



国立大学法人 岩手大学

岩手大・東京大・岩手生工研・夷風凜凜の共同研究が Nature 姉妹誌「Comm. Biol.」に掲載

ビッグデータ解析法の開発 将来気候に適応するイネ品種の選抜法

岩手大学農学部下野裕之教授らの研究グループは、将来の気候条件での持続可能な食料供給の達成に向け、イネをはじめ作物の品種改良を迅速化・効率化するため、ビッグデータ解析法「YpCGM 法」を開発しました。この成果は、英国の Nature 姉妹誌「Communications Biology」2023 年 7 月 21 日付（日本時間午後 10 時）に掲載されました。本研究で開発した手法は世界各国で蓄積されたイネ、コムギ、トウモロコシ、ダイズなど作物のビッグデータへの適応による育種への利用が期待されます。

【発表のポイント】

- ① **新手法の開発：イネ品種が持つ単位土地面積当たりの生産性（収量）^(注1)を数理モデル^(注2)で算出した潜在収量から評価する手法を開発しました（図1）。**

本手法により、これまで比較ができなかった広範の品種群について、生産性について比較が可能になりました。たとえば、北海道のイネ品種と九州のイネ品種の生産性を直接的に比較することができません。北海道に九州の品種を持ってきて比較しても、出穂が遅すぎ秋の低温により登熟まで至らずその能力を発揮できません。また、九州に北海道の品種を持って行って比較しても、北海道の品種は温度感応性が高く、播種後、わずか1か月ですぐに出穂してしまい、北海道で正常に生育した場合に比べて非常に小さな植物体となってしまい同様にその能力を発揮できません。また、例え、同じ地点であっても、同一年次に比較していないとそのデータを比較はできません。数十年前に評価された古い品種の収量データと今年評価した新しい品種の収量データは直接比較できません。本手法は気象条件から算出した潜在収量を基準として標準化することで比較を可能としました。いわば、数理モデルを「橋渡し」として「同じ土俵での比較」を可能としました。
- ② **日本のイネ 237 品種の収量性の評価：北海道から九州までの 110 か所で過去 38 年に 237 品種を栽培した 72,510 セットのイネ収量ビッグデータ^(注3)について解析し、品種特性**



を2つのパラメーターで定量化しました（図2）。

この情報は将来の育種における交配母本の選定の基礎となります。

- ③ 日本のイネ品種の収量性の改良：本手法を用いた品種特性のパラメーター解析から、1920年から2017年までの品種改良によりイネ品種の収量性が10年で10.2kg/haのスピードで向上した点、興味深い点として、実際に利用する生産者は環境がよくなったときに収量を高める性質（可塑性）が高い品種を利用している点を明らかにしました（図3）。
- ④ 前述の可塑性に関わる新規遺伝子領域の特定：収量の可塑性 α を支配する新規遺伝子領域を第10染色体上にみいだしました（図4）。

【背景】

私たちの豊かな生活は、安定した農業生産に支えられています。世界の主要作物（イネ、コムギ、トウモロコシ）の収量（単位土地面積当たりの生産性）は1960年以降に2~3倍以上に増加し、これには品種改良が大きく貢献しています。品種改良といえば、DNAや遺伝子の評価（ジェノタイピング、genotyping）が注目されがちですが、それにも増して重要なのが、その品種が持つ能力の正確な評価（フェノタイピング、phenotyping：ある品種が持つ能力は、これまでの従来品種よりも高いのかの評価）です。前者のDNAや遺伝子の評価（ジェノタイピング）が次世代シーケンサーなど高度の技術革新により効率化されているのに対し、後者の品種が持つ収量性の評価（フェノタイピング）はいまだ野外フィールドでの栽培試験が世界標準のスタンダードの評価方法であり続け、品種改良のボトルネックとなっています。

栽培試験による収量の評価の利点は、導入する地域での気象条件で実際に評価することで品種の能力を正確に予測できる点です。一方で欠点は、得られる収量データは評価した環境に依存する点です。結果、品種を比較した試験結果がビッグデータとして蓄積していても、同じ年次に同じ地点で行ったデータ同士しか比較ができません。1年に1回、多大の労力と時間をかけて貴重な野外フィールドでの栽培試験の結果がうまく利用できず、品種改良のスピードを制約しています。地球温暖化による気候変動が拡大する中、蓄積されたビッグデータの効率的な利用方法の開発が必要です。そこで、開発したのが本手法です。

わが国は、国土が南北に長く、それぞれの気候に適した品種が育成、栽培されています。例えば、北海道は「ななつぼし」、東北地方では「ひとめぼれ」、北陸地方では「コシヒカリ」、九州地方は「ヒノヒカリ」などがありますが、それら主要品種を超える新たな品種育成が日々行われています（岩手県の「銀河のしずく」、山形県の「つや姫」、青森県の「晴天の霹靂」）。イネをはじめとする作物の品種改良には、有用な母本の交配からはじまり、その後、後代の系統を評価、選抜する過程に10年弱の年数を要します。気候変動が拡大する中、10年先を見通した交配母本の選択が必要です。本研究は、蓄積されたイネの収量のビッグデータを再解析することで、貴重なデータをアップサイクル（upcycle）し、気候変動に適応した新たな品種育成のための有望な品種を選抜することで先を見通した有用な母本の選択に役立てる手法を開発しました。



本研究成果は、従来、収量はある基準品種との相対値としてしか表現できなかったものを、本手法によりイネ品種の収量という特性を数理モデルを用いてわずか2つのパラメーターに一般化することに成功しました。ビッグデータを利用することで、時間と労力が必要な野外フィールドでの調査を行わずに、データ解析のみから埋もれていた全国の有用な品種の発掘、また関連する新たな遺伝子領域の発見も可能となりました。

【研究内容】

岩手大学作物学研究室の下野教授は、イネの成長を数理モデルで評価する手法を開発しました（図1）。y軸に実測のそれぞれの品種の収量を取り、それぞれの収量が得られた気象情報から数理モデルで算出した潜在収量をx軸にプロットすると、単純な線形回帰でできることを明らかにしました。この線形回帰の切片と傾きが品種のもつ収量特性として評価することができるという、YpCGM法を提案しました（ポイント①）。切片を収量性（yield-ability, β ; t/ha）、傾きを可塑性（yield-plasticity, α ; dimensionless）と提起しました。なお切片は、 $x=0$ での切片ではなく、多数の品種が経験した潜在収量 $x=8t/ha=SPY$ としました。

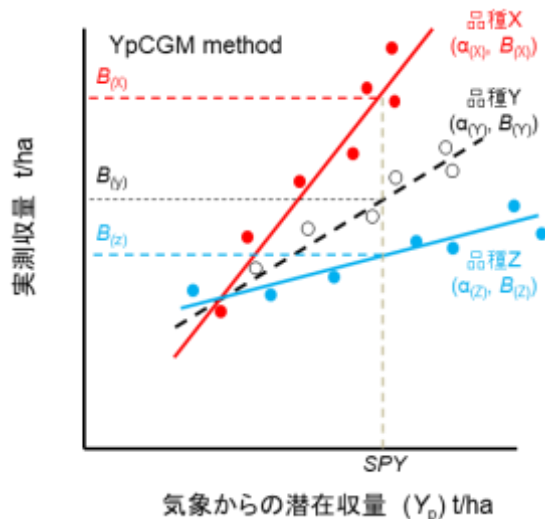


図1. YpCGM法

$$Y_{obs}(i, j) = \alpha(i) Y_p(i, j) + b(i)$$

$$= \alpha(i) (Y_p(i, j) - SPY) + \beta(i)$$

$Y_{obs}(i, j)$ と $Y_p(i, j)$ は、環境 j (年, 地点, 栽培方法)における品種 i の実測収量と潜在収量を示す。 $b(i)$ は x 軸 $Y_p = 0$ のときの線形回帰の切片を示す。

●, ○, ●は、品種 X, Y, Z

開発した手法を用いて、下野教授は、日本のイネ 237 品種（110 か所で過去 38 年）のビッグデータ 72,510 セットについて解析を行った結果、品種特性を2つのパラメーター、収量性（yield-ability, β ; t/ha）には、2.5t/ha～7.3t/ha の変異が、可塑性（yield-plasticity, α ; dimensionless）には-0.23～0.95 までの変異があることを明らかにしました（図2ab）（ポイント②）。さらに、それら品種のパラメーターを品種の育成年次との関係でプロットしてみると、正の相関関係が収量性 β についてみられ、育種過程により、年あたり 10.2kg/ha のスピードで改良が進んでいることを示します（図2c）（ポイント③）。また、その中で生産者が好む主要品種についてみると、その傾きは 21.8kg/ha とほぼ2倍に高くなり、生産者はより収量性の高いものを好むことが示されました。さらに、環境に対する収量の可塑性 α については、育種家は積極的に選抜を行っていないけれど、生産者は、生産性だけでなく、可塑性が高い品種を選んでいることを初めて明らかにしました（図2



d)。このことは将来気候への適応のための育種の目標としての可塑性 α の重要性を示唆します。

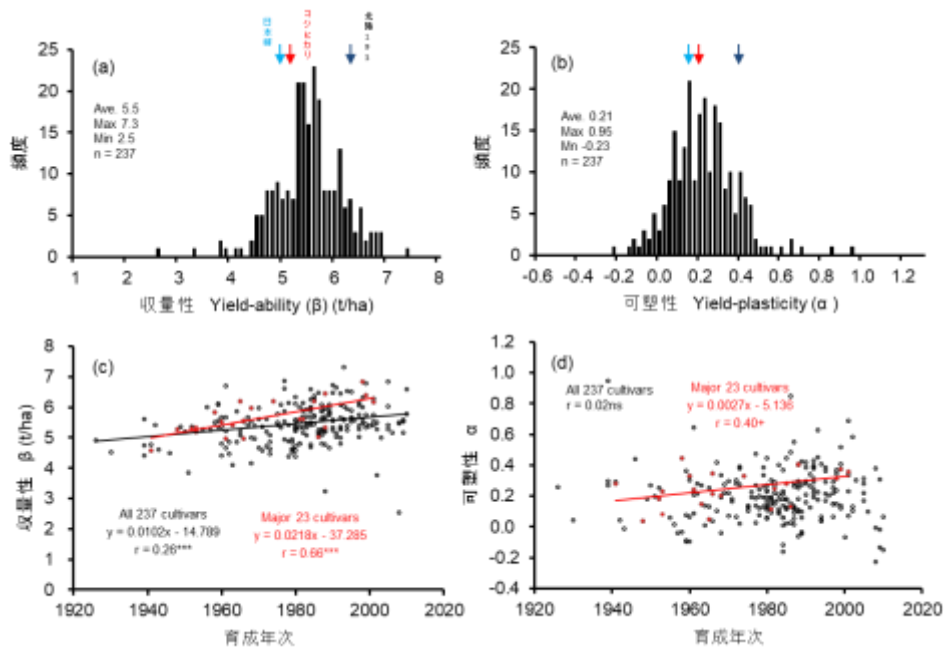


図2. YpCGM法で解析したイネ237品種のパラメーター ((a)収量性 β 、(b)可塑性 α) とそれらの育成年次に対する応答 ((c), (d))。c, dの回帰の黒字は全237品種、赤字は主要23品種を示す。

ゲノム情報を用いて、岩手生物工学研究センターの阿部博士は、これら2つのパラメーターについてゲノムワイド関連解析 (Genome wide association study) を91,800SNPsを用いて行くと、可塑性に関わる新規遺伝子領域を第10染色体上にみいだすことに成功しました (図3a) (ポイント④)。第10染色体の中でも3,157,567~3,160,342bが強くかかわることが明らかになりました (図3de)。

なお、解析を進めるうえで、イネ収量ビッグデータには8,000品種以上のデータが蓄積されていますが、詳細なゲノム情報を用いた解析を行うために、品種数を絞る必要があります。東京大学の岩田教授は、(株) 夷風凜凜の佐藤氏がとりまとめた日本のイネ14,032品種の系譜関係について解析を行い、日本のイネを形作る代表237品種を選定しました。岩手生物工学研究センターの阿部博士は、その237品種について、DNAシーケンスを行い、品種間の違いをDNA上のマーカーとして91,800SNPs^(注4)を特定しました。東京大学の岩田教授また岩手大学の金准教授(当時)はこれらパラメーターが遺伝的な形質である点をゲノミック予測から確認しています。

提案した手法 YpCGM法は、労力と時間を要する収量調査をすることなく、データを再利用 (upcycle) することで、新たな視点を提供でき、今後の品種改良を効率化させることに役立つことが期待されます。



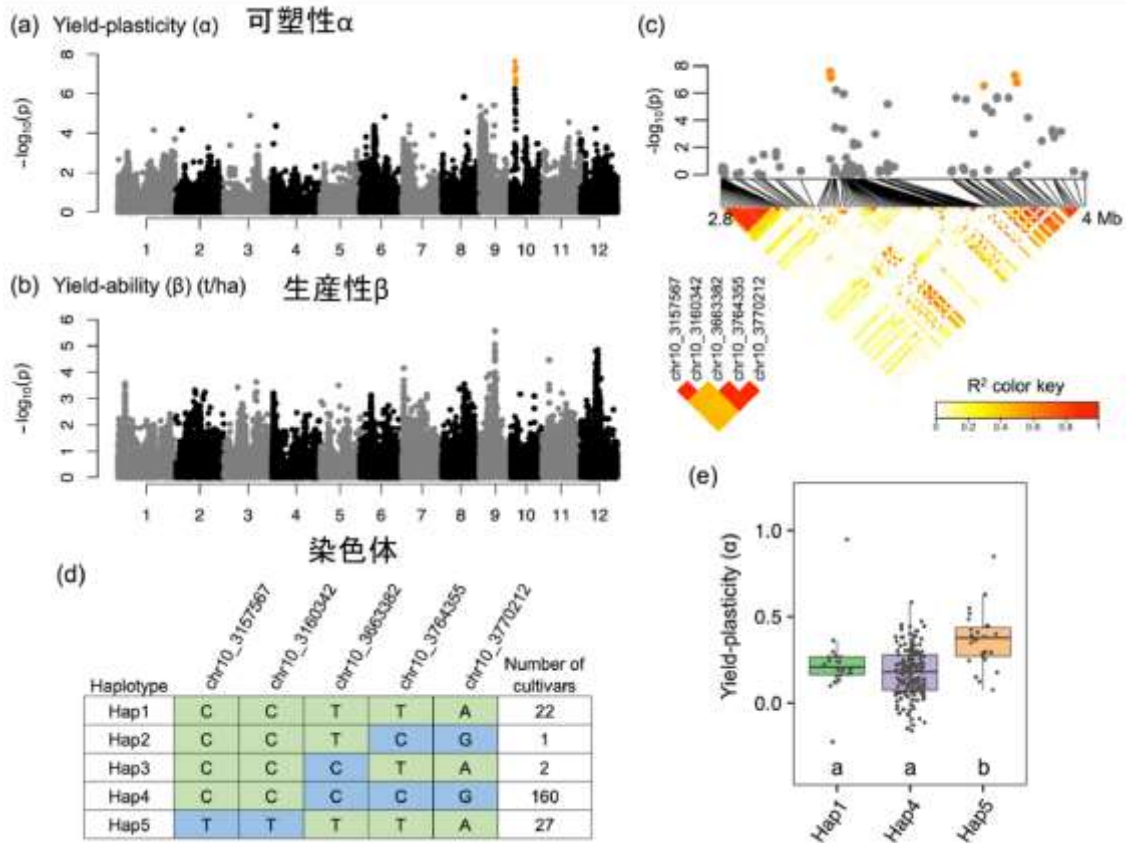


図3. YpCGM 法で得られたパラメータ(a)可塑性 α と(b)収量性 β のゲノムワイド関連解析、(c) 可塑性 α の第10染色体2.8~4 Mb領域の詳細なゲノムワイド関連解析と(d)(e)ハプロタイプ解析により Hap5 が有効なマーカーであることを示す。

【共同研究グループ】

岩手大学： 下野 裕之（農学部 教授）、金 天海（理工学部 准教授、当時）
 東京大学： 岩田 洋佳（農学生命科学研究科 教授）
 岩手生物工学研究センター：阿部 陽（ゲノム育種研究部、主席研究員）
 (株) 夷風凜凜： 佐藤 睦志（代表取締役）

【掲載論文】

掲載誌：Communications Biology: 6, 764 (2023).
 論文名：Upcycling rice yield trial data using a weather-driven crop growth model
 公表日：2023年7月21日付（日本時間午後10時）掲載
 著者：下野 裕之（岩手大学）、阿部 陽（岩手生物工学研究センター）、金 天海（岩手大学）、佐藤 睦志（(株) 夷風凜凜）、岩田 洋佳（東京大学）
 DOI番号：<https://doi.org/10.1038/s42003-023-05145-x>

【研究費】

本研究は科研費・基盤研究 A（「イネの野外ビッグデータを用いた「人工知能」による育種プラットフォームの開発」、2019-21 年、19H00938）の支援を受け実施しました。

【用語説明】

注1 収量：単位土地面積当たりの生産性 t/ha, kg/10a. 収量を高めることは限りある耕地からより多くの収穫物を得ることにつながる。産業革命以降の急激な世界人口の増加は、世界の栽培面積の増加ではなく、収量の増加に強く寄与しており、今後 2050 年に 90 億人に達することが予測される世界人口を扶養するにはさらなる収量の増加が望まれます。

作物育種において、「収量」は最も重要な目標形質ですが、ある品種の収量は、花の色や形などの形質とは異なり、環境より変動します。たとえば、品種「ひとめぼれ」の収量は、岩手県盛岡市で 2021 年に栽培した場合と 2022 年に栽培した収量で異なります。また同一年 2021 年でも盛岡市で得られる「ひとめぼれ」の収量は札幌市や那覇市で栽培した場合とも異なります。その違いを生む要因として、気象条件の違いがあげられます。その中でも水ストレスが少ない湛水条件下で生育するイネは、日射量と温度の違いに強く影響を受けます。収量という形質は、イネは野外のある地点、ある年次に生育させて得られるものであり、「ある品種の収量特性は？」という問いに答えるのが極めて難しく、工業製品のように一律の条件で一定の基準で評価できるものとは一線を画します。

注2 ビッグデータ：全国 110 か所（右図）で 1980 年から 2017 年までの 38 年間に公立農業試験場で得られた **8,524 品種**を用いた**計 207,331 データセット**のイネ収量データ（奨励品種決定試験データベース、農研機構提供）。なお、収量データは全国統一の栽培方法で適切に管理し得られた。数理モデルでの解析には、気象データの入力値が必要で、日平均の気温と日射量、**計 1 億 5300 万データセット**（365 days x 38 years (1980~2017) × 110 locations）
（<https://meteocrop.dc.affrc.go.jp/real/top.php>）
DNA における **91,800 SNPs** の多型情報。



注3 数理モデル＝成長モデル：イネの成長を数式化し、日々の気象データを入力することで、コンピューター上で仮想イネをそれぞれの環境で算出することが可能に、いわば仮想の基準品種の収量を「潜在収量」として、それぞれの収量が得られた環境について算出しました。たとえ話として、ある個人の走る能力を定量化したいとします。走る場所や気温などの天気、また距離によっても、その人の走るタイムは異なります。そこで一定の能力を設定した「仮想の伴走者」をコンピューター上で設定し、その「仮想の伴走者」のタイムに対して、その対象となる人が走るタイムとの関係を見ることで、その人の能力を評価するようなイメージになります。ここではイ

ネの収量、それを得られた環境での数理モデルで予測される「潜在収量」の関係になります。

注4 SNP(スニップ)：一塩基多型(いちえんきたけい)。イネの設計図ゲノム DNA の塩基配列の中で、品種間で異なる塩基の1塩基での違い。これによりイネの収量の品種間差が生じる。

