

令和 5 年 6 月 21 日現在

機関番号：15301

研究種目：若手研究

研究期間：2020～2022

課題番号：20K15503

研究課題名（和文）アフリカ栽培イネ特異的な高温ストレス耐性のゲノムワイド解析とGWASカタログ構築

研究課題名（英文）Genome wide analysis in African rice and establishment of GWAS catalogs

研究代表者

古田 智敬（Furuta, Tomoyuki）

岡山大学・資源植物科学研究所・助教

研究者番号：70774008

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：0. glaberrima品種172系統において全ゲノム関連解析を実施し、種子サイズや到穂日数を含め様々な農業形質の0. glaberrima品種群内の形質多様性に寄与する遺伝子の座上候補領域が検出された。一方で、0. glaberrimaの形質値多様性が各品種の栽培地域ごとに異なっており、土着品種的な性質が強いことが分かった。また、これらの情報を整理し効率的に原因遺伝子探索を行えるようにカタログ化ツールを開発した。上記に加えて、次世代シーケンサーにより判定された遺伝子型データが持つ特有のエラーを修正するツールを開発した。

研究成果の学術的意義や社会的意義

本研究で同定された0. glaberrimaの農業形質に関与する遺伝子座が同定されたことで、0. glaberrimaの遺伝子を0. sativa育種に利用する基盤情報が整備された。また、0. glaberrimaを育種することで新たな食料資源として活用する可能性も開かれた。さらに、GWASカタログ化ツールや遺伝子型データエラー修正ツールなど、遺伝育種研究を支える新たな解析ツールが生み出され、今後の当該分野の発展を支える基盤が構築された。

研究成果の概要（英文）：Whole genome association study was performed on 172 lines of 0. glaberrima cultivars, and candidate loci for genes contributing to trait diversity within the 0. glaberrima cultivars for various agronomic traits including seed size and days to heading were detected. On the other hand, the trait diversity of 0. glaberrima differed among the origin of country of each variety, indicating a strong population structure in the tested population. We also developed a GWAS cataloging tool to organize GWAS results and efficiently search for causal genes. In addition, we developed a tool to correct errors in genotype data determined by next-generation sequencers.

研究分野：遺伝育種

キーワード：アフリカイネ 全ゲノム関連解析 量的形質 バイオインフォマティクス

1. 研究開始当初の背景

栽培イネには、アジアで栽培化された *O. sativa* とアフリカで栽培化された *O. glaberrima* の 2 種が存在する。かつて、*O. glaberrima* と *O. sativa* の交雑によって、アフリカの栽培環境で多収なイネ品種群 New Rice for Africa (NERICA) シリーズが開発された。これは、*O. glaberrima* がもつ様々なストレス耐性を保持しつつ *O. sativa* の多収性を導入し、アフリカ向け品種を開発する試みであった。しかし近年の地球温暖化によりアジア各国でもイネの安定生産が脅かされている状況を考えると、アジア向け品種に対して *O. glaberrima* がもつ高温ストレス耐性を導入することが、打開策の一つとして期待される。高収量かつ良食味であり世界的にも広く普及している *O. sativa* に比べ、*O. glaberrima* の栽培は西アフリカの一部に限定されている。しかしながら、*O. glaberrima* はアフリカの高温環境に適応し、*O. sativa* が持たない病虫害抵抗性や環境ストレス耐性を持つとされている。そのため、環境変動に対応した次世代イネ品種を育種するうえでの重要な育種材料であると期待されている。しかし、*O. glaberrima* は西アフリカの低地から高地に到る様々な地域で栽培されており、環境適応性は多様であり遺伝資源としての適性も多様である。従って、より有用な遺伝子や系統を選抜し *O. glaberrima* のポテンシャルを最大限活用した育種を行なうには、ゲノムワイドな視点から *O. glaberrima* 品種群における表現型の多様性とその遺伝的要因を理解することが求められる。

2. 研究の目的

本研究では、*O. glaberrima* 品種コレクションにおいて全ゲノム関連解析 (GWAS) を行い、高温ストレス耐性遺伝子座の同定および有用アリルに連鎖する SNP マーカーの同定を主な目的とする。さらに本研究の成果を広く育種利用しやすくするために、*O. glaberrima* 版 GWAS カタログの構築を目指す。GWAS カタログとは、GWAS により見出された表現型と遺伝子型の関連情報を蓄積したデータベースであり、ヒト疾患 GWAS カタログが有名である。*O. glaberrima* と *O. sativa* を交配すると不稔や弱勢が現れるため育種利用には従来よりも労力を要する。そのため、効率よく有用遺伝子を集積できる最適な交配組み合わせを見つけ出すことが、育種年限の短縮や労力の低減に繋がる。これまで育種利用を意識した GWAS カタログ構築は世界でも例がなく、本研究成果により育種目標を達成するための品種選定や交配計画を容易に行えるようになると考えられる。さらに本研究で得られたゲノム多型情報を利用すれば、高温ストレス耐性以外の表現型についても GWAS を展開し、その成果をカタログに取り込むことが可能であり、大きな発展性と拡張性を持ったゲノム育種のための情報基盤を構築できる。

3. 研究の方法

GWAS に供試する *O. glaberrima* 品種は、国立遺伝学研究所のイネナショナルバイオリソースプロジェクト (NBRP) から取得した 173 系統を用いる。この品種コレクションに対して、高温ストレス関連形質として、開花期および登熟期に高温に暴露されることで発生する不稔や玄米の白濁化を調査する。高温による不稔や白濁化は、開花期の最高気温が 35 度を超えると発生するとされる。近年、日本各地でも高温による不稔や白濁化によるコメ品質の低下が発生しており、将来的な気温上昇にも耐える有用遺伝子の発見が望まれる。そこで高温区は最高気温 38 度とし、圃場でポット栽培した植物個体を開花期に高温区へと移し 1 週間の高温ストレス処理を行った。そして、高温ストレス処理の有無による不稔率と白濁米比率の変化を調査した。GWAS によって同定された有用遺伝子座について、*O. sativa* 育種における利用可能性を検証した。まず、*O. sativa* 染色体背景で有用遺伝子座に *O. glaberrima* 染色体断片を持つ CSSLs を試験し、高温ストレス耐性を獲得しているかどうかを確認する。*O. glaberrima* の品種 WK21 と CG14 の CSSLs が利用可能で、どちらかの系統が有用アリルを持つ遺伝子については、*O. sativa* に導入した際の効果を証明できる。また、GWAS により得られた統計的なデータをただ一覧にするのではなく、各品種が持つ有用遺伝子の数や効果の大きさを視覚的に把握できるデータベースを構築する。これにより統計学の知識やデータベースの操作に不慣れでも、直感的な操作で必要な情報にアクセス可能となる。

4. 研究成果

GWAS のために NBRP より取得した 172 系統の *O.*

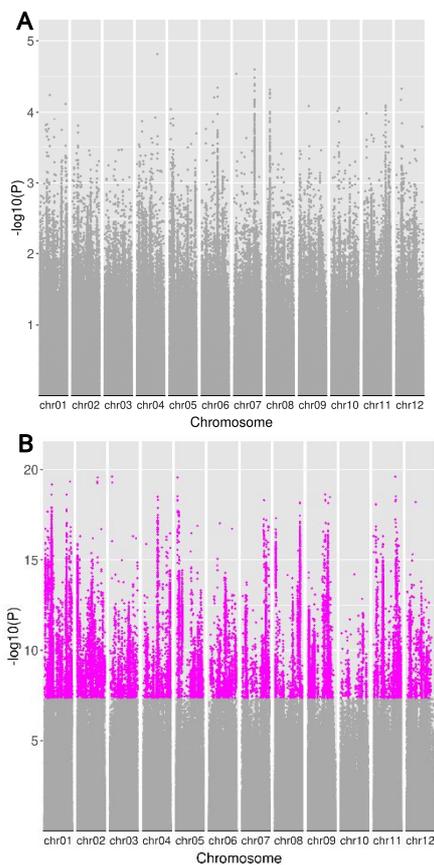


図1 種子長の GWAS 結果
種子長の GWAS を混合線形モデル (A) または一般線形モデル (B) で解析した結果をマンハッタンプロットで示した。マゼンタで示したドットが、統計的に優位な関連を示した多型。

glaberrima の全ゲノム遺伝子型データを解析した結果、*O. sativa* と考えられる系統が 1 系統混ざっていることが判明した。そのため、この 1 系統は解析から除外した。残りの 172 系統を用いて、高温処理による不稔率、種子重量の減少に加えて、種子サイズ、葉サイズ、草丈、穂長、到穂日数など、様々な農業形質について調査を行った。これにより得られた形質データと遺伝子型データを用いて、それぞれの形質に対する遺伝効果の大きさと関与する遺伝子座の同定を試みた。種子サイズ（幅および長さ）の遺伝率（形質が遺伝的に決定される割合）はおよそ 40% であった。アジアイネ *O. sativa* においても種子サイズは比較的遺伝率が高い形質と知られている。しかし、GWAS で一般的に用いられる混合線形モデルによる解析では、種子サイズと統計的に優位な関連を示す多型（DNA 配列の変異）を見出すことができなかった（図 1 A）。一方で、一般線形モデルで解析すると多数の統計的に優位な多型を検出できた（図 1 B）。これは、混合線形モデルでは、解析集団が潜在的に持つ集団構造化の影響を取り除く統計モデルを利用している影響だと考えられる。すなわち、*O. glaberrima* 品種群にみられる種子サイズの多様性が、集団が持つ遺伝構造と相関があり、混合線形モデルを用いると真に種子サイズに影響を与えている多型の遺伝効果までもを過剰に取り除いてしまうためである。実際のところ、*O. glaberrima* 品種群の多型データに基づいて非階層的クラスタリングの一手法である主成分分析を行うと、各品種が採取された地域ごとに遺伝子型に偏りがあることが示唆されている。同様の結果が、種子サイズ以外の形質でも見られており、それぞれの *O. glaberrima* 品種は栽培地域ごとに独立して独自の形質を獲得している可能性が高い。過去の研究においても、*O. glaberrima* の栽培化や品種の成立が地域ごとに独立して同時多発的に起こった可能性が示唆されており、今回の結果はそれを反映したものであると考えられる。一般線形モデルで GWAS を行った場合、真に形質と関連が多型（真陽性）に加えて集団構造化の影響による偽りの関連性（偽陽性）が検出されてしまうという問題がある。図 1 B に示した統計的に優位な関連を示す多型は非常に多くの偽陽性を含んでいるものと考えられる。しかし、一方で真陽性が存在することも事実である。そこで、形質ごとに統計的に優位な関連が見られた多型を整理し原因遺伝子の候補をリスト化・カタログ化するツールを開発した。カタログ化では、まず最も優位な関連を示した多型サイトを選び、その多型と相関係数の二乗が 0.6 以上となる多型サイトの集合を一つの原因遺伝子による効果を示す「ピーク」を形成する多型サイトのセットであると考えた。また、最も優位な関連を示すピークを形成する多型サイトを除外した残りの多型サイトから再び最も優位な関連を示すピークを見つけ出す処理を繰り返し、検出された多型サイトを特定の領域に原因遺伝子があることを示すピークのリストとして整理した。このリストのピークごとに多型サイトの変異が当該領域に座する遺伝子やそこから翻訳されるタンパク質のアミノ酸配列に及ぼす効果を推定した。また、それぞれのピークの遺伝子型パターンと形質値の分離を視覚化するボックスプロットも描画するプログラムも実装した。この GWAS カタログ化ツールにより、効率的に候補遺伝子の絞り込みや遺伝子アノテーション情報に基づく原因遺伝子の推定を行う基盤が構築された（図 2）。今後、このカタログ化された情報に基づき候補遺伝子の中から原因遺伝子の同定を進めることを計画している。

さらに、当初計画に加えて、次世代シーケンサーにより得られた大量の多型データを効率的に可視化しフィルタリングやエラー修正を行うためのツール「GBScleanR」を開発した。これは、本研究における次世代シーケンサーにより得られた多型データを解析する中で、次世代シーケンサーにより得た多型データには想像以上にエラーが含まれており、今後マッピング集団などを出し原因遺伝子の同定を順遺伝学的に進める上で正確な多型データを得るためのツールが必要になると考えられたため開発に至った。GBScleanR は R パッケージであり、各種 OS の R 環境で実行可能である。本パッケージのエラー修正アルゴリズムは、リードカウントデータに基づいて親系統と雑種集団の遺伝子型を同時に推定する。そのため、親系統における欠損値や遺伝子型判定エラーも雑種集団のデータをもとに修正できる。また、親系統がホモ固定された近交系でもヘテロ接合を持つ非近交系でも解析可能である。さらに二系交雑のみならず、四系や八系

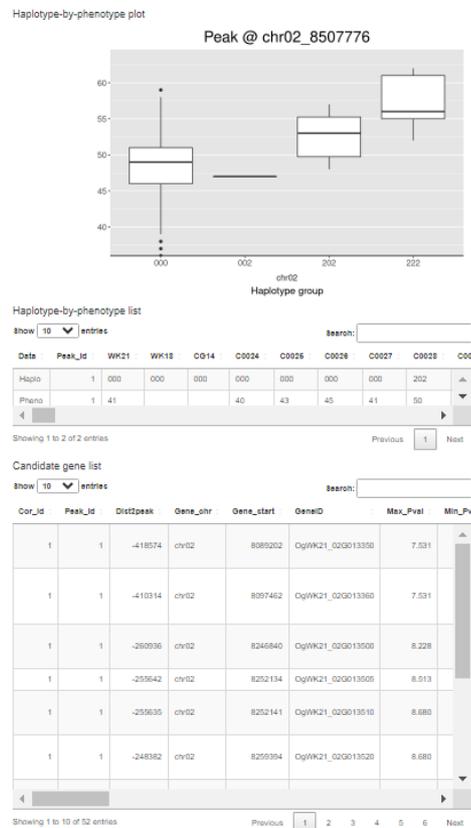


図 2 GWAS カタログ
各ピーク（原因遺伝子座候補領域）における遺伝子型パターンと形質値の関連を視覚化するボックスプロット（A）およびその実測値をまとめた表（B）、さらに候補領域に座する遺伝子の機能アノテーション情報と検出された DNA 配列変異の効果のリストが、HTML 形式のファイルでカタログとして出力される。

といった多系交雑集団にも対応している。エラー修正の性能比較のため、栽培イネと野生イネに由来する F2 集団から得た実データとシミュレーションデータにおけるエラー修正精度を数値化した。シミュレーションデータでは、実データにおいて観測されたアリルリードバイアス (ARB) によるエラーを再現した。ARB とは、ある個体の SNP マーカー座乗領域がヘテロ型である場合、参照アリル型リードと対立アリル型リードがおおよそ半々の数得られるはずが、一方のアリルに偏ってしまう現象である。遠縁交雑集団では、ゲノム構造の違いから PCR バイアスなどの要因により ARB が発生しやすいと考えられる。GBScleanR と既報の類似した機能をもつツールである「LB-Impute」および「magicImpute」との性能比較を実施したところ、ほぼすべてのデータセットにおいて、GBScleanR が最も高い正答率 (正しい遺伝子型を推定した割合) を示した。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 Furuta Tomoyuki、Yamamoto Toshio、Ashikari Motoyuki	4. 巻 -
2. 論文標題 GBScleanR: Robust genotyping error correction using hidden Markov model with error pattern recognition	5. 発行年 2022年
3. 雑誌名 bioRxiv	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1101/2022.03.18.484886	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Furuta Tomoyuki、Yamamoto Toshio、Ashikari Motoyuki	4. 巻 224
2. 論文標題 GBScleanR: robust genotyping error correction using a hidden Markov model with error pattern recognition	5. 発行年 2023年
3. 雑誌名 GENETICS	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/genetics/iyad055	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計4件（うち招待講演 1件/うち国際学会 1件）

1. 発表者名 古田智敬、芦苺基行、山本敏央
2. 発表標題 他殖性作物の多系交雑集団にも対応した遺伝子型データエラー修正ツール「GBScleanR」の開発
3. 学会等名 日本育種学会第141回講演会
4. 発表年 2022年

1. 発表者名 古田 智敬 , 佐藤 豊 , 芦苺 基行
2. 発表標題 アフリカイネ <i>Oryza glaberrima</i> 品種群を用いたゲノム育種基盤構築
3. 学会等名 日本育種学会第 1 3 9 回講演会
4. 発表年 2021年

1. 発表者名 Tomoyuki Furuta, Toshio Yamamoto
2. 発表標題 Pipeline for accurate Genotype-by-Sequencing via eliminating error prone markers
3. 学会等名 The 19th International Symposium on Rice Functional Genomics (国際学会)
4. 発表年 2022年

1. 発表者名 古田智敬
2. 発表標題 Pipeline for accurate genotype calling via eliminating error prone markers.
3. 学会等名 第18回日本ナス科コンソーシアム年会 (招待講演)
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------