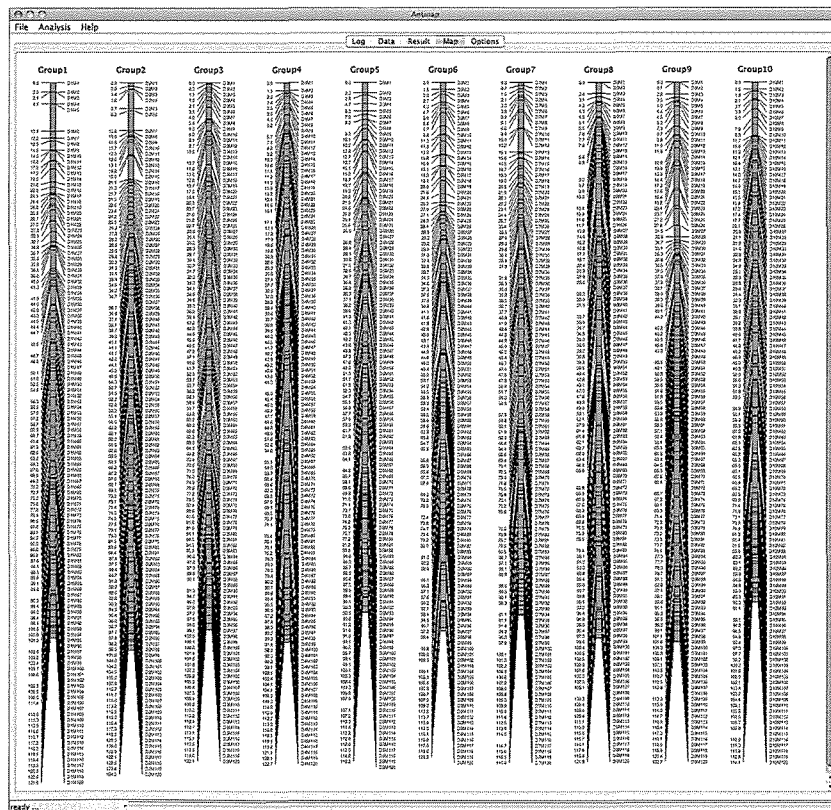
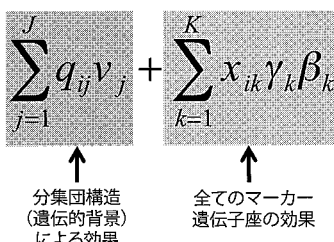


図 5. 遺伝子連鎖地図作製ソフトウェア AntMap の実行画面 実行画面内に表示されているのは、1,200 マーカー遺伝子座が座乗する地図を作成した結果。

が生じやすいという欠点をあわせ持つ。作物では、人為選抜や栽培環境への適応により、品種・系統間に遺伝的分化が進んでいる場合が多く、これを無視して解析すると検出されるアソシエーションが擬陽性や擬発見ばかりになってしまう (Iwata *et al.* 2007)。このような問題を解決するための方法として、Yu *et al.* (2007) は、系統間の遺伝的背景の違いをモデルに取り込む方法を提案した。提案された手法はトウモロコシのアソシエーション解析に適用され、同手法により遺伝的分化による擬陽性や擬発見を抑えられることを示した。Yu *et al.* (2007) の方法は、ソフトウェア TASSEL (Bradbury *et al.* 2007) に実装され、現在、作物の AS に標準的に用いられる手法の一つとなっている。しかし、Yu *et al.* (2007) の方法は、それぞれの DNA 多型について1つずつ検定を行うため、検定している位置以外に存在する QTL によって検定結果が攪乱される可能性がある。

そこで、著者らは、ベイズ的アプローチを用いて、複数の QTL の効果を同時に推定するモデル (図 7) を考案し、作物の GWAS に適用した。イネ遺伝資源で実際に計測された RFLP マーカーの遺伝子型データをもとにシミュレーション解析を行い、QTL 検出精度の評価を行った結果、複数 QTL モデルは、単一 QTL モデルに比べて、擬陽性や擬発見を抑える能力が高いことが示された (Iwata *et al.* 2007)。また、イネ遺伝資源で収集された実データ解析では、イネの玄米長や玄米形について、既知の QTL と一致する位置に有意なマーカーが検出された (Iwata *et al.* 2007, 2009b)。また、オオムギのゲノムワイド SNP データに基づくシミュレーション解析でも、複数 QTL を同時推定する著者らの方法が、マーカーを1つずつ解析する Yu *et al.* (2007) の方法に比べて、擬陽性、擬発見をより低いレベルで抑えられることが示されている

$$y_i = \mu + \sum_{j=1}^J q_{ij} v_j + \sum_{k=1}^K x_{ik} \gamma_k \beta_k + e_i$$



分集団構造
(遺伝的背景)
による効果

全てのマーカー
遺伝子座の効果

$$\gamma_k = \begin{cases} 0 & \text{マーカー } k \text{ の効果をモデルに含めない} \\ 1 & \text{マーカー } k \text{ の効果を QTL としてモデルに含める} \end{cases}$$

図 7. 複数 QTL の効果を同時推定するための遺伝モデル (Iwata *et al.* 2007)

モデルパラメータの推定にはマルコフ連鎖モンテカルロ (Markov chain Monte Carlo: MCMC) 法が用いられる。MCMC により推定された γ_k の事後平均 (MCMC により繰り返し抽出される γ_k の値の平均) の大きさをもとに、表現型多型とのアソシエーションが有意なマーカーを検出する。

(岩田ら未発表データ)。

著者らの手法は、ベイズ的アプローチの拡張性と柔軟性の恩恵をうけ、連続データだけでなく、スコアリングデータや打ち切りデータなど、作物でよく見られる“通常の連続データでないようなデータ”の解析にも応用できる (Iwata *et al.* 2009a)。スコアリングデータや打ち切りデータでは、観察されなかったデータ (スコアリングデータの場合、スコア化される前の連続データ、打ち切りデータの場合、計測が打ち切られなければ観察された連続データ) と実際に観察されたデータ間の関係をモデル化し、そのモデルに基づくアソシエーション解析を行う。イネ遺伝資源で実際に収集された RFLP データに基づくシミュレーション解析の結果、スコアリングデータや打ち切りデータでも妥当な精度で QTL を検出できることが分かった。また、完全なデータ (打ち切られたデータやスコアリングデータの背後にある連続データ) が観察できないにも関わらず、ベイズ的アプローチにより、QTL の遺伝効果を精度良く推定できることも分かった。なお、イネ遺伝資源を用いた実データ解析においても、既知の QTL との一致する位置に有意なマーカーが検出され、同手法の有用性が示された。

作物の遺伝資源は、さまざまな形質について定期的計測が行われ、そのデータがデータベースに蓄積されている。これらデータベースに蓄積された形質データを GWAS による原因遺伝子探索に活用できれば、それら形質の遺伝様式に対する理解が深まるとともに、遺伝資源の育種への活用がよりいっそう促進されることが考えられる。なお、ゲノムワイドマーカーが低密度にしかタイピングされていないマーカー遺伝子型データを、高密度にタイピングされたパネル (コアコレクション等) のマーカー遺伝子型データを用いて補完する方法も考案されており (例えば、Scheet and Stephens 2008)、作物でもその有用性が示されている (Jannink *et al.* 2009)。今後は、このような手法も利用して、既存のデータに含まれる有用な情報を最大限に活用できるような解析システムを構築していく必要がある。

4. ゲノムワイドマーカーを用いて品種・系統の育種価を予測する

ゲノムワイドマーカーが育種にもたらす果実は、GWAS による原因遺伝子の検出だけではない。ゲノムワイドマーカー情報を育種に活用する別のアプローチとして、ゲノミックセレクション (GS) が挙げられる。GS は、ゲノムワイドマーカー多型をもとに品種・系統の育種価を予測し、“表現型を観察することなく” 選抜を行おうとする手法である (図 8)。GS は、Meuwissen *et al.* (2001) によって提案されて以来、動物育種で急速に研究が進み、乳牛育成における雄牛の選抜で実用段階に入っている (Hayes *et al.* 2009)。しかし、植物育種での GS の研究と

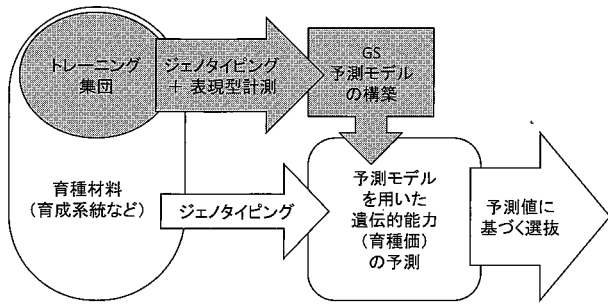


図 8. Genomic selection (GS) のながれ (Heffner *et al.* (2009) より改変)

育種材料のうち代表的なものをトレーニング集団として、マーカー遺伝子型のタイピングと表現型計測を行い、それらデータに基づき、GSのための予測モデルを構築する。その後、選抜対象となる材料についてマーカー遺伝子型のタイピングだけを行い(表現型計測は行わない)、GS予測モデルを用いて、得られたマーカー遺伝子型から予測される遺伝的能力(育種価)を求める。優良系統の選抜は遺伝的能力の予測値に基づいて行われる。

利用は、未だ黎明期にある (Heffner *et al.* 2009)。

著者らは、現在、作物育種におけるGSの有用性を明らかにするために、現在、様々なシミュレーション研究を進めている。オオムギの育種系統について収集されたゲノムワイドSNPデータを用いてシミュレーション研究を行った結果、各QTLの遺伝率が2%以下という極端なポリゾン支配の形質でも、GSの有効性が示された。さらに、GSの表現型選抜に対する相対的な有効性は、遺伝率の低い形質ほど高くなることが明らかにされた (Iwata and Jannink 2009)。また、永年性植物のスキをモデルに、60年間の育種過程のシミュレーションを行った結果、GSを利用することで育種効率を大幅に向上させることが示された (岩田ら 未発表データ)。

従来のMASでは、遺伝率のごく低い形質やポリゾン支配の形質による改良は難しかった。GSは、そのような形質における選抜効率を向上させる可能性が高い。作物の最重要形質である収量性は、まさにそのような形質であり、GSの実用的価値は高いと考えられる。また、選抜時に表現型データを必要としないGSは、表現型データの計測に時間と空間を多く必要とする林木や果樹等の永年性植物において、特に有用と考えられる。植物育種におけるGSの本格的な利用のため、今後は、エピスタシス効果や遺伝子型×環境相互作用を考慮したGSモデルの導入、GSの実証試験、GSを導入した育種プログラムの最適デザインの模索に関する研究等、様々な側面からの研究を進めていく必要がある。

5. おわりに

作物学者 Hugh G. Gauch は、その論文 (Gauch 2006) の中でこう述べている—“At present, worldwide breeding

efforts account for an average yield increase of about 1 percent per year in the major crops. I conservatively estimate that aggressive statistical analysis of the same data would make it possible to increase this average to 1.4 percent per year, at virtually no extra cost. Over a decade, that incremental gain would translate into enough food for millions of persons.” Gauch の言うように、データ解析に要するコストは、圃場実験や分子生物学実験に要するコストに比べて、極めて小さい。育種における費用対効果を上げ、優良な品種・系統をより速く、より効率的に育成していくためには、圃場実験や分子生物学実験で得られた貴重なデータを効果的・効率的に解析する手法を研究開発していくことが非常に重要である。特に、今後の世界的な人口増を考えると、作物収量のよりいっそう速いペースでの改良が必要となる。データ解析手法の開発・改良が、作物の遺伝的改良の加速化に果たせる役割は大きい。

なお、ここで紹介した研究の多くは、まだその緒に就いたばかりである。今後は、更なる精度の向上と、育種における実用化に向けて、より詳細で、より多角的な研究を進めていく必要がある。残念ながら、我が国には、ここで示したような手法研究に携わる研究者がきわめて少ない。今後は、そのような人材を育成していくことも重要な課題となる。本稿が、学生や若い研究者の人たちにとって、データ解析や解析手法の研究に興味をもってもらえるきっかけとなれば幸いである。

謝辞

本研究を遂行するにあたっては、共同研究者をはじめ多くの方々から、多大なるご協力とご支援を賜った。この場をお借りして、皆様に深く感謝の意を表したい。特に、私を統計遺伝学・生物測定学という魅惑的で心躍る世界にいざなって下さった鶴飼保雄先生 (元東京大学) に、深く感謝したい。また、私が研究者としての歩みを始める前からお世話いただいた高野泰先生 (元東京大学)、二宮正士先生 (中央農業総合研究センター)、大澤良博士 (筑波大学)、津村義彦博士 (森林総合研究所)、新倉聡博士 (株式会社トーホク)、現在、GWASとGSに関する研究を共に精力的に進めている江花薫子博士 (生物資源研究所)、林武司博士 (生物資源研究所)、Jean-Luc Jannink博士 (USDA-ARS) に深く感謝する。

引用文献

- Balding, D.J. (2009) A tutorial on statistical methods for population association studies. *Nat. Rev. Genet.* 7: 781–791.
- Dorigo, M., V. Maniezzo and A. Colomi (1996) The ant system: optimization by a colony of cooperating agents. *IEEE Trans. Syst. Man, Cybern.* B26: 29–41.
- Ganal, M.W., T. Altmann and M.S. Röder (2009) SNP identification in crop plants. *Curr. Opin. Plant Biol.* 12: 211–217.

- Gauch, H.G. (2006) Winning the accuracy game. *Amer. Scientist* 94: 133-141.
- Gonzalo, M.J., M.T. Brewer, D. Sullivan, S. Gray and E. van der Knaap (2009) Tomato fruit shape analysis using morphometric and morphology attributes implemented in tomato analyzer software program. *J. Amer. Soc. Hort. Sci.* 134: 77-87.
- Hayes, B.J., P.J. Bowman, A.J. Chamberlain and M.E. Goddard (2009) Invited review: Genomic selection in dairy cattle: progress and challenges. *J. Dairy Sci.* 92: 433-443.
- Heffner, E.L., M.E. Sorrells and J.L. Jannink (2009) Genomic selection for crop improvement. *Crop Sci.* 49: 1-12.
- 岩田洋佳 (2005) コンピュータ画像処理による2次元形態計測. *日本作物学紀事* 74: 371-373.
- Iwata, H., K. Ebana, S. Fukuoka, J.-L. Jannink and T. Hayashi (2009a) Bayesian multilocus association mapping on ordinal and censored traits and its application to the analysis of genetic variation among *Oryza sativa* L. germplasms. *Theor. Appl. Genet.* 118: 865-880.
- Iwata, H., K. Ebana, Y. Uga, T. Hayashi and J.-L. Jannink (2009b) Genome-wide association study of grain shape variation among *Oryza sativa* L. germplasms based on elliptic Fourier analysis. *Mol. Breed.* (in press).
- Iwata, H. and J.-L. Jannink (2009) A simulation study to evaluate the efficacy of genomic selection in barley breeding programs. *Plant and Animal Genome XVII Conference, Final Abstracts Guide.* p. 78.
- 岩田洋佳・新倉 聡 (2005) 植物の根に関する諸問題 [149] —— ダイコン根形の定量的評価と育種 ——. *農業および園芸* 80: 1296-1304.
- Iwata, H., H. Nesumi, S. Ninomiya, Y. Takano and Y. Ukai (2002a) Diallel analysis of leaf shape variations of citrus varieties based on elliptic Fourier descriptors. *Breed. Sci.* 52: 98-94.
- Iwata, H., H. Nesumi, S. Ninomiya, Y. Takano and Y. Ukai (2002b) The evaluation of genotype x environment interactions of citrus leaf morphology using image analysis and elliptic Fourier descriptors. *Breed. Sci.* 52: 243-251.
- Iwata, H., S. Niikura, S. Matsuura, Y. Takano and Y. Ukai (1998) Evaluation of variation of root shape of Japanese radish (*Raphanus sativus* L.) based on image analysis using elliptic Fourier descriptors. *Euphytica* 102: 143-149.
- Iwata, H., S. Niikura, S. Matsuura, Y. Takano and Y. Ukai (2000) Diallel analysis of root shape of Japanese radish (*Raphanus sativus* L.) based on elliptic Fourier descriptors. *Breed. Sci.* 50: 73-80.
- Iwata, H., S. Niikura, S. Matsuura, Y. Takano and Y. Ukai (2004a) Genetic control of root shape at different growth stages in radish (*Raphanus sativus* L.). *Breed. Sci.* 54: 117-124.
- Iwata, H., S. Niikura, S. Matsuura, Y. Takano and Y. Ukai (2004b) Genetic effects by soil type interaction in diallel analysis of root shape and size of Japanese radish (*Raphanus sativus* L.). *Breed. Sci.* 54: 313-318.
- Iwata, H. and S. Ninomiya (2006) AntMap: constructing genetic linkage maps using an ant colony optimization algorithm. *Breed. Sci.* 56: 371-377.
- Iwata, H., Y. Uga, Y. Yoshioka, K. Ebana and T. Hayashi (2007) Bayesian association mapping of multiple quantitative trait loci and its application to the analysis of genetic variation among *Oryza sativa* L. germplasms. *Theor. Appl. Genet.* 114: 1437-1449.
- Iwata, H. and Y. Ukai (2002) SHAPE: a computer program package for quantitative evaluation of biological shapes based on elliptic Fourier descriptors. *Euphytica* 102: 143-149.
- 岩田洋佳・吉岡洋輔・長谷暢一・松浦誠司・大森 宏 (2007) 花の色模様の客観的定量化と人による主観的認識の関係. *育種学研究* 9 (別2) : 116.
- Jannink, J.-L., H. Iwata, P.R. Bhat, S. Chao, P. Wenzl and G.J. Muehlbauer (2009) Marker imputation in barley association studies. *Plant Genome* 2: 11-22.
- Kuhl, F.P. and C.R. Giardina (1982) Elliptic Fourier features of a closed contour. *Comp. Graph. Image Proc.* 18: 236-258.
- Lister, R., B.D. Gregory and J.R. Ecker (2009) Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. *Curr. Opin. Plant Biol.* 12: 107: 118.
- Liu, B.H. (1998) *Statistical Genomics: Linkage, Mapping, and QTL Analysis.* CRC Press, New York.
- McNally, K.L., K.L. Childs, R. Bohnert, R.M. Davidson, K. Zhao, V.J. Ulat, G. Zeller, R.M. Clark, D.R. Hoen, T.E. Bureau, R. Stokowski, D.G. Ballinger, K.A. Frazer, D.R. Cox, B. Padhukasahasram, C.D. Bustamante, D. Weigel, D.J. Mackill, R.M. Bruskiewich, G. Ratsch, C.R. Buell, H. Leung and J.E. Leach (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc. Nat. Acad. Sci. USA* 106: 12273-12278.
- Rohlf, F.J. and J.W. Archie (1984) A comparison of Fourier methods for the description of wing shape in mosquitoes (Diptera: Culicidae). *Syst. Zool.* 33: 302-317.
- Rostoks, N., S. Mudie, L. Cardle, J. Russell, L. Ramsay, A. Booth, J.T. Svensson, S.I. Wanamaker, H. Walia, E.M. Rodriguez, P.E. Hedley, H. Liu, J. Morris, T.J. Close, D.F. Marshall and R. Waugh (2005) Genome-wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress. *Mol. Genet. Genom.* 274: 515-527.
- Sato, K. and K. Takeda (2009) An application of high-throughput SNP genotyping for barley genome mapping and characterization of recombinant chromosome substitution lines. *Theor. Appl. Genet.* 119: 613-619.
- Scheet, P. and M. Stephens (2008) Linkage disequilibrium-based quality control for large-scale genetic studies. *PLoS Genet.* 4: e1000147.
- Shinozaki, K. and H. Sakakibara (2009) Omics and bioinformatics: an essential toolbox for systems analyses of plant functions beyond 2010. *Plant Cell Physiol.* 50: 1177-1180.
- White, R., H.C. Rentice and M.B. Verwilt (1988) Automated image acquisition and morphometric description. *Can. J. Bot.* 66: 450-459.
- Yoshioka, Y., H. Iwata, R. Ohsawa and S. Ninomiya (2004) Quantitative evaluation of flower colour pattern by image analysis and principal component analysis of *Primula sieboldii* E. Morren. *Euphytica* 139: 179-186.
- Zhu, C., M. Gore, E.S. Buckler and J. Yu (2008) Status and prospects of association mapping in plants. *Plant Genome* 1: 5-20.