

これまでに蓄積されてきたイネ育種データをゲノム-表現型関係モデリングに活用する

岩田洋佳¹・佐藤睦志²・阿部陽³・金天海^{4,5}・師田郷太⁶・下野裕之⁴

(¹ 東京大学・² 夷風凜凜・³ 岩手生物工学研究センター・⁴ 岩手大学・

⁵ Sky Ocean Technology Co., Ltd.・⁶ バージニア工科大学)

Utilization of accumulated rice breeding data for genotype-phenotype relationship modeling

Hiroyoshi Iwata³, Chikashi Sato², Akira Abe³, Chyon Hae Kim^{4,5}, Gota Morota⁶,

Hiroyuki Shimono⁴

(¹The University of Tokyo, ²Ifu-Rinrin, ³Iwate Biotechnology Research Center, ⁴Iwate University,

⁵Sky Ocean Technology Co., Ltd., ⁶Verginia Tech.)

【目的】水稲奨励品種決定基本調査成績データセット(以下、奨決データ)と、品種/系統の系譜情報、気象情報、ゲノム情報を組み合わせて、ゲノムと表現型の関係をモデル化する。

【材料および方法】イネ品種・特性データベース(農研機構, <https://ineweb.narcc.affrc.go.jp/>)から得られた系譜情報をもとに、戻し交配の回数も考慮して 15,145 品種/系統について分子血縁行列を計算した。次に、207,304 レコードの奨決データから、分子血縁行列に含まれる品種/系統の 200,604 レコードを抽出した。さらに、水田栽培であり、かつ基肥量・栽植密度情報がある 196,557 レコードを抽出した。これにより、3,196 環境(試験地×年次)、8,024 遺伝子型を含むデータが得られた。同データをもとに、到穂積算温度、登熟積算温度、稈長、穂長、穂数、全重、玄米収量、玄米千粒重、玄米品質、倒伏程度、収穫係数の 11 形質について、遺伝子型、環境(試験地×年次)を変量効果、基肥量、追肥量、栽植密度を母数効果とした混合モデルを当てはめ、遺伝子型値の最良線形不偏予測値(BLUP)を得た。得られた遺伝子型値について、ゲノム情報がある 266 品種/系統を対象として、ゲノム情報に基づく実現血縁行列、系譜情報に基づく分子血縁行列を共分散構造として仮定した混合モデルを適用して予測を行い、5 分割交差検証でその精度を評価した。さらに、Legarraら(2009, J. Dairy Sci. 92:4656)および Christensenら(2012, Animal 6:10)の方法に従い、実現血縁行列と分子血縁行列を混合した血縁行列を作成し、上述した 266 品種/系統に加え、ゲノム情報の無い約 4,400 品種/系統の遺伝子型値も用いた場合の予測精度についても評価した。

【結果および考察】遺伝子型値の BLUP より、品種の配布年にとまなう変化、育成地間の違いを確認することができた。ゲノム情報(実現血縁行列)のみを用いた予測では、系譜情報(分子血縁行列)を用いた予測に比べ、精度が高かった。しかし、ゲノム情報の無い品種/系統の遺伝子型値も用いて分子血縁行列をもとに予測を行った場合、大幅に予測精度が向上した。また、いずれの形質も、実現血縁行列と分子血縁行列を混合した血縁行列を用いて予測を行なった場合に最も精度が高かった。以上のように、これまでに蓄積された育種データを系譜、ゲノム、気象データと併せて解析することにより、有用な情報を引き出すことができることが示唆された。今後は、実現血縁行列および分子血縁行列を用いたゲノムワイドアソシエーション解析を行ない、品種/系統の遺伝子型値の変異と関連の強いゲノム領域を明らかにする予定である。

謝辞:本研究は科研費・基盤研究(A)(19H00938)の支援を受けて行った。