

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 6 月 3 日現在

機関番号：12601

研究種目：基盤研究(B)

研究期間：2010～2012

課題番号：22380010

研究課題名（和文） ゲノミックセレクションを活用した革新的作物育種システムの構築

研究課題名（英文） Development of an innovative crop breeding system utilizing genomic selection

研究代表者

岩田洋佳 (HIROYOSHI IWATA)

東京大学・大学院農学生命科学研究科・准教授

研究者番号：00355489

研究成果の概要（和文）：

ゲノムワイドマーカーをもとに個体や系統の遺伝的能力を予測して選抜するゲノミックセレクション（GS）について、(1) モデリング手法の開発、(2) イネ品種・系統における実データの収集・解析、(3) 育種シミュレーションを行い、それを活用した作物育種効率化の可能性を検討した。(1)では、複数形質の同時予測、交配後代の分離の予測、遺伝子型×環境交互作用パターンの予測のための手法開発を行った。(2)では、イネ遺伝資源についてゲノムワイド SNP データを収集するとともに、それを利用して様々なタイプの形質の予測精度を明らかにした。(3)では、他殖性植物と自殖性植物を想定したシミュレーションをそれぞれ行った。研究の結果、ゲノム情報に基づく予測のための種々の手法が開発された。また、作物育種に適した GS の利用形態を明らかにできた。

研究成果の概要（英文）：

Genomic selection (GS) is a new breeding technology for selecting individuals/lines based on their genetic potential predicted based on genome-wide markers. We evaluated the potential of GS in crop plant breeding through (1) the development of novel methods for building a prediction model, (2) the collection and analysis of real marker and phenotypic data of rice varieties/lines, (3) simulation studies of breeding programs using GS. In (1), we conducted the development of novel methods for predicting multiple traits simultaneously, trait segregation in a progeny population, and the pattern of genetic by environmental interaction. In (2), we collected genome-wide SNP data and phenotypic data for rice germplasm collection, and used the collected data to assess the prediction accuracy of GS through simulations and cross-validations. In (3), we performed studies simulating breeding programs in allogamous and autogamous plants. Through the researches, we developed various methods for prediction based on genomic information, and revealed proper ways to utilize GS in a crop plant-breeding program.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	6,000,000	1,800,000	7,800,000
2011年度	3,600,000	1,080,000	4,680,000
2012年度	2,900,000	870,000	3,770,000
総計	12,500,000	3,750,000	16,250,000

研究分野：生物測定学

科研費の分科・細目：農学・育種学

キーワード：植物育種・遺伝・ゲノム育種・ゲノミックセレクション・量的遺伝子座（QTL）・塩基多型（SNP）・育種シミュレーション

1. 研究開始当初の背景

ゲノムワイドマーカーをもとに、未知の系統の育種価を予測し選抜を行うゲノミックセレクション（genomic selection: GS）が、現在、ポリジーン支配の量的形質の効率的選抜法として高い注目を浴びている。シミュレーション解析などを通して作物育種におけるGSの有用性が示されているが、具体的な活用法については明らかにされていない。今後、実際の作物育種でGSを活用するには、GSの関わる様々な疑問点を解決し、最適な活用法を探る研究が不可欠である。本研究では、GSを活用した革新的な作物育種システムの実現のための基盤的知見を得る。

2. 研究の目的

本研究では、以下の知見を統合し、作物におけるGSを活用した高効率な育種システムのデザインを新たに提言する。

（1）作物育種のためのゲノミックセレクションモデルの構築

現在用いられているGSの統計モデルは、QTLの相加効果のみを考慮した単純なモデルであり最適なモデルとはいえない。本研究では、ベイズ的アプローチを含め、様々な統計手法を用いてモデリングを行い、これら問題に対処するための手法について検討する。

（2）イネ品種コレクションのゲノムデータの拡充と、実データに基づくGS精度評価

作物におけるGSの精度評価は、実際の形質における精度は明らかにされていなかった。本研究では、イネ等で収集されたゲノムデータと形質データを用いて、交差検証法を行い、実際の形質で期待されるGSの選抜精度を明らかにする。同時に、将来、様々な実用形質におけるGSやゲノムワイドアソシエーション解析を可能にするために、イネ品種コレクションのゲノムワイドSNPデータを拡充する。

（3）GSを活用した高効率作物育種システムの探索

これまで、シミュレーション研究によって、作物育種でのGSの潜在的な有効性は示されてきたが、作物育種における具体的かつ効率

的なGSの利用法については明らかにされていない。本研究では、まず、様々な条件下で様々な育種操作を試行できるようなシミュレーションシステムを構築し、同システムを用いて、様々な条件下における育種シミュレーションを行う。

3. 研究の方法

（1）作物育種のためのゲノミックセレクションモデルの構築

（1-1）複数形質を同時に予測して選抜するためのモデリング

作物や家畜における実際の育種では、単一形質ではなく複数形質を同時に改良対象とする場合が普通である。したがって、GSを用いた選抜においても複数形質を同時に考慮した手法が必要となる。そこで、複数形質を選抜対象としたGS予測モデルの構築法の開発を行った。

i 番目の個体の T 個の形質の表現型値を $T \times 1$ のベクトル \mathbf{y}_i ($i=1,2,\dots,n$)で表し、 \mathbf{y}_i に対して以下のような線形モデルを適用する。

$$\mathbf{y}_i = \mathbf{X}_i \mathbf{b} + \sum_{l=1}^N \gamma_l u_{il} \mathbf{g}_l + \mathbf{e}_i, \quad (1)$$

ここで、 \mathbf{b} はモデルの切片を含む非遺伝的な効果を表すベクトルであり、 \mathbf{X}_i は \mathbf{b} を i 番目の個体に関係づける計画行列である。また、 u_{il} は i 番目の個体の l 番目のSNPの遺伝子型を表す変数、 \mathbf{g}_l は l 番目のSNPが T 個それぞれの形質に及ぼす効果を表す $T \times 1$ のベクトル、 γ_l は l 番目のSNPを予測モデルに含めるか否かを示す変数、 \mathbf{e}_i は環境の効果を含む残差項を表す $T \times 1$ のベクトルである。

(1)のモデルにおける \mathbf{g}_l と γ_l ($l=1,2,\dots,N$)を推定するために、ベイズ推定を用いてパラメータの推定を行った。ベイズ推定では、計算機による乱数発生を利用したMCMC法（MCBayes法）と反復計算によってパラメータの近似的な事後分布を効率的に構築する変分ベイズ法による推定法（varBayes法）を用いた。シミュレーションによって生成したデータを解析することで予測精度の評価を行った。

（1-2）最適な交配組合せを選択するためのモデリング

目的形質の分離を交配前にあらかじめ予測できれば、それに基づき最適な交配組合せを選択できる。Iwataら（投稿中）は、GSモ

デルに基づいて、交配後代における目的形質の分離パターンを予測する方法を考案した。同方法は、交配計画の立案に極めて有用と考えられるが、その精度検証は十分に行われていない。本研究では、同方法によりイネ F2 集団における出穂期分離を予測し、予測結果を実際の観察データと比較することで実証的に精度検証を行った。

(2) イネ品種コレクションのゲノムデータの拡充と、実データに基づく GS 精度評価

(2-1) イネ品種コレクションのゲノムデータの拡充

イネの品種コレクションのゲノムデータとしては、これまでに、ゲノムワイドにランダムに増幅したアンプリコン配列の情報や、日本晴ゲノム配列とコシヒカリゲノムの配列比較から得られた SNP 候補を検証しての SNP データがあった。しかしながら、アンプリコン配列解析は、多くの品種で同程度に増幅する領域の変異に限られるため、多型データが得られにくい。また、日本晴とコシヒカリとのゲノム比較では、イネ全体の多様性を反映したゲノムデータではなく、日本晴およびコシヒカリに近縁の品種における多型に偏った結果となる。そこで、dbSNP などの既存のデータベースを活用するとともに、コーネル大学で作成した 44KSNP アレイ、個別に解析した次世代シーケンサー配列間の比較等から、SNP 候補を選び、NIAS のコアコレクション品種セットおよび日本品種を供試して、GS に利用可能と考えられるゲノム情報を拡充することとした。

(2-2) 実データに基づく GS の精度評価

(2-2-1) 実データに基づくゲノミックセレクションの精度評価

日本のイネ品種計 114 品種について、3072SNPs のマーカー遺伝子型データと、それらをつくば、つくばみらい、福山、加西の複数環境下で栽培した時の出穂期、穂長、穂数、稈長、籾長、籾幅の計測値を形質データとして利用した。本研究で用いた回帰手法は 2 つのカテゴリに大別でき、1 つは正則化最小二乗法と呼ばれる方法で、2 つめはカーネル関数を用いた機械学習により回帰を行う方法である。本研究では機械学習法として、サポートベクターマシン (SVM) と、関連ベクターマシン (RVM) を用いた。予測精度を評価するために、交差検証を行い、全環境におけ

る観測値の平均 (遺伝子型値の期待値) と予測値との相関係数を計算した。

(2-2-2) 高次元データとして計測される形質の予測

作物育種の対象となる形質には、1 形質でも数十や数百といった高次元の特徴として計測されるものがある。本研究では、イネの玄米形を例として、高次元データとして計測される形質の予測のためのモデリング手法の検討を行った。解析には、アジア栽培イネ 179 アクセッション (S1) および 386 アクセッション (S2) の 2 セットを用い、ゲノムワイドマーカーからコメ粒形を予測する GS モデルを構築した。モデル構築には正則化最小二乗法と機械学習法を用いた。精度評価はクロスバリデーションにより行った。

(3) GS を活用した高効率育種システムの探索

(3-1) 連鎖不平衡の程度が低い他殖性植物における GS 育種のシミュレーション

集団としての改良を目的とする他殖性植物の育種では、一般的に集団選抜法が用いられる。しかし、個体単位での表現型の評価が難しいため、改良効率は低く、品種作成までに多大な時間を必要とする。さらに、選抜の目的形質が子実形質の場合、表現型を評価して選抜を行うのが交配後となり、花粉側の親は選抜することができず、改良効率がさらに低下する。そこで本研究では、集団内の優良個体を高精度に選抜する新技術としての GS の可能性について、育種シミュレーションをもとに明らかにした。シミュレーションでは、まず、受粉前に優良な花粉親を選抜できる場合を想定し、次に、受粉前に優良な花粉親を選抜できない場合を想定して行った。

(3-2) 系統育種法や集団育種法など、自殖後代の選抜が行われている植物における GS 育種のシミュレーション

主要穀物の多くは自殖性である。このような作物の改良には、一般に集団育種法や系統育種法が用いられ、両親系統間の交雑以降は自殖を繰り返しながら選抜を行うため、組換えによって新しい遺伝子の組み合わせが生じる可能性が限られている。本研究では、日本のイネ 112 品種について収集された 3, 102 SNP の実データを用いて育種シミュレーショ

ンを行い、GS で循環選抜を行う育種法と従来の表現型選抜に基づく系統育種法を比較した。

4. 研究成果

(1) 作物育種のためのゲノミックセレクションモデルの構築

(1-1) 複数形質を同時に予測して選抜するためのモデリング

シミュレーションでは、各 SNP をモデルに含めるか否かを示す変数 γ を 1 に固定した場合 ($\pi=0$ 、すべての SNP がモデルに含まれる) と可変にした場合 ($0 < \pi < 1$ 、選択された SNP のみがモデルに含まれる) について予測精度を評価した (表 1)。

表 1 各モデル構築法における育種価予測精度

Method		Trait A	Trait B	Trait C
MCBayes	$\pi=0$	0.902 ± 0.032	0.706 ± 0.103	0.519 ± 0.097
	$0 < \pi < 1$	0.868 ± 0.047	0.731 ± 0.120	0.401 ± 0.182
varBayes	$\pi=0$	0.859 ± 0.049	0.656 ± 0.110	0.438 ± 0.074
	$0 < \pi < 1$	0.838 ± 0.061	0.678 ± 0.140	0.330 ± 0.157
singleBayes	$\pi=0$	0.884 ± 0.039	0.485 ± 0.086	0.493 ± 0.089
	$0 < \pi < 1$	0.843 ± 0.044	0.597 ± 0.120	0.601 ± 0.109

その結果、多形質を考慮した MCBayes 法および varBayes 法と個々の形質を単一形質として扱う singleBayes 法の比較から、遺伝率が 0.8 と高い形質 A については、複数形質の解析と単一形質の解析で、予測精度の差はないが、遺伝率が 0.1 と小さい形質 A との遺伝相関が 0.72 と高い形質 B については、複数形質の解析によって予測精度は顕著に高くなった。一方、遺伝率が形質 B と同様に 0.1 と低く、形質 A とは独立な形質 C については、複数形質の解析による予測精度の増加は見られなかった。この結果は、複数形質の解析によって、QTL の多面発現の効果を効率的に予測モデルに取り入れることが可能となったためである。一般に、遺伝率の高い形質ほど予測精度は高くなるが、遺伝率の低い形質についても遺伝的に相関を持つ遺伝率の高い形質と同時に解析することにより、予測精度の向上が可能となることが明らかとなった。この結果は、実際の育種場面において、複数形質を同時に考慮した GS が、従来の単一形質を対象とした GS に比べてより有用であることを示唆している。

(1-2) 最適な交配組合せを選択するためのモデリング

112 品種のデータをもとに構築された GS モデルを用いて F2 集団の出穂期分離を予測した結果、各集団における出穂期の分離パターンを精度良く予測できることが分かった (図 2)。例えば、いくつかの集団では大きな超越分離が見られたが、それらを的確に予測することが可能であった。より具体的には、集団内の 5%、50%、95% の個体が出穂する日を予測した結果、平均誤差はそれぞれ 3.6 日、2.9 日、4.5 日であった。以上のように、イネの出穂期分離はかなり高精度に予測できることが分かった。同方法は、親品種の表現型値とマーカー遺伝子型のデータさえあれば適用できる。同方法を交配計画の立案時に用いることにより、交配育種の大幅な効率向上が期待できる。

(2) イネ品種コレクションのゲノムデータの拡充と、実データに基づく GS 精度評価

(2-1) イネ品種コレクションのゲノムデータの拡充

日本型イネ品種において多型を検出できる可能性がある SNP を、在来品種である亀の尾および旭の SNP 候補から 384 個を選び、日本水稲品種 112 品種に供試した。これによって、これまでの日本型 SNP で十分ではなかった、第 5、6、9 染色体上の多型を十分に評価できるようになった (図 1)。

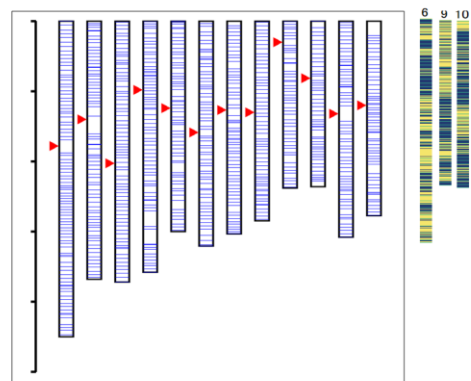


図 1. 日本型イネ品種において検出された SNP 位置。右の 3 本の染色体は、新規に抽出された日本型イネ用 SNP の位置。これらのデータを用いることでゲノム全体をカバーした。

また、インド型・日本型を含む世界のイネ品において多型が検出できる SNP 候補 768SNP を選び、96 品種で SNP 特性の評価を行った。この結果、これまでに遺伝子型データが得られていなかった領域を狭めることはできたが、用いた SNP の品種群内での多型頻度は低く、インド型品種間の遺伝的差異を明らかにするにはいたらなかった。これは、現在行われている SNP 検出が、日本晴のゲノム情報をリファレンスとして、この日本晴ゲノムと異なる塩基を SNP と認識しているためと考えられる。インド型品種間、日本品種以外の日本型品種間で十分に多型を検出するためには、インド型品種のゲノムリファレンスを作成して SNP を検出する、あるいは、数多くの SNP 候補について評価を行う等の対策が考えられる。多様な品種を対象とする GS を実現させるためには、品種群内で多型を示す SNP を整備する必要があり、さらにゲノム情報を充実させる必要がある。

(2-2) 実データに基づく GS の精度評価

(2-2-1) 実データに基づくゲノミックセレクションの精度評価

交差検証による精度評価の結果、リッジ回帰と RVM の 2 手法が安定して相関係数が高かった。要因としてはリッジ回帰の場合、正則化最小二乗法である LASSO が少数の変数による効率的な回帰を目指しているのに対し、回帰係数の 2 乗でペナルティを科すリッジ回帰では効果の小さな遺伝子の影響を忠実に反映できたためと思われる。また、RVM ではカーネル関数を用いた非線形回帰を行うことで、他の手法より高い精度を示した可能性がある。ただし、今回低精度であった手法でもパラメータの設定方法によっては予測精度を向上させることも可能である。今後、様々な工夫を施す余地は残されている。

(2-2-2) 高次元データとして計測される形質の予測

画像解析により玄米の輪郭の座標情報を抽出し、輪郭形状を 77 個の楕円フーリエ記述子として記述し、フーリエ記述子の主成分得点に基づく予測モデル構築を行った結果、多くの品種で比較的精度良く玄米形を予測できることが分かった (図 2)。なお、有意なアソシエーションが検出されたマーカーだけでは予測精度が低いことも示された。この結果は、効果の大きな QTL だけでは、玄米形変異を十分正確に説明できないことを示唆している。

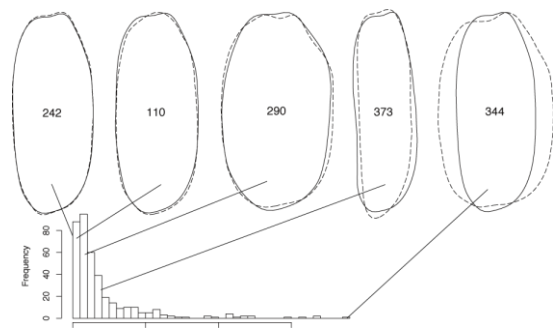


図 2. 予測された玄米形 (実線) と観察された玄米形 (破線)、および、予測誤差の平方和 (predicted residual sum of squares: PRESS)

(3-1) 連鎖不平衡の程度が低い他殖性植物における GS 育種のシミュレーション

シミュレーションの結果、GS は年あたりの選抜回数が多いほど年あたりの遺伝的獲得量が向上した。また、GS は、すべての選抜回数において、従来の育種法である表現型選抜より高い遺伝的獲得量を示した。このことから、GS は LD の程度が極めて低い植物の育種においても有効であることがわかった。花粉親が選抜できない場合は、表現型選抜と年 1 回の GS では、花粉親を選抜できる場合に比べて改良効率が大きく低下した。これは、表現型選抜と年 1 回の GS では選抜時に必ず表現型を計測するために、常に花粉親が無選抜となることが原因である。一方、年あたり 2 回以上 GS を行う場合、花粉親を選抜できる場合とできない場合の間で改良効率に大きな差が見られなかった (図 3)。これは、選抜モデルを更新する世代に花粉親を選抜できなくても、次の世代の GS では花粉親由来の染色体を高精度に選抜できることによることがわかった。

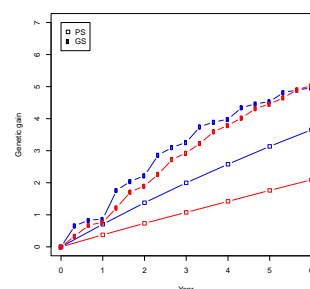


図 3. LD の低い他殖性植物の集団選抜のシミュレーション。表現型選抜 (□) と GS (○) による遺伝的獲得量 (縦軸) の変化。横軸は選抜年数。受粉後選抜が必要な形質 (赤) では、受粉前選抜ができる形質 (青) に比べて GS を用いる意義が大きい。

(3-2) 系統育種法や集団育種法など、自殖後代の選抜が行われている植物におけるGS育種のシミュレーション

シミュレーションの結果、GSでは5回程度の循環選抜で、表現型選抜によるよりも同等かそれ以上の能力を持つ系統を育成できた(図4)。なお、GSでは表現型計測の必要がないため世代促進を行いながら選抜することも可能であり、表現型選抜に比較して大幅に育種時間を短縮できる。また、GSを用いた循環選抜では個体選抜が可能であり、系統あたり複数個体を評価する必要がある系統育種法に比べて、栽培に要する圃場の面積やコストを大幅に節約できる。以上から、イネ育種において、GSは従来の育種法に比較して有効であることが示された。

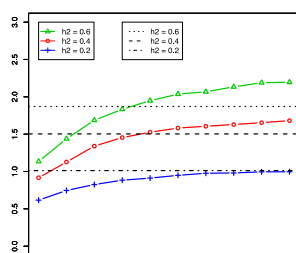


図4. イネの育種にGSを用いた場合の効率評価。折れ線がGSによる遺伝的獲得量(縦軸)を表し、水平線(点線、破線、一点破線)は表現型選抜を用いた系統育種法で達成される遺伝的獲得量。横軸はGSを用いた循環選抜育種の選抜回数を表す。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計7件)

- ① Yabe S, Ohsawa R, Iwata H, Potential of genomic selection for mass selection breeding in annual allogamous crops, *Crop Science*, 査読有, 53, 2013, 95-105
 - ② Hayashi T, Iwata H, A Bayesian method and its variational approximation for prediction of genomic breeding values in multiple traits, *BMC Bioinformatics*, 査読有, 14, 2013, 34
 - ③ Iwata H, Hayashi T, Terakami S, Takada N, Sawamura Y, Yamamoto T, Potential assessment of genome-wide association study and genomic selection in Japanese pear *Pyrus pyrifolia*, *Breeding Science*, 63, 2013, 125-140
- ④ Iwata H, Jannink JL, Accuracy of genomic selection prediction in barley breeding programs: a simulation study based on the real SNP data of barley breeding lines. *Crop Science*, 査読有, 51, 2011, 1915-1927
- [学会発表] (計39件)
- ① Iwata H, Hayashi T, Yoshioka T, Yamasaki M, Ebana K, Terakami S, Takada N, Saito T, Yamamoto T, Genomic prediction of trait segregation in progeny populations: examples in Japanese pear and rice, *Plant and Animal Genome XXI*, 2013, 1月12日～1月16日, アメリカ合衆国、サンディエゴ
 - ② Iwata H, Ebana K, Uga Y, Hayashi T, Integration of Genomics and Morphometrics Approaches Toward Systematic Plant Shape Improvement, 2nd ISBSA, 2011, 9月9日, 沖縄県男女共同参画センター(沖縄県)
 - ③ Hayashi T, Iwata H, Non-MCMC methods for effective Bayesian analysis of high-dimensional genomic data. *Plant and Animal Genome XIX*, 2011, 1月16日, アメリカ合衆国、サンディエゴ

6. 研究組織

(1) 研究代表者

岩田 洋佳 (HIROYOSHI IWATA)

東京大学・大学院農学生命科学研究科・准教授

研究者番号: 00355489

(2) 研究分担者

江花 薫子 (KAWORU EBANA)

独立行政法人農業生物資源研究所・遺伝資源センター多様性活用研究ユニット・主任研究員

研究者番号: 00370643

林 武司 (TAKESHI HAYASHI)

独立行政法人農業・食品産業技術総合研究機構・中央農業総合研究センター・上席研究員

研究者番号: 70370674