

厚生労働科学研究費補助金
食品の安全確保推進研究事業

新たなバイオテクノロジーを用いて得られた食品の
安全性確保とリスクコミュニケーションのための研究

令和4年度 総括・分担研究報告書
【21KA1002】

研究代表者 近藤 一成

令和5（2023）年 6月

目 次

I. 総括研究報告書

近藤 一成	1
-------	---

II. 分担研究報告書

1. リスクコミュニケーションに関する研究

小泉 望	5
------	---

2. ゲノム上の意図しない変化の網羅的解析手法の開発と環境整備

柴田 識人	19
-------	----

3. ゲノム編集食品のトランスクリプトーム変化の検証

吉場 聡子	59
-------	----

4. メタボロームインフォマティクスによる未知化合物推定

早川 英介	75
-------	----

5. アレルゲン性予測、食物アレルゲン解析、リスク評価と AI

安達 玲子、為広 紀正	79
-------------	----

6. 新規アレルゲン性評価手法開発のための基盤的研究

富井 健太郎	89
--------	----

III. 研究成果の刊行に関する一覧表

	101
--	-----

厚生労働科学研究費補助金（食品の安全確保推進研究事業）
新たなバイオテクノロジーを用いて得られた食品の安全性確保と
リスクコミュニケーションのための研究

総括研究報告書

研究代表者 近藤 一成 国立医薬品食品衛生研究所

研究要旨：

リスクコミュニケーション分野では、細胞培養を用いた製品（肉、魚、乳など）、昆虫食等の代替タンパクについて、その理解度や受容度を調査し、日本国内では、植物性由来代替肉や培養肉などの細胞性食品への要求があまりないことから、理解や受容は進んでいないことが分かった。

ゲノム解析分野では、全ゲノムシーケンスデータをアセンブリ解析することが挿入配列の検出や配列解析に有効であること、及び必要なデータ量を示した。意図しないゲノム由来の変化を捉える手法として、トランスクリプトーム解析（遺伝子発現変化）からの検討も実施した。ゲノム編集前後のサンプルの解析から標的遺伝子産物以外の遺伝子にも発現の増減が認められた。さらに、ゲノム解析の弱点を補完できるかどうかを含めてさらに検討する必要があると考えられた。意図しない変化を捉えるための主要な手法としての代謝成分変化の解析では、これまでに確立したスペクトル類似度による未知化合物の同定、あるいは、推定を行うシステムを、ゲノム編集食品に応用した。目的の遺伝子以外の多数の代謝化合物が変動することが示唆された。今後それら化合物の同定を行い、手法の有用性を検証していく必要がある。解析環境として、網羅的な代謝物変化を捉えて解析する手法を Web ベースで行う環境を作製した。今後活用されることが期待される。タンパクのアレルゲン性予測では、深層学習を用いた手法（protein BERT）が最も優れていることを見いだした。また、食物由来タンパクの主要組織適合性複合体である HLA（MHC）への結合性を予測する手法を検討して、前者と併せて異なる 2 つの観点からのアレルゲン予測手法を通じて、精度の高い予測システムを構築していく。運営維持しているアレルゲンデータベース ADFS の内容を更新するために、エピトープ情報とアレルゲン情報を追加した。さらに、人工知能（AI）の各研究分野への応用が活発であることから、今回リスク分野への応用について文献等調査を行った。規制機関での活用は、規制プロセスの効率化、データの深い洞察への活用、その他、食品偽装で応用できる可能性を有していることが示唆された。

本研究班は、研究代表者を含む以下の7名から構成され各分担課題について研究を行った。

研究分担者	安達 玲子	（国立医薬品食品衛生研究所）
研究分担者	柴田 識人	（国立医薬品食品衛生研究所）
研究分担者	吉場 聡子	（国立医薬品食品衛生研究所）
研究分担者	小泉 望	（大阪公立大学）
研究分担者	早川 英介	（沖縄科学技術大学院大学）
研究分担者	富井 健太郎	（産業技術総合研究所）

以下に、研究目的、方法、および研究成果の概要を記載する。研究内容の詳細については、各分担報告書に記載した。

A. 研究目的

遺伝子改変技術を応用した食品開発は、技術的には外来遺伝子導入による遺伝子組換え食品（GM 食品）から生物自身が持つ内在性遺伝子改変で新たな形質を生み出すゲノム編集技術応用食品（ゲノム編集食品）へ、また、その生物が持たない多数の遺伝子を導入した酵母などから新規食品機能成分を産生させる合成生物学利用も諸外国を中心に進んでいる。ゲノム編集技術では DNA 2 本鎖切断を誘導するオリジナル手法から、1 本鎖切断から 1 塩基編集を行う塩基置換編集（Base editing）、これを発展させ数塩基の自由な組合せの塩基編集（prime editing）、さらに標的配列への制限をなくした PAM レス（PAM 配列を要求しない）編集、RNA 編集など非常に多様な手法が生み出され、そこから想定される意図しない変化も一様でないことが明らかになりつつある。したがって、配列に依存しない意図しない塩基変化やそこから生じる代謝成分の変化を網羅的に検出または予測し、その変化が与える影響を正確に評価することは、食品の安全性確保において急務の課題である。また、これらゲノム編集食品や合成生物学利用食品に加えて、細胞培養によって作成される肉、魚、乳、チョコレートなど細胞性食品の研究開発が急速に進んでいる。ますます、リスクコミュニケーションの重要性が増している。国民受容とともに製品開発や普及が並行して進むことで、規制精度の整備、国民理解と受容、イノベーション推進が進むことが望まれる。

本研究では、（１）多様な遺伝子改変技術と開発に関する情報収集、（２）一様でない意図しない変化の影響解析のための手法開発（ゲノム、代謝成分、アレルゲン性）、（３）ゲノム編集食品の理解の前段階として不可欠な国内 GM 食品利用の現状と審査届出制度の理解に重点したリスクコミュニケーション（若手研究と連携）、（４）リスク評価側の最新技術理解と能力向上、人材育成を柱に若手研

究代表者とも連携して実施する。

B. 研究方法

本研究班構成では、意図しないゲノム DNA 配列の変化の解析手法開発と標準化を柴田が、意図しないタンパクの生成に伴うアレルゲン性の評価手法開発と実用化およびアレルゲンデータベース ADFS の維持更新を、深層学習も取り入れながら安達、富井が、また、意図しない代謝物変化の網羅的開発手法の開発と Web 環境で利用できるよう実用化を早川が担当した。リスクコミュニケーションについては、ゲノム編集食品、合成生物学利用食品、特に代替タンパク質に重点を置きながら小泉が担当した。

C. 研究成果

本研究では、新しいタイプの食品（ゲノム編集食品、細胞性食品等）に対するリスクコミュニケーション、および、意図しない変化を検出するための解析手法の開発を行っている。

リスクコミュニケーションに関して、ゲノム編集食品や細胞培養によって作成される肉や乳、昆虫食等の中で「代替タンパク」にフォーカスして前年度の 5,000 人意識調査結果を解析するとともにグループインタビューを実施した。その結果、日本国内では、植物性由来代替肉や培養肉などの細胞性食品への要求があまりないことから、理解や受容は進んでいない。認知度は、植物由来代替肉＞昆虫食＞微細藻類＞培養肉＞代替乳の順に高く、食べたいかどうかという設問については、植物由来肉＞微細藻類＞人工乳＞培養肉＞昆虫食の順であり、昆虫食の嫌悪感が強い。この分野（いわゆるフードテック）における用語が統一されていない現状で、リスクコミュニケーション推進にはそこから解決していくことも必要である。

意図しない変化の検出方法の検討では、ゲノム、

代謝成分、タンパクアレルギーを標的に手法開発を行っている。ゲノム上の変化では、昨年度の検討において全ゲノムシーケンスデータをアセンブリ解析することで、ゲノム上での残存・挿入箇所・挿入された配列を明らかにできることを示した。そこで、今回解析可能な残存配列の長さやデータ量を検討した。その結果、25×以上のシーケンスカバレッジデータがあれば、40–4,000 bp 程度の長さの標的配列を 100%の確率で再現できることを、そして、本アセンブリ解析がゲノム編集食品の外来遺伝子の有無を調べる手法として妥当であることを示唆していた。一方で、意図しない変化をゲノム解析だけで行うのは難しい。そのため、トランスクリプトーム解析（遺伝子発現変化）からの検討も実施した。ゲノム編集前後のサンプルの解析から標的遺伝子産物以外の遺伝子にも発現の増減が認められた。これ自体は珍しい現象ではないが、意図しない変化の解析という観点から、ゲノム解析の弱点を補完できるかどうかを含めてさらに検討する必要がある。

意図しない変化を捉えるための主な手法としての代謝成分変化の解析では、これまでに確立したスペクトル類似度による未知化合物の同定、あるいは、推定を行うシステムを、ゲノム編集食品に応用した。ゲノム編集トマトと未編集トマト 9 種の LC/MS のノンターゲット解析から主成分分析することで、目的の GABA 以外の多数の化合物が変動することが示唆された。ただし、一つの遺伝子の変化が多数の成分含量変化に影響することは生物システムでは珍しいことではない。これらの独自に開発しているシステムを、Web 環境として公開していくための検討も行った。タンパクのアレルギー性予測では、これまでに機械学習手法からサポートベクターマシン（SVM）を用いた予測法を開発して、既存モデルよりも高い精度であることを示した。一方で、絶対的性能ではまだ改善の余地が残されていたため、予測に用いる機械学習の分類器を種々検討して、深層学習を用いた

手法（protein BERT）が最も優れていることを見いだした。また、運営維持しているアレルギーデータベース ADFS の内容を更新するために、エピトープ情報とアレルギー情報を追加した。タンパクのアレルギー性予測では異なる視点からも検討している。食物アレルギー発症に至る最初の反応段階である、食物由来タンパクの主要組織適合性複合体である HLA（MHC）への結合性を予測する。ここで親和性のないものは IgE 抗体も生成しないため食物アレルギーに至らないと考えられる。現在までに、既存 DeepSeqPanII を基本に改良とデータセットを変えながら予測性能を比較検討した。これまでに既存よりも高い性能を確認できた。さらに、人工知能（AI）の各研究分野への応用が活発であることから、今回リスク分野への応用について文献等調査を行った。規制機関での活用は、規制プロセスの効率化、データの深い洞察への活用、その他、食品偽装で応用できる可能性を有していることが示唆された。

情報収集と解析は、今年度から「食品等試験検査費」で遺伝子組換え食品とゲノム編集食品などを一元的に行う事になったため、そちらで報告済みである。最近活発な研究開発がされている細胞農業（細胞性食品、いわゆる培養肉等）について、規制当局が発行したものを検索して、FAO/WHO が 2023 年に公表したレポートの解析を行っている（つい最近の発行のため、内容の把握が完了していないことから、詳細内容や課題などは次年度の研究報告書で報告する）。

Food safety aspects of cell-based food (FAO/WHO, 2023)

<https://www.fao.org/documents/card/en/c/cc4855en>

D. 健康危険情報

該当なし

研究業績、知的財産権の出願などは、各分担報告書を参照。

厚生労働科学研究費補助金（食品の安全確保推進研究事業）
「新たなバイオテクノロジーを用いて得られた食品の安全性確保と
リスクコミュニケーションのための研究」
分担研究報告書（令和4年度）

リスクコミュニケーションに関する研究

研究分担者 小泉 望 大阪公立大学 教授

研究要旨：

遺伝子組換え作物やゲノム編集食品とは別に合成生物学やフードテックと呼ばれる方法で作られた食品が注目を集めている。本研究ではこうした食品を「新たなコンセプトで作られた食品」と呼び、中でも代替タンパク質に着目した。より具体的には昆虫食、植物由来代替肉、培養肉、代替乳製品（人工乳、精密発酵乳）、微細藻類を対象とした。国内外の研究開発動向の調査と消費者の量的、質的意識調査を行った。海外では目まぐるしく代替タンパク質の研究開発が進んでいるが、日本では2022年度になって産業界を中心とした取り組みが活性化しつつあるものの諸外国と比べると限定的である。2021年度末に行った5,000人を対象とした大規模アンケートの解析を行った。昆虫食についてはネガティブな言葉がより選ばれて、悪いイメージが連想されやすい傾向にあった。培養肉は昆虫食に次いで「気持ちが悪い」という言葉が連想されていたが、ネーミングが影響した可能性もある。科学的関心が代替タンパク質への嗜好と関連していることも示された。欧米諸国の調査結果と比較すると大まかには同様の結果であったが、例えば昆虫食への拒絶感は日本では若年層がより高いのに比べ、海外では高齢層で高かった。また微細藻類に関する受容度は日本では欧米と比べて高いなどの違いが見られた。代替タンパク質が注目される背景と5つのタンパク質をとり上げ、消費者（子供を持つ主婦）を対象とするグループインタビューを6回行った。

A. 研究目的

新たなバイオテクノロジーによって作られる食品としては、除草剤耐性や害虫抵抗性を持つ遺伝子組換え農作物に由来する食品（遺伝子組換え食品）が主流であり、1996年に実用化が始まり既に25年以上利用されている。遺伝子組換え微生物由来の添加物（ビタミンや精製酵素など）も2001年から使用されている。遺伝子組換え技術を用いた食品と添加物を比較すると前者ではリスクコミュニケーションが困難な状況が続いている。諸外国でも同様で、単なる科学的な説明に加えてELSI（倫理的・法的・社会的課題）が関与し、コミュニケーションを複雑にしている。一方、添加物については今のところ、そうした混乱は見られない。

2021年秋に我が国においてゲノム編集技術応用食品（ゲノム編集食品）の実用化が始まった。具体的にはGABA高蓄積トマト、可食部増量マダイ、高成長トラフグである。ゲノム編集食品に関するリスクコミュニケーションはその実用化の数年前から複数の機関により始まった。行政、国の研究機関、大学、NPOなどである。開発者により設立されたベンチャー企業もコミュニケーション活動に積極的に関与した。遺伝子組換え食品では市場に導入されてからコミュニケーションが始まったのに対してゲノム編集食品では実用化の前に対応がとられたこともあり、現状では比較的冷静なリスクコミュニケーションが行われている。

既存の遺伝子組換え食品、ゲノム編集食品に続き新しいバイオテクノロジーを用いた食品が登場

しつつある。その製造法や性質、用途が多岐に渡るため一般的な呼称は定着していないが「合成生物学 (Synthetic biology : Synbio)」、「フードテック」と言った用語が使われることが多い。しかし、Synbio、フードテックの概念はかなり漠然としており、分類のされ方も様々である。両方の要素を持つ食品も少なくないがイコールではない。遺伝子組換え技術あるいはゲノム編集食品技術が使われることもあるが、一概に遺伝子組換え食品と同列に扱うのは適当でない場合が多い。Synbio とフードテックの両方の概念を併せ持ち遺伝子組換え技術を使った食品の例として、米国の Perfect Day 社が開発した乳製品（実際に乳製品と呼べるかどうかは議論の余地があるが）が挙げられる。同社では、牛乳の主要なタンパク質を遺伝子組換え技術により微生物で生産し、そのタンパク質を混合することで牛乳（代替乳：これも定着した用語ではない）あるいはアイスクリームを模した食品を製造することに成功している。このような製造方法は「精密発酵」と呼ばれることが多い。精密発酵由来食品はすでに米国では実用化されており、動物由来しないのでビーガンの人にも受け入れられている。

この乳製品に限らず「ポスト新たなバイオテクノロジーによって作られた食品」とでもいうべき食品を本研究では「新たなコンセプトで作られた食品」と呼ぶ（世間一般に認知された用語ではないことに注意）。しかし、前述のように「新たなコンセプトで作られた食品」は本研究で扱おうとする食品は幅広いことから明確な定義づけは容易ではない。また海外では Novel food、Innovative food といった呼ばれ方をすることもある。今後の日本での呼称については検討の余地があろう。

新たなコンセプトで作られた食品の特徴として、全てでは無いが「アニマルフリー」、「生物資源の保護」などのいわゆる遺伝子組換え作物の特徴である効率性とは違うコンセプトが挙げられる。さらにアニマルフリーは環境負荷の低減、動物愛

護などの異なる観点からとらえられる。環境負荷の低減は家畜の飼育による温室効果ガス排出量、水使用量、エネルギー使用量の軽減を意味する。例えば牛のげっぶは温室効果ガスのかなりの部分を占める。生物資源の保護の例としては微生物でのバニリン（バニラの香気成分）生産などがあげられる。この場合、バニリンの合成に関わる複数の酵素の遺伝子が導入されており新たな代謝系が構築されたといえる。こうした方法は、合成生物学、Synbio と呼ばれるが狭義な合成生物学の厳密な定義には当てはまらない。

また、世界人口の増加と生活レベルの向上に伴う肉食の増加によるタンパク質クライシスが懸念されている。そのため代替肉の研究開発も盛んである。代替肉は大きくは主に大豆を原料としたダイズミート（エンドウなどが使われる場合もある）に代表される植物由来代替肉（人工肉）と、細胞培養で牛などの細胞を培養し、それを成型する培養肉に大別される。植物由来代替肉は特にハンバーガーなどを中心にすでに国内外で実用化されている。米国では Impossible foods と Beyond meat が有名である。前者は遺伝子組換え技術を使いダイズのレグヘモグロビンを酵母で生産し、添加物として使用している。レグヘモグロビンの添加により、動物肉（主に牛肉）が含む血液の風味が加えられているとされる。国外（主に米国）ではすでに実用化されているが、国内では遺伝子組換え技術を用いていることから安全性審査が求められ実用化のハードルは高いと考えられる。

以上、例を挙げてきたような新たなコンセプトで作られた食品が主として国外で次々と開発され実用化されているが、その国内における認知度は低いと考えられる。近い将来こうした食品が国内でも流通する可能性は十分予想され、ホライズン・スキニングの考え方に基づき新たなコンセプトで作られた食品の安全性について効果的なリスクコミュニケーション手法を開発することは円滑な厚生労働行政に資すると考えられる。

2021 年度は①新たなコンセプトで作られた食品の国内外の事例調査、②新たなコンセプトで作られた食品に対する多様なステークホルダーの受け止め方の調査、③代替タンパク質に対する 5,000 人規模の意識調査をおこなった。

2022 年度はフードテック、Synbio 全てを対象とすると扱う食品の種類が膨大になることや現在世間で注目を集めていること、2021 年度で大規模意識調査の対象としたことなどから「代替タンパク質」を対象を絞り、そのリスクコミュニケーションに関する調査研究を行った。5,000 人規模の定量的調査を基に定性的調査を行い深掘りすることが目的である。具体的には①代替タンパク質に関する国内外の事例調査、②代替タンパク質に対する 5,000 人の意識調査結果の解析（意識調査自体は 2021 年度に実施）、③代替タンパク質に関する消費者を対象としたグループインタビュー（2 月 27 日、3 月 1 日、2 日、3 日、13 日、14 日に実施するので報告書作成時点では未実施）を行うこととした。④ゲノム編集食品のリスクコミュニケーションに関してもマスメディアの果たした役割を考察するとともに、2021 年度に引き続きその効率的な情報提供方法を検討することとした。

B. 研究方法

①代替タンパク質に関する事例調査

書籍等による調査に加えて、フードテックに関するサイト「Foovo」（<https://foodtech-japan.com/>）に登録し、各種の情報を取得するとともに、Foovo が主催する複数のセミナーに参加した。またバイオインダストリー協会（JBA）の Food Bio Plus の会員になり、やはり複数のセミナー、研究会に参加した。自身でも日本植物バイオテクノロジー学会において「植物バイオテクノロジーがもたらすフードテック&アグリテック・イノベーション」（2022 年 9 月 11 日）、「植物バイオテクノロジーと代替タンパク質」（2022 年 12 月 13 日）を企画し、情報提供と情報共有を行った。培養肉

の培地成分の開発を行った日本ハム、国内有数の乳業メーカー明治の研究者にインタビューを行うとともに③で記載するグループインタビューでの専門家からも聞き取り調査を行った。

②代替タンパク質に対する 5,000 人の意識調査結果の解析

5 つの代替タンパク質（昆虫食、培養肉、植物由来代替肉、代替乳（培養乳、人工乳、精密発酵乳などと呼ばれるが、本報告書では原則代替乳とする）、微細藻類）に関する一般的なイメージや受容、その期待などを比較するために、それぞれの短い説明文を事前に読んだ上で、モニターの属性も含む全 17 問の質問項目に選択肢を選んで回答してもらった。対象は日本全国の 20 歳から 65 歳の男女 5,000 人の登録モニターとした（2021 年度末に実施）。質問、選択肢は海外の先行研究をもとに設計した。具体的には「認知度」、「食経験」、「食べたかどうか」、「社会に受けられると思うか」等について尋ねた。設問中ではそれぞれの代替タンパク質について 24 のキーワードから印象を尋ねた。得られたデータは多重対応分析（Multiple Correspondence Analysis ; MCA）に供した。主として 5 つの代替タンパク質に対する印象の違いを解析した。

③代替タンパク質に関する消費者を対象としたグループインタビューとコンテンツ作成

5,000 人を対象とした意識調査で取り上げた 5 種類のタンパク質と代替タンパク質が注目されている背景に関してグループインタビューを行うこととした。グループインタビューは毎回異なる代替タンパク質の専門家 1 名とファシリテーター（大阪公立大学・山口夕准教授）、20 代から 50 代の 3 名の主婦（楽天インサイトを通じて募集）で行うこととした。毎回主婦は異なるモニターを選んだ。

具体的には以下の 6 回を設定した。

1. 2月27日「新食材“代替タンパク質”はなぜ求められているのか」
専門家：五十嵐圭介（東北大学大学院 助教／日本細胞農業協会 代表理事）
2. 3月1日「細胞農業と培養肉」
専門家：杉崎麻友（Forsea Foods（イスラエル）研究員／日本細胞農業協会 理事）
3. 3月2日「タンパク質源としての微細藻類」
専門家：佐々木俊弥（（株）タベルモ COO）
4. 3月3日「消費者目線の精密発酵」
専門家：橋詰寛也（（株）Kinish CEO）
5. 3月13日「植物由来代替肉の現状と今後」
専門家：穴井豊昭（九州大学 教授）
6. 3月14日「地域課題解決と養殖昆虫食の共進化」
専門家：佐伯真二郎（NPO 法人食用昆虫科学研究会 理事長／JICA 草の根技術協力事業 プロマネ）

グループインタビューの構成は

- ・趣旨説明、アイスブレイク（5分程度）
- ・専門家による簡単な代替タンパク質の紹介（5分程度）
- ・グループインタビュー（前半：70分）
- ・休憩（10分）
- ・専門家による情報提供（50分）
- ・グループインタビュー（後半：40分）

の計180分（3時間）と設定した。

尚、事前に代替タンパク質に関する簡単な動画（YouTubeとして配信）の視聴を求め、アンケート記入をインタビューの前後に行う。

6回のグループインタビューを通して得られた情報をもとにリスクコミュニケーションに使うための平易なコンテンツ（冊子）を作成する。2022年度中にコンセプトをまとめ2023年度の6月までを目標にコンテンツを仕上げ、多くのステーキ

ホルダーが入手可能な方法での配布を行う。

④ゲノム編集食品に関するリスクコミュニケーション

ゲノム編集食品の取り扱いルールが出来た経緯と新聞報道、多様なステーキホルダーの情報共有によるリスクコミュニケーションについて考察した。新聞報道については大阪公立大学の図書館で閲覧可能な大手4紙（朝日、毎日、読売、日経）から見出しに「ゲノム編集食品」を含む記事を抽出した。

お米の研究会（市民グループ）、灘高等学校、日本ハム、奈良県、伊丹サイエンスカフェで講演ならびに意見交換を行った。各対象に応じて内容を変更し、適宜資料を配布した。また高GABAトマト、肉厚マダイ、高成長トラフグの上市に応じて「ゲノム編集食品について話す」と題した冊子の改訂を行った。

C. 研究結果および考察

①代替タンパク質に関する事例調査

代替タンパク質が注目されている背景や実用化、消費者の動向の概要について述べる。

1) 海外の研究開発動向

欧米では数年前から代替タンパク質に関する関心が高まっていた。コロナ禍、異常気象、さらにはウクライナ危機もあって、環境問題、食料問題への懸念から欧州では肉食を避ける傾向がさらに強まっている。オランダはフードテックで世界をリードしており、世界20大食品企業のうち15社がオランダに拠点を置いている。産学官連携、先端技術を国家が支援している。

米国はそれ以前からフードテックが進んでおり、目的でも述べた Impossible foods や Beyond meat による植物由来代替肉の消費が盛んである。また、精密発酵で作られた Perfect Day 社などによる代替乳製品の導入が進んでいる。クリームチーズ Kiri など有名な Bel は Perfect Day と提携

し 2023 年には精密発酵由来代替乳を使ったクリームチーズを発売した。

欧米に限らず複数の国で国策として代替タンパク質中心とするフードテックが推進されている。例えばシンガポールは食料自給率が低い（10%）こともあり、2030 年までに国民に必要な栄養の 30%を国産にする「30 by 30」が国策として進められている。最初に培養肉（チキンナゲット）が販売されたのもシンガポールである。培養（鶏）肉は世界で初めて 2020 年 12 月に承認されている。シンガポールでは培養鶏肉に留まらず培養魚脂肪、培養魚肉、代用卵、代替乳など様々なスタートアップ企業による研究開発が進められている。

イスラエルもまた代替タンパク質の研究開発が盛んである。官民の投資でフードテック 400 社が設立されている。培養肉に加えて昆虫養殖のコンソーシアムなども設立されている。培養魚肉、培養（代替）乳製品などの研究開発も盛んである。この中には培養鰻の開発を手がける Forsea foods 社なども含まれる。③で実施予定のグループインタビューでは専門家として当社の研究員が参画する。イスラエルの Remilk 社は Perfect Day に続く精密発酵による代替乳製品の開発企業である。また Pigmentum は微生物では無く遺伝子組換えレタスでの乳タンパク質生産を目指している。

韓国は昆虫食を国家として 2014 年から推進している。カイコ蛹を食用としていることから②で述べるような日本での昆虫食に対する忌避感はいささか小さいと考えられる。また、培養肉に関しても産官学の R&D が支援されている。

上述の事例は世界のフードテックのごく一部に過ぎない。フードテックのカオスマップは数カ月単位で更新必要なほど研究開発動向は目まぐるしく変化している。

2) 国内の研究開発動向

日本においても政府が後押しする形で官民の連携が図られている。具体的には農林水産省が事務局となり 2020 年に「フードテック官民協議会」が

設立された。また産を中心とした動きとしては 2022 年にはバイオインダストリー協会（JBA）が Food Bio Plus を立上げ勉強会を開始している。

国内における研究開発動向については②の意識調査の対象とした 5 種類の代替タンパク質について述べる。他にも菌類由来の代替タンパク質等を実用化する動きがある。

植物由来代替肉（大豆ミートなど）については、すでに複数のメーカーから各種の商品が販売されている。ベジバーガーを食せる店も増えつつある。加工面での品質改善に加えて青臭さを減らした大豆の育種が行われ、品質向上に貢献している。料理番組でも食材として取り上げられる等、市民権を得ている一方で、店頭で商品が目につきにくいケースも多い。

微細藻類については、主に健康食品として利用されている。日本人にとっては馴染みの深いクロレラやユーグレナ、スピルリナなどがある。尚、このうちスピルリナは海外で古くから食されてきた経験もあり、欧米では認知度はクロレラより高い。

昆虫食については、日本では特定の地方では蜂の子、イナゴなどが食されてきた歴史がある。ここ数年はコオロギに注目が集まり食用にコオロギを養殖するスタートアップも出てきており、すでに「コオロギせんべい」として商品化されている。また、様々な昆虫を食品用に販売している企業もある。メディアへの露出も多い。しかし、②で述べるように昆虫に対するモニターの受容度は低く、どの程度消費量が増えて行くかは未知数の部分が多い。

培養肉についてはインテグリカルチャー社が独自の細胞培養技術を用いて日本ハムやマルハニチロ、日清食品などと提携して研究開発を進めている。細胞農業と称されることが多く、NPO 等もあるが、その実用化については現時点では不透明である。

精密発酵による代替乳製品に関しては手がける企業が現状では国内には無いと思われる。

3) 国内外の社会受容とその背景

代替タンパク質の社会受容を一概に論ずることは勿論できない。植物由来代替肉、微細藻類、昆虫食に関しては既に流通している。特に遺伝子操作などを行っているわけでは無いので、昆虫を食べることに対する忌避感などはあっても危険視する傾向は見られない。培養肉に関しては2018年に日清食品と弘前大学による調査が、2019年には帯広畜産大による調査が行われているが、未だ消費者の意識が十分に把握されているとは言えない。一番の理由は実用化されていないこともあり、その実態が消費者に分からないことが理由と考えられる。精密発酵による代替乳製品については、さらにイメージすることが難しく社会受容についての評価は現状では現実的ではない。

国内では代替タンパク質の導入あるいは肉食を避けることへの意識が欧米と比べて進んでいないように思われる。ビーガンやアニマルウェルフェアの観点も異なるが、地球環境問題への影響についての認識が大きく異なる。欧米の方が肉食に対する思い入れが強く肉の代替品を必要とするという説もある。また、環境問題への意識が強く、よく言われる牛のげっぶなどへの対応についての危機意識が強いとされる。こうした背景から代替タンパク質（主として植物由来代替肉）の消費がオランダ等では盛んである。環境問題への対応につながる肉食削減は政策的に進められていて、例えばオランダでは食肉広告を禁止する自治体がある。また、意外であるが酪農国であるニュージーランドでは牛や羊を飼育する酪農家にげっぶ税を課することが提案されるなど、世界の多くの国で地球温暖化を中心とする地球環境の悪化に対する懸念が高まっている。国内ではこうした代替タンパク質が注目されている背景が明確に伝えられていない。他の先進国ではいわゆる Planetary Health Diet に対する考え方が進んでいる。その普及にはビル・ゲイツの言動やポール・マッカートニーの Meat Free Monday への賛同といったインフルエ

ンサーの役割が大きい。日本では代替タンパク質のリスクコミュニケーションを行う際に、いわゆる食品安全だけを議論の対象とするのか、環境問題あるいはタンパク質危機への対応の概念も含めた上での議論をするかにより結果は異なると推測される。遺伝子組換え食品の場合は食品の安全性以外の論点がリスクコミュニケーションに持ち込まれ冷静なコミュニケーションを困難とした。代替タンパク質におけるコミュニケーションについても慎重な対応が求められる。例えば培養肉に関してはやはり科学的な安全性評価に基づくコミュニケーションが行われるかどうかに関しては懸念がある。すでに培養肉の安全性に疑問を呈する書籍も出版され、懐疑的な論評が週刊誌等で見られる。

4) 精密発酵による代替乳製品について

上述の様に精密発酵による代替乳製品の研究開発が日本で行われている気配はない。しかし、今後、研究開発が行われる可能性や海外からの参画も充分予想される。代替乳製品リスクコミュニケーションには意識調査が求められるが、培養肉に関しては先行調査があるのに対して代替乳製品に関しては、情報が圧倒的に不足していることもあり、意識調査は行われていない。そこで、2022年に意識調査を行うことを念頭に異なる観点から情報収集を行った。

前述のように Perfectday（米国）、Remilk（イスラエル）といったスタートアップ企業により代替乳製品の研究開発が進んでいる。世界のスタートアップ企業数は2022年12月現在で25社に上る。企業秘密にかかわることもあり、詳細は明らかとされていないが、基本的に牛乳成分のホエイタンパク質（ β -ラクトグロブリン）を微生物（酵母）で生産し、精製したものを原料にアイスクリーム、チーズなどが作られている。しかし、カゼインの効率的な生産方法が確立されておらずこのある製品の開発には至っていない。

精密発酵による代替乳製品は当然遺伝子組換え

技術を用いる。海外の例では目的のタンパク質（ホエイ）を精製するので、遺伝子組換え生物は含まれておらず米国では遺伝子組換え食品としては扱われない。日本での規制については明確では無い。精製タンパク質としては例えば乳製品の凝固に使われる組換え微生物由来のレンネットが遺伝子組換え添加物として認可されている。その他、複数の酵素が認可されているが、あくまで添加物で食品ではない。表示に関しては法的拘束力は付されない可能性が高いが透明性の確保という点ではメーカーの自主的な表示がかえって受容度を高める可能性がある。

②代替タンパク質に対する 5,000 人の意識調査結果の解析

認知度に関しては図1に示すように、植物由来代替肉＞昆虫食＞微細藻類＞培養肉＞代替乳の順に高かった。培養肉、代替乳は流通しておらず妥当な結果となった。食べたいかどうかという設問については、植物由来肉＞微細藻類＞人工乳＞培養肉＞昆虫食の順に受容度が高い結果となった（図2）。社会に受け入れられるかどうかについても同様の結果であった。食経験については流通していない培養肉、代替乳を除いて質問したところ、植物由来肉はそれなりに食べられていることが示された（図3）。表示についての質問では、代替タンパク質に関しては、大きな違いは見られなかった。

言葉との関連性については、表1および図4に示す。ここでもやはり昆虫食についてはネガティブな言葉がより選ばれて、悪いイメージが連想されやすい傾向にあった。また、培養肉は昆虫食に次いで「気持ちが悪い」という言葉が連想されていた。ここで、ネーミングについて注意が必要である。今回の調査では肉は培養肉としたが、微生物でタンパク質を生産して調整した代替乳を設問では代替乳製品とした。例えば培養乳という言葉を使っていたら結果に影響していた可能性が考え

られる。

培養肉に関しては、海外では cultured meat、lab meat、clean meat 等、様々な呼称が使われている。今後日本でどのようなネーミングを採用するかで社会受容に影響することが考えられる。一方で、「不味そう」、「不安な」といったネガティブなイメージは培養肉、代替乳ともに高かった。また、「信頼できる」を選んだモニターは限られていた。

栄養価については「栄養価が高い」という言葉は、モニターの4割以上が昆虫食、植物由来肉、微細藻類のイメージとして選んでいた。多重対応分析の結果では図4に示すように昆虫食はかなり別のグレードに入りネガティブな語句と関連が強い結果となった。また技術と伝統（自然）、動物と植物の二つの大まかな軸がみられた。植物由来肉はポジティブな言葉を連想しやすく、微細藻類のイメージと最も近い傾向にあった。培養肉はネガティブな言葉を連想しやすく、代替乳のイメージと最も近い傾向にあった。昆虫食は他の代替タンパク質とはイメージが大きく異なる結果となった。また科学的関心が代替タンパク質への嗜好と関連していることも示された。科学的関心が高いと培養肉や代替乳製品といった新しいバイオテクノロジーに対して寛容である傾向が示された。「このような科学的関心が、5つの代替タンパク質に対する態度を決める共通の重要な属性であることが分かった。

こうした結果を欧米諸国の調査結果と比較すると大まかには同様の結果であったが例えば昆虫食への拒絶感は日本では若年層がより高いのに比べ、海外では高齢層で高い。また微細藻類に関する受容度は日本では欧米と比べて高いなどの違いが見られた。各代替タンパク質に対する受け止めが多様である一方で政府が表示に介入することについては昆虫食を除き政府による表示規制の必要性については、受け止め方や受容性に比べて、代替タンパク質による違いはあまりみられなかった。このことは各代替タンパク質の社会受容や受

け止めか方の違いは必ずしも規制に必要性和直接的に結びつくものではないことを示唆している。

③代替タンパク質に関する消費者を対象としたグループインタビューとコンテンツ作成

グループインタビューは2023年2月27日～3月14日にかけて6回に分けて実施するため結果は2023年2月25日時点では出ていない。結果については2023年度の報告書で述べる。

④ゲノム編集食品に関するリスクコミュニケーション

ゲノム編集食品の取り扱いルールの策定が具体的に始まった2018年6月から2022年10月までの「ゲノム編集食品」を見出しとした新聞の報道数を調べた。届出、表示の概要示された2019年3月、正式に決まった同年9月、ゲノム編集トマトの届出が行われた2020年12月、トマトの上市、タイ、トラフグの届出、上市が行われた2021年秋に明らかに報道件数が増えている(図5)。こうした報道によりゲノム編集食品に関する議論が活性化されたと考えられ、マスメディアはリスクコミュニケーションにおいて一定の役割を果たした。

多様なステークホルダーとの対話を通して対象により情報提供のあり方を変える必要があることが改めて明らかになった。また、相手が頑なにコミュニケーションを拒めば適切な実施が難しいことも再認識した。今後改訂したコミュニケーション用冊子(図6)を有効活用する。

D. まとめと来年度の計画

5,000人を対象とした意識調査の解析結果では昆虫食がネガティブな言葉を連想させ、国内で実用化に至っていない培養肉、代替乳製品の認知度が低かった。代替タンパク質を巡る環境は開発動機も含め特に国外でドラスティックに変化している。遺伝子組換え食品やゲノム編集食品のように、代替タンパク質に関してもおそらく純粋な(科学

的な)食品安全以外の要因もリスクコミュニケーションに関与すると予想される。従って、リスクコミュニケーションにおける重要ポイントを抑える必要がある。2022年度末に実施するグループインタビューを有効に活用したい。ゲノム編集食品に関するリスクコミュニケーションの成果が代替タンパク質の場合に活かすことが出来ると考えている。

2023年度は、

- ①引き続き事例調査の実施
 - ②グループインタビューの結果のとりまとめ
 - ③インタビューに基づくコンテンツ(冊子)作成
 - ④インタビューに基づく中規模意識調査
 - ⑤精密発酵による代替乳製品の事例調査
 - ⑥意識調査
- を行い、代替タンパク質に関するリスクコミュニケーションにおける重要ポイントを明らかにする。特に代替乳製品について重点を置く。
- ⑦ゲノム編集食品についても引き続き研究する。

E. 研究発表・業績

1. 論文発表
 - 1) Takeda, K.F., Yazawa, A., Yamaguchi, Y., Koizumi, N., & Shineha, R. (2023) Comparison of public attitudes toward five alternative proteins in Japan. Food Quality and Preference, 105, 104787
2. 学会発表
 - 1) 2022年11月12日、リスク学会「ゲノム編集食品のリスクコミュニケーション」(招待講演)
 - 2) 2022年12月2日、日本食品衛生学会「ゲノム編集食品のリスクコミュニケーション」(招待講演)
 - 3) 2022年12月4日、日本サイエンスコミュニケーション協会「ゲノム編集食品に関するコミュニケーションの振り返りと今後」

3. 書籍

- 1) 小泉 望 (2023) リスクコミュニケーション
のために求められること:ゲノム編集技術 ～
実験上のポイント／産業利用に向けた研究開
発動向と安全性周知、p295-p301、情報機構
- 2) 小泉 望、山口 夕、標葉隆馬 (2022) OECD
加盟国におけるゲノム編集作物に関するパブ
リック・エンゲージメント事例集、科学コミ
ュニケーション研究所

F. 健康危険情報

該当なし

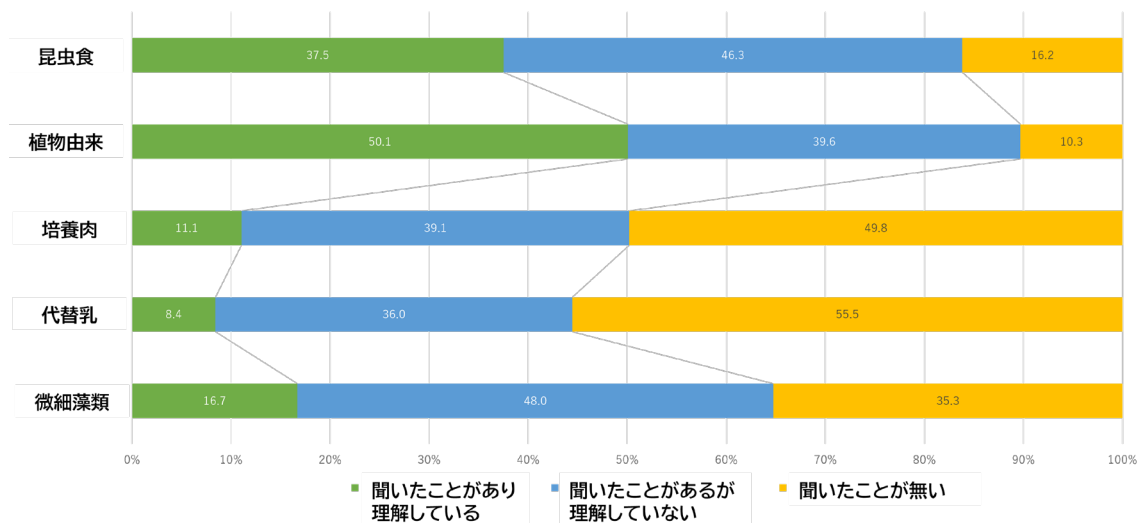
G. 知的財産権の出願・登録状況

該当なし

表1 各代替タンパク質へのイメージ

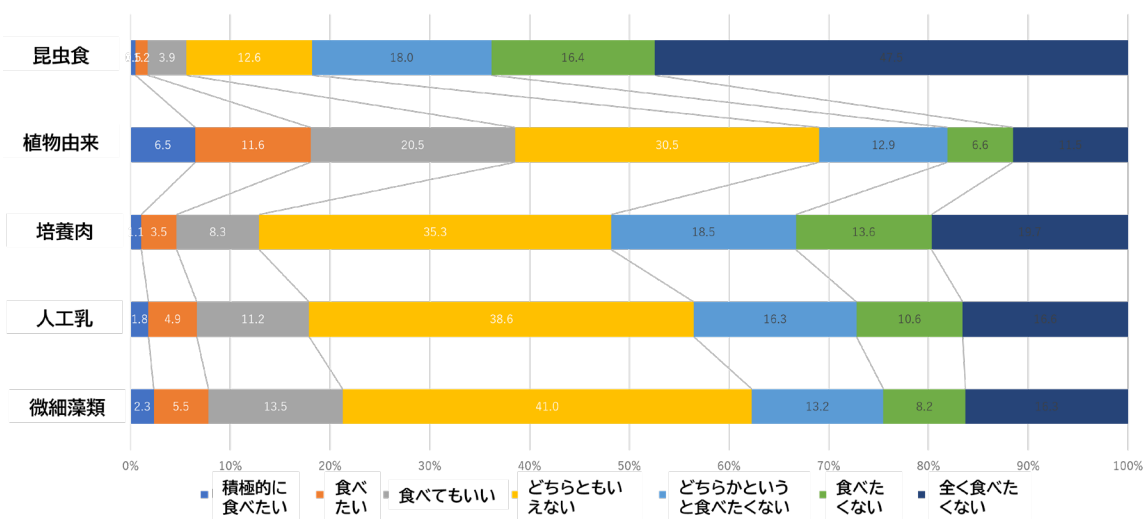
	健康に良い	環境にやさしい	持続可能な	気持ち悪い	贅沢な	庶民的な	地域に特有の	洗練された	伝統的な	自然な	美味しそう
昆虫食	21.3	23.4	19.2	74.3	4.5	11.3	29.6	4.7	18.7	40.8	3.6
植物由来	55.0	38.7	28.4	14.1	6.1	14.6	7.6	9.6	5.4	33.2	16.1
培養肉	12.5	33.3	33.4	49.8	10.3	10.9	11.0	12.5	4.2	11.6	6.7
代替乳	25.5	34.9	32.9	33.6	7.8	14.0	11.6	13.0	5.0	17.6	9.8
微細藻類	52.7	37.1	27.9	24.2	7.5	8.5	9.9	12.3	7.1	42.0	6.3

不味そう	安全な	評判の良い	匂いがある	親しみやすい	最先端の	便利な	不安な	信頼できる	倫理的な	ベジタリアンな	栄養価が高い	安心できる
68.6	6.5	4.3	43.2	4.4	18.6	7.9	53.4	4.1	8.8	12.1	44.4	3.8
26.7	25.3	14.9	14.8	18.8	31.8	21.3	18.7	16.5	12.3	54.8	44.6	23.2
57.9	9.8	5.6	28.0	6.3	58.5	23.8	63.8	7.0	24.8	13.5	18.7	6.4
48.6	12.9	7.1	24.6	10.3	51.8	26.2	50.4	8.8	17.4	19.8	29.8	10.0
38.0	17.6	10.4	20.6	9.0	39.2	16.4	27.1	12.0	10.9	31.6	51.5	13.6



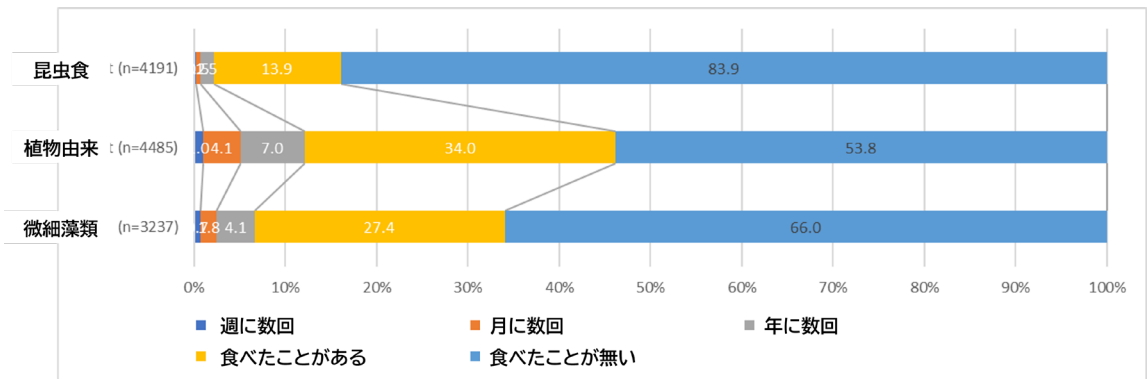
植物由来＞昆虫食＞微細藻類＞培養肉＞代替乳の順に認知度が高い

図1 代替タンパク質の認知度



植物由来＞微細藻類＞人工乳＞培養肉＞昆虫食の順に受容度が高い

図2 代替タンパク質の受容度



植物由来はそれなりに食されている(昆虫食の頻度は低い)

図3 食べる頻度、経験（培養肉、人工乳は対象とせず）

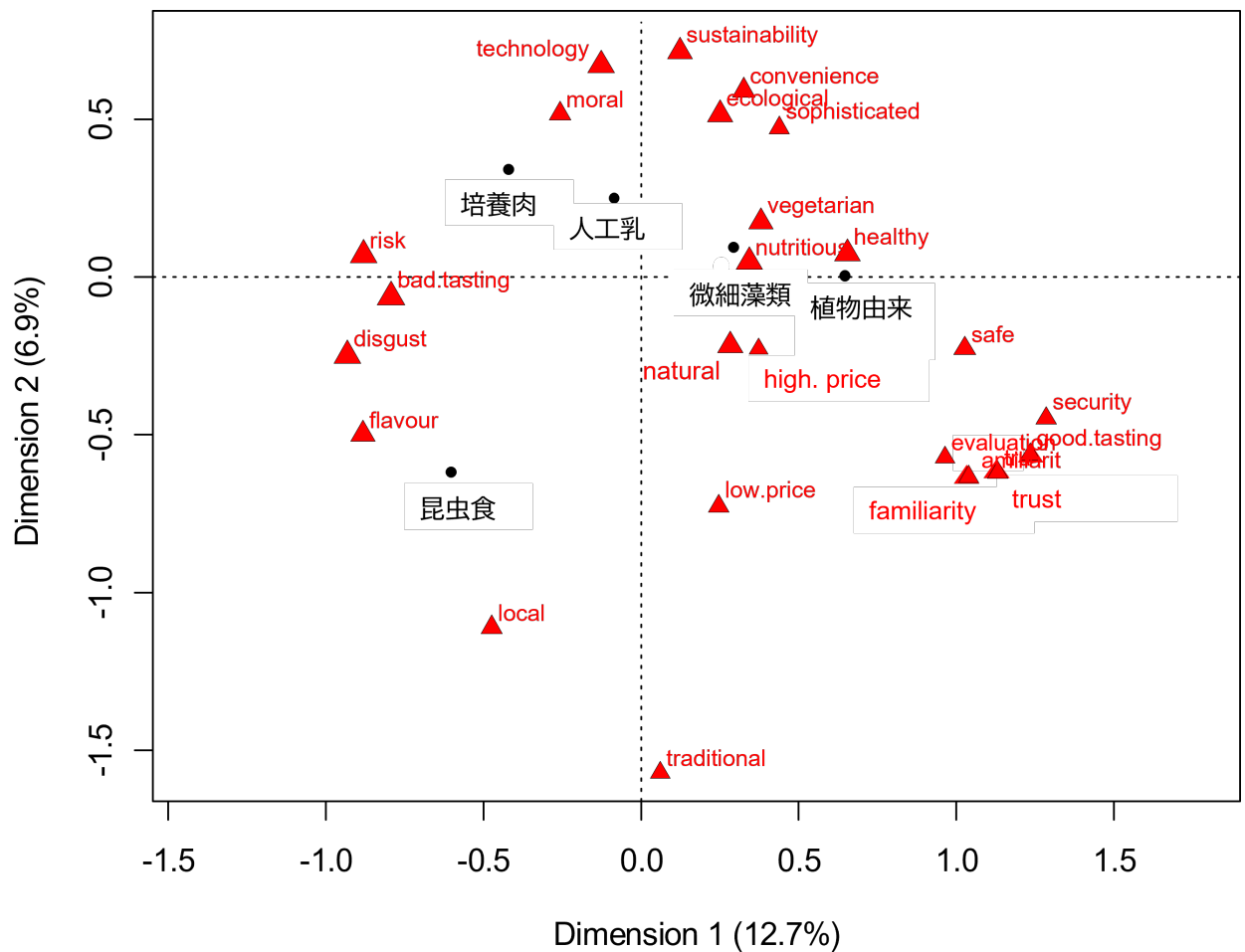


図4 代替タンパク質とイメージの多重対応分析の結果

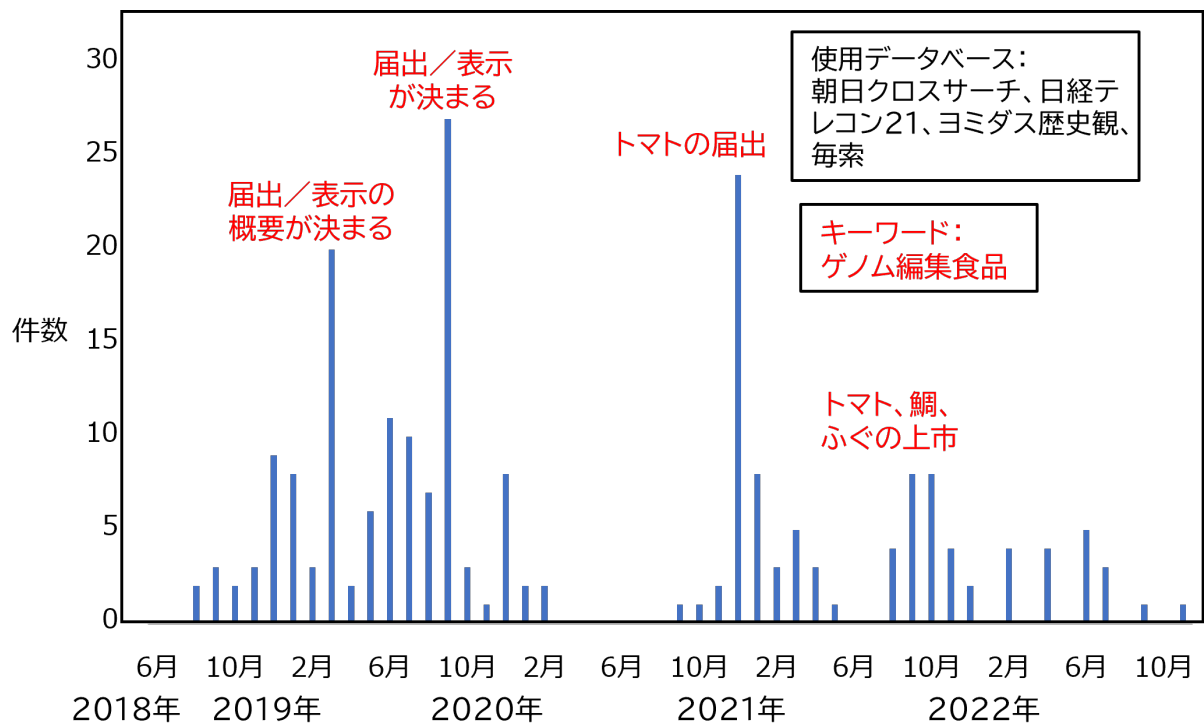


図5 見出しに「ゲノム編集食品」を含む新聞報道数

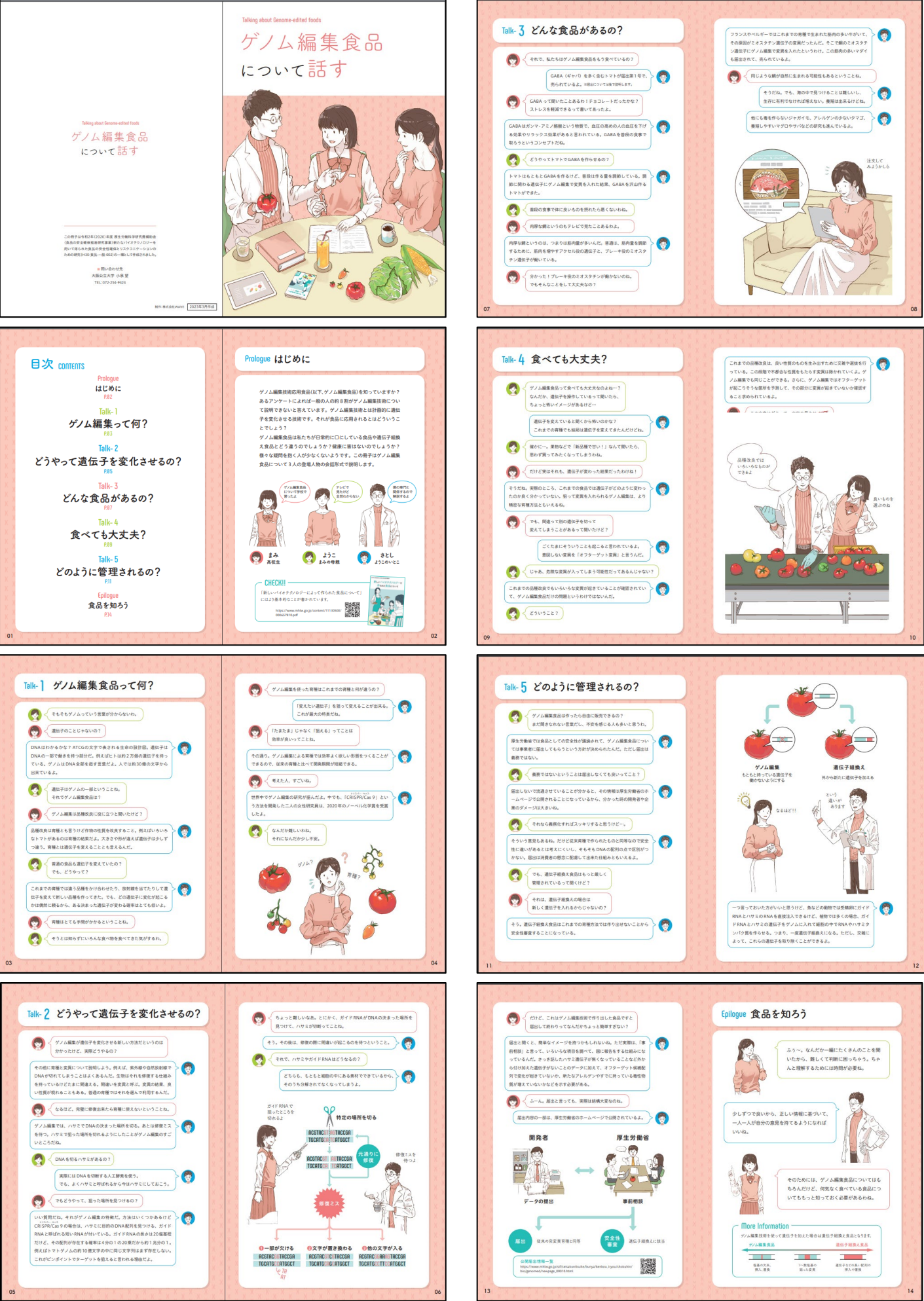


図6 ゲノム編集食品に関する冊子(改訂版)

厚生労働省科学研究費補助金（食品の安全確保推進研究事業）
「新たなバイオテクノロジーを用いて得られた食品の安全性確保と
リスクコミュニケーションのための研究」
分担研究報告書（令和4年度）

多様な遺伝子改変技術から生じる意図しない変化の網羅的解析手法の開発と環境整備に関する研究

研究分担者 柴田 識人 国立医薬品食品衛生研究所

研究要旨：

本研究では、ゲノム編集食品の安全性評価の一つである外来遺伝子の残存の有無を網羅的に調べる方法として、全ゲノムシーケンスによって得られたデータを用いた標準的解析手法の開発に取り組んでいる。これまでに、残存が想定される配列が予め分かっている場合、全ゲノムシーケンスデータをアセンブリ解析に供することで、残存の有無・挿入箇所・挿入された配列の内容を明らかにできることが示唆された。今年度は、外来遺伝子の残存性評価における本解析の妥当性を検討すると共に、解析可能な残存配列の長さや、解析で必要とされる全ゲノムシーケンスにて取得すべきデータ量（シーケンスカバレッジ）について検討した。こうした本手法の備えるべき必要要件を明確にすることは、全ゲノムシーケンスによる外来遺伝子の残存性評価法を標準的な安全性評価の一つにする上で必要不可欠なものであり、ひいてはゲノム編集食品の安全性評価の精緻化・向上に役立つと期待される。

研究協力者

曾我 慶介 （国立医薬品食品衛生研究所）
成島 純平 （国立医薬品食品衛生研究所）
杉野 御祐 （国立医薬品食品衛生研究所）

A. 研究目的

我が国では 2019 年よりゲノム編集食品の事前相談・届出制度が開始されており、これまでに3品がゲノム編集食品として届出・公表され、流通されている。この制度では、届出の対象となるゲノム編集食品の備えるべき要件の一つとして、ゲノム中に外来遺伝子を全く含まないことを求めている。厚生労働省「ゲノム編集技術応用食品等の取扱いに関する留意事項」ではこうした外来遺伝子およびその一部が残存していないことを調べる方法として、次世代シーケンサーによる全ゲノムシーケンス(WGS)、サンガーシーケンス、サザンブロット法などを挙げている。このうち、得られるデータの網羅性や、ゲノム編集技術に

よって生じる意図しない遺伝子変化全般の解析にも適用できる汎用性などを考え、次世代シーケンサーによる解析が着目されている。但し次世代シーケンシング解析の標準的手順や必要要件が明確になっていないことから、解析結果の再現性や信頼性の面で問題が起こる可能性がある。従って次世代シーケンシング解析を通知試験法に加えるには、この点は検討すべき課題である。

本研究ではゲノム編集食品の外来遺伝子の有無を調べる標準的な次世代シーケンシング解析手法の確立を試み、昨年度アセンブリ解析の有用性を報告した。そこで今年度は本アセンブリ解析の必要要件などを検討した。

B. 研究方法

1. サンプル

外来遺伝子残存モデルサンプルとして遺伝子組換えダイズ・トウモロコシを、ゲノム編集サンプ

ルとしてゲノム編集 GABA トマトを実験に供した。

遺伝子組換えダイズ

除草剤耐性遺伝子組換えダイズ RRS2 系統の認証標準物質である粉碎試料 Monsanto MON8978 8 Soybean Powder 994.0 g 超/kg (AOCS; 0906-B)、コントロールサンプルとして Monsanto Non-Modified Soybean Powder 1.0 g 未満/kg (AOCS; 0906-A) を購入した。

遺伝子組換えトウモロコシ

チョウ目害虫抵抗性トウモロコシ MIR162 系統の認証標準物質である粉碎試料 Syngenta MIR162 Maize Powder 998.80 g 超/kg (AOCS; 1208-A)、コントロールサンプルとして Syngenta Non-Modified Maize Powder 1.0 g 未満/kg (AOCS; 0407-A) を購入した。

ゲノム編集トマト

ゲノム編集トマト (シシリアンルージュハイギャバ、以後 GABA トマトと表記) はサナテックシード社より購入した。コントロールサンプルとして野生型シシリアンルージュを神奈川県内のスーパーにて購入した。

2. ゲノム DNA 抽出

ダイズ

各ダイズ粉碎試料 0.3 g から以下に示す Cetyl trimethyl ammonium bromide (CTAB) 法により、ゲノム DNA を抽出。

- 1) 1.5 mL の CTAB 抽出 buffer (1.5% CTAB, 7.5mM Tris-HCl pH8.0, 15mM EDTA, 1.05 M NaCl) を加え、56°C で 30 分間、振盪しながら加温。
- 2) 1.25 mL のイソアミルアルコール：クロロホルム = 1:24 を加え混合し、Hula mixer (Thermo Fisher Scientific) を用いて 30 rpm で 15 分間混合。

- 3) 遠心分離し、上清を回収。上記工程を 2 回繰り返す。
- 4) 上清に 100 μ L の 10%CTAB を加え混合。
- 5) 遠心分離し、上清を回収。CTAB 沈殿 buffer (1%CTAB, 50mM Tris-HCl pH8.0, 10mM EDTA) を 1.5 mL 加え、混合して、DNA と CTAB の複合体の沈殿を形成させる。
- 6) 遠心後に、上清を除去。沈殿に対し、1 M NaCl を 500 μ L および RNase A (ニッポンジーン) を 1.5 μ L 加え、55°C で 30 分間処理。
- 7) エタノール沈殿を行い、脱塩後に超純水にて DNA ペレットを溶解。

トウモロコシ

各トウモロコシ粉末試料 (MIR162 粉末 1.1 g、コントロール粉末 1.0 g) から NucleoBond® HMW DNA (MACHEREY-NAGEL) によりゲノム DNA を抽出。150 μ L の HE Buffer (5mM Tris-HCl pH8.5) で溶出し、ゲノム DNA 溶液を得た。

トマト

- 1) 各トマト青果試料は数 mm 間隔で輪切りにし、-80°C にて 24 時間予備凍結した後、24 時間凍結乾燥処理をした。凍結乾燥には EYELA 製 FDU-1200 を使用した。
- 2) 凍結乾燥したトマト青果試料 0.5 g を液体窒素中で乳鉢・乳棒にて破碎。
- 3) 10 mL の CTAB 抽出 buffer および RNase A を加え、56°C で 30 分間、振盪しながら加温。
- 4) 5 mL のイソアミルアルコール：クロロホルム：フェノール = 1:24:25 を加え、転倒混和。
- 5) 8,000 $\times$ g で 10 分間遠心分離し、上清を回収。
- 6) 5 mL のイソアミルアルコール：クロロホルム = 1:24 を加え、転倒混和。
- 7) 8,000 $\times$ g で 10 分間遠心分離し、上清を回収。
- 8) イソプロパノール沈殿を行い、脱塩後に超純水にて DNA ペレットを溶解。

3. トウモロコシサンプルの純度確認

トウモロコシサンプルについて、コントロール試料に MIR162 試料がコンタミネーションしていないか、リアルタイム PCR 法にて確認した。装置は ABI 7500 Fast (Thermo Fisher Scientific) を使用。

反応液の組成は以下の通り。TaqMan Fast Advance 12.5 μ L、各対象プライマー溶液 (50 μ mol/L) 各 0.25 μ L、対象プローブ溶液 (10 μ mol/L) 0.5 μ L を混合し、純水で全量 20 μ L に調製後、各 DNA 試料液 (10 ng/ μ L) 5 μ L を添加して全量 25 μ L とした。

PCR 反応条件は以下の通りである。95°C で 20 秒間加温し、ホットスタート法で反応を開始。95°C で 6 秒間、59°C で 30 秒間を 1 サイクルとして、45 サイクルの増幅反応を行った。

使用したプライマー・プローブ配列は以下の通り。

SSIb3-5': CCAATCCTTTGACATCTGCTCC

SSIb3-3': GATCAGCTTTGGGTCCGGA

SSIb-Taq: FAM-AGCAAAGTCAGAGCGCTG
CAATGCA-TAMRA

MIR162-f1: GCGCGGTGTCATCTATGTTAC
TAG

MIR162-r1: TGCCTTATCTGTTGCCTTCAGA

MIR162-p1: FAM-TCTAGACAATTCAGTACA
TTAAAAACGTCCGCCA-TAMRA

4. DNA マーカーを用いたトマトの系統解析

トマト葉かび病抵抗性遺伝子 *Cf-9* の DNA マーカーとして、Kuroyanagi らの文献 (Kuroyanagi *et al.* Res. Bull. Aichi Agric. Res. Ctr., 42:15-22, 2010) にある DNA マーカーを使用。トマト黄化葉巻病 (TYLCV) 抵抗性遺伝子 *Ty-3/Ty-3a* の DNA マーカーとして、Ji らの文献 (Y. Ji, *et al.* Tomato Genetics Cooperative, 57: 25-28, 2007) にある DNA マーカーを使用。PCR 反応液の組成は以下の通り。KOD FX Neo buffer 12.5 μ

L、2mM dNTPs 溶液 5 μ L、各対象プライマー溶液 (50 μ mol/L) 各 0.2 μ L、KOD FX Neo ポリメラーゼ 0.2 μ L、各トマトゲノム DNA 100 ng を混合し、純水で全量 25 μ L とした。

PCR 反応条件は以下の通りである。94°C で 2 分間加温し、ホットスタート法で反応を開始。98°C で 10 秒間、58°C で 30 秒間、68°C で 20 秒間を 1 サイクルとして、30 サイクルの増幅反応を行った。

使用したプライマーの配列は以下の通り。

CF9-8_Fwd: GCATCTCGATTTGTTCGCAT

CF9-8_Rev: GCCGTTCAAGTTGGGTGTTA

P6-25-F2: GGTAGTGGAAATGATGCTGCTC

P6-25-R5: GCTCTGCCTATTGTCCCATATAT
AACC

5. ゲノム編集 GABA トマトにおける標的部位編集の確認

ゲノム編集 GABA トマトの標的遺伝子である *GAD3* での編集を確認するべく、サンガーシーケンスを実施した。*GAD3* 遺伝子領域に設計した特異的プライマーで PCR 法にて対象領域を増幅し、その増幅産物に対するサンガーシーケンスを受託企業へ外注した。PCR 反応液の組成は以下の通り。PrimeSTAR GXL 10 μ L、2.5mM dNTPs 溶液 4 μ L、各対象プライマー溶液 (50 μ mol/L) 各 0.25 μ L、PrimerSTAR GXL ポリメラーゼ 1 μ L、各トマトゲノム DNA 100 ng を混合し、純水で全量 50 μ L とした。

PCR 反応条件は以下の通りである。98°C で 10 秒間、60°C で 15 秒間、68°C で 2 分間を 1 サイクルとして、30 サイクルの増幅反応を行った。

PCR 増幅に使用したプライマーは以下の通り。

SlGAD3_Fwd1: CCAAGGTTTGAGATGCTAG
AGAAGTC

SlGAD3_Rev1: CCTTCAAACAACCATTAAT
CCTTCCC

サンガーシーケンスに使用したプライマーは以下の通り。

SlGAD3_seq_Fwd: GAGACCCCTCCGTAGGTT
TGG

6. ショートリードシーケンサーによる WGS

MagAttract HMW DNA kit (QIAGEN) を用いて、抽出したゲノム DNA を精製および低分子 DNA を除去。精製した DNA は 200 ng ずつを用いて Illumina DNA Prep, (M) kit (Illumina) により、シーケンスライブラリを調製。Qubit Fluorometer (Thermo Fisher Scientific) および Bioanalyzer 2100 (Agilent) により定量・定性を行い、それぞれ 15 nM に希釈の後、等量ずつ混合して DNA ライブラリとした。シーケンスは HiSeq X Ten (Illumina) による 150 bp×2 のペアエンドシーケンスを受託企業へ外注した。

得られたリードデータのクオリティは「FastQC (ver. 0.11.9)」で確認し、リードデータに含まれるシーケンスアダプター配列および低クオリティリードデータは「Trim Galore! (ver. 0.6.7)」を用いて除去した。

7. ロングリードシーケンサーによる WGS

Short read eliminator XS (Circulomics) を用いて、抽出したゲノム DNA より低分子 DNA を除去。精製した DNA 1 µg を用いて Ligation Sequencing Kit (Oxford Nanopore Technologies; SQK-LSK109 [ダイズ・トウモロコシゲノム試料], SQK-LSK110 [トマトゲノム試料]) によりシーケンスライブラリを調製。Qubit Fluorometer により二本鎖 DNA 濃度を測定し、調製したライブラリを PromethION フローセル (R9.4.1) に供し、PromethION (Oxford Nanopore Technologies) により 72 時間のシーケンスを実施した。

ナノポアシーケンサーにおいて、解析ソフト「MinKNOW (20.06.18)」にて取得したポア通過時の電流変化データ (fast5 format) を、ダイズお

よびトウモロコシではベースコーラ「Guppy (ver. 5.0.11)」の high accuracy mode によりクオリティスコア 9 以上のリードを塩基配列 (fastq format) へと変換した。トマトでは同じく「Guppy (ver. 4.0.11)」の high accuracy mode によりクオリティスコア 7 以上のリードを塩基配列 (fastq format) へと変換した。一般的にナノポアシーケンサーのデータは読み初めの数十 bp の正確性が低いと言われているため、リードデータのトリミングツール「NanoFilt (ver. 2.8.0)」の --headcrop オプションを用いて、リードデータから 5' 側の 50 bp を除去した。トリミングされたシーケンスデータは「NanoPlot (1.38.0)」を用いてクオリティを確認した。なおトウモロコシおよび野生型トマトのリードデータは 4 回分のデータを合算した。

8. WGS データを用いた外来遺伝子の再構成

遺伝子組換えダイズ (RRS2) および遺伝子組換えトウモロコシ (MIR162) に挿入された外来遺伝子の配列情報は、データベース Nexplorer (<https://bioit-webapp-prod.sciensano.be/nexplorer/>) より取得した。ゲノムに残存した外来遺伝子をアセンブリ解析によって検出する際にその配列長がどう影響するか検討するため、RRS2 (図 1 参照) では挿入全長 (4314 bp) および各エレメントである① *L_TSF1_ARATH* (46 bp)、② *TS_CTP2_ARATH* (228 bp)、③ *CS_EPSPS_RHIRD* (1368 bp)、④ *TS_CTP2_ARATH* + *CS_EPSPS_RHIRD* + *T_rbcS-E9_PEA* (2246 bp)、MIR162 (図 2 参照) では挿入全長 (8302 bp) および各エレメントである① *T35S* (70 bp)、② *TNOS* (256 bp)、③ *pmi* (1176 bp)、④ *vip3A* (2370 bp) のそれぞれ 5 種類の長さの参照配列を用意した。各参照配列に対して、ショートリードシーケンスデータでは「BWA (ver. 0.7.17-r1188)」の MEM アルゴリズムを、ロングリードシーケンスデータでは「Minimap2 (ver. 2.21-r1071)」を用いてマッピングを行った。その後「Samtools (ver. 1.1

3) view コマンドの -F オプションを 4 に指定してアンマップデータを除くと同時に BAM 変換を行い、sort コマンドでソート、bam2fq コマンドで BAM ファイルから FASTQ ファイルへ変換することで参照配列にマッピングしたリードデータの抽出を行った。ショートリードシーケンスデータのアセンブリは、「Velvet (ver. 1.2.10)」または「SPAdes (ver. 3.15.5)」にて行い、両者の性能を比較した。なお Velvet でのアセンブリでは、まず「KmerGenie (ver. 1.7051)」を用いてアセンブリに最適な k 値を推定した。なお、KmerGenie で推定した最適 k 値でコンティグが 1 本にまとまらない場合は、目視にて 2 番目に最適と思われる k 値を使って再度アセンブリ解析を実施している。一方 SPAdes ではアセンブリに適した k 値はデフォルトで Auto となっているため、特に入力していない。

ロングリードシーケンスデータのアセンブリには「Flye (ver2.9)」を用いた。Flye ではナノポアシーケンサーのエラー率 5% 以下を想定した --nano-hq オプション、およびポリッシング回数を指定するオプション -i 5 にて実施した。

9. シーケンスリードのランダムサンプリング

シーケンスリードデータのランダムサンプリングには「Seqkit (ver. 2.0.0)」の Sample コマンドを使用した。なお、ゲノムに挿入された外来遺伝子の全長領域またはその一部を含むリードがランダムサンプリング後にどの程度残るかによって結果にブレが生じることが想定される。そのため、ランダムサンプリングをする際にはアルゴリズムを変化させる Seqkit sample コマンドの -s オプションを使用し、12 種類 (s=1~12 を指定) のランダムサンプリングデータを用意した。12 種類のリードデータを各長さの参照配列にマッピングさせた後、ショートリードシーケンスデータについてはマップしたリード数のばらつきを調べる

ため、Smirnov-Grubbs 検定による外れ値検定を行った。ロングリードシーケンスデータについては各リードの長さが異なるため、外れ値検定は行わなかった。

10. ゲノム解析情報の視覚化

ゲノムのマッピング状況等を確認する場合は、Broad Institute が開発した Integrative Genomics Viewer (IGV; <https://software.broadinstitute.org/software/igv/>) を用いて、目視により行った。

C. 研究結果

1. 遺伝子組換えダイズを用いた外来遺伝子の再構成

WGS データの概要

RRS2 の WGS データは前年度の検討で使用したものを再利用した。表 1 に取得したシーケンスデータの統計値を示す。なお推定シーケンスカバレッジはダイズゲノムを 1.1 Gb として算出した。

挿入された外来遺伝子へのマッピングとそのアセンブリ

図 1 に RRS2 に挿入された外来遺伝子の概略を、表 2、図 3 に挿入配列の全長および各エレメントに対してマッピングした結果を示す。得られたマッピングリードについて、ショートリードシーケンスデータでは Velvet、ロングリードシーケンスデータでは Flye でアセンブリを行った。

ショートリードシーケンスデータを用いた場合、全長および各エレメントいずれも 1 本のコンティグ (アセンブリにより生成された配列) で元の配列を完全に再現した。

ロングリードシーケンスデータを用いた場合、エレメント① (46 bp) の挿入配列に対してはマッピングされたリードを得ることができなかったため、アセンブリを行うことができなかった。全長およびその他の各エレメントはいずれも 1 本のコ

ンティグで元の配列をほぼ再現したが、昨年度示した通り、配列内にあるチミン塩基の 8 塩基ポリマー配列を正確に再現することができなかった。

ショートリードシークエンスデータのランダムサンプリング

ショートリードシークエンサーによる WGS データを用いて残存した外来遺伝子を検出する際に、どの程度のシークエンスカバレッジが必要となるか検討するため、Seqkit の sample コマンドを使用してリードデータを 50×、40×、30×、25×、20×、15×、10×、5×にダウンサイズした。ダウンサイズの方法としては、まず Seqkit の stat コマンドでシークエンスカバレッジを算出する。今回の場合シークエンスにより得られた総塩基数 (66,929,724,803 b) をダイズのゲノムサイズ (1.1 Gb) で割ると、シークエンスカバレッジは約 60.8×であった (表 2)。10×のリードデータを生成する場合は、10 (目標のシークエンスカバレッジ) / 60.8 (実際のシークエンスカバレッジ) = 0.1644...であるため、Seqkit sample コマンドの -p オプションの値を 0.16 とした。各シークエンスカバレッジのリードデータを 12 試行分生成した後、それらを全長および各エレメントの配列にマッピングした時のマッピングリード数および標準偏差を図 4 に示す。なお 12 試行分マッピングリード数を使って Smirnov-Grubbs 検定を実施したところ、エレメント② (228 bp) における 10×およびエレメント④ (2246 bp) における 20×でそれぞれ外れ値が見つかったため、外れ値データを除き、以後の解析を行った。

ショートリードシークエンスデータを用いた各シークエンスカバレッジでの標的配列のアセンブリ (Velvet)

各シークエンスカバレッジにダウンサイズし、全長および各エレメントに対してマッピングすることで抽出したリードデータを、Velvet によって

アセンブリする。12 試行分の結果を表 3 に示す。アセンブリ解析の成否については、(1) 元の挿入配列全体を複数のコンティグでもれなく再現できているか、(2) 元の挿入配列を一本のコンティグで 100%正確に再現できたか、2つの評価基準で判断している。上記アセンブリを 12 試行で実施したが、12 試行とも成功 (外れ値検定によりデータを除いた部分は 11 試行中 11 試行成功) した部分は再現性の高い解析と言える。その結果、全体的な傾向としては挿入配列が短いほど低カバレッジでも配列を再現しやすい傾向にあり、特に 46 bp (エレメント①) および 228 bp (エレメント②) に関しては 15×のシークエンスカバレッジでも正確に挿入配列を再現できることが示唆された。一方で 2 kb を超える挿入配列を、1 本で正確に再現する成功率はやや低く、50×のシークエンスカバレッジデータであっても 12 試行中 1、2 回はアセンブリに失敗した。

ショートリードシークエンスデータを用いた各シークエンスカバレッジでの標的配列のアセンブリ (SPAdes)

Velvet によるアセンブリでは k 値を入力する必要があるが、予め KmerGenie により最適な k 値を推定していたが、これは実効性の面で問題点となり得る。そこで k 値を入力する必要がある別のアセンブラーとして SPAdes にて同様の検討を行った。結果は表 4 に示す。全体的な傾向としては Velvet とは対照的に挿入配列が短いほど高いシークエンスカバレッジが必要となる一方で、2 kb を超える挿入配列に対しては 15~25×と比較的低いシークエンスカバレッジで正確に再現できることが示唆された。

ロングリードシークエンスデータを用いた各シークエンスカバレッジでの標的配列のアセンブリ

ショートリードシークエンス同様に、各シークエンスカバレッジにおけるロングリードシークエ

ンスデータを用いた外来遺伝子挿入配列のアセンブリにおいて、再現の正確性とシークエンスカバレッジの要件を検討した。ロングリードシークエンスデータ (47.3×) からシークエンスカバレッジ 40×、30×としたランダムサンプリングを 12 回試行で行い、それらを全長および各エレメントの配列にマッピングした時のマッピングリード数やリード長を表 5 に示す。全般的に RSD% で 10% 程度となっており、併行精度として良好であると判断した。これらのデータを Flye に供してアセンブリ解析を行なった (表 5)。なお Flye によるアセンブリでは成功すればコンティグが 1 本以上生成されるが、失敗した場合は途中のエラーで止まりコンティグが生成されない。また、ロングリードシークエンスの特徴としてホモポリマー配列を正確に読めない特徴を有するため、ベースコーラの正確性が 95% 以上であることを鑑みて、アセンブリ解析の成否は「元の挿入配列を含むコンティグで 95% 以上正確に再現できたか」で判断している。エレメント① (46 bp) の短い配列は最大カバレッジ (47×) の時同様、ダウンサンプリングした場合はコンティグが得られなかった (0/12 回)。エレメント②～④および全長では、シークエンスカバレッジ 40×で 10～11 回、シークエンスカバレッジ 30×カバレッジでは 9～11 回でコンティグが得られた。ホモポリマー配列で一塩基欠失が見られ、完全に標的配列を再現することは本法では難しかったが、228 bp 以上の配列であれば高確率で挿入された外来遺伝子の配列を再構成できることが示唆された。

2. 遺伝子組換えトウモロコシを用いた外来遺伝子の再構成

ゲノム DNA 抽出

ゲノム抽出・精製の結果、High Molecular Weight (HMW) で十分な量のゲノム DNA を得られた (図 5)。抽出したコントロールトウモロコシゲノム DNA に遺伝子組換えトウモロコシ

MIR162 ゲノムがコンタミネーションしていないか調べるためにリアルタイム PCR を実施した。その結果、トウモロコシ内在性遺伝子 SSIIb 検知法ではコントロールおよび MIR162 ゲノムサンプルにおいて SSIIb 遺伝子の増幅が確認されたが、MIR162 系統特異的検知法ではコントロールサンプルでは MIR162 系統特異的配列の増幅が見られず、MIR162 サンプルのみで増幅が確認された (図 6)。よって、コントロールサンプルに MIR162 ゲノム DNA のコンタミネーションはないと考えられる。以後の評価ではこれらゲノム DNA を用いた。

WGS データの概要

WGS に供したライブラリの QC 結果を図 7 に示す。150 -1500 bp 間の平均塩基長はコントロールサンプルライブラリが 596 bp、MIR162 ライブラリが 580 bp であり、どちらも良好であった。これらサンプルの WGS により得られたリードデータについては表 6 に示した。得られたリードデータはクオリティ、カバレッジともに良好であった。なお推定シークエンスカバレッジはトウモロコシゲノムを 2.4 Gb として算出した。

挿入された外来遺伝子へのマッピングとそのアセンブリ

図 2 に MIR162 に挿入された外来遺伝子の概略を、表 7、図 8 に挿入配列の全長および各エレメントに対してマッピングした結果を示す。得られたマッピングリードについて、ショートリードシークエンスデータでは Velvet、ロングリードシークエンスデータでは Flye でアセンブリした。

ショートリードシークエンサーおよびロングリードシークエンサーいずれの WGS データも MIR162 に挿入された全長の外来遺伝子を正確に再現することはできなかった。ショートリードシークエンスデータから得られたマッピングリードを詳細に解析したところ、(1) 標的外来遺伝子内

に同一の配列として Maize *ubiquitin promoter* が 2 コピー存在するため、この部位に対するマッピングリードが重複していた (Mapping Quality: MAPQ 値が低下する)、および (2) 標的外来遺伝子内にトウモロコシ内在性配列 (Maize *ubiquitin promoter* および *PEPC intron*) が含まれるため、外来遺伝子由来でない、トウモロコシゲノム内在性のマッピングリードもこの標的外来遺伝子に重複してマッピングされていた。従って、これらの要因により、本検討でのアセンブリによる方法では MIR162 に挿入された全長の外来遺伝子を WGS データから正確に再現することはできなかったものと考えられる。ロングリードシーケンスデータの場合でも、Flye によって得られたコンティグは 3 本であったが、それぞれ上述したトウモロコシ内在性配列が混在していた。さらに MIR162 およびコントロールサンプルから得られたリードで全長の外来遺伝子に対してマッピングしたリード数はそれぞれ 274 リード、94 リードと、コントロールサンプル由来のリードにおいても一定のリードが外来遺伝子配列にマッピングされたことから、全長の外来遺伝子に特異的な再現が困難であったことが示唆される。以上より、標的配列にマルチコピーな配列や対象生物に内在性の配列が含まれる場合は、本アセンブリ解析で標的配列を再現することは困難と考えられる。

各エレメント配列については、ショートリードシーケンスデータではエレメント③ (*vip3A* 配列、2,370 bp) を一本のコンティグにアセンブリすることはできなかったが、それ以外はショートリードシーケンスデータ、ロングリードシーケンスデータともに各エレメント配列を正しく再現することができた。

ショートリードシーケンスデータのランダムダウンサンプリング

シーケンスカバレッジの必要要件を検討するため、MIR162 のショートリードシーケンサー

による WGS データから各シーケンスカバレッジのリードデータを 12 試行分生成した後、それらを全長および各エレメントの配列にマッピングした時のマッピングリード数および標準偏差を図 9 に示す。なお 12 試行分のマッピングリード数を使って Smirnov-Grubbs 検定を実施したところ、挿入全長 (8302 bp) における 5×、エレメント① (70 bp) における 50×および 10×、エレメント② (253 bp) における 30×、エレメント④ (2370 bp) における 25×および 10×で外れ値が見つかったため、外れ値のデータは除き、以後の解析を行った。

各シーケンスカバレッジにおけるアセンブリ結果

各シーケンスカバレッジにダウンサイズし、全長および各エレメントに対してマッピングすることで抽出したリードデータを、Velvet または SPAdes によってアセンブリする。12 試行分の結果を表 8 (Velvet)、表 9 (SPAdes) に示す。傾向としては RRS2 の場合と同じように、Velvet はエレメント① (70 bp) のように短い配列に対しては 20×以上のシーケンスカバレッジであれば正確に再現できた。一方で SPAdes を用いる場合、エレメント①を正確に再現しようとする、50×のシーケンスカバレッジが必要であったが、エレメント②～④のアセンブリでは、標的となる挿入配列全体を少なくとも複数のコンティグでもれなく再現できているかという判断基準で考えると、25×以上のシーケンスカバレッジが必要であった (表 9)。なお、SPAdes を用いたエレメント③ (*pmi*) 配列のアセンブリにおいて、コンティグ数が一本にならなかったが、詳細に調べたところ、ショートリードシーケンサーによる WGS データにおいて *pmi* 領域にマッピングされるリードの一部に変異が多く蓄積していた (図 10)。ロングリードシーケンサーによる WGS データにおいて当該領域にマッピングされたリードには同じ変

異は見られないことから、この変異は MIR162 ゲノムに内在的ではないことが示唆される。今回のショートリードシーケンサーによる WGS データに用いた Illumina ライブラリは PCR 増幅を行うキットを使用して調製したが、何らかの理由で PCR エラーが生じてしまった可能性がある。よりアンバイアスなシーケンスが必要な場合、PCR フリーのライブラリ調製キットの使用を検討する必要がある。

ロングリードシーケンサーによる WGS データ (40.7×) から 12 回のダウンサンプリングを行い、それらを全長および各エレメントの配列にマッピングした時のマッピングリード数やリード長を表 10 に示す。全般的に RSD% で 10~20% 程度となっており、ややサンプリング間のバラツキが大きい結果となっていた。これらのデータを用いて、Flye でアセンブリを実施した。標的配列の再現をシーケンスカバレッジの観点から評価し、結果を表 10 に示した。RRS2 での結果 (表 5) と同様に、標的配列が短いエレメント① (70 bp) では、コンティグが得られないことが多かった (0/12) が、エレメント②~④では、シーケンスカバレッジ 30× で再現率は高く、エレメント③、④では 12 データ全てで再現した。いずれのエレメント配列も 20× データでは再現率は下がったことから、ロングリードシーケンサーによる WGS データでは現状 30× 以上のデータ量が必要と考えられる。

3. ゲノム編集トマト

ゲノム編集食品のモデルサンプルとして GABA トマトに外来遺伝子が残存していないか、上記で検討したアセンブリ解析を用いて検証した。

ゲノム DNA 抽出

DNA 抽出・精製の結果、トマト青果からでも HMW なゲノム DNA を十分量得られた (図 11)。なお、シシリアンルージュには CF 系統と TY 系統の 2 種類が販売されている。厚生労働省ホーム

ページにて公開されている GABA トマトの届出資料によると、ゲノム編集トマトのバックグラウンド品種はシシリアンルージュ CF 系統である。まず購入した野生型シシリアンルージュの系統を調べた。文献情報に従い、*Cf-9* (Kuroyanagi *et al.* Res. Bull. Aichi Agric. Res. Ctr., 42:15-22, 2010) および *Ty-3/Ty-3a* (Y. Ji, *et al.* Tomato Genetics Cooperative, 57: 25-28, 2007) の DNA マーカーを用いて系統の解析を行った結果、野生型シシリアンルージュおよび GABA トマトはいずれも *Cf-9* 陽性/*Ty-3a* 陰性であった (図 12)。よって購入したシシリアンルージュは CF 系統であると確認された。

サンガーシーケンスによる GABA トマトのゲノム編集部位の確認

抽出したゲノム DNA を用いて野生型トマトおよび GABA トマトの *GAD3* 標的部位の配列の比較を行った。届出資料 (<https://www.mhlw.go.jp/content/11120000/000828873.pdf>) によると、GABA トマトは *GAD3* の配列にチミンが一塩基挿入されている。実際にサンガーシーケンスで *GAD3* 配列を解析したところ、GABA トマトの *GAD3* の 7 番エクソン配列においてチミンが一塩基挿入しており、この部位を境に波形が二重になることが確認された (図 13)。トマトは 2n=24 の染色体を有する二倍体の種である。従ってここが Cas9 による double strand break 部位であり、これ以降はチミンが一塩基挿入した配列と野生型の配列が重なっていると推察された。本研究に使用した GABA トマトは市場に出回っている F1 品種であるため、ゲノム編集によってチミンの一塩基挿入が生じているアレルと野生型のアレルがヘテロ接合体で存在する結果、このような二重の波形が観察されたと考えられる。

WGS データの概要

WGS に供したライブラリの QC 結果を図 14 に示す。150 - 1500 bp 間の平均塩基長はコントロールトマトライブラリが 621 bp で GABA トマトライブラリが 600 bp でどちらも良好であった。これらトマトサンプルについて、ショートリードシーケンサーおよびロングリードシーケンサーによる WGS を実施した。得られたリードデータは表 11 に示したが、そのクオリティ、カバレッジともに良好であった。なお推定シーケンスカバレッジはトマトゲノムサイズを 950 Mb として算出した。

WGS によるゲノム編集部位の解析

ショートリードシーケンサーおよびロングリードシーケンサーによる WGS のリードデータをトマトリファレンスゲノム SL4.0 (https://solgenomics.net/organism/Solanum_lycopersicum/genome/) へマッピングし、*GAD3* のゲノム編集標的部位を確認した。その結果、どちらの WGS データにおいてもチミン挿入が示唆されるリードとそうでないリードが混在していることが確認された (図 15)。また IGV 上でショートリードシーケンスリードではチミンの一塩基挿入を「Insertion」と正しく変異判定する一方で、ロングリードシーケンスリードでは「Insertion」と変異判定されず、チミン挿入塩基の一塩基前のグアニンがチミンに「Substitution」し、さらにその一塩基前のアデニン塩基がグアニンに「Substitution」したといった変異判定がされていた。これはチミン挿入塩基の 2 塩基前のアデニンの 6 塩基ポリマー配列の 6 塩基目を読み飛ばし、本来は一塩基欠失と表示されるはずが、チミンの一塩基挿入により塩基数が揃ったことによる誤表示であると考えられる (図 16)。いずれにせよ、どちらの WGS 結果ともに明らかに GABA トマトゲノムが想定通りにゲノム編集されたことが確認できたため、これらのリードデータを用いて外来遺伝子の残存性確認を実施することとした。

残存が想定される配列へのマッピング結果

届出資料にはゲノム編集時に使用したベクター情報の記載はないため、開発者らの以前の論文 (Nonaka, S. *et al.* Sci. Rep. 7, 7057, 2017) を参考に、*SpCas9* 配列 (4104 bp) およびネオマイシン耐性遺伝子 *NptII* 配列 (795 bp) の 2 種類の配列に対してショートリードシーケンサーおよびロングリードシーケンサーによるそれぞれの WGS リードデータのマッピングを実施した。その結果、いずれもマッピングリード数はゼロであった。

D. 考察

本研究ではゲノム編集食品の外来遺伝子の有無を調べる標準的な次世代シーケンシング解析手法の確立を目的としており、これまでにアセンブリ解析が有用であることを見出している。そこで今年度はモデル試料として遺伝子組換えダイズ RRS2 ばかりでなく遺伝子組換えトウモロコシ MIR162 も用いたショートリードシーケンサーおよびロングリードシーケンサーによる WGS を実施し、挿入配列の全長および各エレメント配列に対するアセンブリ解析、さらにはランダムにダウンサンプリングした WGS データを用いたアセンブリ解析を行うことで、シーケンスカバレッジや標的配列長に関する必要要件を検討した。その結果、ショートリードシーケンサーによる WGS では、25×以上のシーケンスカバレッジデータがあれば、Velvet と SPAdes を併用することで 40 - 4000 bp 程度の長さの標的配列を 100% の確率で再現できた (但し標的配列にマルチコピーな配列や対象生物に内在性の配列が含まれない場合に限る) (表 3, 4, 8, 9)。これは RRS2 および MIR162 どちらの場合でも同様の傾向であったことから、ある程度一般性のある傾向と考えられる。一方、ロングリードシーケンサーによる WGS では、RRS2 なら 40×、MIR162 なら 30×

のシーケンスカバレッジのデータがあれば、200 - 4000 bp 程度の長さの標的配列を高確率で再現できた（表5,10）。これらの各 WGS データによる結果は、対象食品や残存外来遺伝子の種類などによらずある程度同様な傾向であったことから、本アセンブリ解析がゲノム編集食品の外来遺伝子の有無を調べる手法として妥当であることを示唆しており、本手法の必要要件や測定機器の特徴を考える重要な結果である。しかしながら以下に挙げるようにいくつかの課題が残っている。

まず一つ目の課題は、今回のモデルサンプルで挿入されている配列は除草剤耐性遺伝子や害虫殺虫活性タンパク質をコードする遺伝子などであり、ゲノム編集食品の作製において残存する可能性のある遺伝子ではないという点である。残存する可能性のある配列としては Cas9 や抗生剤耐性遺伝子などが挙げられ、実際にゲノム編集豚では抗生剤耐性遺伝子の残存が報告されている（Norris, A.L. et. al. Nat. Biotechnol. 38, 163, 2020）。従って今回見出されたシーケンスカバレッジに関する必要要件が Cas9 や抗生剤耐性遺伝子など実際に残存する遺伝子の検出にも当てはまるか、検討する必要がある。今回ゲノム編集サンプルとして GABA トマトを入手し、本アセンブリ解析を実施するべく、ショートリードシーケンサーおよびロングリードシーケンサーによる WGS を実施した。各々79.2×および 55.1×というシーケンスカバレッジをもつリードデータを取得したものの、Cas9 配列や NptII 配列にマッピングするリードは見つからなかった。これはゲノム編集された GABA トマトには、Cas9 配列や NptII 配列が残存していないことを強く示唆する結果である一方、Cas9 配列や NptII 配列をアセンブリ解析で検出するには上述したシーケンスカバレッジでは不足しているという可能性も否定できない。今後何らかの食用植物のゲノムに Cas9 や抗生剤耐性遺伝子などの配列を挿入させたシミュ

レーションサンプルを作成し、シーケンスカバレッジの必要要件の検討を実施することで、Cas9 や抗生剤耐性遺伝子の検出に必要なシーケンスカバレッジの要件を定めると共に、ゲノム編集された GABA トマトに外来遺伝子が残存していないことを本アセンブリ解析でも確認したい。

二つ目の課題として、今回の実施したシーケンスカバレッジに関する必要要件の検討は、ランダムなダウンサンプリングデータによるシミュレーション実験である。従って例えばショートリードシーケンサーによる WGS データについて得られた「残存配列の検出には 25×以上のシーケンスカバレッジデータが必要」という結果の妥当性を、実サンプルすなわち 25×程度のシーケンスカバレッジをもつ WGS データにて検証する必要がある。今後 RRS2 または MIR162 といったモデルサンプルにてそのような WGS データを取得し、本課題を検討する。

最後の課題として、残存する可能性のある配列の一部のみがゲノムに残存している場合でも、その残存部位に WGS で得られたリードはきちんとマッピングできるのか確認する必要がある。今回の検討ではアセンブリ解析における残存配列の長さの影響についてもいくつか検討したが、これはモデルとした遺伝子組換え食品の挿入配列の全長またはその一部である各エレメントを参照配列として、この配列にマッピングされるリードを集め、アセンブリ解析に供していた。すなわちどの長さの参照配列であれ、残存していると考えられる配列の全体にわたってリードデータが存在するという状況であった。しかしながら実際のゲノム編集サンプルで外来遺伝子の残存を検証する場合、何らかの外来遺伝子がゲノム中に残存していて、かつ残存する可能性のある配列が分かるケースでも、実際に残存している配列はその可能性のある配列の一部であるケースがほとんどである。特に残存する可能性のある配列に対して、実際に残存して

いる配列が極めて短いケースでは、正しくマッピングリードを集められるか不明であるが、今年度実施したここまでの検討ではこうした点を十分に検討できていない。ただ、今回のショートリードシーケンサーによる WGS データにおいて、各エレメントにマッピングしたリードを挿入配列全長に対してマッピングした結果が図 3, 8 にあるが、46 bp や 70 bp といった比較的短い配列に対してもマッピングリードを得ることができている。ショートリードシーケンサーのリード長は約 150 bp であるが、そのリード長より短い 100 bp 以下の短い配列がゲノム中に仮に残存した場合でも、この配列にマッピングするリードを正しく抽出できると推察される。従って、残存する可能性のある配列の一部として 20 - 100 bp 程度が残存するようなシミュレーションサンプルを作成し、本アセンブリ解析に供することで、この課題を検討すると共に、残存配列長に関する検出限界についても明らかにしたい。WGS データを用いた外来遺伝子の残存を解析する別の方法として、k-mer 法を利用したものも報告されている (Itoh et al. Sci. Rep. 10, 4914, 2020) が、この方法では 20 塩基以上の残存を検出できるとされている。我々のアセンブリ解析と k-mer 法を利用したものとの比較検討についても今後議論したい。

E. 結論

本年度の検討より、アセンブリ解析による WGS データを用いた外来遺伝子残存に関する解析について、本手法の妥当性が示唆されたと共に、解析可能な残存遺伝子配列の長さや解析に必要なとされるシーケンスカバレッジを明らかにすることができた。今後こうした必要要件の妥当性をさらに検証すると共に、この解析手順のコントロールになるような標準サンプルの作成などに取り組む必要がある。

F. 研究発表・業績

1. 論文発表

- 1) Takabatake R, Egi T, Soga K, Narushima J, Yoshiba S, Shibata N, Nakamura K, Kondo K, Kishine M, Mano J, Kitta K: Development and interlaboratory validation of a novel reproducible qualitative method for GM soybeans using comparative Cq-based analysis for the revised non-GMO labeling system in Japan. *Anal. Chem.* 2022; 94: 13447-13454
- 2) Soga K, Nakamura K, Egi T, Narushima J, Yoshiba S, Kishine M, Mano J, Kitta K, Takabatake R, Shibata N, Kondo K: Development and validation of a new robust detection method for low-content DNA using DDCq-based real-time PCR with optimized standard plasmids as a control sample. *Anal. Chem.* 2022; 94:14475-14483.
- 3) Shibata N, Soga K, Sugino M, Nakamura K, Narushima J, Yoshiba S, Egi T, Takabatake R, Kondo K: Evaluation of conversion factor for rapid quantification of authorized genetically modified maize and soybean in Japan. *BPB Reports.* 2022; 5:115-120.
- 4) Narushima J, Kimata S, Shiwa Y, Gondo T, Akimoto S, Soga K, Yoshiba S, Nakamura K, Shibata N, Kondo K: Unbiased prediction of off-target sites in genome-edited rice using SITE-Seq analysis on a web-based platform. *Genes to Cells.* 2022; 27: 706-718.

2. 学会発表

- 1) 曾我慶介、江木智宏、成島純平、吉場聡子、真野潤一、高畠令王奈、橘田和美、中村光亮、柴田識人、近藤一成：新遺伝子組み換え表示制度に向けた試験法開発 ～「遺伝子組み換えでない」表示の今後～、第 8 回次世代を担う

- 若手のためのレギュラトリサイエンスフォーラム、東京、2022年8月26日
- 2) 近藤一成、曾我慶介、成島純平、吉場聡子、柴田識人、田口千恵、坂田こずえ、加藤怜子：ゲノム編集によって発生する意図しない変異はどこから来るのか、NGS EXPO 2022、大阪、2022年10月18日
 - 3) 成島純平、木俣真弥、志波優、榎藤崇裕、秋元智、曾我慶介、吉場聡子、中村公亮、柴田識人、近藤一成：ゲノム編集におけるオフターゲット予測法 SITE-Seq の Galaxy ベース新規解析パイプラインの開発とその作物への応用、NGS EXPO 2022、大阪、2022年10月18日
 - 4) 曾我慶介、吉田光範、成島純平、吉場聡子、柴田識人、近藤一成：ナノポアリードで構築したスギヒラタケアセンブリ結果からみたアセンブラーツールの選択について、NGS EXPO 2022、大阪、2022年10月19日
 - 5) 吉場聡子、成島純平、曾我慶介、杉野御祐、柴田識人、近藤一成：遺伝子組換え食品の同定に資する新たなゲノム解析技術の検討、第59回全国衛生化学技術協議会年会、東京、2022年10月31日-11月1日
 - 6) 柴田識人、成島純平、曾我慶介、吉場聡子、近藤一成：ゲノム編集食品における外来性遺伝子の残存を評価する全ゲノムシーケンスデータ解析の標準化に向けた取り組み、第59回全国衛生化学技術協議会年会、東京、2022年10月31日-11月1日
 - 7) 成島純平、杉野御祐、曾我慶介、吉場聡子、柴田識人、近藤一成：NGSを用いた網羅的なオフターゲット変異候補部位予測法の高GABA産生ゲノム編集トマトへの適用と検証、第59回全国衛生化学技術協議会年会、東京、2022年10月31日-11月1日
 - 8) 柴田識人、成島純平、曾我慶介、吉場聡子、近藤一成：ゲノム編集食品における外来性遺伝子の安全性評価における全ゲノムシーケンスデータを用いた解析の標準化に向けた課題、第118回日本食品衛生学会学術講演会、長崎、2022年11月10日-11日
 - 9) Soga K, Yoshida S, Narushima J, Shibata N, Kondo K: Genome analysis of deadly poisonous mushroom *Amanita virosa* using nanopore sequencing technology, Cell Bio 2022-An ASCB/EMBO Meeting, Washington DC, USA, 2022/12/5
 - 10) Kondo K, Fukuda N, Soga K, Yoshida S, Narushima J, Shibata N, Taguchi C, Sakata K, Kato R: G2/M synchronization with CDK1 inhibitor suppresses genome-wide mutations and genome rearrangements during genome editing, Cell Bio 2022-An ASCB/EMBO Meeting, Washington DC, USA, 2022/12/5-6
 - 11) 高島令王奈、大西真理、峯岸恭孝、布藤聡、曾我慶介、柴田識人、中村公亮、近藤一成、真野潤一、橘田和美：アグロバクテリウム法によって作出された遺伝子組換え植物中に存在する超短配列 25 bp の検出法の開発、第64回日本植物生理学会年会、仙台、2023年3月15日-17日
 - 12) 曾我慶介、成島純平、吉場聡子、柴田識人、近藤一成：猛毒キノコドクツルタケのドラフトゲノムを用いた新規毒性ペプチドの探索、日本薬学会第143年会、札幌、2023年3月27日

G. 知的財産権の出願・登録状況

該当なし

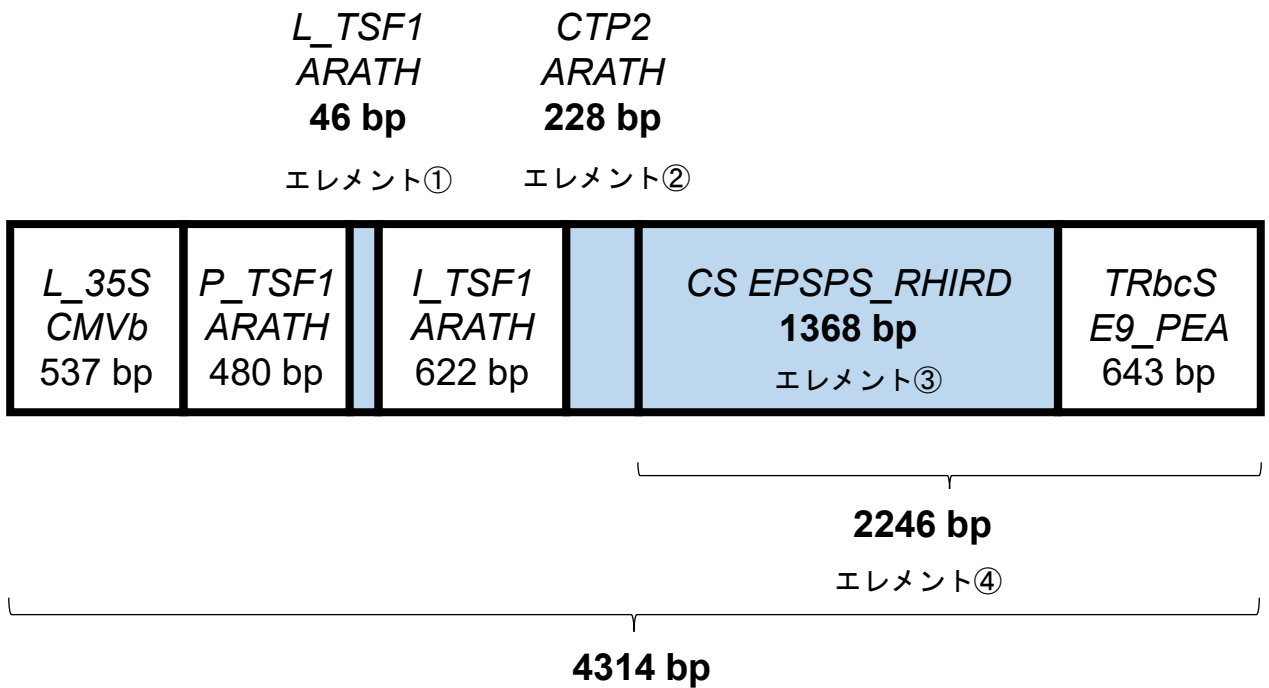


図1. 遺伝子組換えダイズRRS2に挿入されている外来遺伝子配列の概要

遺伝子組換えダイズRRS2系統は1番染色体に約4314 bpの外来遺伝子が挿入されている。
 本研究ではこの全長および各エLEMENT①～④について検討を行った。

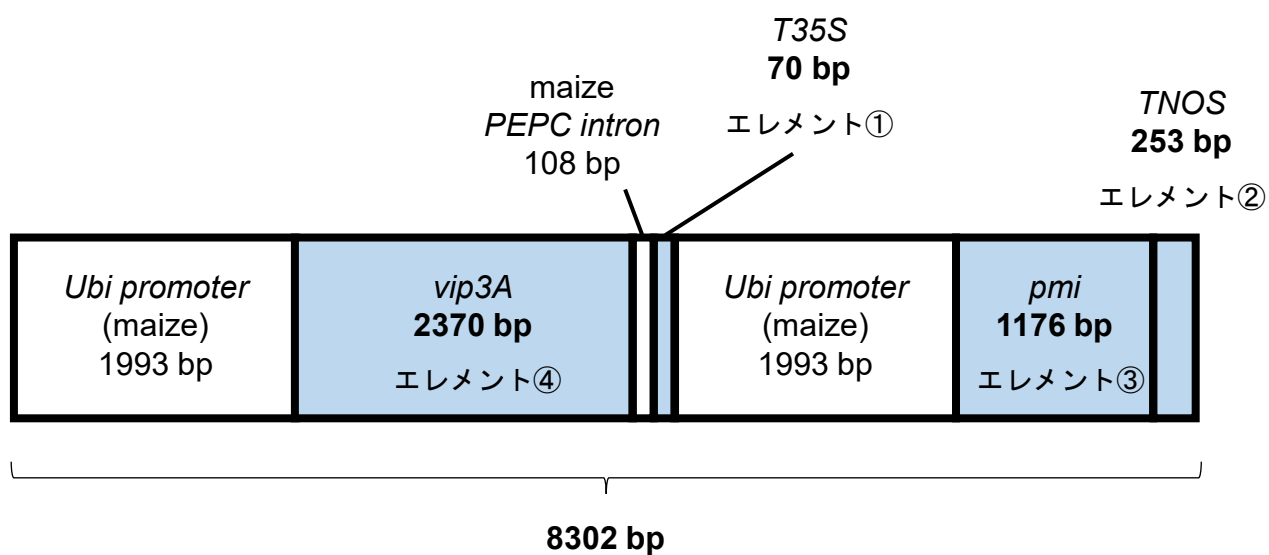


図2. 遺伝子組換えトウモロコシMIR162に挿入されている外来遺伝子配列の概要

遺伝子組換えトウモロコシMIR162系統は5番染色体に約8302 bpの外来遺伝子が挿入されている。本研究ではこの全長および各エレメント①～④について検討を行った。



図3. RRS2に挿入された外来遺伝子にマッピングしたショートリードシーケンスデータのIGV画像

各エレメント配列または全長にマッピングしたリードを再度全長配列にマッピングした。

	Target (bp)	50×	40×	30×	25×	20×	15×	10×	5×
Mapped read count (Average ± SD, n=12)	46	35.8 ± 4.4	28.6 ± 4.3	19.8 ± 5.1	16.5 ± 4.3	12.8 ± 3.9	9.3 ± 3.1	6.3 ± 2.9	2.2 ± 2.1
	228	89.6 ± 2.4	71.8 ± 8.1	52.3 ± 8.2	42.3 ± 6.7	33.9 ± 7.2	25.4 ± 5.0	14.1 ± 3.3	7.4 ± 2.2
	1,368	390.0 ± 11.1	316.2 ± 15.7	235.3 ± 17.8	203.0 ± 13.8	161.2 ± 10.6	123.8 ± 9.3	81.1 ± 9.6	41.3 ± 9.4
	2,246	620.4 ± 14.7	502.3 ± 20.4	374.2 ± 18.5	251.4 ± 5.8	251.4 ± 5.8	192.1 ± 5.9	127.7 ± 11.9	65.3 ± 11.9
	4,314	1,226 ± 22.8	1,002 ± 31.1	742.3 ± 28.1	623.9 ± 29.7	495.3 ± 25.3	373.1 ± 23.4	241.3 ± 19.2	119.0 ± 14.0

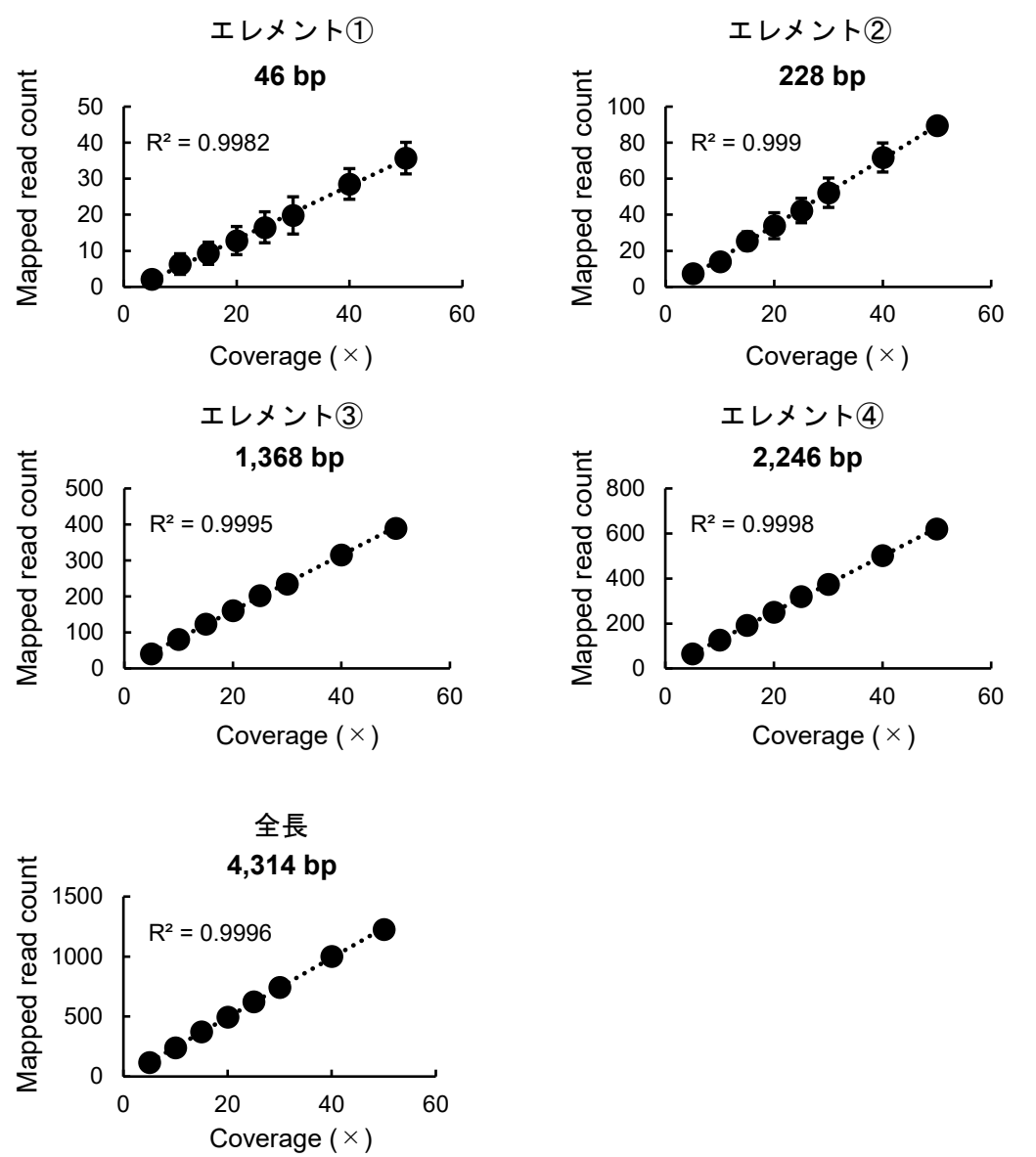


図4. Seqkitランダムサンプリングの併行精度 (RRS2)

各シークエンスカバレッジになるよう12回試行でランダムサンプリングを行い、リードデータを全長および各エレメントにマッピングした。マッピングしたリード数、その標準偏差、シークエンスカバレッジに対する相関を示す。

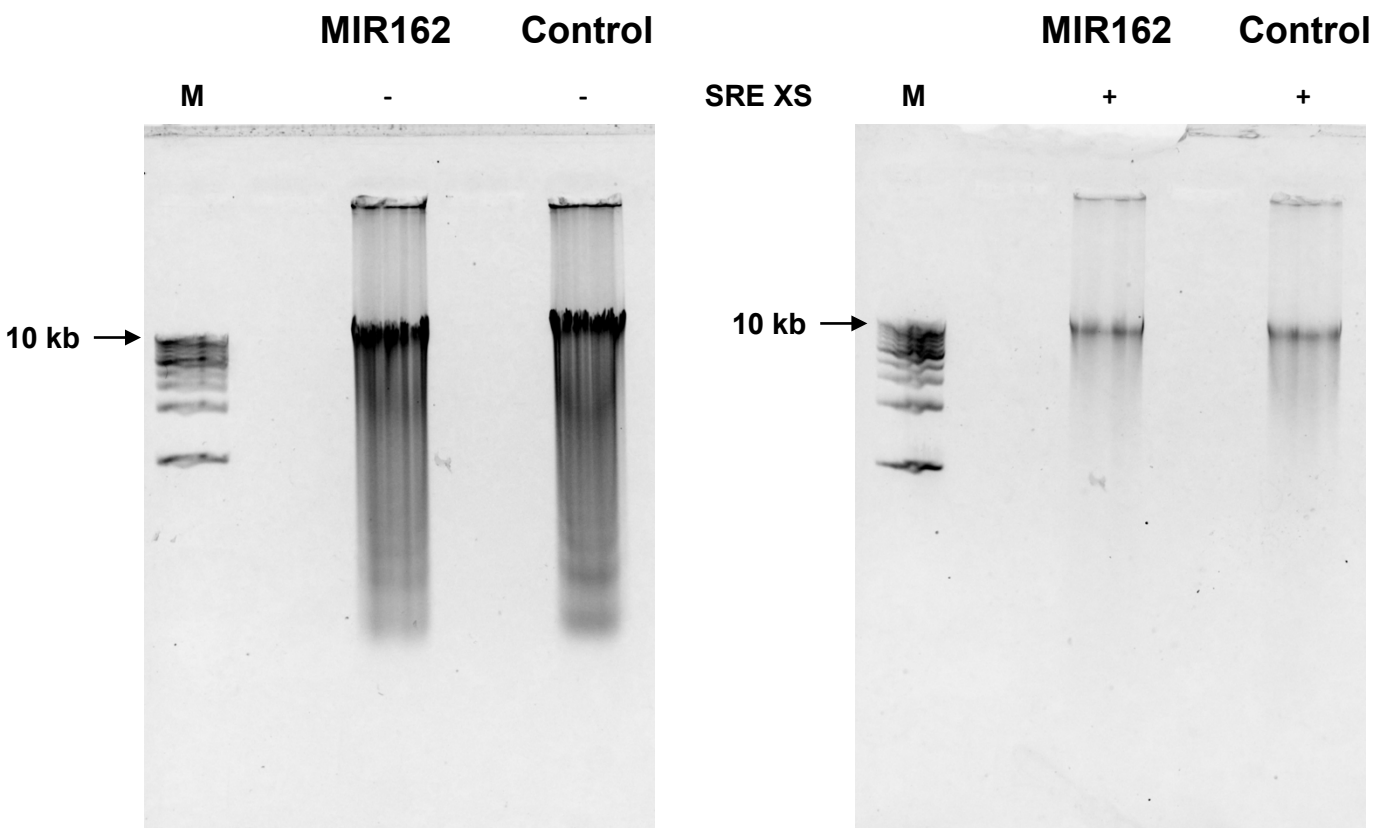


図5. トウモロコシから抽出したゲノムDNAの1%アガロース電気泳動図

コントロールトウモロコシおよびMIR162から抽出したゲノムDNAの電気泳動図を示す。

-はSRE XS (Circulomics社) 処理前、+は処理後を表す。

M: 10 kb Ladder Marker

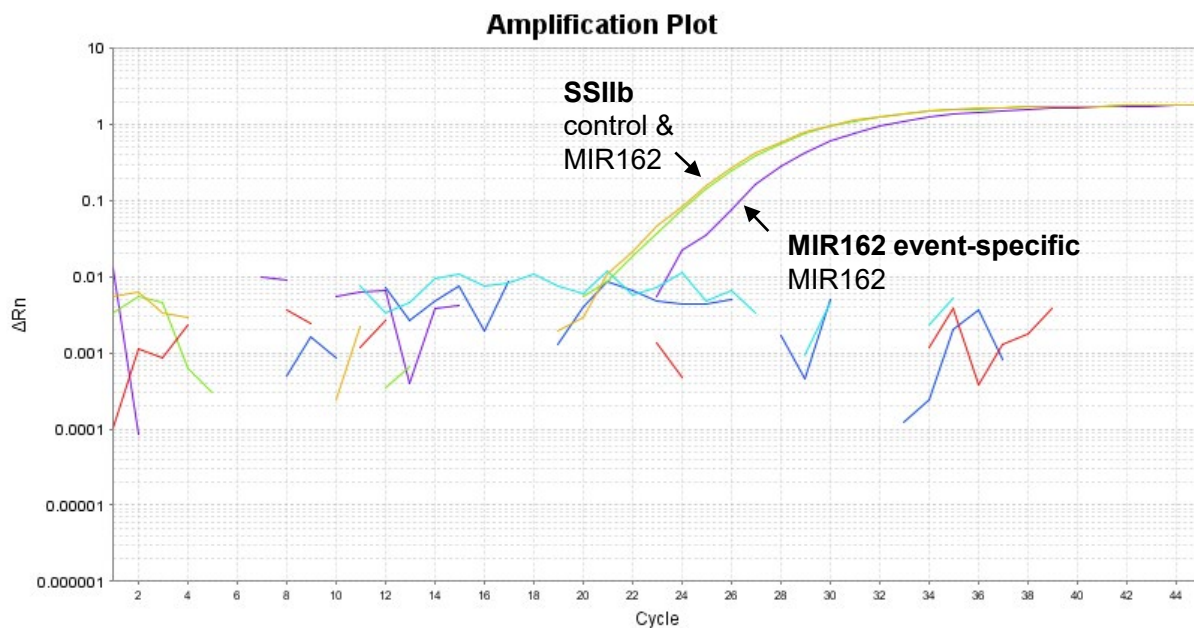
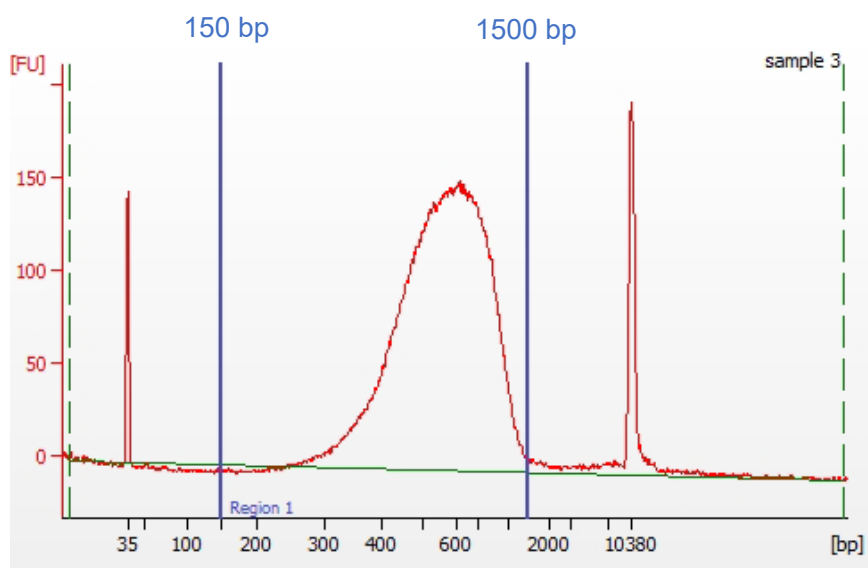


図6. SSIIbおよびMIR162系統特異的検知法結果

トウモロコシ内在性遺伝子SSIIb検知法ではコントロールおよびMIR162の増幅が確認されたが、MIR162系統特異的検知法ではコントロールは増幅せず、MIR162のみで増幅が確認された。

Control



MIR162

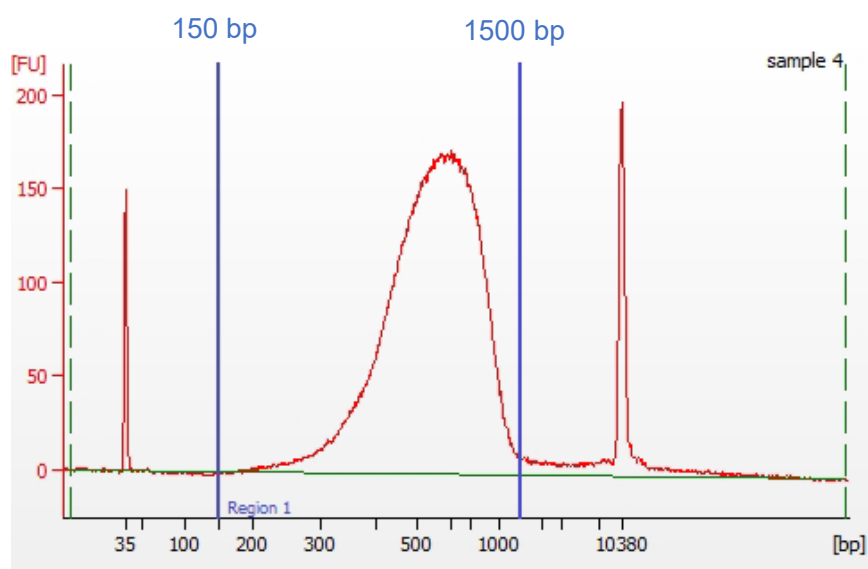


図7. コントロールトウモロコシおよびMIR162におけるWGS用ライブラリのQC結果

コントロールトウモロコシ（上図）およびMIR162（下図）のWGSライブラリはAgilent Bioanalyzer 2100による品質確認を行い、サイズやアダプターオリゴのコンタミネーションがないことを確認した後、シーケンスを行った。

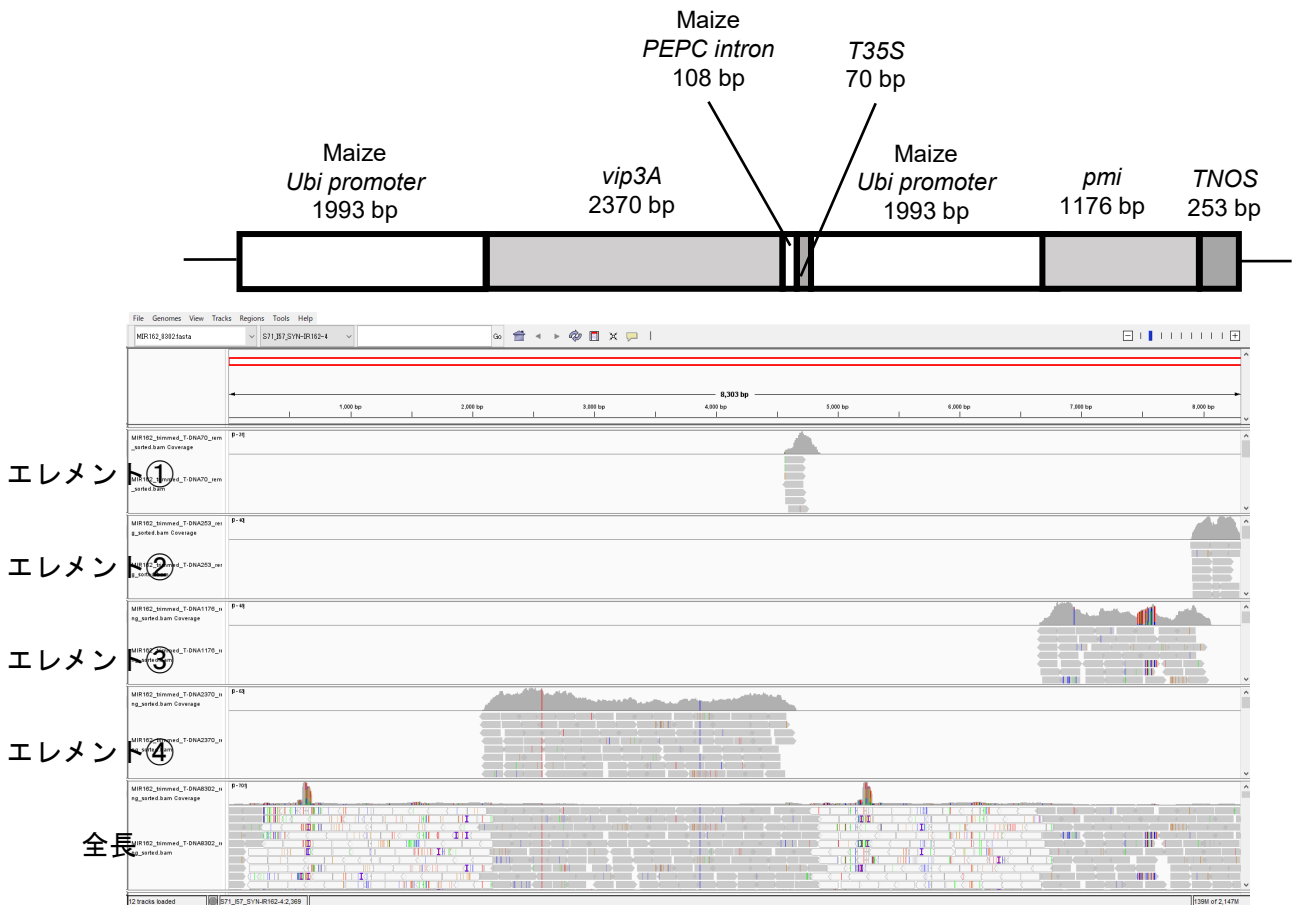


図8. MIR162に挿入された外来遺伝子にマッピングしたショートリードシークエンスデータのIGV画像

各エレメント配列または全長配列にマッピングしたリードを再度全長配列にマッピングした。白抜きになっているリードはMAPQが0であることを示す。

	Target (bp)	50×	40×	30×	25×	20×	15×	10×	5×
Mapped read count (Average ± SD, n=12)	70	28.3 ± 1.5	22.2 ± 3.4	16.6 ± 2.8	13.5 ± 3.1	10.2 ± 3.6	7.7 ± 3.1	4.8 ± 2.6	2.9 ± 1.9
	253	74.1 ± 4.8	60.8 ± 6.9	46.6 ± 5.8	39.3 ± 5.1	30.7 ± 4.6	22.5 ± 5.2	14.7 ± 4.7	6.8 ± 3.0
	1,176	220.8 ± 7.4	182.2 ± 14.8	139.3 ± 14.9	118.7 ± 10.4	93.4 ± 11.1	69.8 ± 8.2	46.9 ± 8.3	24.8 ± 4.5
	2,370	542.7 ± 16.5	430.8 ± 29.1	327.9 ± 25.0	281.1 ± 11.0	222.7 ± 13.6	163.5 ± 14.1	106.2 ± 16.1	49.8 ± 16.8
	8,302	3,915.3 ± 112.0	3,144.2 ± 209.2	2,378.1 ± 209.4	1,996.6 ± 113.3	1,567.7 ± 117.8	1,176.3 ± 104.3	787.3 ± 99.0	394.3 ± 97.9

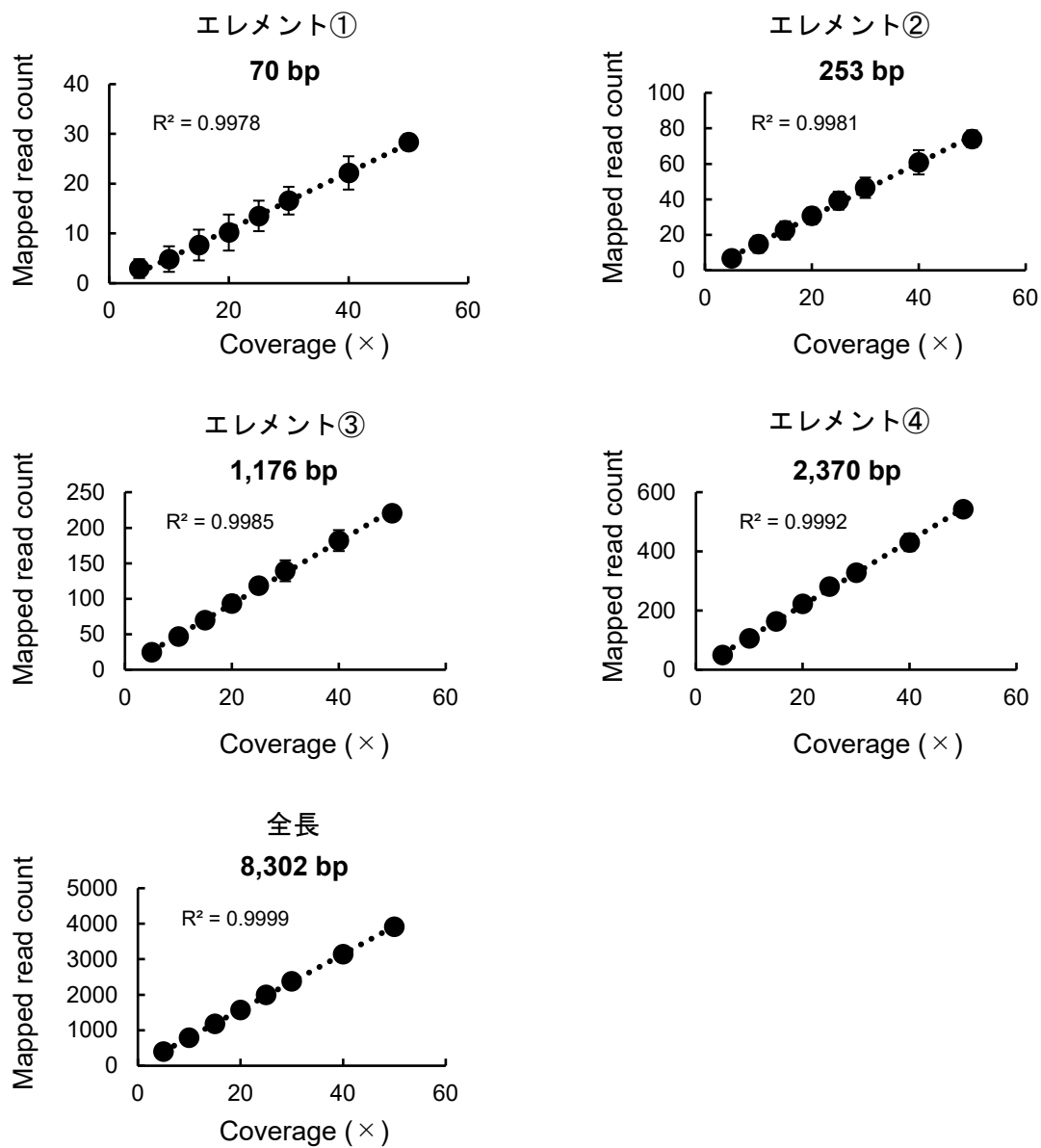


図9. Seqkitランダムサンプリングの併行精度 (MIR162)

各シークエンスカバレッジになるよう12回試行でランダムサンプリングを行い、リードデータを全長および各エレメントにマッピングした。マッピングしたリード数、その標準偏差、シークエンスカバレッジに対する相関を示す。

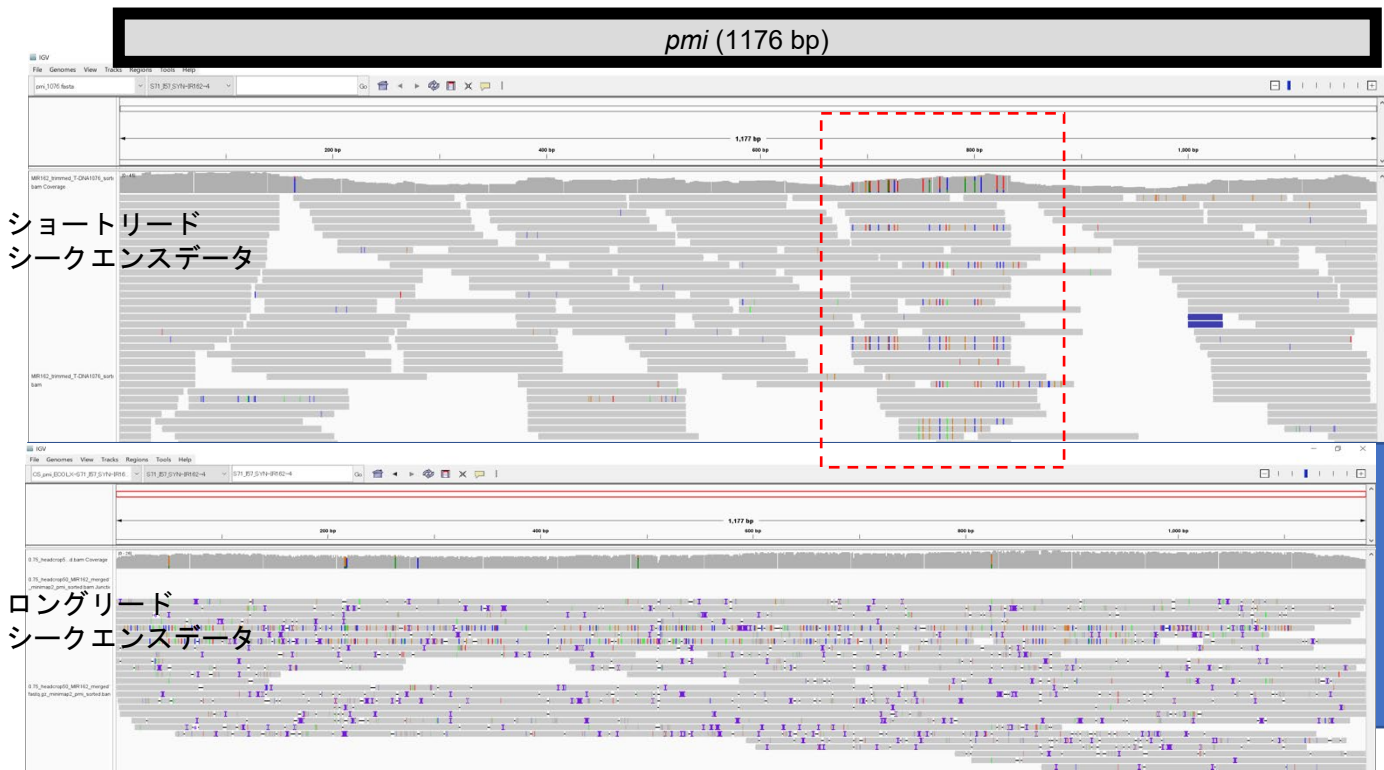


図10. *pmi*領域に各WGSリードデータをマッピングしたIGV画像 (MIR162)

ショートリードシーケンスデータ（上段）およびロングリードシーケンスデータ（下段）について、*pmi*領域にマッピングした結果を示す。赤点線枠はショートリードシーケンスデータで変異が蓄積していた領域。

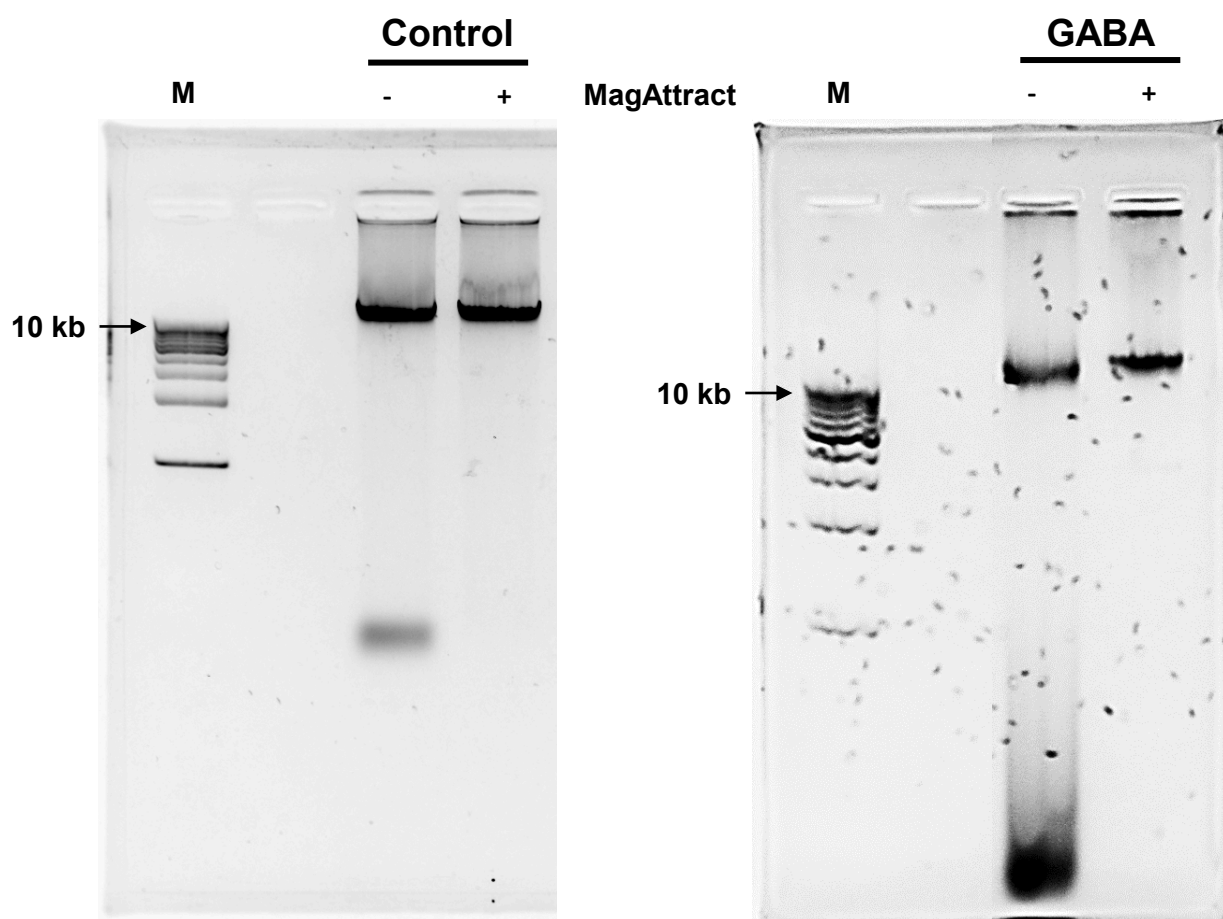


図11. トマト青果から抽出したゲノムDNAの1%アガロース電気泳動図

左図は野生型シシリアンルージュ、右図はゲノム編集GABAトマトから抽出したゲノムDNAを表す。
-はMagAttract (QIAGEN社) 処理前、+は処理後を表す。

M: 10 kb Ladder Marker

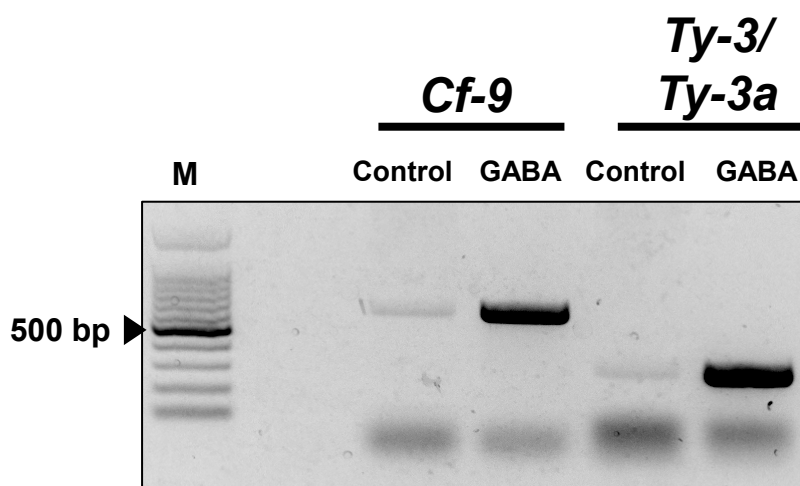


図12. DNAマーカーを用いたトマトの系統確認

DNAマーカーを用いて野生型シシリアンルージュの系統確認を行った。

Cf-9マーカーは陽性であれば650 bp、陰性であればバンドなし。

Ty-3/Ty-3aマーカーはTy-3陽性であれば450 bp、Ty-3a陽性であれば630 bp、陰性であれば320 bpにバンドが確認される。

M: 100 bp Ladder Marker

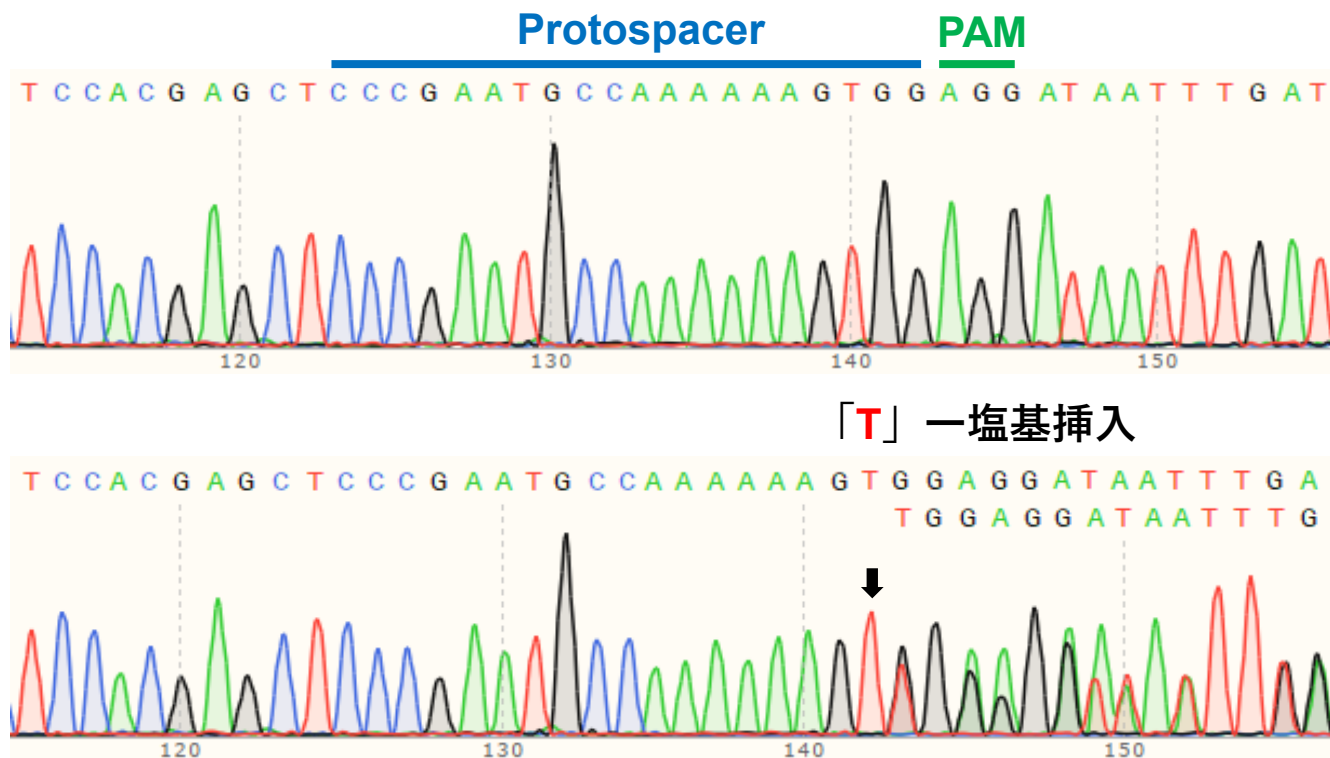
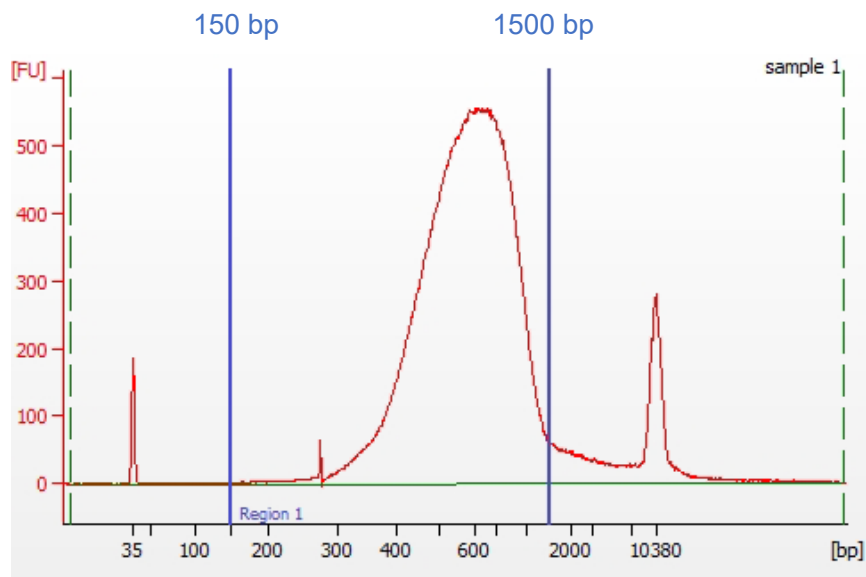


図13. 野生型トマトおよびゲノム編集GABAトマトにおけるGAD3配列の比較

野生型トマト（上段）およびゲノム編集GABAトマト（下段）におけるGAD3配列を比較した。図中のPAM（Protospacer Adjacent Motif）およびProtospacerは、シーケンス結果より推定されるCas9の結合配列、GAD3-gRNA配列を表す。

Control



GABA

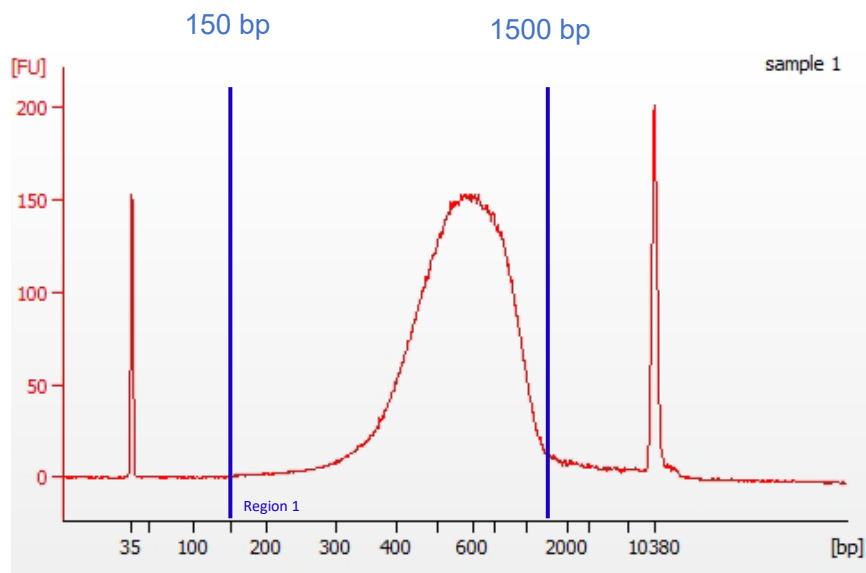
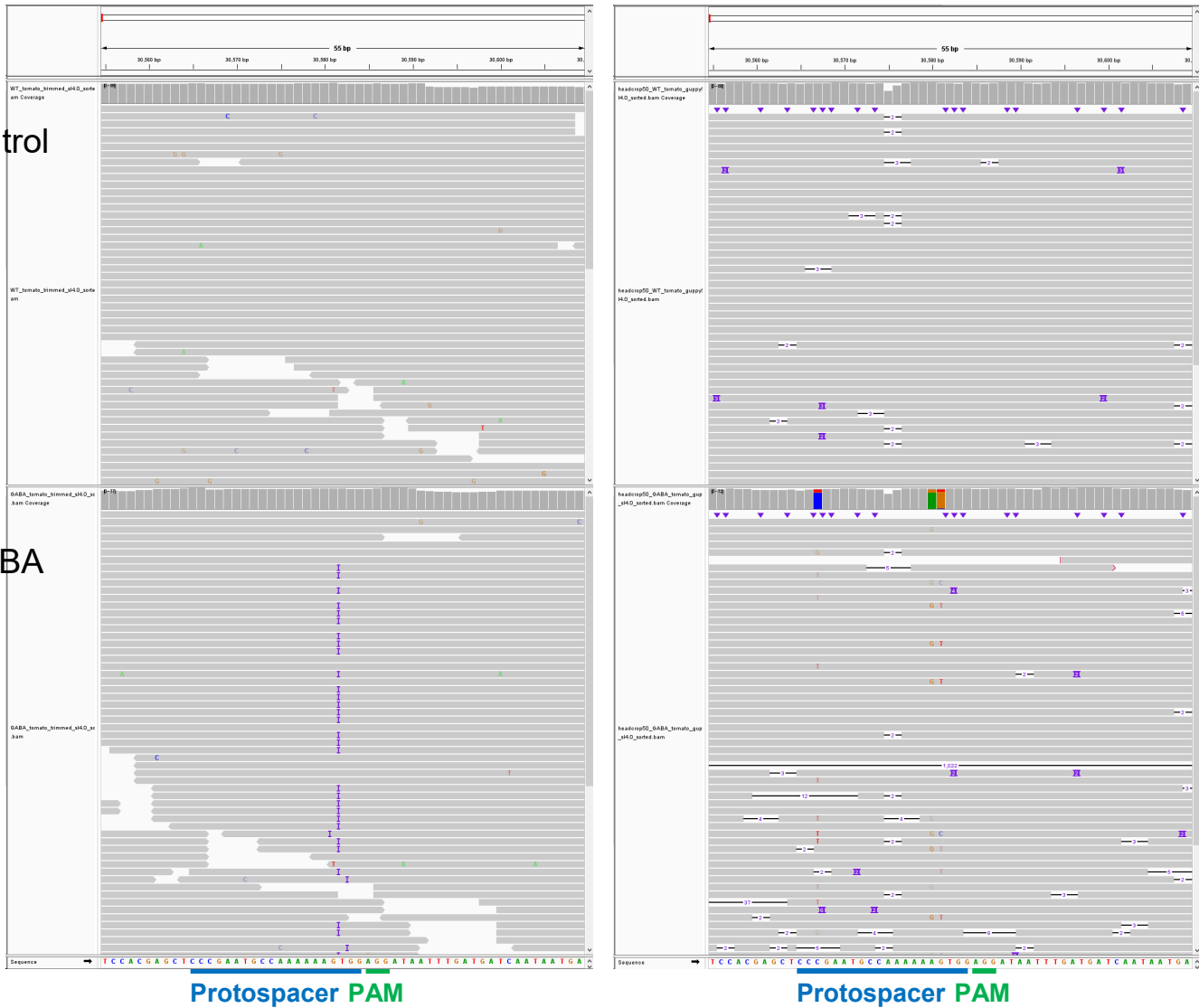


図14. 野生型トマトおよびGABAトマトのWGS用ライブラリのQC結果

野生型トマト（左図）およびGABAトマト（右図）のWGSライブラリはAgilent Bioanalyzer 2100による品質確認を行い、サイズやアダプターオリゴのコンタミネーションがないことを確認した後、シーケンスを実施した。

Control

GABA



ショートリード
シークエンスデータ

ロングリード
シークエンスデータ

図15. ショートリードシークエンス、ロングリードシークエンスにおけるGAD3標的部位のマッピングリードの比較

全ゲノムシークエンスデータにおけるGAD3 標的部位のIGV画像。左図がショートリードシークエンスリードデータ、右図がロングリードシークエンスリードデータ。いずれも上段が野生型トマト由来のリードを表し、下段がGABAトマト由来のリードを表す。図中のPAM（Protospacer Adjacent Motif）およびProtospacerは、シークエンス結果より推定されるCas9の結合配列およびGAD3-gRNA配列を表し、紫のマークはInsertionを表す。

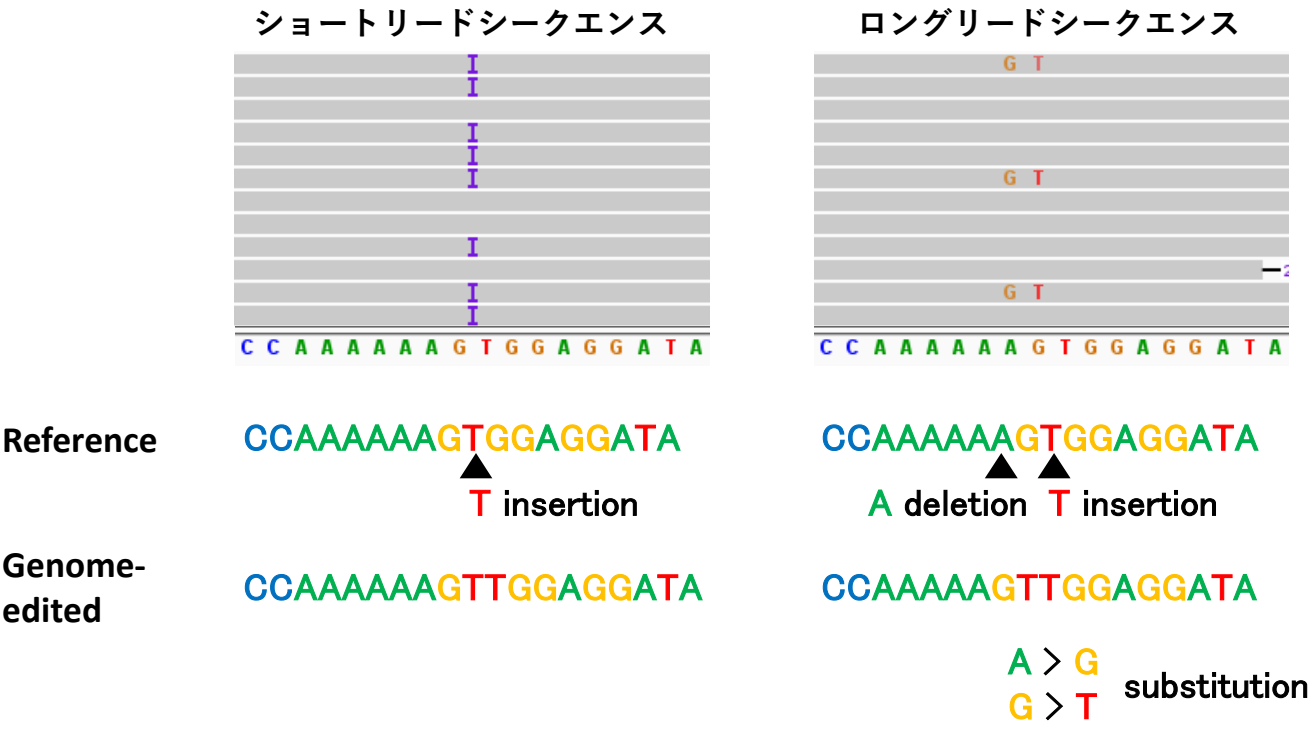


図16. ゲノム編集トマトにおけるGAD3標的部位のショートリードシーケンスとロングリードシーケンスのマッピングリードの比較

ショートリードシーケンス（左図）ではゲノム編集によるTの一塩基挿入を正しく「insertion」と表示するが、ロングリードシーケンス（右図）ではAのホモポリマー配列のエラーによりAが一塩基「deletion」し、結果A > GおよびG > Tの「substitution」と表示されている。

表1. RRS2およびコントロールサイズのWGSにより得られたリードデータの統計値と解析に使用したソフトウェア

		ショートリードシーケンス		ロングリードシーケンス	
		Control	RRS2	Control	RRS2
ミ ニ マ ル コ ン ト ロ ール	Total reads	474.8M	454.7M	9.89 M	3.9 M
	Total bases	70.0 G	66.9 G	26.4 G	52.1 G
	N50	150 b	150 b	4.26 kb	25.0 kb
	Coverage	63.6 x	60.8 x	23.5 x	47.2 x
Trimming		Trim Galore (ver. 0.6.7)		NanoFilt (ver. 2.8.0)	
Mapping		BWA-MEM (ver. 0.7.17-r1188)		Minimap2 (ver. 2.21-r1071)	
Assembly		Velvet (ver. 1.2.10) SPAdes (ver. 3.15.5)		Flye (ver. 2.9)	

表2. RRS2に挿入された外来遺伝子全長および各エレメントにマッピングしたリードの統計値とアセンブリ結果

Target (bp)	ショートリードシーケンス						ロングリードシーケンス				
	Mapped reads	Mapped bases (b)	Average length (b)	>Q30 (%)	Contigs	Accuracy (%)	Mapped reads	Mapped bases (b)	Contigs	Assembly length(b)	Accuracy (%)
46	46	6,882	149.6	94.55	1	100	0	0	-	-	-
228	106	15,654	147.7	91.54	1	100	41	1,380,543	1	90,947	100
1,368	475	70,498	148.4	93	1	100	44	1,408,707	1	121,913	100
2,246	762	112,971	148.3	94.03	1	100	49	1,498,178	1	121,893	99.9
4,314	1,481	218,718	147.4	93.85	1	100	58	1,662,854	1	91,213	99.9

-: 未検討

表3. Velvetによるアセンブリ結果 (RRS2)

	Target (bp)	50 ×	40 ×	30 ×	25 ×	20 ×	15 ×	10 ×	5 ×
Contig 複数で全体を カバー	46	12/12	12/12	12/12	12/12	12/12	12/12	10/12	5/12
	228	12/12	12/12	12/12	12/12	12/12	12/12	9/11	9/12
	1368	12/12	11/12	12/12	11/12	11/12	12/12	9/12	1/12
	2246	12/12	12/12	12/12	12/12	11/11	11/11	10/12	1/12
	4314	12/12	12/12	11/12	10/12	10/12	8/12	7/12	0/12
Contig 1本で全体を カバー	46	12/12	12/12	12/12	12/12	12/12	12/12	10/12	5/12
	228	12/12	12/12	12/12	12/12	12/12	12/12	9/11	9/12
	1368	12/12	10/12	10/12	8/12	6/12	3/12	4/12	0/12
	2246	10/12	10/12	9/12	8/12	3/11	1/11	2/12	0/12
	4314	11/12	9/12	7/12	5/12	2/12	0/12	1/12	0/12

表4. SPAdesによるアセンブリ結果 (RRS2)

	Target (bp)	50 ×	40 ×	30 ×	25 ×	20 ×	15 ×	10 ×	5 ×
Contig 複数で全体を カバー	46	12/12	12/12	9/12	6/12	3/12	1/12	0/12	0/12
	228	12/12	12/12	12/12	11/12	9/12	5/12	0/11	0/12
	1368	12/12	12/12	12/12	12/12	12/12	12/12	4/12	0/12
	2246	12/12	12/12	12/12	12/12	11/11	11/11	6/12	0/12
	4314	12/12	12/12	12/12	12/12	12/12	12/12	5/12	0/12
Contig 1本で全体を カバー	46	12/12	12/12	9/12	6/12	3/12	1/12	0/12	0/12
	228	12/12	12/12	12/12	11/12	9/12	5/12	0/11	0/12
	1368	12/12	12/12	12/12	12/12	12/12	11/12	4/12	0/12
	2246	12/12	12/12	12/12	12/12	11/11	10/11	4/12	0/12
	4314	12/12	12/12	12/12	12/12	11/12	10/12	4/12	0/12

表5. ロングリードシーケンスデータのランダムサンプリング、およびマッピングリード数とFlyeによるアセンブリ結果 (RRS2)

	Target (bp)	40 ×	30 ×	25 ×
Mapped read count (Average ± SD)	46	0	0	-
	228	31.3 ± 2.2	24.2 ± 2.5	-
	1368	33.8 ± 2.7	26.2 ± 3.1	-
	2246	37.3 ± 2.7	28.9 ± 2.7	24.1 ± 2.3
	4314	45.1 ± 2.4	34.8 ± 2.6	
リード長 N50 (kb) (Average ± SD)	46			-
	228	42.0 ± 2.9	41.6 ± 3.2	-
	1368	41.2 ± 2.4	40.7 ± 3.5	-
	2246	40.9 ± 2.3	40.2 ± 3.6	41.3 ± 3.1
	4314	39.3 ± 2.1	38.6 ± 3.0	
残存配列の再現 (Contig1本で全体をカバー)	46	-	-	-
	228	10/12	11/12	-
	1368	10/12	9/12	-
	2246	10/12	9/12	8/12
	4314	11/12	10/12	-

-: 未検討

表6. MIR162およびコントロールトウモロコシのWGSにより得られたリードデータの統計値と解析に使用したソフトウェア

		ショートリードシーケンス		ロングリードシーケンス	
		Control	MIR162	Control	MIR162
トリミング後	Total reads	405.9 M	447.1 M	34.5 M	57.2 M
	Total bases	118.6 G	131.7 G	72.7 G	97.7 G
	N50	150 b	150 b	3.93 kb	2.78 kb
	Coverage	49.4 x	54.9 x	30.3 x	40.7 x
Trimming		Trim Galore (ver. 0.6.7)		NanoFilt (ver. 2.8.0)	
Mapping		BWA-MEM (ver. 0.7.17-r1188)		Minimap2 (ver. 2.21-r1071)	
Assembly		Velvet (ver. 1.2.10) SPAdes (ver. 3.15.5)		Flye (ver. 2.9)	

表7. MIR162に挿入された外来遺伝子全長および各エレメントにマッピングしたリードの統計値とアセンブリ結果

Target (bp)	ショートリードシーケンス						ロングリードシーケンス				
	Mapped reads	Mapped bases (b)	Average length (b)	>Q30 (%)	Contigs	Accuracy (%)	Mapped reads	Mapped bases (b)	Assembly length (b)	Contigs	Accuracy (%)
70	31	4,553	146.9	94.33	1	100	15	58,373	11,834	1	100
253	83	12,206	147.1	90.35	1	100	24	110,259	14,919	1	100
1,176	243	35,650	146.7	89.73	1	100	42	143,444	18,379	1	100
2,370	594	87,886	148	88.57	3	-	51	129,286	4,593	1	100
8,302	4,316	636,362	147.4	89.62	11	-	274	1,001,196	-	3	-

-: 未検討

表8. Velvetによるアセンブリ結果 (MIR162)

	Target (bp)	50×	40×	30×	25×	20×	15×	10×	5×
Contig 複数で全体を カバー	70	11/11	12/12	12/12	12/12	12/12	11/12	5/11	3/12
	253	-	-	-	-	-	-	-	-
	1176	-	-	-	-	-	-	-	-
	2370	-	-	-	-	-	-	-	-
	8302	-	-	-	-	-	-	-	-
Contig 1本で全体を カバー	70	11/11	12/12	12/12	12/12	12/12	11/12	5/11	3/12
	253	-	-	-	-	-	-	-	-
	1176	-	-	-	-	-	-	-	-
	2370	-	-	-	-	-	-	-	-
	8302	-	-	-	-	-	-	-	-

-: 未検討

表9. SPAdesによるアセンブリ結果 (MIR162)

	Target (bp)	50 ×	40 ×	30 ×	25 ×	20 ×	15 ×	10 ×	5 ×
Contig 複数で全体を カバー	70	11/11	11/12	7/12	2/12	1/12	0/12	0/11	0/12
	253	12/12	12/12	11/11	12/12	9/12	3/12	1/12	0/12
	1176	12/12	12/12	12/12	12/12	10/12	1/12	0/12	0/12
	2370	12/12	12/12	12/12	11/11	12/12	12/12	1/11	0/12
	8302	0/12	0/12	0/12	0/12	0/12	0/12	0/12	0/11
Contig 1本で全体を カバー	70	11/11	11/12	7/12	2/12	1/12	0/12	0/11	0/12
	253	12/12	12/12	11/11	12/12	9/12	3/12	1/12	0/12
	1176	0/12	3/12	7/12	7/12	8/12	1/12	0/12	0/12
	2370	12/12	12/12	12/12	11/11	10/12	9/12	0/11	0/12
	8302	0/12	0/12	0/12	0/12	0/12	0/12	0/12	0/11

表10. ロングリードシーケンスデータのランダムサンプリング、およびマッピングリード数とFlyeによるアセンブリ結果 (MIR162)

	Target (bp)	30×	20×
Mapped read count (Average ± SD)	70	11.8 ± 1.5	7.3 ± 1.8
	253	18.5 ± 1.2	12.2 ± 2.9
	1176	31.9 ± 1.4	21.2 ± 2.9
	2370	37.6 ± 2.7	25.3 ± 3.4
リード長 N50 (kb) (Average ± SD)	70	8.1 ± 0.4	11.9 ± 5.6
	253	8.4 ± 1.3	8.0 ± 1.5
	1176	6.5 ± 1.3	6.4 ± 2.0
	2370	4.6 ± 1.3	4.6 ± 1.8
残存配列の再現 (Contig 1本で全体をカバー)	70	2/12	1/12
	253	11/12	4/12
	1176	12/12	5/12
	2370	12/12	9/12

表11. GABAトマトおよび野生型トマトのWGSにより得られたリードデータの統計値と解析に使用したソフトウェア

		ショートリードシーケンス		ロングリードシーケンス	
		Control	GABA	Control	GABA
ト マ ト 後 ミ ニ マ ピ ン グ	Total reads	191.4 M	258.0 M	3 M	5.78 M
	Total bases	55.9 G	75.3 G	58.5 G	52.4 G
	N50	150 b	150 b	35.5 kb	13.8 kb
	Coverage	58.9 x	79.2x	61.6 x	55.1 x
Trimming		Trim Galore (ver. 0.6.7)		NanoFilt (ver. 2.8.0)	
Mapping		BWA-MEM (ver. 0.7.17-r1188)		Minimap2 (ver. 2.21-r1071)	

厚生労働省科学研究費補助金（食品の安全確保推進研究事業）
「新たなバイオテクノロジーを用いて得られた食品の安全性確保と
リスクコミュニケーションのための研究」
分担研究報告書（令和4年度）

ゲノム編集食品のトランスクリプトーム変化の検証

研究分担者 吉場 聡子 国立医薬品食品衛生研究所

研究要旨：

本分担研究では、ゲノム編集による意図しない変化について検証する上で、トランスクリプトーム（遺伝子発現）変化に着目した。RNA-seq による遺伝子発現変化の解析を行うことで、mRNA レベルの変化を予測や先入観なく網羅的に検証し、ゲノムのオフターゲット解析やアレルゲン予測では捕捉できない変化を検出できる可能性がある。本研究ではモデルサンプルとして、届出済みゲノム編集食品から「グルタミン酸脱炭酸酵素遺伝子の一部を改変し GABA 含有量を高めたトマト；GABA トマト」を対象とし、未編集トマトと GABA トマトの青果のトランスクリプトから遺伝子発現の比較解析を試みた。ゲノム編集前後の遺伝子発現の変化の検証結果及び今後の課題について報告する。

A. 研究目的

現在ゲノム編集食品の安全性評価においては、ゲノム編集による意図しない変化の影響、すなわちオフターゲット（意図しないゲノム配列の改変）リスクや新規アレルゲン産生及び毒性物質増加の可能性、代謝系への影響などが主な評価の対象となっている。これらの検証は基本的に予測に基づいており、例えばゲノムの変化は、*in silico* で予測された特定の遺伝子配列に対して、変異の有無を確認する方法が用いられており、非常に精度が高い一方で、予測されなかった箇所の変異は検証されない。アレルゲンや代謝成分に関しても同様であり、遺伝子配列の変化による新規アレルゲンタンパク質の産生や、オンターゲットの編集により想定される代謝物の変化について検証されているが、想定していない新規アレルゲンや代謝経路の変化については、評価の対象となっていない。

一方で、本研究課題（全体）の分担研究において、質量分析による網羅的な未知化合物の推定が

行われ、ゲノム編集トマト（GABA トマト；グルタミン酸脱炭酸酵素遺伝子の一部を改変し GABA 含有率を高めたトマト）のゲノム編集前後で、未知化合物を含む複数の化合物が有意に増減していることが示されている。現在ゲノム編集技術は、オリジナルの Cas9 による二本鎖 DNA の切断と修復による変異の導入に加えて、base editing や prime editing などの塩基編集、RNA をターゲットとした RNA editing など多様化しており、今後「想定しない」変化を網羅的に検出し、その変化による影響を評価することは重要になる可能性がある。

本分担研究では、ゲノム編集による意図しない変化のうち、トランスクリプトーム変化（遺伝子発現の変化）に着目した（図1）。編集前後の遺伝子発現の変化を調べることで、mRNA レベルの変化を予測や先入観なく網羅的に検証し、ゲノム解析やアレルゲン予測では捕捉できない可能性のある変化の掘り出しとこれらの変化とのつながりについて検証しようと試みた。

モデルサンプルとして、3つの届出済みのゲノム編集食品、GABA トマト、可食部増大マダイ、高成長トラフグのうち、GABA トマトの解析を行った。その理由として、マダイとトラフグの作出方法においては、Cas9 の導入に mRNA が使用されており、ゲノムに挿入されず作用が一時的である、すなわちオフターゲットの可能性が低い一方で、GABA トマトは T-DNA による Cas9 の導入により、比較的持続的に作用している可能性がある、意図しない変化が起こりやすいと考えられるためである（参考文献 1,2）。また、前述の通り GABA トマトのメタボローム解析により、ゲノム編集により未知化合物の量が変化していることが示されており、トランスクリプトーム解析によりメタボロームとのつながりについても検証できる可能性がある。

研究開始時点では、トランスクリプトーム比較解析で新たに得られる可能性のある情報として、下記を想定した。1.オンターゲット変異による代謝経路の変化に伴う間接的な遺伝子発現の変化、2.オフターゲット予測から漏れたオフターゲット変異による遺伝子発現の変化、3.Cas9 の持続的な作用による遺伝子発現の変化（DNA 損傷シグナルの不活化等）（図 1）。

本研究は、ゲノム編集食品の編集前後の想定しない変化とその影響について、トランスクリプトーム比較解析により、網羅的に検証し評価することで、ゲノム編集技術の安全性評価において新たな問題提起につながる事象を探ること、さらにゲノム編集食品のトランスクリプトーム解析に必要な要件及び課題について議論・検証することを目とした。

B. 研究方法

概要

ゲノム編集トマト（GABA トマト）をモデルサンプルとして、野生型および GABA トマトから

total RNA を抽出し、RNA-seq による遺伝子発現比較解析を行う。

サンプル

GABA トマト青果はサナテックシード社より購入した。野生型トマト（cv. シシリアンルージュ CF）は神奈川県内のスーパーで購入した。

Total RNA の抽出

生鮮サンプルを-30℃で凍結した後、凍結乾燥機（EYELA FDU-1200; 東京理化機械株式会社）により凍結乾燥した。凍結乾燥サンプルは、液体窒素を入れた乳鉢で粉碎後、Fruit-mate for RNA purification（タカラバイオ株式会社）による前処理と NucleoSpin RNA plant（Macherey-nagel GmbH & Co.KG）を組み合わせ total RNA を回収した。回収した total RNA は分光光度計（NanoDrop; Thermo Fisher Scientific）および TapeStation（Agilent Technologies Japan, Ltd.）を用いて、濃度測定及び品質検定を行った。RNA-seq に供する total RNA の品質は、次の値を目安とした。

OD260/280: ≥ 1.6

OD260/230: ≥ 1.6

RINe (RNA integrity number): ≥ 7.0

28S/18S (rRNA ratio): ≥ 1.0

RNA-seq

シーケンスは、次世代シーケンス（NGS）解析サービス（タカラバイオ株式会社）により行った。作業内容は下記の通りである。

SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing を用いて、RNA 増幅、Nextra XT DNA Library Prep Kit を用いて DNA ライブラリ作製。ライブラリ作製は、Smart-Seq2 法（参考文献 3）を用いている。シーケンスは、NovaSeq システム（illumina）を用いて、150base 両末端解析（6Gb/4,000 万リード；2,000 万リードペア）を

野生型 (WT) トマト及び GABA トマトそれぞれ 3 サンプルずつ (合計 6 サンプル) 実施。得られたリード配列をリファレンスゲノム配列にマッピング後、遺伝子ごとに正規化された発現量、リードカウント値及び TPM (transcript per million) 値を算出している。なお、マッピングに用いたトマトのリファレンスゲノム配列および遺伝子アノテーション情報はそれぞれ、SL3.0、ITAG2.3、遺伝子数の合計は 35,825 である。結果のリード配列は FASTQ 形式、マッピング結果は bam 形式、遺伝子発現量データは csv ファイルで納品された。

遺伝子発現解析

再現性の確認 (technical replicate) 及び 3 反復 2 群間比較による遺伝子発現量変動解析を行う。発現解析スキームを図 2 で示す。A. リードカウント値からトランスクリプトームデータ分析ウェブプラットフォーム iDEP (integrated Differential Expression and Pathway analysis; iDEP.96, 参考文献 4) を用いて解析する方法と、B. TPM 値から手動で発現変動遺伝子を抽出する方法を平行して行った。

A では、遺伝子発現量データからリードカウントデータを作成し、iDEP を用いて、データの前処理と検証後、発現量変動解析、Gene Ontology (GO) エンリッチメント解析などを行った。iDEP のワークフローを図 3 に示す。

B では、正規化された各サンプルの TPM 値を用いて t-検定 (関連なし二群比較) を行い、下記の基準で WT トマトと GABA トマトで発現量が有意に異なる遺伝子を抽出しリスト化した。

FoldChange:

$$|\log_2 TPM_{GABA} - \log_2 TPM_{wt}| \geq 1$$

t-test p-value ≤ 0.05

t-test q-value ≤ 0.05

さらに、抽出した遺伝子リストを用いて、gProfiler (参考文献 5) による GO 解析を行った。

C. 研究結果と考察

1. total RNA の品質検定の結果

解析に供した total RNA 品質検定の結果を図 4 に示す。RNA-seq 解析サービスへのサンプル送付前と送付後に品質検定が行われ、ここでは解析サービスによるライブラリ作成直前の品質検定結果のみを示した。

2. マッピング及び遺伝子アノテーション

マッピング結果の情報を表 1 で示す。6 つすべてのサンプルにおいて、input reads に対し、ペアードリード (ペアでマッピングされたリード) の合計はそれぞれ 4,000 万リード以上 (99%以上)、また MAPQ が 10 以上のリードについても 4,000 万リード以上 (88%以上) だった。これらのリードに対して、EnsembleID を用いて 35,825 遺伝子がアノテーションされた。

3. 遺伝子発現解析の結果

3-1. リードカウント値を用いた iDEP による解析

遺伝子発現量のリードカウントデータを iDEP に入力、最小 CPM (count per million) 値を 0.5 に設定したところ、EnsembleID でアノテーションされた 35,825 遺伝子のうち、17,597 遺伝子がパスした。この 17,597 遺伝子に対して、発現量変動解析を行った。方法は DESeq2 を選択し、FDR cutoff: 0.1、fold-change (GABA/WT): ≥ 2 で発現変動遺伝子の絞り込みを行ったところ、3,210 遺伝子で発現量が上昇 (Upregulated)、3,525 遺伝子で減少 (Downregulated) していた。結果を volcano plot で示す (図 5)。FDR 値の高かった遺伝子の情報をそれぞれ上から順に 10 ずつ示す。さらに抽出された遺伝子を用いて、GO エンリッチメント解析を行った。結果を表 2 で示す。Up-regulated genes では、ribonucleoprotein の生合成や chitin の活性に関わる GO、一方 down-regulated genes では、Carbohydrate metabolic/

catabolic process, plastid, chloroplast に関わる GO などが抽出された。

3-2.TPM 値を用いた t-検定

マッピングの結果から算出された TPM 値に対して t-検定を行い、WT トマトと GABA トマトで発現量が有意に異なる遺伝子を抽出しリスト化した。結果の一部を表 3 で示す。EnsemblID でアノテーションされた 35,825 遺伝子のうち、研究方法に示した基準で、発現量が有意に上昇していた遺伝子 (up-regulated genes) は 932、減少していた遺伝子 (down-regulated genes) は 4,154、合計 5,086 遺伝子が抽出された。

3-3.gProfiler による GO 解析

上記で抽出されたそれぞれの遺伝子リストを用いて、GO 解析を行ったところ、up-regulated genes では、chitin や cell wall の代謝プロセス、ribonucleoprotein の生合成など、down-regulated genes では、catalytic activity, small molecule metabolic process, plastid に関係する GO が抽出された。結果を図 6 で示す。これらの結果は、iDEP の結果と傾向が一致する。

3-4.GAD 遺伝子の発現量変化

GABA トマトでは、グルタミン酸脱炭酸酵素 (glutamate decarboxylase; GAD) をコードする遺伝子の一つ、SIGAD3 が改変されている。当該酵素は、グルタミン酸のカルボキシル基を除去することで GABA を合成する。トマトは 5 つの GAD 遺伝子を有しているが、GABA 合成に寄与するのは GAD2 と SIGAD3 であり、特に SIGAD3 が主要な働きをすることがわかっている (参考文献 6,7)。GABA トマトは、CRISPR/Cas9 による変異導入 (一塩基挿入) により、SIGAD3 の C 末端の自己阻害領域が除去されており、その結果 SIGAD3 の活性が上昇することで、GABA の蓄積量が増大する。

SIGAD3 を含む 5 つの GAD のゲノム編集前後の遺伝子発現量の変化を表 4 に示す。SIGAD3 及び GAD2, GAD4 の遺伝子発現量に有意な差は見られなかった。一方 GAD の発現量は 1/3 以下 (logFC: -1.8) を示し、有意に減少が認められた。SIGAD3 の活性の上昇により間接的に影響した可能性はあるが、関連は不明である。

なお、SIGAD3 のマッピングリードを IGV 上で確認したところ、WT と GABA どちらにおいても、3'側にリードの偏りが見られたが、その他の点において特に相違は見られなかった (図 7)。

3-5.その他 GABA 代謝系に関わる遺伝子の発現量変化

GABA はグルタミン酸を基質として GAD により合成された後、GABA アミノ基転移酵素 (GABA-T) により Succinic semialdehyde (SSA)へと代謝される。SSA は SSA 脱水素酵素 (SSADH) により Succinate に代謝された後、TCA 回路へ流入する (GABA shunt, 図 8)。これらの GABA 代謝経路に関わる可能性のある分子について、遺伝子の発現量を比較した。結果を表 5 に示す。WT と比較して GABA トマトでは、GABA-TP1 の発現量が有意に増加 (logFC: 1.7)、GABA-TP3 と SSADH の発現量が有意に減少 (logFC: -4.9, -1.3) していた ($p < 0.05$)。これらに関しても、SIGAD3 の活性の上昇により遺伝子発現が間接的に影響した可能性はあるが、関連は不明である。

4. 考察と今後の課題

本研究では、ゲノム編集による意図しない変化のうち、トランスクリプトーム変化に着目し、編集前後の遺伝子発現の変化を調べることで、mRNA レベルの変化を予測や先入観なく網羅的に検証することを試みた。

GABA トマトをモデルサンプルとして、WT と GABA トマトから mRNA を抽出し、RNA-seq 及び遺伝子発現の比較を行った。2 つの方法で遺伝

子発現変動解析及び GO 解析を行ったところ、WT に比べて GABA トマトで発現量が上昇していたのは、chitin や cell wall の代謝プロセス、ribonucleoprotein の生合成などに関係する遺伝子であり、一方で発現量が減少していたのは、catalytic activity, small molecule metabolic process, plastid に関係する遺伝子だった。また、GAD 遺伝子や GABA 代謝に関わる遺伝子についても、いくつかの遺伝子で発現量の変化が見られた。

このように、WT と GABA トマトでは、有意に遺伝子発現のパターンの違いが見られたが、その原因として、ゲノム編集による影響が考えられるが、トマトの生育環境、収穫時期、生果の保存状態の違いなどその他の影響についても可能性は無視できない。特に当初に想定した中で、オンターゲット変異による代謝経路の変化に伴う間接的な遺伝子発現の変化、オフターゲット予測から漏れたオフターゲット変異による遺伝子発現の変化を明確にするためには、今回得られた遺伝子発現の変化のデータを、遺伝子配列や発現制御領域などゲノムの変化に落とし込む必要がある。3 つめに想定した、Cas9 の持続的な作用による遺伝子発現の変化に関して、特に DNA 損傷シグナルに関連する遺伝子は抽出されなかった。

本研究は、ゲノム編集食品の編集前後の想定しない変化とその影響について、遺伝子発現の変化に着目し、これまでと異なる切り口で検証を試みた。今後の課題としては、編集の前後で発現変化のあった遺伝子について、ゲノム変化やメタボローム変化を参照すること、その方法を含めて検証することが必要であると考ええる。

参考文献：

- 1) ゲノム編集技術応用食品及び添加物の食品衛生上の取扱要領に基づき届出された食品及び添加物一覧 (https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryuu/shokuhin/bio/genomed/newpage_00010.html)

- 2) グルタミン酸脱炭酸酵素遺伝子の一部を改変し GABA 含有量を高めたトマトに関する届出情報 (<https://www.mhlw.go.jp/content/11120000/000828873.pdf>)
- 3) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Picelli et al., Nat Methods, 10 1096-1098, 2013 doi:10.1038/nmeth.2639.
- 4) iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. Ge SX et al., BMC Bioinformatics, 19:534, 2018 doi: 10.1186/s12859-018-2486-6
- 5) g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update) Nucleic Acids Research 2019; doi:10.1093/nar/gkz369
- 6) Tomato Glutamate Decarboxylase Genes SlGAD2 and SlGAD3 Play Key Roles in Regulating γ -Aminobutyric Acid Levels in Tomato (*Solanum lycopersicum*). Takayama et al., Plant Cell Physiol. 56:1533-45, 2015 doi: 10.1093/pcp/pcv075
- 7) Activating glutamate decarboxylase activity by removing the autoinhibitory domain leads to hyper γ -aminobutyric acid (GABA) accumulation in tomato fruit. Takayama et al., Plant Cell Rep. 36:103-116, 2017 doi: 10.1007/s00299-016-2061-4

D. 研究発表・業績

1. 論文発表

なし

2. 学会発表

- 1) 曾我慶介、江木智宏、成島純平、吉場聡子、真野潤一、高畠令王奈、橘田和美、中村公亮、

柴田識人、近藤一成：新遺伝子組み換え表示制度に向けた試験法開発 ～「遺伝子組み換えでない」表示の今後～、第8回次世代を担う若手のためのレギュラトリーサイエンスフォーラム、東京、2022年8月26日

- 2) 近藤一成、曾我慶介、成島純平、吉場聡子、柴田識人、田口千恵、坂田こずえ、加藤怜子：ゲノム編集によって発生する意図しない変異はどこから来るのか、NGS EXPO 2022、大阪、2022年10月18日
- 3) 成島純平、木俣真弥、志波優、権藤崇裕、秋元智、曾我慶介、吉場聡子、中村公亮、柴田識人、近藤一成：ゲノム編集におけるオフターゲット予測法 SITE-Seq の Galaxy ベース新規解析パイプラインの開発とその作物への応用、NGS EXPO 2022、大阪、2022年10月18日
- 4) 吉場聡子、成島純平、曾我慶介、杉野御祐、柴田識人、近藤一成：遺伝子組換え食品の同定に資する新たなゲノム解析技術の検討、第59回全国衛生化学技術協議会年会、東京、2022年10月31日-11月1日
- 5) 柴田識人、成島純平、曾我慶介、吉場聡子、近藤一成：ゲノム編集食品における外来性遺伝子の残存を評価する全ゲノムシーケンスデータ解析の標準化に向けた取り組み、第59回全国衛生化学技術協議会年会、東京、2022年10月31日-11月1日
- 6) 成島純平、杉野御祐、曾我慶介、吉場聡子、柴田識人、近藤一成：NGSを用いた網羅的なオフターゲット変異候補部位予測法の高GABA産生ゲノム編集トマトへの適用と検証、第59回全国衛生化学技術協議会年会、東京、2022年10月31日-11月1日

- 7) 柴田識人、成島純平、曾我慶介、吉場聡子、近藤一成：ゲノム編集食品における外来性遺伝子の安全性評価における全ゲノムシーケンスデータを用いた解析の標準化に向けた課題、第118回日本食品衛生学会学術講演会、長崎、2022年11月10日-11日

E. 知的財産権の出願・登録状況

該当なし

図1 ゲノム編集による意図しない変化

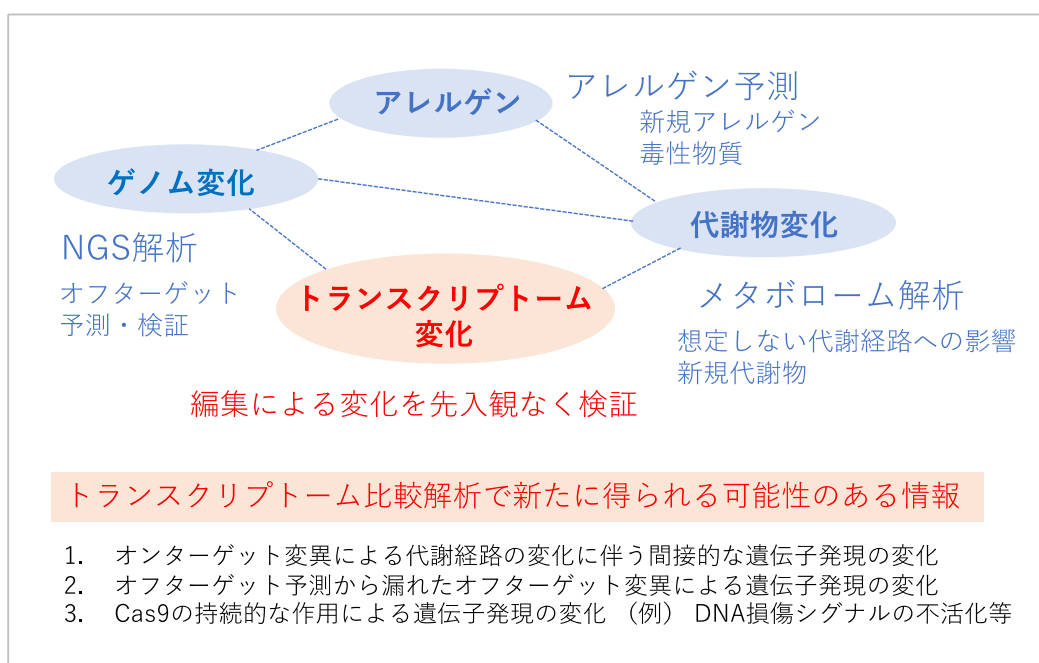


図2 遺伝子発現解析スキーム

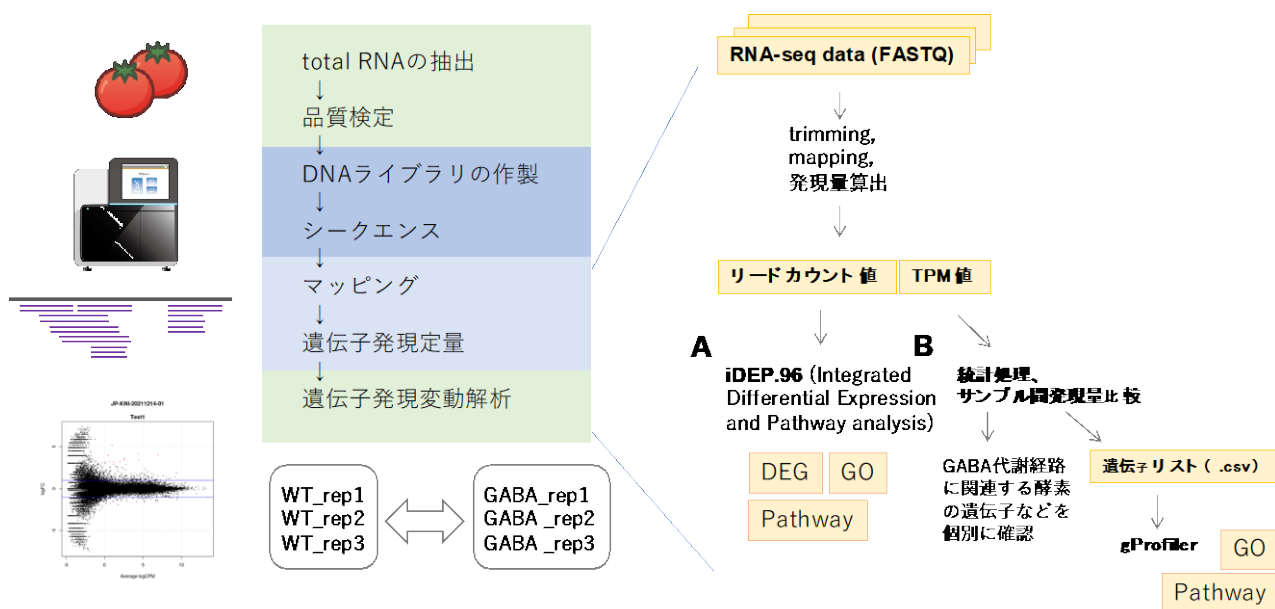


図3 iDEP workflow
(<http://bioinformatics.sdstate.edu/idep96/>)

iDEP: Integrated Differential Expression and Pathway analysis

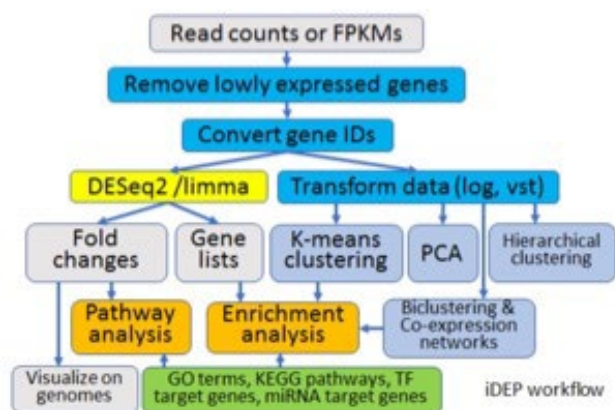


図4 total RNA品質検定結果

品質検定結果

サンプル ID	残量 (ul)	Agilent TapeStation		
		濃度 (ng/ul)	総量 (ng)	RIN [®]
PR3188_01.a	29.0	22.9	664.1	6.5
PR3188_02.a	28.0	13.0	364.0	5.8
PR3188_03.a	28.0	16.5	460.6	7.5
PR3188_04.a	28.0	18.3	512.4	6.3
PR3188_05.a	28.0	16.1	450.8	7.5
PR3188_06.a	28.0	16.0	446.6	6.2

サンプル情報

サンプル ID	依頼書記載名称
PR3188_01.a	wt1
PR3188_02.a	mt1
PR3188_03.a	wt2
PR3188_04.a	mt2
PR3188_05.a	wt3
PR3188_06.a	mt3

*各数値は小数点以下 1 桁で表記しています。未検出の場合は N.D. と表記しています。

Agilent TapeStation 分析結果

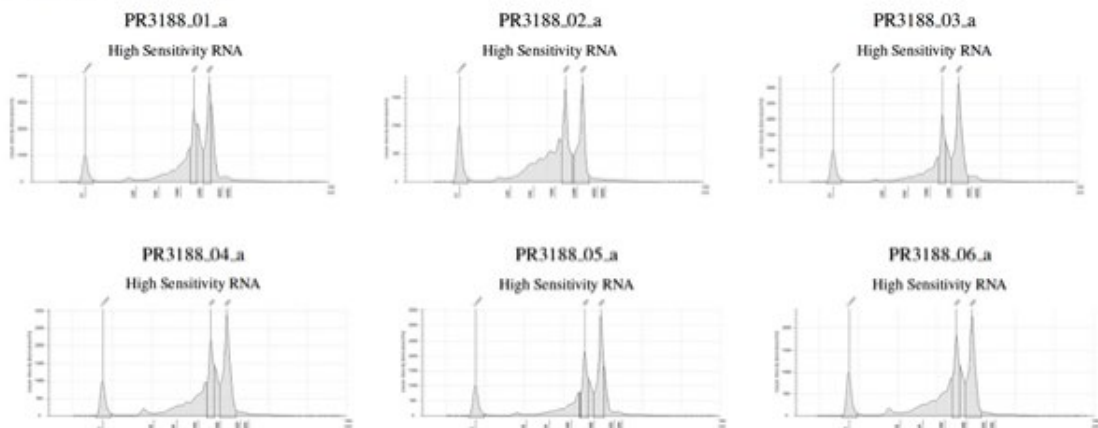


図5 発現変動遺伝子のプロット (GABA/WT)

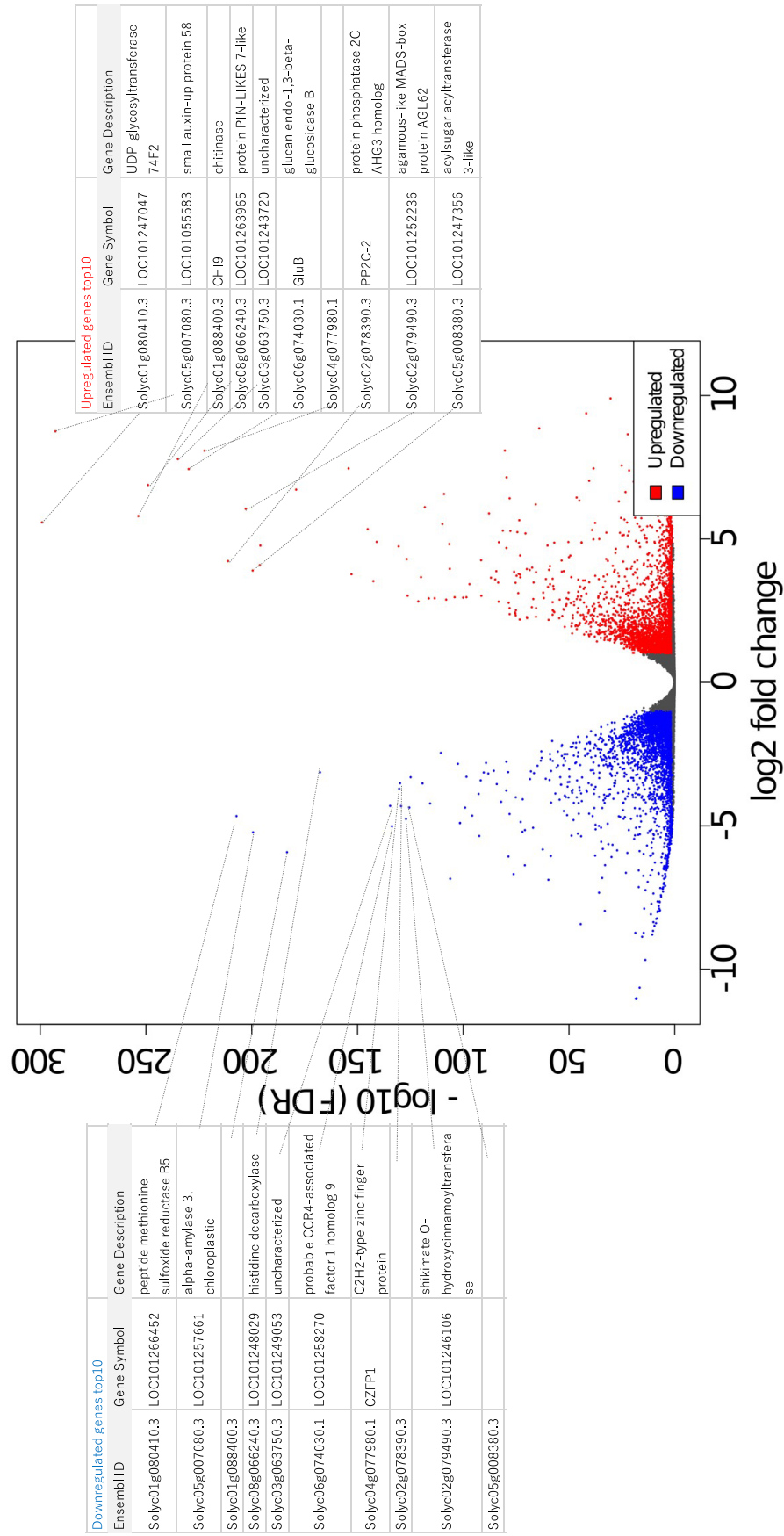
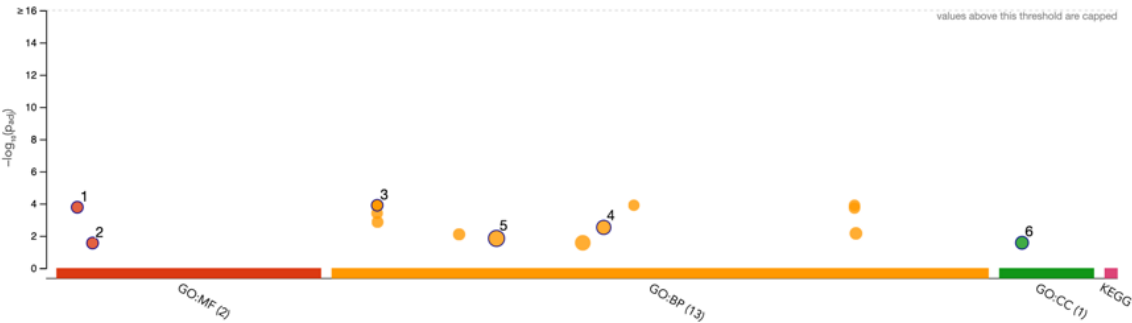


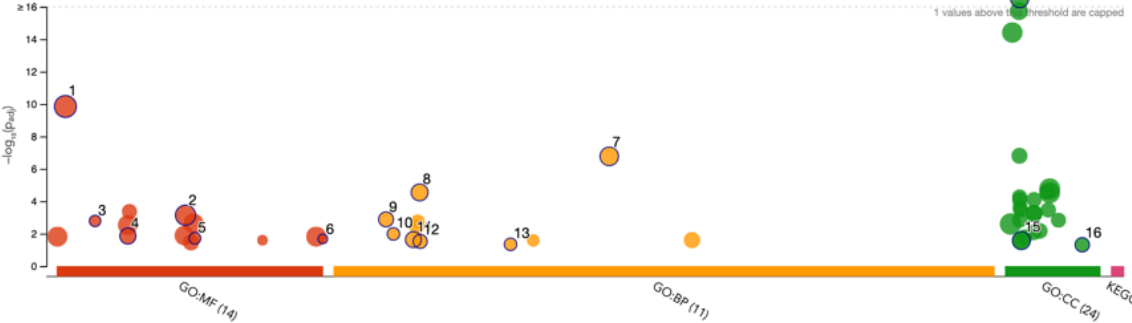
図6 gProfilerによるGO解析結果

Upregulated



ID	Source	Term ID	Term Name	Padj (query_1)
1	GO:MF	GO:0004568	chitinase activity	1.672×10 ⁻⁴
2	GO:MF	GO:0008061	chitin binding	2.795×10 ⁻²
3	GO:BP	GO:0006026	aminoglycan catabolic process	1.260×10 ⁻⁴
4	GO:BP	GO:0044036	cell wall macromolecule metabolic process	2.970×10 ⁻³
5	GO:BP	GO:0022613	ribonucleoprotein complex biogenesis	1.443×10 ⁻²
6	GO:CC	GO:0030684	preribosome	2.678×10 ⁻²

Downregulated



ID	Source	Term ID	Term Name	Padj (query_1)
1	GO:MF	GO:0003824	catalytic activity	1.413×10 ⁻¹⁰
2	GO:MF	GO:0036094	small molecule binding	7.243×10 ⁻⁴
3	GO:MF	GO:0008235	metalloexopeptidase activity	1.633×10 ⁻³
4	GO:MF	GO:0016791	phosphatase activity	1.397×10 ⁻²
5	GO:MF	GO:0043621	protein self-association	1.923×10 ⁻²
6	GO:MF	GO:2001070	starch binding	2.095×10 ⁻²
7	GO:BP	GO:0044281	small molecule metabolic process	1.689×10 ⁻⁷
8	GO:BP	GO:0009628	response to abiotic stimulus	2.834×10 ⁻⁵
9	GO:BP	GO:0006399	tRNA metabolic process	1.306×10 ⁻³
10	GO:BP	GO:0006779	porphyrin-containing compound biosynthetic p...	1.036×10 ⁻²
11	GO:BP	GO:0009117	nucleotide metabolic process	2.318×10 ⁻²
12	GO:BP	GO:0009657	plastid organization	2.995×10 ⁻²
13	GO:BP	GO:0031146	SCF-dependent proteasomal ubiquitin-depend...	4.519×10 ⁻²
14	GO:CC	GO:0009536	plastid	5.904×10 ⁻¹⁸
15	GO:CC	GO:0012505	endomembrane system	2.693×10 ⁻²
16	GO:CC	GO:0098791	Golgi apparatus subcompartment	4.825×10 ⁻²

図7 SIGAD3 read mapping

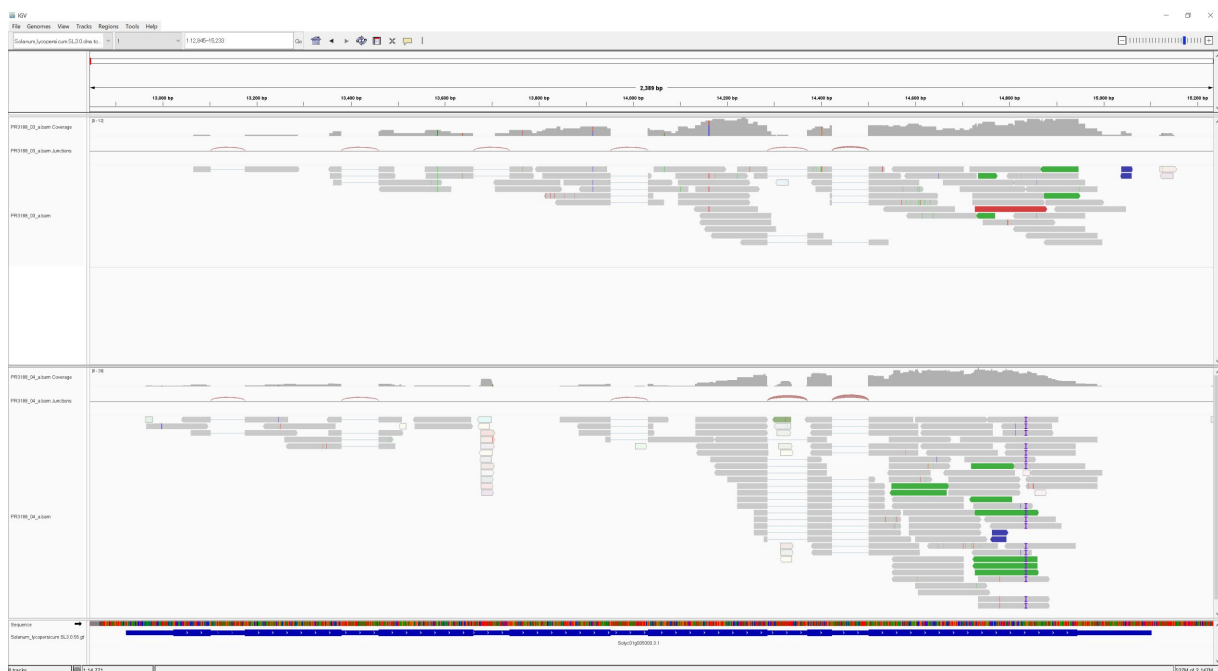


図8 高等動物におけるGABA代謝経路（参考文献2より抜粋）

GABAトマトは、ゲノム編集によりグルタミン酸脱炭酸酵素(GAD)遺伝子の自己阻害領域の除去を行うことでGADの活性を上昇させGABA蓄積量を向上させている。

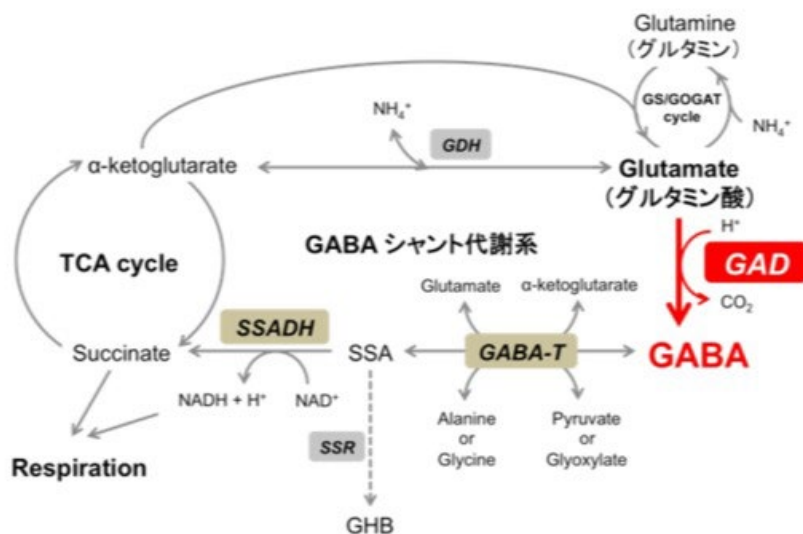


表1 マッピング結果

Sample ID	Total input reads	Mapped reads	Mapped reads (%)	Mapped reads R1	Mapped reads R1 (%)	Mapped reads R2	Mapped reads R2 (%)	Unmapped reads	Unmapped reads (%)	Singleton reads (itself mapped; mate unmapped)	Singleton reads (itself mapped; mate unmapped) (%)
wt1	46082366	45903657	99.61	22960132	99.65	22943525	99.58	178709	0.39	45361	0.1
mt1	47732748	47398604	99.3	23705867	99.33	23692737	99.27	334144	0.7	65088	0.14
wt2	45115786	44944938	99.62	22478825	99.65	22466113	99.59	170848	0.38	36536	0.08
mt2	96484460	95702762	99.19	47862501	99.21	47840261	99.17	781698	0.81	138076	0.14
wt3	48528112	48334153	99.6	24173643	99.63	24160510	99.57	193959	0.4	32997	0.07
mt3	52680576	52263148	99.21	26141942	99.25	26121206	99.17	417428	0.79	75776	0.14

Paired reads (itself & mate mapped)	Paired reads (itself & mate mapped) (%)	Properly paired reads	Properly paired reads (%)	Not properly paired reads (discordant)	Not properly paired reads (discordant) (%)	Reads with MAPQ [10:inf)	Reads with MAPQ [10:inf) (%)	Covered bases (%)	Total alignments	Insert length: mean	Insert length: median	Insert length: standard deviation
45858296	99.51	44992084	97.63	866212	1.88	42790760	92.86	6.77	51275540	192	192	81.66
47333516	99.16	46463430	97.34	870086	1.82	42427878	88.89	6.53	57886941	173	173	74.83
44908402	99.54	44228140	98.03	680262	1.51	42414296	94.01	7.51	49890905	205	205	86.42
95564686	99.05	93782610	97.2	1782076	1.85	86095587	89.23	8.14	115471297	173	173	71.04
48301156	99.53	47775146	98.45	526010	1.08	45878385	94.54	6.96	54556178	158	158	62.16
52187372	99.06	51340186	97.46	847186	1.61	47345311	89.87	7.4	62631102	136	136	51.92

表2 iDEPによるGO解析結果

GO: Biological Process

Enriched pathways in DEGs for the selected comparison:

Direction	adj.Pval	nGenes	Pathways
Down regulated	1.6e-06	178	Carbohydrate metabolic process
	8.6e-03	46	Carbohydrate catabolic process
Up regulated	2.0e-09	94	Ribonucleoprotein complex biogenesis
	7.2e-08	166	Peptide metabolic process
	1.4e-07	75	Ribosome biogenesis
	3.6e-07	149	Translation
	4.4e-07	149	Peptide biosynthetic process
	1.1e-06	175	Cellular amide metabolic process
	2.5e-06	153	Amide biosynthetic process
	3.7e-06	44	Ribonucleoprotein complex assembly
	4.6e-06	44	Ribonucleoprotein complex subunit organization
	5.7e-06	33	Ribosomal large subunit biogenesis
	1.6e-05	166	Cellular component biogenesis
	3.2e-04	26	Ribosome assembly
	4.4e-04	23	Cytoplasmic translation
	6.2e-04	104	Protein-containing complex subunit organization
	1.5e-03	9	Aminoglycan catabolic process

GO: Molecular Function

Enriched pathways in DEGs for the selected comparison:

Direction	adj.Pval	nGenes	Pathways
Down regulated	2.8e-03	69	Hydrolase activity, hydrolyzing O-glycosyl compounds
	2.8e-03	49	Vitamin binding
	2.8e-03	45	Carbohydrate binding
	2.8e-03	524	Small molecule binding
	2.8e-03	37	ADP binding
	3.4e-03	74	Hydrolase activity, acting on glycosyl bonds
	6.6e-03	231	Kinase activity
	6.6e-03	491	Anion binding
	6.6e-03	34	Pyridoxal phosphate binding
	6.6e-03	34	Vitamin B6 binding
Up regulated	7.9e-03	264	Transferase activity, transferring phosphorus-containing groups
	7.9e-03	215	Phosphotransferase activity, alcohol group as acceptor
	5.3e-13	124	Structural constituent of ribosome
	9.5e-12	138	Structural molecule activity
	2.0e-04	11	Chitin binding
	7.7e-03	9	Chitinase activity

GO: Cellular Component

Enriched pathways in DEGs for the selected comparison:

Direction	adj.Pval	nGenes	Pathways
Down regulated	1.0e-06	218	Chloroplast
	1.0e-06	223	Plastid
	1.4e-03	61	Photosynthetic membrane
	1.4e-03	54	Thylakoid membrane
	2.4e-03	60	Plastid stroma
	2.4e-03	54	Chloroplast thylakoid
	2.4e-03	57	Chloroplast stroma
	2.4e-03	54	Plastid thylakoid
	3.2e-03	71	Thylakoid
	3.6e-03	47	Chloroplast thylakoid membrane
	3.6e-03	57	Plastid membrane
	3.6e-03	47	Plastid thylakoid membrane
	5.4e-03	77	Plastid envelope
Up regulated	2.1e-19	83	Cytosolic ribosome
	9.4e-17	92	Ribosomal subunit
	2.8e-15	56	Cytosolic large ribosomal subunit
	2.8e-15	140	Ribonucleoprotein complex
	3.7e-15	129	Ribosome
	1.8e-13	62	Large ribosomal subunit
	5.8e-12	228	Non-membrane-bounded organelle
	5.8e-12	228	Intracellular non-membrane-bounded organelle
	1.5e-05	57	Extracellular region
	1.5e-05	27	Cytosolic small ribosomal subunit
	2.9e-05	167	Cytosol
	8.1e-05	21	Preribosome
	3.2e-04	30	Small ribosomal subunit
	2.0e-03	22	DNA packaging complex
	6.3e-03	12	Mitochondrial outer membrane

GO: KEGG

Enriched pathways in DEGs for the selected comparison:

Direction	adj.Pval	nGenes	Pathways
Down regulated	8.4e-03	23	Starch and sucrose metabolism
	1.0e-02	14	Photosynthesis
Up regulated	2.2e-15	78	Ribosome

表3 発現変動遺伝子リスト (一部)

	A	B	C	D	E	F	G	N	O	P	Q	R	S	T
1	TPM値							log2TPM_FoldChange		Variance				
2	Gene ID	wt_rep1	gaba_rep1	wt_rep2	gaba_rep2	wt_rep3	gaba_rep3	FoldChang	ABS	wt	gaba	t-test	p-value	t-test
								e				p-v	rank	q-v
3	ENSRNA050030304	0	113.437	0	64.827	0	110.85	16.511379	16.51138	0	0.208613	0.0003	505	0.0095
4	Solyc04g012140.1	29.806	0	61.834	0	88.26	0	-15.73633	15.73633	0.637435	0	0.0009	1116	0.0095
5	Solyc04g081655.1	30.867	0	66.012	0	61.593	0	-15.6116	15.6116	0.367662	0	0.0005	792	0.0095
6	Solyc01g091590.3	65.582	0	106.792	0.058	92.052	0	-14.43768	14.43768	0.129997	11.53516	0.0171	5938	0.0095
7	Solyc11g032190.1	13.351	0	58.277	0	13.431	0	-14.41627	14.41627	1.500395	0	0.0024	2198	0.0095
8	Solyc02g088595.1	22.325	0	16.301	0	19.186	0	-14.22235	14.22235	0.051478	0	8E-05	214	0.0095
9	Solyc09g061793.1	16.087	0	19.492	0	21.593	0	-14.20757	14.20757	0.046474	0	8E-05	201	0.0095
10	ENSRNA050029545	0	13.222	0	8.974	0	41.72	14.056981	14.05698	0	1.329127	0.0022	2108	0.0099
11	Solyc05g006145.1	14.205	0	19.523	0	15.61	0	-13.99248	13.99248	0.055515	0	9E-05	236	0.0099
12	Solyc09g082480.2	0	7.577	0	18.362	0	26.041	13.906889	13.90689	0	0.842725	0.0014	1590	0.0101
13	ENSRNA050030085	0	9.562	0	8.621	0	42.789	13.894013	13.89401	0	1.672833	0.0029	2424	0.0101
14	ENSRNA050029941	0	20.187	0	12.66	0	9.995	13.738816	13.73882	0	0.26628	0.0005	747	0.0101
15	Solyc09g066400.2	0	20.6	0	10.08	0	8.804	13.577962	13.57796	0	0.434183	0.0008	1051	0.0101
16	ENSRNA050029543	0	19.832	0	2.991	0	29.8	13.561861	13.56186	0	3.131344	0.0056	3486	0.0101
18	Solyc03g019690.1	0.123	44.369	0	59.489	0	63.101	13.429619	13.42962	16.12028	0.074063	0.0281	7272	0.0101
19	Solyc02g087215.1	0	6.68	0	14.224	0	12.47	13.369426	13.36943	0	0.339263	0.0006	918	0.0101
20	Solyc01g005290.3	15.934	0	8.645	0	7.785	0	-13.32147	13.32147	0.311422	0	0.0006	876	0.0101
21	Solyc03g093550.1	43.78	0	35.463	0.047	32.017	0	-13.30457	13.30457	0.052986	10.39727	0.0186	6170	0.0101
22	ENSRNA050029561	0	8.13	0	3.673	0	29.23	13.222522	13.22252	0	2.278952	0.0043	3042	0.0101
23	Solyc01g095500.3	0	4.341	0	12.609	0	15.586	13.211493	13.21149	0	0.976562	0.0019	1849	0.0101
24	Solyc06g065890.1	0	9.562	0	4.31	0	17.116	13.120066	13.12007	0	0.997349	0.0019	1900	0.0101
25	Solyc05g054650.1	0	3.712	0	10.767	0	13.585	12.994222	12.99422	0	0.995739	0.002	1922	0.0101
27	Solyc10g076380.2	0	10.928	0	4.58	0	9.523	12.931558	12.93156	0	0.454656	0.0009	1159	0.0101
28	Solyc02g085510.1	9.933	0	14.384	0	3.209	0	-12.91293	12.91293	1.270681	0	0.0025	2270	0.0101
29	Solyc01g011093.1	6.321	0	8.597	0	7.878	0	-12.87992	12.87992	0.05226	0	0.0001	258	0.0101
31	Solyc09g060160.2	8.164	0	5.108	0	8.346	0	-12.78037	12.78037	0.160019	0	0.0003	596	0.0104
32	Solyc01g095995.1	6.774	0	5.474	0	7.529	0	-12.67436	12.67436	0.054852	0	0.0001	272	0.0104
33	Solyc06g083130.3	0	10.081	0	4.796	0	5.495	12.650526	12.65053	0	0.325499	0.0007	961	0.0105
34	Solyc09g065770.1	0	5.194	0	7.417	0	5.667	12.556116	12.55612	0	0.071766	0.0002	337	0.0105
35	Solyc03g117130.3	0	6.425	0	4.937	0	6.781	12.54897	12.54897	0	0.060002	0.0001	299	0.0105
36	Solyc10g050510.1	0	4.076	0	5.527	0	7.339	12.422466	12.42247	0	0.179968	0.0004	669	0.0144
37	Solyc08g067640.1	0	9.982	0	3.181	0	5.013	12.404241	12.40424	0	0.689736	0.0015	1615	0.0195
38	Solyc01g010310.3	0	5.778	0	7.986	0	3.267	12.378078	12.37808	0	0.426075	0.0009	1175	0.0195
41	Solyc09g014390.1	0	8.224	0	3.117	0	4.949	12.295139	12.29514	0	0.489937	0.0011	1306	0.0113
42	ENSRNA050030331	0	9.25	0	2.426	0	5.149	12.250236	12.25024	0	0.936459	0.0021	2008	0.0113
44	Solyc02g062770.2	0	10.079	0	2.512	0	3.101	12.064462	12.06446	0	1.16652	0.0027	2333	0.0113
45	Solyc10g080840.1	15.103	0	19.94	0	18.625	0.082	-11.99203	11.99203	0.043628	13.54704	0.0297	7440	0.0113
46	Solyc06g062520.1	6.832	0	3.137	0	3.004	0	-11.96903	11.96903	0.444815	0	0.001	1274	0.0113
47	Solyc06g074750.1	6.444	0	3.088	0	3.077	0	-11.94489	11.94489	0.377098	0	0.0009	1138	0.0113
49	Solyc07g054550.1	0	4.475	0	8.573	0	1.464	11.903477	11.90348	0	1.662243	0.0039	2860	0.0113
50	Solyc07g038137.1	2.915	0	3.554	0	5.126	0	-11.87644	11.87644	0.170596	0	0.0004	684	0.0113
51	Solyc09g057520.2	2.42	0	3.727	0	4.326	0	-11.72824	11.72824	0.189323	0	0.0005	727	0.0113
52	Solyc02g077570.2	0	5.8	0	2.576	0	2.518	11.710733	11.71073	0	0.46995	0.0011	1355	0.0113
53	Solyc06g072193.1	2.369	0	3.985	0	3.862	0	-11.69563	11.69563	0.176903	0	0.0004	705	0.0113
54	Solyc03g098125.1	3.95	0	3.177	0	2.713	0	-11.6627	11.6627	0.074008	0	0.0002	397	0.0113
55	Solyc03g020050.3	0	1.907	0	3.435	0	5.195	11.662515	11.66252	0	0.527544	0.0013	1469	0.0218
57	Solyc03g095860.3	8.897	0	8.978	0.027	10.5	0	-11.60083	11.60083	0.018052	7.703554	0.0184	6137	0.0176
58	Solyc05g014830.2	0	1.896	0	3.599	0	4.228	11.583129	11.58313	0	0.37433	0.0009	1180	0.0176
59	Solyc07g051990.2	2.287	0	2.453	0	4.832	0	-11.55316	11.55316	0.355025	0	0.0009	1142	0.0176
60	Solyc03g020035.1	3.623	0	1.958	0	3.758	0	-11.54513	11.54513	0.279069	0	0.0007	982	0.0176
61	Solyc08g062515.1	7.881	0	10.681	0.025	7.918	0	-11.52597	11.52597	0.063136	7.364711	0.0173	5973	0.0344
62	Solyc06g061110.2	0	7.746	0	1.448	0	2.078	11.480646	11.48065	0	1.620374	0.0041	2941	0.0344
65	Solyc08g082490.2	2.148	0	3.193	0	2.755	0	-11.37965	11.37965	0.083489	0	0.0002	448	0.0076
66	Solyc09g018550.1	2.71	0	2.842	0	2.433	0	-11.37564	11.37564	0.013183	0	3E-05	102	0.0076
67	Solyc07g006670.1	0	6.394	0	1.508	0	1.83	11.346841	11.34684	0	1.27896	0.0033	2618	0.0076
68	Solyc03g095870.2	2.627	0	2.348	0	2.787	0	-11.3342	11.3342	0.015765	0	4E-05	116	0.0076
69	Solyc01g009700.2	1.837	0	3.281	0	2.778	0	-11.32154	11.32154	0.185494	0	0.0005	762	0.0076
70	Solyc05g010367.1	2.563	0	1.63	0	3.775	0	-11.29279	11.29279	0.367431	0	0.001	1209	0.0076
72	Solyc09g072655.1	2.469	0	2.675	0	2.202	0	-11.2538	11.2538	0.019889	0	5E-05	140	0.0076
73	ENSRNA050030350	0	4.18	0	1.867	0	1.844	11.248772	11.24877	0	0.457387	0.0012	1396	0.0076
74	Solyc03g063460.1	0	1.895	0	0.847	0	8.372	11.216064	11.21606	0	2.808811	0.0074	3986	0.0076
76	Solyc08g067360.3	0	0.777	0	3.689	0	3.993	11.138883	11.13888	0	1.771022	0.0047	3189	0.0076
77	ENSRNA050029620	0	2.357	0	2.095	0	2.123	11.096446	11.09645	0	0.008663	2E-05	77	0.0076

表 4 GAD遺伝子の発現変動

EnsemblID	GeneID	Symbol	Aliases	description	TPM値					log2TPM_FoldChange		Variance		t-test p-value	p-value rank	t-test q-value
					wt_rep1	gaba_rep1	wt_rep2	gaba_rep2	wt_rep3	gaba_rep3	FoldChange	ABS	wt			
Solyc01g005000.3	100147723	GAD3	LeGAD3, SIGAD3	glutamate decarboxylase isoform3	0.333	0.334	1.079	0.722	1.129	0	-3.572526	3.5725261	0.99381152	26.966626	0.3554154	16660
Solyc05g054050.3	101251185	LOC101251185	SIGAD4	glutamate decarboxylase-like	0	0	0	0	0	0.069	2.04309434	2.04309434	0	12.5227034	0.4226497	19864
Solyc11g011920.2	100147722	GAD2	LeGAD2	glutamate decarboxylase isoform2	3.116	0.205	6.311	2.09	4.221	2.448	-2.09969333	2.09969329	0.26073984	3.99771095	0.205406	14584
Solyc03g098240.3	544313	LOC544313	GAD, LeGAD1	glutamate decarboxylase	191.211	65.349	222.49	48.684	220.927	70.643	-1.7953497	1.79534966	0.01521832	0.08024551	0.0030812	2523
Solyc04g025530.3	101259235	LOC101259235		glutamate decarboxylase 4	7.653	0	7.389	0.738	6.225	0.671	-6.4785662	6.47856624	0.02540926	29.8407455	0.1762243	13168

表5 GABA代謝系に関わる遺伝子の発現変動

EnsemblID	GeneID	Symbol	description	TPM値						log2TPM_FoldChange		Variance		t-test p-value	p-value rank	t-test q-value
				wt_rep1	gaba_rep1	wt_rep2	gaba_rep2	wt_rep3	gaba_rep3	FoldChange	ABS	wt	gaba			
Solyc09g090700.1	100147726	SSADH	Succinic semialdehyde dehydrogenase	84.863	42.671	99.919	30.189	94.095	44.535	-1.26590903	1.26590903	0.01420315	0.09460575	0.01084191	4873	0.03642051
Solyc03g094010.3	101246116	LOC101246116	glutamate dehydrogenase	2.619	0.977	2.167	2.031	2.89	2.813	-0.51802369	0.51802369	0.04453302	0.60973497	0.3701068	16825	
Solyc12g006470.2	543874	GABA-TP2	gamma aminobutyrate transaminase 2	0.302	0.043	0	0.019	0	0	0.512728577	0.51272858	22.6499724	8.29639505	0.88246467	23630	
Solyc12g006450.2	544258	GABA-TP3	gamma-aminobutyrate transaminase subunit precursor isozyne 3	21.02	1	40.991	0.866	41.31	1.651	-4.8665748	4.8665748	0.31308473	0.23841537	0.00037936	661	0.01423556
Solyc07g043310.3	543873	GABA-TP1	gamma-aminobutyrate transaminase subunit precursor isozyne 1	70.52	307.092	93.803	226.602	85.441	284.238	1.709694533	1.70969453	0.04403772	0.05193769	0.00069136	975	0.01758812

厚生労働科学研究費補助金（食品の安全確保推進研究事業）
「新たなバイオテクノロジーを用いて得られた食品の安全性確保と
リスクコミュニケーションのための研究」
分担研究報告書（令和4年度）

メタボロームインフォマティクスによる未知化合物推定

研究分担者 早川 英介 沖縄科学技術大学院大学

研究要旨：

本研究はゲノム編集等の新しい技術で作られた食物に含まれる未知化合物の迅速な検出・解析を可能にする解析手法の開発および一般利用可能なツールとしての実装である。本年度はゲノム編集トマトと多品種トマトとのノンターゲットでの網羅的な比較定量解析および多変量解析に関して、本解析技術の適用を行うことで化合物群の変動を明らかにした。本解析技術に関してデータ解析ツールを公開可能な Web アプリとしてサーバー上に構築し、各種解析機能を実装した。本研究により開発されたツールは食品分析に従事する研究者・分析者に広く利用される解析プラットフォームとなり得る。

A. 研究目的

近年では CRISPR 等の技術によるゲノム編集などの新しい技術を用いて食品の品種改良が行われている。一方でこれまで一般ではなかった技術により作られた食品（生物）に想定外の変化が起きる可能性とそれに伴う食品衛生上のリスクの懸念も高まっている。生物に含まれる化合物・代謝物は生体内の複雑な代謝パスウェイの連鎖により生成されるため、僅かの代謝パスウェイの変化が想定外の化合物・代謝物の質的・量的変化につながることも考えられる。そのような想定外の化学的なリスクを迅速に検出・解析する技術が今後重要になることは明らかである。

食品分析において広く用いられる質量分析のアプローチはあらかじめ分析対象の化合物を決めておく“ターゲット分析”が主である。しかし、ゲノム編集等で危惧される想定外の未知化合物はその“ターゲット”の外であり検出が不可能である。一方事前にターゲットを設定しない“ノンターゲット分析”のアプローチでは未知化合物に関して構

造情報を得る解析技術が整っておらず未だ困難である。そこで本研究ではゲノム編集食品を含む食品のノンターゲット質量分析データから未知化合物・代謝物を網羅的に迅速に検出および構造解析するデータ解析プラットフォームの開発を行った。

B. 研究方法

本研究では一般にも広く用いられている液体クロマトグラフィー連結型質量分析（LC-MS）によるノンターゲット分析を機器分析の手法として採用し、得られたデータを解析するための技術開発・アプリケーション・解析システムの構築を行っている。本年度は昨年度までにすでに構築したフラグメントスペクトル類似度による未知化合物解析を行う解析システム、スペクトルライブラリを用いて、ゲノム編集トマトの網羅的な多品種間の比較定量解析および多変量解析を行い、それを基にした変動化合物の特定及び構造解析を実施した。

データ解析システムの構築に関しては昨年度ま

でローカルなツールとして実装されていたものを一般公開可能な Web アプリとして発展させるための開発を行うとともに、広範囲なユーザーを想定した各種機能とその GUI の開発及び実装を行った。

C. 研究結果および考察

昨年までに実装していたデータ解析機能およびスペクトルライブラリのアプリケーションとしてゲノム編集トマト（GABA シェント代謝系 GAD を編集）とその他非ゲノム編集トマトの多品種での比較解析を行った。現在市販されているゲノム編集トマトに加え9種の非ゲノム編集トマトの計10種をLC-MS分析を行い、ノンターゲット分析データを取得した。10種のデータに関して検出された全未知化合物ピークを比較定量解析および主成分解析を行った。その結果、ゲノム編集トマトは主成分プロットにおいて非ゲノム編集トマト群とは主成分1で明確に区別され（図1）、またゲノム編集トマトで増加しているとされているGABA以外の多数の化合物が量的に変動し主成分に寄与していることが分かった。

特に大きく変動し主成分1に寄与している化合物に関してスペクトル類似度計算に基づくスペクトルライブラリとのネットワーク解析を行ったところ Carboxylic acids and derivatives, organooxygen compounds, flavonoids に属する化合物であることが推定された。この結果から、ゲノム編集食品における想定外（編集対象の代謝パスウェイ外）の化合物・代謝物に量的な変動が生じていることが強く示唆された。

このデータ解析技術に関して、昨年までローカルで稼働していたプログラムを Web アプリケーションとしてオンライン公開可能なシステムとしての開発が行われた。前述のような多試料分析の多変量解析の結果を可視化ツールに取り入れたことに加え、化合物クラス・部分構造・ネットワーク解析といった機能も Web アプリ上に実装された。

さらに質量範囲の指定、化合物クラスの指定、化合物構造のポップアップでの表示などのユーザビリティが向上する各種機能が実装された。

D. 結論

今年度は昨年度までに開発されたデータ解析システムおよびスペクトルライブラリを実際のゲノム編集食品を含む多試料比較解析への適用を行うことで、多変量解析への対応とさらにその結果を反映した未知化合物解析ワークフローの確立が行われた。

すでに述べたように、ゲノム編集トマトでは他品種との比較で複数の未知化合物が量的に変動しており、その未知化合物の構造（化合物クラス）推定を本解析手法で迅速に行うことが可能であることが実証された。一方で、より厳密な化学構造を提示する機能に関しては今後の課題であり、すでに実装したプログラムを解析システムに組み込むことで来年度に実現可能である。

この解析システムは広く一般の食品分析に関わる研究者・分析者に利用されるプラットフォームとして成立することを目指しており、今年度に整えた Web アプリとしての実装（図2）を基盤としつつ、今後ユーザーのフィードバックを取り入れながら拡張をしていく予定である。

E. 研究発表・業績

1. 論文発表

無し

2. 学会発表

早川英介、渡邊 寛、近藤一成：質量分析インフォマティクスによる 食品試料の迅速・ノンターゲットな未知化合物解析プラットフォーム、日本食品衛生学会第118回学術講演会、長崎、令和4年11月10日－11日

F. 健康危険情報

該当なし

G. 知的財産権の出願・登録状況

該当なし

厚生労働科学研究費補助金（食品の安全性確保推進研究事業）
「新たなバイオテクノロジーを用いて得られた食品の安全性確保と
リスクコミュニケーションのための研究」
分担研究報告書（令和4年度）

新規アレルギー性予測手法開発のための基盤的研究

研究分担者 安達 玲子 国立医薬品食品衛生研究所生化学部 室長

研究要旨：

本研究では、遺伝子改変技術応用食品のアレルギー性について、より高い精度での評価・予測を可能とすることを目的として、アレルギー性予測手法（allerStat）の機能拡充を行った。分類器として、これまで使用してきた線形 SVM に加えて、非線形 SVM、ロジスティック回帰、lightGBM、LSTM、proteinBERT を実装し、その性能を検証した。今後、バイオテクノロジー技術応用食品のリスク評価手法としての実用化を目指す。また、国立医薬品食品衛生研究所にて運用・公開しているアレルギーデータベース（Allergen Database for Food Safety, ADFS）に関して、令和3年6月から令和4年5月までの1年間に NCBI PubMed に掲載された論文から、エピトープ配列決定に関する18報のピアレビューを行い、14種のアレルギーについて、総数52のエピトープ情報を ADFS に追加した。また HESI が運営するアレルギーデータベース COMPARE の登録アレルギーに関するアップデートを ADFS に反映させた。これらの情報更新により、ADFS のアレルギー及びイソアレルギーのアミノ酸配列情報は2,403、エピトープ既知のアレルギー数は274となり、遺伝子改変技術応用食品のアレルギー性評価に有用なデータベースである ADFS を充実させることができた。また、医療・食品分野やリスク評価分野等における AI の活用実態の調査を実施した。現状では医療分野への活用や規制の取り組みが多いことが示され、食品分野やリスク評価分野への活用は今後進んでいくものと考えられた。

研究協力者

爲廣 紀正 国立医薬品食品衛生研究所生化学部
協力研究員

を検出または予測し、その変化が与える影響を正確に評価することは、食品の安全性確保において急務の課題である。

A. 研究目的

遺伝子改変技術を応用した食品開発は、技術的には、外来遺伝子導入による遺伝子組換え食品から、内在性遺伝子の改変を行うゲノム編集技術応用食品へ、また、酵母等に多数の外来遺伝子を導入し新規食品機能成分を産生させる合成生物学の利用へと変化している。現在、ゲノム編集技術では多様な手法が生み出されており、これらの手法による意図しない塩基変化も一様ではないことが明らかになりつつある。従って、このような意図しない変化、及びそこから生じる代謝成分の変化

バイオテクノロジー技術を用いて開発された食品のリスクの1つに、アレルギー性増大の可能性がある。本研究では、国立医薬品食品衛生研究所生化学部にて管理・公開している、アレルギー性予測機能（FAO/WHO 法等）を装備したアレルギー・エピトープ情報データベース（Allergen Database for Food Safety, ADFS）について、新規アレルギー及びエピトープ情報の収集・解析等によりアレルギー性評価に関する検討を行い、遺伝子改変技術応用食品のリスク評価に資するデータベースとなるよう、情報を更新し内容を充実させる。

また、人工知能（AI）を活用した新規高精度アレルギー性予測手法の開発を進める。令和2年度

までの先行研究班では、アレルゲン及び非アレルゲンタンパク質から抽出した特徴的なアミノ酸配列パターンを利用して機械学習によりアレルゲン性を予測する手法 (allerStat) を開発してきた。本研究班では、この予測システムの機能を拡充し、高精度アレルゲン性予測法としての実用化を進める。

また、医療・食品分野やリスク評価分野等における AI の活用実態を調査し、課題を整理する。

B. 研究方法

アレルゲン性予測手法-1 (allerStat) の機能拡充

現システムの分類器である線形 SVM (Support Vector Machine) に加えて、非線形 SVM、ロジスティック回帰、LightGBM (Gradient Boosting Machine)、LSTM (Long short-term memory)、proteinBERT (Bidirectional Encoder Representations from Transformers) を分類器として追加実装し、それぞれの分類器を用いた場合の予測性について検証した。

ADFS 登録アレルゲンのアップデート

昨年度までは米国ネブラスカ大学リンカーン校が運営しているアレルゲンデータベース (AllergenOnline) における登録アレルゲンのアップデート内容を ADFS に反映させていたが、2021 年 2 月を最後にデータ更新が行われなくなっている。そのため今年度は、HESI (the Health and Environmental Sciences Institute、国際生命科学協会 ILSI のグローバルブランチ、非営利組織) が運営しているアレルゲンデータベースである COMPARE (Comprehensive Protein Allergen Resource) における登録アレルゲンのアップデート内容を、ADFS に反映させた。

ADFS エピトープ情報の追加

令和 3 年 6 月から令和 4 年 5 月までの 1 年間に NCBI PubMed に収載された論文から、キーワード検索により、エピトープ決定に関するものを抽

出した。キーワードとしては、IgE、epitope、linear、conformational、sequence、recognition 等々のワードを使用し、これらを複数組み合わせ 6 通りの検索式を作成して検索を行った。この検索により抽出されてきた論文についてピアレビューを行った。その結果エピトープ情報を報告していると判断された論文について、そのエピトープ情報を整理し、ADFS のデータに追加した。

AI のリスク評価分野等への応用に関する調査

Web of Science にて、AI (artificial intelligence)、machine learning、deep learning、neural network のキーワードを用いて、2020 年以降に出版された総説を検索し、ヒットした 14,626 報の中から、引用数及び内容を考慮して食品・農業分野、医療分野、生物学分野の合計 168 報を入手した。入手した総説からさらに 48 報 (食品分野 18 報、医療分野 20 報、生物学分野 10 報) を選択し、その内容を精査した。

続いて、精査した総説において出現頻度及び重要度の高い機械学習・深層学習関連の 15 種の用語 (Neural network、Machine learning、artificial intelligence、Random forest、Support Vector Machine、Deep Learning、Logistic regression、Ensemble、Principal component analysis、Boosting、k-nearest neighbor、Decision tree、Naive Bayes、Long Short Term Memory、Natural language processing) のいずれかを含み、かつ “risk assessment” または “safety assessment” をキーワードとして含む 2020 年以降に出版された論文を、Web of Science にて検索したところ、5,181 報がヒットした。その中から、引用数及び内容を考慮して食品・農業分野、医療分野の合計 474 報について検討し、さらに 65 報 (食品分野 7 報、医療分野 58 報) を選択して、その内容を精査した。

また、国際機関や各国規制当局 (世界保健機関 (WHO)・国際連合食糧農業機関 (FAO)・コーデックス委員会・医薬品規制調和国際会議 (ICH)・薬事規制当局国際連携組織 (ICMRA)・国際医療機

器規制当局フォーラム (IMDRF)・米国食品医薬品局 (FDA)・米国国農務省 (USDA)・欧州食品安全機関 (EFSA)・欧州医薬品庁 (EMA)・英国食品基準庁 (FSA)・英国医薬品・医療製品規制庁 (MHRA)・中国国家衛生健康委員会 (NHC)) における AI 関連の取り組みについても調査を行った。

C. 研究結果

アレルギー性予測手法 (allerStat) の機能拡充

allerStat について、線形 SVM に加えて新たに非線形 SVM、ロジスティック回帰、LightGBM、LSTM、proteinBERT を分類器として追加実装し、それぞれの予測性について比較検討した。20 品目の食品のアレルゲン及び非アレルゲンデータを用いて leave-category-out cross-validation を行い、予測性の指標となる ROC (Receiver Operatorating Characteristic) 曲線の AUC (Area Under Curve) を算出した。各分類器における AUC を表 1 及び図 1 に示す。ロジスティック回帰、LightGBM、及び LSTM では、線形 SVM とほぼ同程度の AUC が得られた。非線形 SVM では線形 SVM と比較して AUC が小さい品目が多く、また品目ごとの AUC の変動が大きかった。一方 proteinBERT では線形 SVM と比較して AUC が大きい品目が多く、品目ごとの AUC の変動が小さかった。

これまで分類器として使用してきた線形 SVM、及び、上記クロスバリデーションにおいて大きな AUC を示した proteinBERT について、allerStat の学習データ (アレルゲン及び非アレルゲンデータ) の分布図を図 2 に示す。線形 SVM では、アレルゲンデータのほとんどは高値の領域に存在していたが、非アレルゲンデータは比較的広い領域に分布していた。proteinBERT では、アレルゲンデータのほとんどは高値の領域に、非アレルゲンデータのほとんどは低値の領域に分布しており、線形 SVM と比較して分類性能が改善されていることが示された。このように、線形 SVM 以外の分類器を使用することにより予測性能を向上させること

が可能であることが示された。

ADFS 登録アレルゲン (アミノ酸配列情報) のアップデート

米国ネブラスカ大学リンカーン校が運営しているアレルゲンデータベースである AllergenOnline は、登録アレルゲンの全てが国際的なアレルギーの専門家チームによるピアレビューを経ており、登録データの信頼性が非常に高いデータベース (但しエピトープ情報は含まない) である。ADFS における登録アレルゲンは平成 20 年度に AllergenOnline の登録アレルゲンと統合し、その後も、昨年度まで AllergenOnline のアップデートに伴って ADFS 登録アレルゲンのアップデートを行ってきた。しかし、2021 年 2 月を最後に AllergenOnline のデータ更新が行われなくなっている。そのため今年度は、AllergenOnline と同様に専門家によるピアレビューにより登録データの信頼性が非常に高いデータベースである、COMPARE (ILSI のグローバルブランチである HESI が運営している) における登録アレルゲンのアップデート内容を、ADFS に反映させた。

ADFS エピトープ情報の追加

令和 3 年 6 月から令和 4 年 5 月までの 1 年間で、キーワード検索により抽出された論文は 23 報であった。その中からエピトープ情報が記載されていると思われる 18 報を選択し (表 2)、ピアレビューを行った。その結果、11 報の論文から 14 種のアレルゲンについて、総数 52 種のエピトープ情報を新たに追加した (表 3)。

上記のアレルゲン及びエピトープ情報更新作業により、最終的に、ADFS のアレルゲン及びイソアレルゲンのアミノ酸配列情報は 2,403、エピトープ既知のアレルゲン数は 274、構造既知のアレルゲン数は 194、糖鎖付加アレルゲン数は 127 となった。

AI のリスク評価分野等への応用に関する調査

各分野において AI 技術の活用が進められる中で、特に医療分野においては、画像解析やプログラム医療機器 (Software as Medical Device: SaMD) をはじめとする様々な方面で AI 技術の導入及び検討が進められている。各国規制当局は、この革新的ツールの重要性を認識しており、多くの国が、これらの手法や技術の研究、開発、採用を促進するために、国家的な AI 戦略や政策を策定している (後述)。食品 (農業) 分野における AI の活用は医療分野ほどには進んでいないが、これまでに、土壌特性や気象パターンの予測、作物収量予測、病気や雑草の検出、スマート灌漑、家畜の生産・管理、インテリジェント収穫、生産・流通管理、品質管理、需要・購買行動予測、食品表示からの成分・栄養価予測等に関する報告がある。生物学分野においては、マイクロバイオーム研究、バイオマーカー選択、タンパク質コード配列同定、メタゲノム解析データの活用等に関する報告がある。一方、AI 活用における課題としては、データセットの入手・内容・質、サンプルサイズ、モデルの性能評価・実用化・説明可能性等が挙げられている。

また、医療分野では、様々な疾病関連のリスク予測への AI 活用事例の論文報告がいくつかあった。食品分野では、不確実状況下でのサプライチェーンにおけるリスク予測、家畜の病気に関するリスク予測、乳製品の品質に関するリスク評価、病原菌・重金属・化学物質による汚染に関するリスク予測等の論文報告があった。

国際機関や各国規制当局の取り組みやリスク評価への活用については、次のような調査結果が得られた。

医療・健康分野に関して、WHO では、デジタルヘルスとして AI 技術の活用について多くの検討が進められている。ITU/WHO Focus Group on artificial intelligence for health (FG-AI4H) では ガイダンス文書“Ethics and governance of artificial intelligence for health”が取りまとめられており、健康分野における AI 活用状況、AI に適用される法律・政策及び原則、AI の倫理原則や倫理的課題

等がまとめられている。

ICH では医薬品承認申請におけるデジタル関連技術活用のための取り組みが進められている。

ICMRA では、Informal Innovation Network が AI に関するホライズン・スキニングとして、2 件のケーススタディ (中枢神経系アプリ、ファーマコビジランスにおけるシグナル管理) を実施し、その結果を踏まえて提言を行っている。

IMDRF では、AI 関連の医療機器に関するワーキンググループが 2013 年に結成され、イノベーションを支援するガイダンスの開発と、安全で効果的な SaMD へのアクセスを世界規模でタイムリーに行うことを目指して活動が行われている。

FDA では、デジタルヘルスとして AI 技術の規制の検討が進められている。デジタルヘルスイノベーション行動計画、デジタルヘルスソフトウェア事前認証プログラム (Pre-Cert プログラム) の試験運用、AI/ML ベースの SaMD に関する行動計画の公表、デジタルヘルス専門組織 (DHCoE) の設立、医療機器開発のための指針原則、リアルワールドエビデンス (RWE) の活用ガイダンス、画期的なデバイスプログラム (BDP) 関連の取り組み等が行われている。

EMA では、AI と新しいデジタル技術を活用して医薬品規制プロセスの効率を改善し、データに対する洞察を深める目的でプロジェクト・勧告を運営・発表している。

MHRA では、ソフトウェアと AI に対する規制要件を明確にし、患者を保護することを目的とした医療機器規制の改革プログラムを 2021 年 9 月に公表している。

食品・農業分野に関して、FAO では、農業情報・知識へのグローバルアクセスの強化、リモートセンシングと GEO-AI を用いた農作物季節学と農事暦の生成、衛星リモートセンシングデータを用いた水ストレス (乾燥期や干ばつ) の検出、FAO データラボによるツールの実装、写真データからの害虫 (ツマジロクサヨトウ) の検出、FAO デジタルポートフォリオ、ヒレの写真からのサメの種類

の識別、水生産性（作物収量を生産に使われた水量で除した値）の向上を目的としたポータルサイトの公開等、AI 技術を活用した多くの取り組みが進められている。

Codex では、AI を食品不正に対抗する革新的な手段の 1 つであるにとらえ、2022 年 9 月に、食品偽装を検出するための AI に関する国際会議を開催している。

USDA では、進行中または計画されている AI の使用事例の目録を作成し公表している。2022 年 12 月時点で、USDA の組織内外における AI の使用事例 26 件が登録されている。

EFSA では、リスク評価プロセスのエビデンス管理段階において AI 手法を実装するためのアプローチを開発するため、「リスク評価におけるエビデンス管理のための AI 関連活動のためのロードマップ」プロジェクトを 2021 年 5 月に発足させ、その報告書を 2022 年 5 月に公開している。2027 年までに人間の専門知識と密接に共存する人間中心の AI を適用することによってエビデンスのアクセス性と範囲を拡大し、リスク評価プロセスの信頼性を高めることを目標に掲げ、ロードマップを作成している。

FSA では、2016 年にデータ戦略アクションプランを公表しデータの活用の検討等を進めている。AI を食品衛生評価制度（Food Hygiene Rating Scheme：FHRS）に基づく地方自治体による食品施設の衛生検査を効率化するためのスキームに活用している。

なお、NHC に関しては、2022 年 12 月時点で AI 関連の取り組みに関する英語で記された情報は確認できなかった。

一方、課題としては、デジタルデバインド（情報通信技術を利用して恩恵を受ける者と、利用できずに恩恵を受けられない者との間に生ずる知識・機会・貧富などの格差、情報格差）が AI の導入に影響を与える可能性、デジタル化が想定されていない書類様式等のデジタル化への対応、法令や規制体制の整備（枠組み作成、規制側の人材確保等）の

必要性、患者や公衆の安全の確保（AI 導入におけるリスクの考慮、要件の明確化、市販後調査等）、サイバーセキュリティの確保、人間の解釈可能性の程度／人間の解釈が安全性や有効性に影響する可能性、AI を導入したシステムの継続的な学習及び性能確保、公的機関が AI を利用する場合の透明性・公正性の確保、データ利用に関する責任（合法的、安全、公正、倫理、持続性、説明可能な方法でのデータ利用）等が挙げられている。

D. 考察

本研究では、AI を活用した新規アレルゲン性予測手法開発に向けて、アレルゲン性予測手法（allerStat）の機能拡充を行った。分類器として、これまで使用してきた線形 SVM に加えて、非線形 SVM、ロジスティック回帰、lightGBM、LSTM、proteinBERT を実装し、その性能を検証した。現在細かな点の修正等プログラムの整備を進めており、今後、バイオテクノロジー技術応用食品のリスク評価手法としての実用化を目指す。また、ADFS に、アレルゲン及びイソアレルゲンのアミノ酸配列情報を 24 種追加、また、14 種のアレルゲンについて総数 52 のエピトープ情報を追加し、アレルゲンデータベースとしての継続的な充実を進めた。AI の活用状況の調査では、現状では医療分野への活用や規制の取り組みが多いこと、食品・農業分野やリスク評価分野への活用は今後進んでいくと考えられることが示された。また AI 導入の問題点としては、各分野に共通のものとして、データセットの入手・内容・質、サンプルサイズ、モデルの性能評価・実用化・説明可能性等が挙げられる他、規制当局側の問題点としては、法令や規制体制の整備、患者や公衆の安全確保、透明性・公正性の確保、デジタルデバインドへの対応等が挙げられる。

E. 結論

本研究では、遺伝子改変技術応用食品のアレル

ゲン性について、より高い精度での評価・予測を可能とすることを目的として、アレルギー性予測手法（allerStat）の機能拡充を行った。また、令和3年6月から令和4年5月までの1年間に NCBI PubMed に収載された論文から、エピトープ配列決定に関する18報のピアレビューを行い、14種のアレルゲンについて、総数52のエピトープ情報を ADFS に追加した。また、COMPARE の登録アレルゲン（アミノ酸配列情報）に関するアップデートを ADFS に反映させた。これらの情報更新により、遺伝子改変技術応用食品のアレルゲン性評価に有用なデータベースである ADFS を充実させることができた。AI の活用状況の調査では、現状では医療分野への活用や規制の取り組みが多いことが示され、食品分野やリスク評価分野への活用は今後より進むものと考えられた。

3. その他
なし

F. 研究発表

1. 論文発表

- 1) Goto K, Tamehiro N, Yoshida T, Hanada H, Sakuma T, Adachi R, Kondo K, Takeuchi I. AllerStat: Finding Statistically Significant Allergen-Specific Patterns in Protein Sequences by Machine Learning. J Biol Chem, submitted.

2. 学会発表

- 1) 安達玲子. 医療分野等における AI 技術の利用状況とリスク評価分野への適用性. 日本薬学会第143年会（2023年3月26日、札幌）
- 2) 爲廣紀正. 機械・深層学習を利用した新たなアレルギー性予測手法の開発. 日本薬学会第143年会（2023年3月26日、札幌）

H. 知的財産権の出願・登録状況

1. 特許取得

なし

2. 実用新案登録

なし

表1 leave-category-out cross-validationによる評価結果：ROC-AUC

		SVM_ linear	SVM_ non-linear	logistic- regression	lightGBM	LSTM	proteinBERT
1	apple	0.998	0.851	0.998	0.999	1.000	1.000
2	bovine	0.927	0.909	0.932	0.940	0.837	0.945
3	buckwheat	0.874	0.532	0.872	0.827	0.842	0.985
4	carrot	0.947	0.794	0.942	0.898	0.968	0.976
5	chicken	0.797	0.492	0.795	0.829	0.927	0.963
6	corn	0.972	0.866	0.971	0.975	0.966	0.984
7	crab	1.000	1.000	1.000	1.000	1.000	1.000
8	kiwi	0.866	0.758	0.874	0.926	0.883	0.896
9	mustard	0.835	0.451	0.844	0.844	0.842	0.918
10	olive	0.980	0.107	0.988	0.996	0.863	1.000
11	oyster	1.000	1.000	1.000	1.000	0.988	0.992
12	peach	0.998	0.848	0.998	1.000	0.998	0.998
13	peanut	0.699	0.672	0.678	0.401	0.524	0.856
14	potato	0.674	0.685	0.680	0.749	0.711	0.960
15	rice	0.840	0.444	0.837	0.831	0.815	0.934
16	salmon	0.924	0.755	0.927	0.866	0.810	0.978
17	shrimp	1.000	0.778	1.000	0.833	1.000	1.000
18	soybean	0.766	0.836	0.762	0.780	0.757	0.961
19	tomato	0.871	0.669	0.861	0.825	0.885	0.841
20	wheat	0.855	0.567	0.852	0.839	0.809	0.927
21	AVERAGE	0.891	0.701	0.891	0.868	0.871	0.956

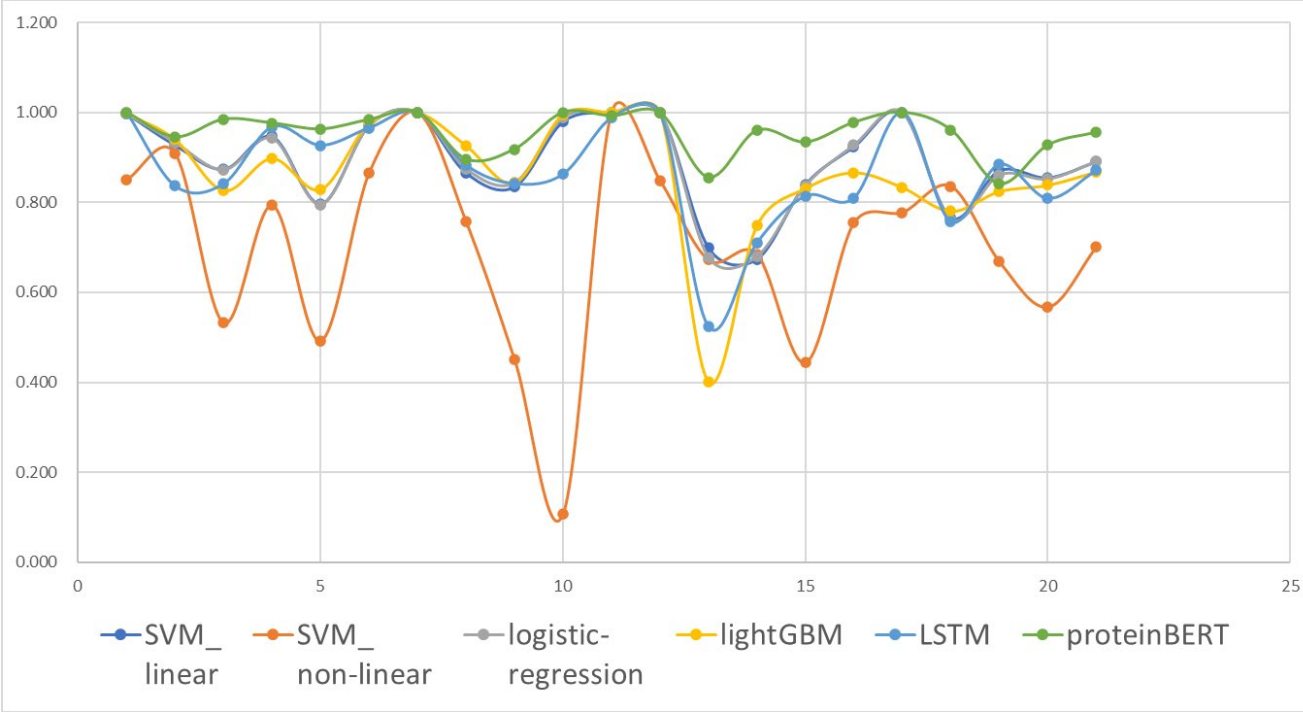


図1 leave-category-out cross-validationによる評価結果：ROC-AUC
※横軸の数字は表1の番号に対応

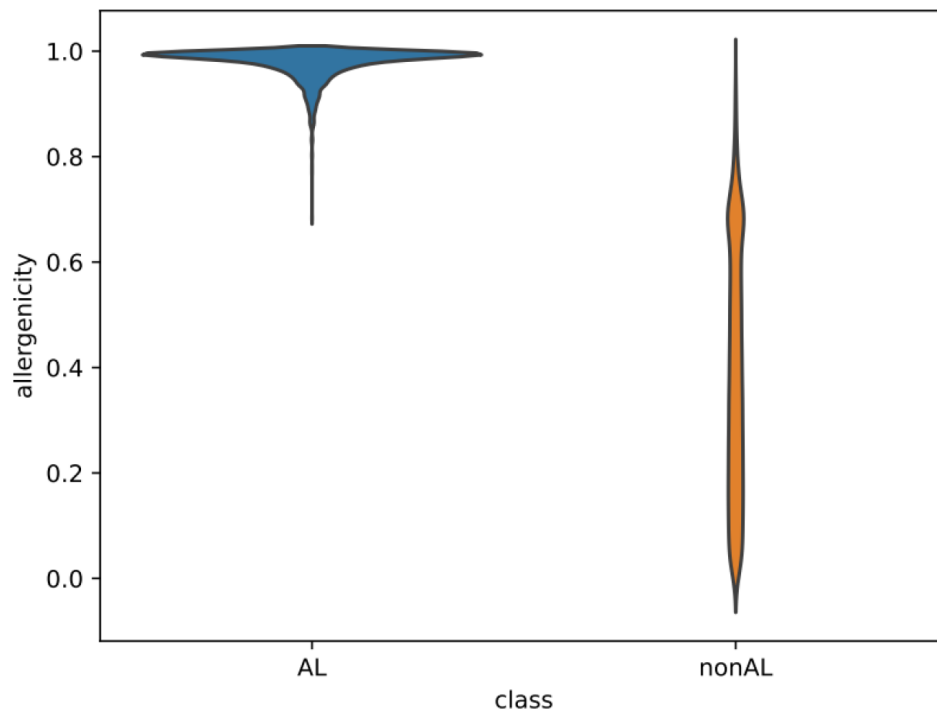
表2 令和4年度にエピトープ情報収集のためピアレビューした論文

	Múnera M, Martínez D, Wortmann J, Zakzuk J, Keller W, Caraballo L, Puerta L. Structural and allergenic properties of the fatty acid binding protein from shrimp <i>Litopenaeus vannamei</i> Allergy. 2022 May;77(5):1534-1544.
1.	PMID : 34695231 He XR, Yang Y, Kang S, Chen YX, Zheng PY, Chen GX, Chen XM, Cao MJ, Jin T, Liu GM. Crystal Structure Analysis and IgE Epitope Mapping of Allergic Predominant Region in <i>Scylla paramamosain</i> Filamin C, Scy p 9 J Agric Food Chem. 2022 Feb 2;70(4):1282-1292.
2.	PMID : 35040643 Schmalz S, Mayr V, Shosherova A, Gepp B, Ackerbauer D, Sturm G, Böhle B, Breiteneder H, Radauer C. Isotype-specific binding patterns of serum antibodies to multiple conformational epitopes of Bet v 1 J Allergy Clin Immunol. 2022 May;149(5):1786-1794.
3.	PMID : 34740603 Han TJ, Huan F, Liu M, Li MS, Yang Y, Chen GX, Lai D, Cao MJ, Liu GM. IgE epitope analysis of sarcoplasmic calcium-binding protein, a heat-resistant allergen in <i>Crassostrea angulata</i> Food Funct. 2021 Sep 20;12(18):8570-8582.
4.	PMID : 34338271 Moten D, Kolchakova D, Todorov K, Mladenova T, Dzhambazov B. Design of an Epitope-Based Peptide Vaccine Against the Major Allergen Amb a 11 Using Immunoinformatic Approaches Protein J. 2022 Apr;41(2):315-326.
5.	PMID : 35362839 Li M, Xia F, Chen Y, Liu M, Liu Q, Yang Y, Chen G, Wang L, Cao M, Liu G. Two hypo-allergenic derivatives lacking the dominant linear epitope of Scy p 1 and Scy p 3 Food Chem. 2022 Mar 30;373(Pt B):131588.
6.	PMID : 34801289 Curin M, Huang HJ, Garmatiuk T, Gutfreund S, Resch-Marat Y, Chen KW, Fauland K, Keller W, Zieglmayer P, Zieglmayer R, Lemell P, Horak F, Hemmer W, Focke-Tejkl M, Flicker S, Vrtala S, Valenta R. IgE Epitopes of the House Dust Mite Allergen Der p 7 Are Mainly Discontinuous and Conformational Front Immunol. 2021 Jun 15;12:687294.
7.	PMID : 34220841 Li WY, Cai ZL, Zhang BP, Chen JJ, Ji K. Identification of an immunodominant IgE epitope of Der p 39, a novel allergen of <i>Dermatophagoides pteronyssinus</i> World Allergy Organ J. 2022 May 6;15(5):100651.
8.	PMID : 35600837 Chen G, Shrock EL, Li MZ, Spergel JM, Nadeau KC, Pongracic JA, Umetsu DT, Rachid R, MacGinnitie AJ, Phipatanakul W, Schneider L, Oettgen HC, Elledge SJ. High-resolution epitope mapping by AllerScan reveals relationships between IgE and IgG repertoires during peanut oral immunotherapy Cell Rep Med. 2021 Oct 19;2(10):100410.
9.	PMID : 34755130 Sharma S, Vashisht S, Gaur SN, Lavasa S, Arora N. Effects of the Maillard reaction on the epitopes and immunoreactivity of tropomyosin, a major allergen in <i>Chlamys nobilis</i> . Immunobiology. 2021 Nov;226(6):152146.
10.	PMID : 34717182 Huan F, Han TJ, Liu M, Li MS, Yang Y, Liu QM, Lai D, Cao MJ, Liu GM. Identification and characterization of <i>Crassostrea angulata</i> arginine kinase, a novel allergen that causes cross-reactivity among shellfish Food Funct. 2021 Oct 19;12(20):9866-9879.
11.	PMID : 34664604 Cheng Q, Feng X, Zhao X, Gu R, Lu J, Liu W, Li G. Physicochemical characterization and identification of major linear epitopes of sarcoplasmic calcium-binding protein (SCP) allergen from Pacific oyster (<i>Crassostrea gigas</i>) J Sci Food Agric. 2021 Dec 1.
12.	PMID : 34854091 Calzada D, Aranda T, M Gallego G, Escutia MR, Balsa D, Álvarez J, Mayorga C, Salas M, Odena MA, Oliveira E, Pascal M, Carnés J. Immunological mechanisms involved in the human response to a dog dander allergoid Mol Immunol. 2022 May;145:88-96.
13.	PMID : 35306358 Xu LL, Gasset M, Lin H, Yu C, Zhao JL, Dang XW, Li ZX. Identification of the Dominant T-Cell Epitopes of Lit v 1 Shrimp Major Allergen and Their Functional Overlap with Known B-Cell Epitopes J Agric Food Chem. 2021 Jul 7;69(26):7420-7428.
14.	PMID : 34170668 Brassee-Estandarte HA, Martínez-Cruz O, Cárdenas-López JL, García-Orozco KD, Ochoa-Leyva A, López-Zavala AA. Identification of arginine kinase as an allergen of brown crab, <i>Callinectes bellicosus</i> , and in silico analysis of IgE-binding epitopes Mol Immunol. 2022 Mar;143:147-156.
15.	PMID : 35131595 Kumagai A, Nara T, Uematsu M, Kakinuma Y, Saito T, Masuda K. Development and characterization of a unique anti-IgE mouse monoclonal antibody cross-reactive between human and canine IgE Immun Inflamm Dis. 2021 Dec;9(4):1740-1748.
16.	PMID : 34533288 Saetang J, Tipmanee V, Benjakul S. In Silico Prediction of Cross-Reactive Epitopes of Tropomyosin from Shrimp and Other Arthropods Involved in Allergy Molecules. 2022 Apr 21;27(9):2667.
17.	PMID : 35566021 Zettl I, Ivanova T, Strobl MR, Weichwald C, Goryainova O, Khan E, Rutovskaya MV, Focke-Tejkl M, Drescher A, Böhle B, Flicker S, Tilib SV. Isolation of nanobodies with potential to reduce patients' IgE binding to Bet v 1 Allergy. 2022 Jun;77(6):1751-1760.
18.	PMID : 34837242

表3 令和4年度に新たにADFSに追加したエピソード情報

	Name	start	end	Sequence	Method	CTYPE	Reference	UniProt acc.No
001	Lit v 13	40	85	VEITKDGDTYTMKTTTTTKTEIKFKLGEFEETTADGRVVKSTIT	Peptide array/ELISA	L	PMID 34695231	E2IH93
	Lit v 13	107	136	ELLREFTDDKMLMECKVDDVCKRVYSRL	Peptide array/ELISA	L	PMID 34695231	E2IH93
002	Scy p 9	336	353	GTNRQTERIKRQREAVPL	Dot blot/bioinformatics/ELISA	L	PMID 35040643	A0A5J6X3F8
	Scy p 9	363	377	TFKLPGISPFDLGAT	Dot blot/bioinformatics/ELISA	L	PMID 35040643	A0A5J6X3F8
	Scy p 9	418	433	EMHIPGSPFQFTVGPL	Dot blot/bioinformatics/ELISA	L	PMID 35040643	A0A5J6X3F8
	Scy p 9	446	462	PGLERGEQGMPEFNWV	Dot blot/bioinformatics/ELISA	L	PMID 35040643	A0A5J6X3F8
	Scy p 9	490	505	DGSCYVSYVVAEPGEY	Dot blot/bioinformatics/ELISA	L	PMID 35040643	A0A5J6X3F8
	Scy p 9	512	528	NDKHIPDSPYKVYITPS	Dot blot/bioinformatics/ELISA	L	PMID 35040643	A0A5J6X3F8
	Scy p 9			338 – 343,449 – 456,476 – 482,498 – 504,506,522,524 – 528	Dot blot/bioinformatics/ELISA	C	PMID 35040643	A0A5J6X3F8
	Scy p 9			345,347,349 – 359,379 – 389,403 – 408,430 – 434,467 – 470,512 – 514	Dot blot/bioinformatics/ELISA	C	PMID 35040643	A0A5J6X3F8
	Scy p 9			366 – 375,390 – 397,415 – 420	Dot blot/bioinformatics/ELISA	C	PMID 35040643	A0A5J6X3F8
003	Bet v 1				ELISA	C	PMID 34740603	O23746
004	Cra a 4	22	33	KISIEDVEESRN	ELISA	L	PMID 34338271	A0A6G7MAZ4
	Cra a 4	64	75	TGAGKEISESEF	ELISA	L	PMID 34338271	A0A6G7MAZ4
	Cra a 4	80	90	TEAYKKDKVGF	ELISA	L	PMID 34338271	A0A6G7MAZ4
	Cra a 4	107	116	TNKDRTIDED	ELISA	L	PMID 34338271	A0A6G7MAZ4
	Cra a 4	144	159	NKHVPLKDIVSEWVKF	ELISA	L	PMID 34338271	A0A6G7MAZ4
	Cra a 4	131	150	G48, D99, I101, D103, I105, D110, R111, D114, D116, A124, A136, W156	ELISA	C	PMID 34338271	A0A6G7MAZ4
	Cra a 4	170	188	D114, R134, A136, S138, L139, N141, N144, L149, D151, V153, S154, V157, F159	ELISA	C	PMID 34338271	A0A6G7MAZ4
005	Amb a 11	173	186	GKLVKFSEQQLVDC	bioinformatics/ELISA	L	PMID 35362839	V5LU01
006	Scy p 1	39	48	TEEEIRATQK	ELISA	L	PMID 34801289	A7L5V2
	Scy p 1	153	162	FLAEEADRKY	ELISA	L	PMID 34801289	A7L5V2
007	Scy p 3	35	47	DCLRALNLNPTLA	ELISA	L	PMID 34801289	A0A514C9K9
	Scy p 3	66	80	DDFLPIFAQVKKDKD	ELISA	L	PMID 34801289	A0A514C9K9
	Scy p 3	109	120	HILLSLGERLEK	ELISA	L	PMID 34801289	A0A514C9K9
008	Der p 7	18	47	DPIHYDKITEINKAVDEAVAAIEKSETFD	immunoblot/inhibition ELISA	L	PMID 34220841	P49273
	Der p 7	37	67	VAAIEKSETFDPMKVPDHSKDFERHIGIIDL	immunoblot/inhibition ELISA	L	PMID 34220841	P49273
	Der p 7	67	97	LKGELDMRNIQVRGLKQMKRVGDANVKSEGD	immunoblot/inhibition ELISA	L	PMID 34220841	P49273
	Der p 7	107	142	VHDDVVSMEYDLAYKLDLHPNTHVISIDIQDFVEL	immunoblot/inhibition ELISA	L	PMID 34220841	P49273
	Der p 7	140	165	VELSLEVSEEGNMTLTSFEVRQFANV	immunoblot/inhibition ELISA	L	PMID 34220841	P49273
	Der p 7	166	193	VNHIGLSILDPIFAVLSDVLTAFQDT	immunoblot/inhibition ELISA	L	PMID 34220841	P49273
	Der p 7	187	215	TAIFQDTVRAEMTKVLAPAFKKELERNNQ	immunoblot/inhibition ELISA	L	PMID 34220841	P49273
	Der p 7				immunoblot/inhibition ELISA	C	PMID 34220841	P49273
	Der p 7				immunoblot/inhibition ELISA	C	PMID 34220841	P49273
009	Der p 39	87	97	QEELREAFRMY	Western Blot/dot blot/ELISA	L	PMID 35600837	A0A6P6YD30
	Der p 39	107	117	TSALREILRAL	Western Blot/dot blot/ELISA	L	PMID 35600837	A0A6P6YD30
	Der p 39	123	133	NDELDEMAIEI	Western Blot/dot blot/ELISA	L	PMID 35600837	A0A6P6YD30
010	Ara h 2.02	21	40	ARQQWELQGDRRRCQQLERA	AllerScan	L	PMID 34755130	Q6PSU2
011	Ara h 3.01	301	320	ERDPYSPSQDPYSPSPYDRR	AllerScan	L	PMID 34755130	O82580
012	Ara h 7.01	71	90	QEQDEYPYSRRGSRGRQPGE	AllerScan	L	PMID 34755130	Q9SQH1
013	Per a 5	32	47	VTNLMAGEHLTPEFLK	bioinformatics/ELISA	L	PMID 34717182	A0A2K9YV03
	Per a 5	77	92	QYGKDDSLYPKDAKKR	bioinformatics/ELISA	L	PMID 34717182	A0A2K9YV03
	Per a 5	114	135	YYPIYFAKQAADPEKMKKLEEA	bioinformatics/ELISA	L	PMID 34717182	A0A2K9YV03
014	Cra a 2	56	73	SGAGVYACDPGEYEVFKE	dot blot/inhibition ELISA	L	PMID 34664604	
	Cra a 2	77	92	PVIMDYHKVKVEHPP	dot blot/inhibition ELISA	L	PMID 34664604	
	Cra a 2	93	107	CDFGPQDKLGFDPD	dot blot/inhibition ELISA	L	PMID 34664604	
	Cra a 2	121	133	GRSHEGYPPFPVS	dot blot/inhibition ELISA	L	PMID 34664604	
	Cra a 2	134	146	TDEQRKEMENKTI	dot blot/inhibition ELISA	L	PMID 34664604	
	Cra a 2	165	176	TMTPEENQQLID	dot blot/inhibition ELISA	L	PMID 34664604	
	Cra a 2	185	200	DKMLGDAGGYNGWPKA	dot blot/inhibition ELISA	L	PMID 34664604	
	Cra a 2	227	238	QKGGDVGVEYKRR	dot blot/inhibition ELISA	L	PMID 34664604	
	Cra a 2	306	323	GEHTESVGGYDISNKRR	dot blot/inhibition ELISA	L	PMID 34664604	

SVM



proteinBERT

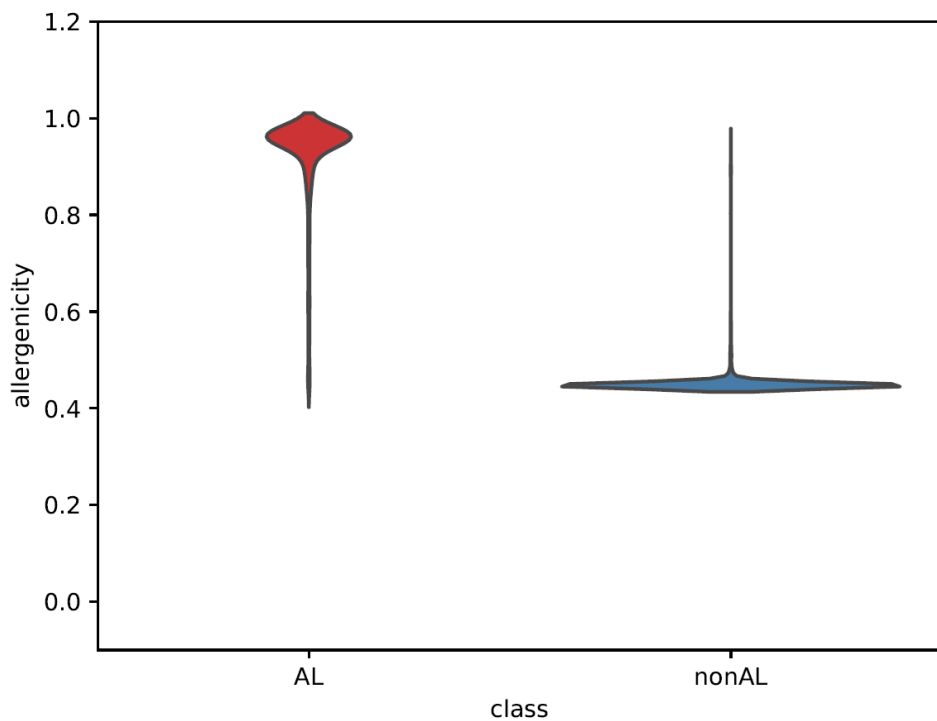


図2 アレルゲン及び非アレルゲンデータの分布
AL: allergen data, nonAL: non-allergen data

厚生労働科学研究費補助金（食品の安全確保推進研究事業）
「新たなバイオテクノロジーを用いて得られた食品の安全性確保と
リスクコミュニケーションのための研究」
分担研究報告書（令和4年度）

新規アレルギー性評価手法開発の基盤研究と AI のリスク評価への応用

研究分担者 富井 健太郎 産業技術総合研究所

研究要旨：

HLA クラス II 分子-ペプチド間の結合は、残基間相互作用が本質的に鍵を握っていると考え、既存の深層学習モデルの DeepSeqPanII をベースモデルに、特徴量として（デフォルトの one-hot encoding と BLOSUM62 に加えて）アミノ酸物理化学的インデックス、PSSM、アミノ酸残基間ポテンシャルインデックスを組み合わせるとトレーニングを行い、LOAO 交差検証テストによってモデルの予測性能を比較し、性能向上の可能性とどのような特徴量が有効であるかを検証した。一次元配列情報データセットに対する性能評価結果では、物理化学的インデックスを特徴量に加えた場合、DeepSeqPanII のデフォルトによる AUC 値よりも平均 2.1%の向上が確認でき、立体構造情報データセットに対する性能評価結果では、特徴量に物理化学的インデックスと PSSM を組み合わせることによって平均 6.9%の向上が得られた。

研究協力者

池田 修己 産業技術総合研究所
坂無 英徳 産業技術総合研究所

A. 研究目的

第二次産業革命を契機として、世界の人口増加、温暖化による砂漠化などによる農地面積の減少、作物収穫量の低下、農業従事者の減少などの諸問題によって世界的な食糧不足への懸念が強まっている。こうした問題を背景に、害虫抵抗性や除草剤耐性をもたせることで収量増を見込める遺伝子組み換え作物（GMO）の実用化が進んでいる。日本では遺伝子組み換え作物規制条例で栽培を規制しているが、家畜飼料用がほとんどではあるものの輸入に依存しているトウモロコシ、ダイズ、菜種などは半量が既に遺伝子組換え作物であると推定されている。

GMO に対してはアレルギー性が遺伝子改変食品の安全性が問われているが、すべての遺伝子組

換え食品のアレルギー性を実験的に評価するのはコストの面から現実的ではない。このため、科学的根拠をもつ信頼性のある評価方法の確立が求められている。適切なリスク管理対策の適用により、遺伝子改変食品のアレルギー性リスクを低減することが出来るかもしれない。

組換え DNA 技術で導入した新規遺伝子産物（タンパク質）や形質転換による意図しない新規タンパク質のアレルギー性予測方法としては、FAO（国連食糧農業機関）/WHO（世界保健機関）が提唱しているデータベースに登録済みのアレルギータンパク質との相同性比較（[1] 80 個の連続したアミノ酸配列について 35%以上の相同性、[2] 6～8 個の連続したアミノ酸配列の完全一致）が標準的に使用されている。

しかし、配列長が短い既知アレルギー性ペプチドとの類似性に基づくため偽陽性が高いことが指摘されている。また、進化に基づくアミノ酸置換行列を用いる配列類似性比較は、オフターゲット効果による変異をもつ新規タンパク質に対するア

レルゲン性の判定には十分ではない可能性がある。

機械学習は使用する学習データに強く依存する。アレルギー発症の機序において、HLA 分子は種々のペプチドを結合し、T 細胞に抗原提示を行い、T 細胞受容体との結合によってサイトカインの産生や細胞傷害性活性が起こる。近年、必ずしも HLA 分子とペプチドとのアフィニティ(結合親和性) およびオフレートと T 細胞の活性化が一致していないことが明らかとなっている。したがって、T 細胞活性化/不活性化の実験的根拠をもつ正例/負例データの選定およびデータセットの準備には特段の注意を要する。

そこで本研究では、標的配列と類似した配列のオフターゲット検索しかできない点を克服すべく、人工知能を活用して相同性がないアレルギー性タンパク質由来ペプチド-HLA クラス II 分子間結合予測法の開発を行うことを目的に取り組んでいる。

研究計画初年度の昨年度では、既存予測手法が使用した機械学習法およびデータセットについて調査を行なった。さらに、HLA クラス II 分子-アレルギー性ペプチド-T 細胞受容体複合体の立体構造データを用いて、ペプチド結合クレフトにおける HLA クラス II 分子-アレルギー性ペプチド間の相互作用状態について解析した。

今年度は、まず、既存のデータセット以外に新たにデータセットを 2 種類構築した。次に、これを用いて既存手法の深層学習による予測法に、アミノ酸物理化学的インデックス、位置特異的スコア行列 (PSSM)、アミノ酸残基間ポテンシャルインデックスを特徴量として用いたモデルを構築した。LOAO (leave-one-allele-out) 交差検証テストによってモデルの予測性能を比較し、性能向上の可能性とどの特徴量が有効であるかを検証した。

B. 研究方法

(1) データセットの構築

前述のとおり、以下の 3 種類のデータセットを用意した。

①既に提案されている予測法 DeepSeqPanII の学習に使用した「BD2016 データセット」を公開サイトからダウンロードし、そのまま使用した。

②オリジナルのデータセットとして「IEDB2022 データセット」を構築した。構築フローを図 2 に示す。HLA クラス II 分子の配列情報およびアレルギー情報を収めた IPD-IMGT/HLA データベース (release 3.47, 2022-01) からデータファイルをダウンロードした。アレルギー名に接尾辞「N」、「Q」、「L」、「S」の記載をもつエントリを取り除いた。一方、IEDB (2022-04) から結合ペプチド分子名、HLA 遺伝子名・アレルギーグループ・アレルギー型、結合ペプチド配列、アフィニティ、アッセイなどの情報を含むデータをダウンロードした。データの信頼性を確保するために、以下の条件を設けフィルタリング処理によるデータ選別を行った。

- ・ IC50 (50%阻害濃度) の値をもち、カラム「Units」の値が「nM」の単位として与えられており、カラム「Quantitative measurement」の値が「null」でないデータ。
- ・ 対象とするアレルギー名に「/」が含まれるデータ (α 鎖タイプ/ β 鎖タイプ)、および「HLA-DRB」であるデータ。
- ・ 同じアレルギーかつ同じペプチドが存在する場合、その重複のペプチドが正例と負例の両方に存在する場合、それら全てを除外する。正例または負例に重複のペプチドが存在する場合は、その中で IC50 が最小のものを採用し、それ以外は削除する。
- ・ アレルギー名に接尾辞「N」、「Q」、「L」、「S」の記載をもたないデータ。

③高分子立体構造データベース PDB に登録されているペプチド結合状態の HLA クラス II 分子を、ウェブツール IMGT を用いて検索・抽出した。さらに、以下の条件でデータの選別を行った。

- ・ CD4、TCR エントリを除去。

- ・ HLA-DM タイプのエントリを除去。
- ・ HLA のペプチド結合領域、結合ペプチドに非標準アミノ酸／unknown を含むエントリを除去。
- ・ HLA のペプチド結合領域に欠失／置換のあるエントリを除去

同じく IMGT を用いて、HLA クラス II 分子と結合ペプチド間において水素結合を形成する残基情報、位置、結合数の情報を抽出し、「PDB-HB データセット」として構築した。構築フローを図 3 に示した。

(2) 特徴量

予測法に組み込む特徴量は以下の 5 種類を用いた（ただし、One-hot encoding と BLOSUM62 に関しては、DeepSeqPanII において採用されている特徴量である）。

- ・ One-hot encoding: 20 種類のアミノ酸を 20 次元の bit ベクトルで表現。
- ・ BLOSUM62: アミノ酸置換行列の値を 23 次元のベクトルで表現。
- ・ AAindex 物理化学的インデックス: 20 種類のアミノ酸に KEGG AAindex データベースに登録されている 566 種類のインデックスの値を割り当てる。566 種類の全てを次元圧縮した場合と、類似したインデックスを相関係数に基づいて非冗長な 62 種類に削減したうえで次元圧縮した場合とを別々に特徴量として追加。
- ・ PSSM: マルチプルアラインメント処理によって得られたカラム重複部分の重み付きアミノ酸出現頻度として、20 次元のベクトルで表現。
- ・ AAindex 残基間ポテンシャルインデックス: KEGG AAindex データベースに登録されている 20 種類のアミノ酸残基間におけるポテンシャルのスコア行列 (47 種類) を用いて、立体構造既知の HLA クラス II 分子とペプチド間で 3.5 オングストローム以内に近接する残基ペアに対してスコアを割当。

(3) 深層学習ベースモデル

既に提案されている予測法の DeepSeqPanII をベースモデルとして、前節で述べた 5 種類の特徴量を組み合わせて、3 種類のデータセットごとに LOAO 交差検証テストを行った。

ここで、DeepSeqPanII をベースモデルに用いた利点は、第一に、DeepSeqPanII が再帰的ニューラルネットワーク (Recurrent Neural Network; RNN) の一種である LSTM (long short-term memory) を用いていることである。LSTM の強みは、時系列データの学習や予測 (回帰・分類) にあり、HLA クラス II の α 鎖、 β 鎖、ペプチドで構成される 3 組の一次元配列間における線形での結合状態を考慮する上で、モデルが複雑にならないことが期待できる (図 1)。第二に、注意機構 (attention mechanism) を採用しており、RNN が記憶しきれない過去の情報を記憶にキャッシュすることによって、ニューラルネットワークの内部を可視化することができる長所をもつ。第三に、一連のプログラムが GitHub にて公開されており、MIT License として再利用が認められていることである。したがって、新たに別の特徴量を組み込んで利用することは問題にならない。

(4) LOAO 交差検証テストと AUC の算出

LOAO (leave-one-allele-out) 交差検証では、データセットに含まれるアレル分子群から 1 つのエントリだけを抜き出してテスト事例とし、残りをトレーニングとする。この処理を全てのエントリが一回ずつテスト事例となるよう検証を繰り返す。

LOAO 交差検証テストの結果から、各 HLA クラス II 分子に対し、

$$\text{FPR} = \text{FP} / (\text{FP} + \text{TN})$$

$$\text{TPR} = \text{TP} / (\text{TP} + \text{FN})$$

を算出できる。横軸に FPR (偽陽性率)、縦軸に TPR (真陽性率) をとった時に描かれる ROC 曲

線下側の面積（AUC：area under the curve）の値に基づいて、予測性能を評価する。

C. 研究結果および考察

（1）構築したデータセット

LOAO 交差検定テストに使用する「IEDB2022 データセット」および「PDB-HB データセット」の構築フローを、それぞれ図 2 と図 3 に示した。信頼性の高い学習効果を得ることを目的に、データ選別あるいはクリーニングの条件を設けた。

3 種類のデータセットにおけるアレルタイプ別の配列数を表 1 にまとめた。また、「BD2016 データセット」と「IEDB2022 データセット」における正例と負例ペプチドデータ数を表 2 にまとめた（「PDB-HB データセット」は立体構造既知データに由来しているため、結合ペプチドデータは全て正例として扱う）。

「IEDB2022 データセット」は「BD2016 データセット」に比べて新しいデータが含まれており、また、曖昧な実験情報あるいは矛盾した実験結果が記載されているエントリを含めないデータ選別処理を行ったため、「BD2016 データセット」よりもやや少ない構成となっている。「PDB-HB データセット」に関しては立体構造データそのものが限られており、「IEDB2022 データセット」と同様に HLA-DP タイプが含まれていない。

（2）LOAO 交差検証テスト結果

①多くの予測法が学習用データセットとして利用している「BD2016 データセット」に対する AUC による LOAO 交差検証テスト結果を図 4 に示す。One-hot encoding と BLOSUM62 を特徴量に用いているオリジナルの予測法 DeepSeqPanII よりも、AAindex を特徴量に加えた場合が全体的に高い AUC 値を示した。特に、AAindex（次元圧縮）が 54 アレル中 50 アレルで DeepSeqPanII よりも高く、2 アレルは同じ AUC 値であった。AAindex（次元圧縮）のほうが AAindex（非冗長 62 インデック

ス＋次元圧縮）よりも、2 アレルを除いて、優れていることが分かった。

②本研究で構築したオリジナルデータセットである「IEDB2022 データセット」に対する LOAO 交差検証結果を図 5 に示す。「BD2016 データセット」と同様に AAindex（次元圧縮）は DeepSeqPanII よりも AUC 値が高い傾向にあり（15/19 アレル）、AAindex（非冗長 62 インデックス＋次元圧縮）よりも全アレルで優れていることが分かった。正例／負例の閾値（500 [nM]、1000 [nM]）の違いは AUC 値に顕著な影響を与えなかった。図には示していないが、AAindex を次元圧縮せずに 566 インデックスを特徴量に加えても AAindex（次元圧縮）の AU 値を全アレルで超えられていなかった。

③HLA クラス II 分子－ペプチドの立体構造データから選別した「PDB-HB データセット」に対する LOAO 交差検証結果を図 6 に示す。AAindex（次元圧縮）＋PSSM による AUC 値が 8 アレル中 5 アレルにおいて最も高い AUC 値を示した。1 アレル（DRA*01:01－DRB1*01:01）のみ、DeepSeqPanII のデフォルトと同値であるが、それ以外の 7 アレルでは顕著な差（AUC：6.9±3.2）があった。この特徴量の単独と比較すると、組み合わせによる効果が 5/8 アレルで有意であることが分かった。アミノ酸残基間ポテンシャルを加えた結果は、AAindex（次元圧縮）＋PSSM よりも高い AUC が得られた一方で、1 アレル（DQA1*03:01－DQB1*03:02）では他の特徴量に比べて顕著に低い値となった。

D. 結論

深層学習を用いた DeepSeqPanII をベースモデルに、特徴量（デフォルトの one-hot encoding と BLOSUM62 に加えて、アミノ酸物理化学的インデックス、PSSM、アミノ酸残基間ポテンシャルインデックス）を組み合わせでトレーニングさせた。

2 種類の一次元配列情報データセットに対する LOAO 交差検定による予測性能評価結果では、物理化学的インデックスを用いた場合、DeepSeqPanII の AUC 値よりも平均 2.1%向上した。ペプチド結合残基位置が明らかな立体構造情報データセットに対する性能評価結果からは、特微量に PSSM と残基間ポテンシャルインデックスを組み合わせることによって平均 6.9%の予測精度を向上できることを明らかにできた。

計画 3 カ年中の 2 年目において概ね予定通りに進捗している。最終年度となる次年度では、実際に食物アレルギーの発症が確認されているアレルギーペプチドデータへの予測と評価を行う。また、他の既存の予測法との性能比較を行う予定である。論文誌上ならびに学会や研究会などでの成果発表を行い、研究に使用したデータセットや開発した一連のプログラムをコミュニティ向けに限定公開し、改良を重ねつつ利用促進活動を行う。

E. 研究発表・業績

1. 論文発表

無し

2. 学会発表

無し

F. 健康危険情報

該当なし

G. 知的財産権の出願・登録状況

該当なし

表 1. LOAO 交差検証テストに用いた各データセットにおけるアレルタイプ別配列数

アレルタイプ	Dataset		
	BD2016 ^{※1}	IEDB2022 ^{※2}	PDB-HB
HLA-DR	25	17	7
HLA-DQ	20	2	1
HLA-DP	9	0	0

※1 BD2016 の学習にはペプチド数が 10 以上のアレルに対して LOAO を実行

※2 IEDB2022 の学習にはペプチド数が 1000 以上のアレルに対して LOAO を実行

※3 PDB-HB の学習には BD2016 に含まれる同一のアレルに対して LOAO を実行

表 2. LOAO 交差検証テストに用いた各データセットにおけるアレルタイプ別の正例と負例ペプチド数

アレルタイプ	BD2016 ^{※1}		IEDB2022 (500 [nM]) ^{※1}		IEDB2022 (1000 [nM]) ^{※2}	
	正例	負例	正例	負例	正例	負例
HLA-DR	87,363	40,756	22,177	32,195	26,647	26,825
HLA-DQ	15,564	5,135	3,753	9,294	7,978	5,067
HLA-DP	28,081	9,098	801	3,771	3,433	1,139

※1 正例 : $IC_{50} < 500$ [nM]、負例 : $IC_{50} \geq 500$ [nM]

※2 正例 : $IC_{50} < 1000$ [nM]、負例 : $IC_{50} \geq 1000$ [nM]

図1. DeepSeqPanIIのモデルアーキテクチャ (出典：Liu et al. (2022) IEEE/ACM Transactions on Computational Biology and Bioinformatics)

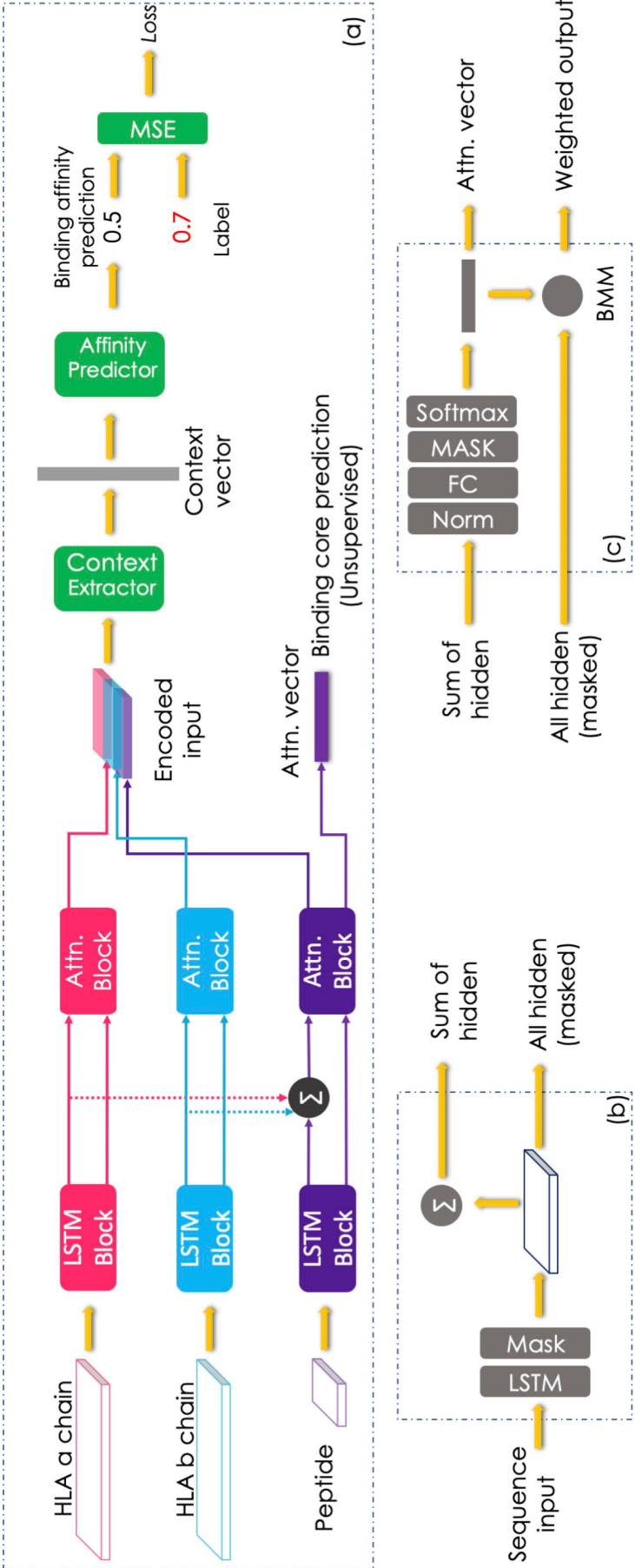


図2. 「IEDB2022データセット」構築フロー

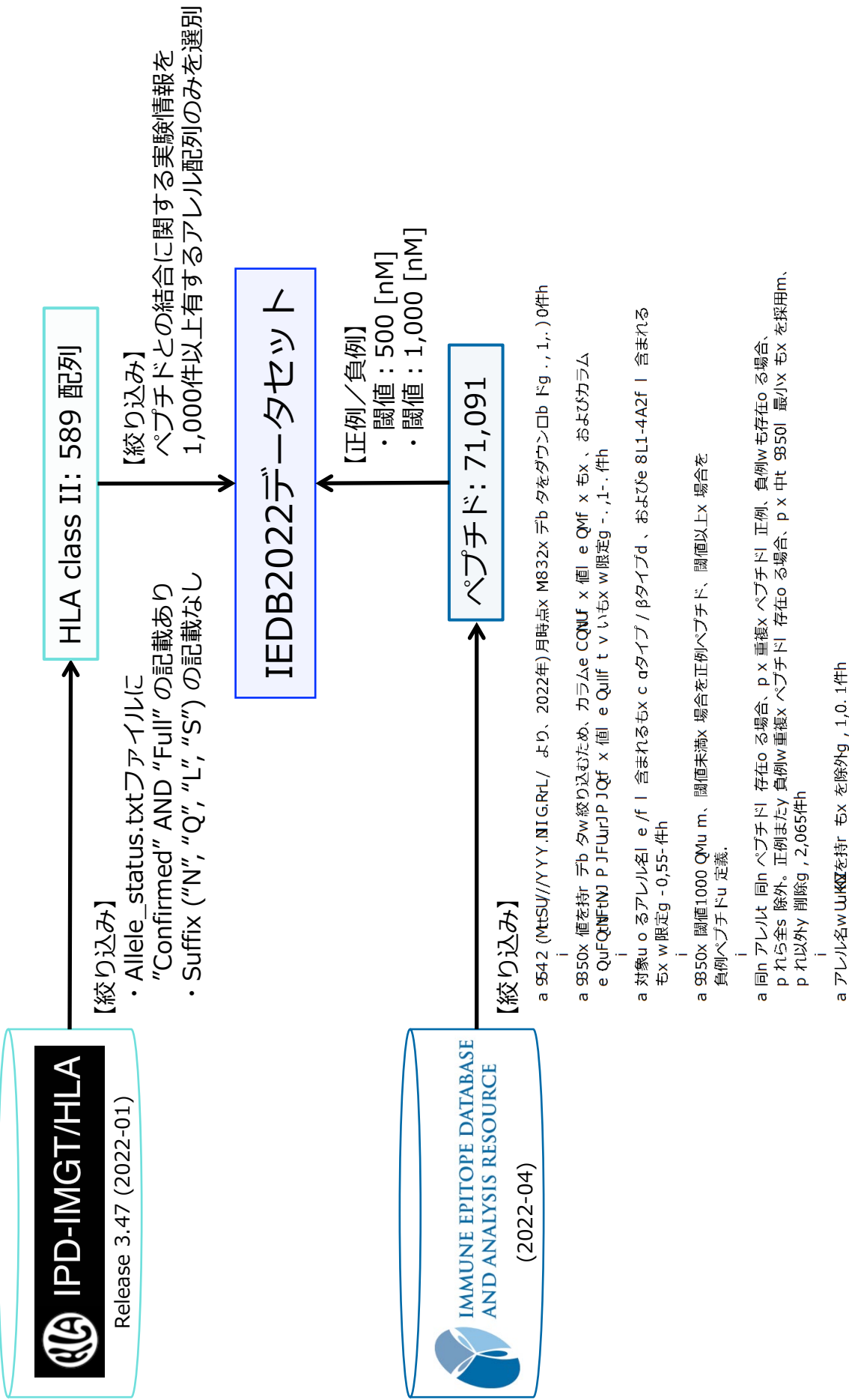


図3. 「PDB-HBデータセット」構築フロー

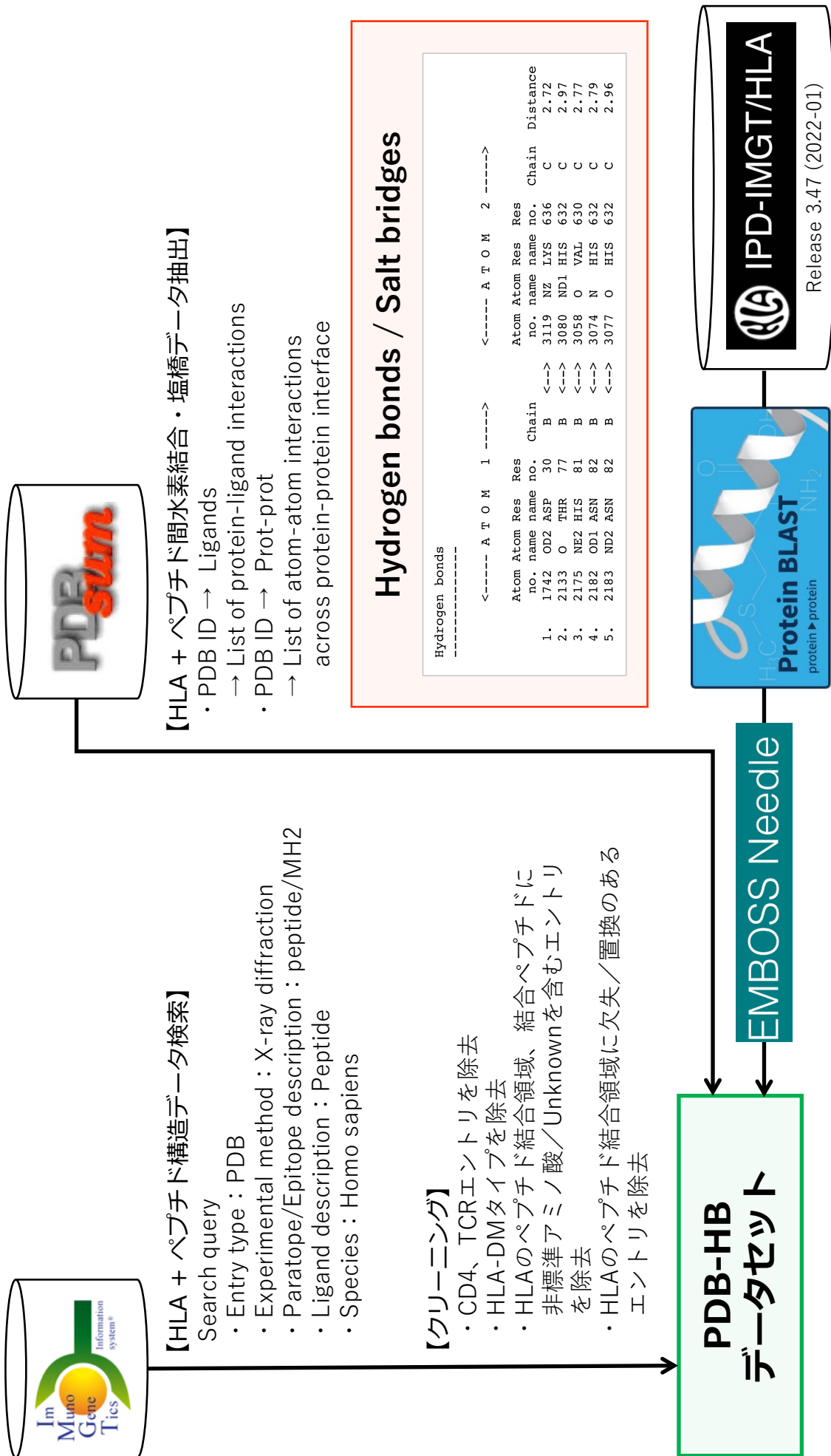


図4. 「BD2016データセット」に対するLOAO交差検証テストの結果（結合ペプチド数10以上のアレル：54）

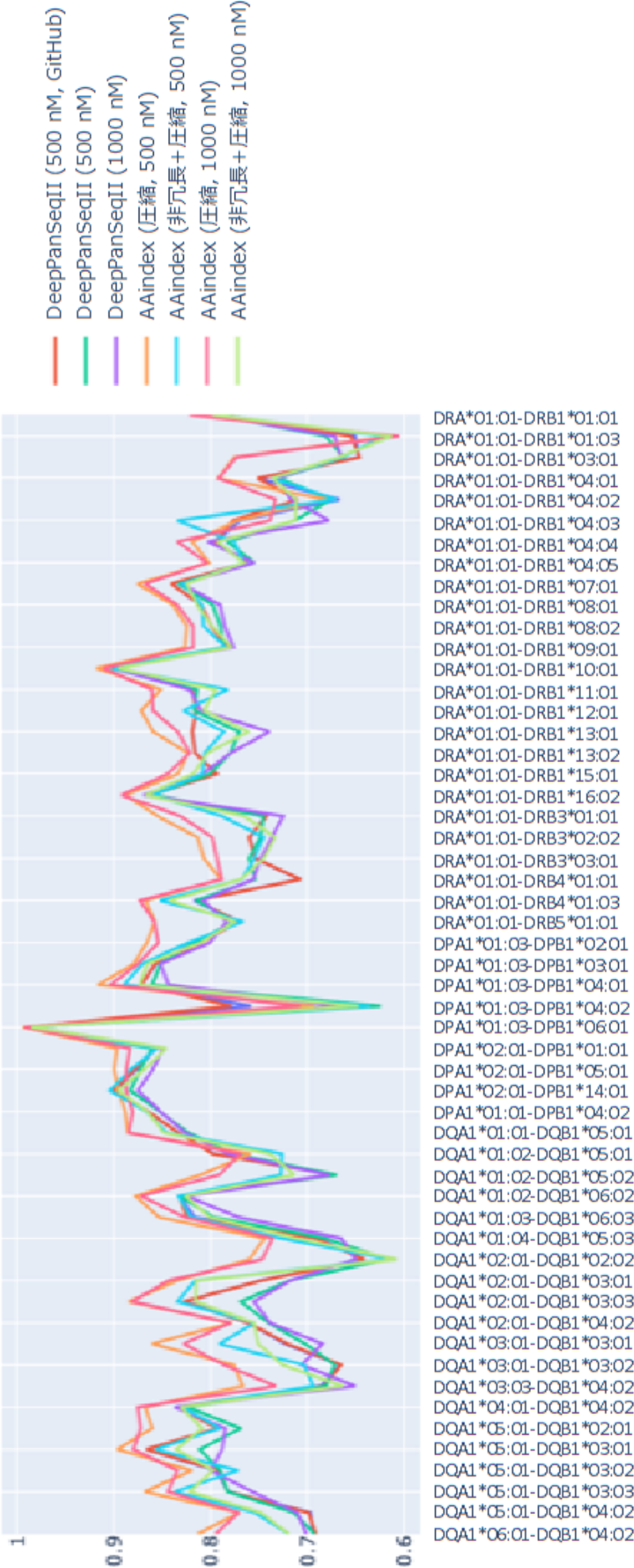


図5. 「IEDB2022データセット」に対するLOAO交差検証テストの結果（結合ペプチド数1000以上のアレル：19）

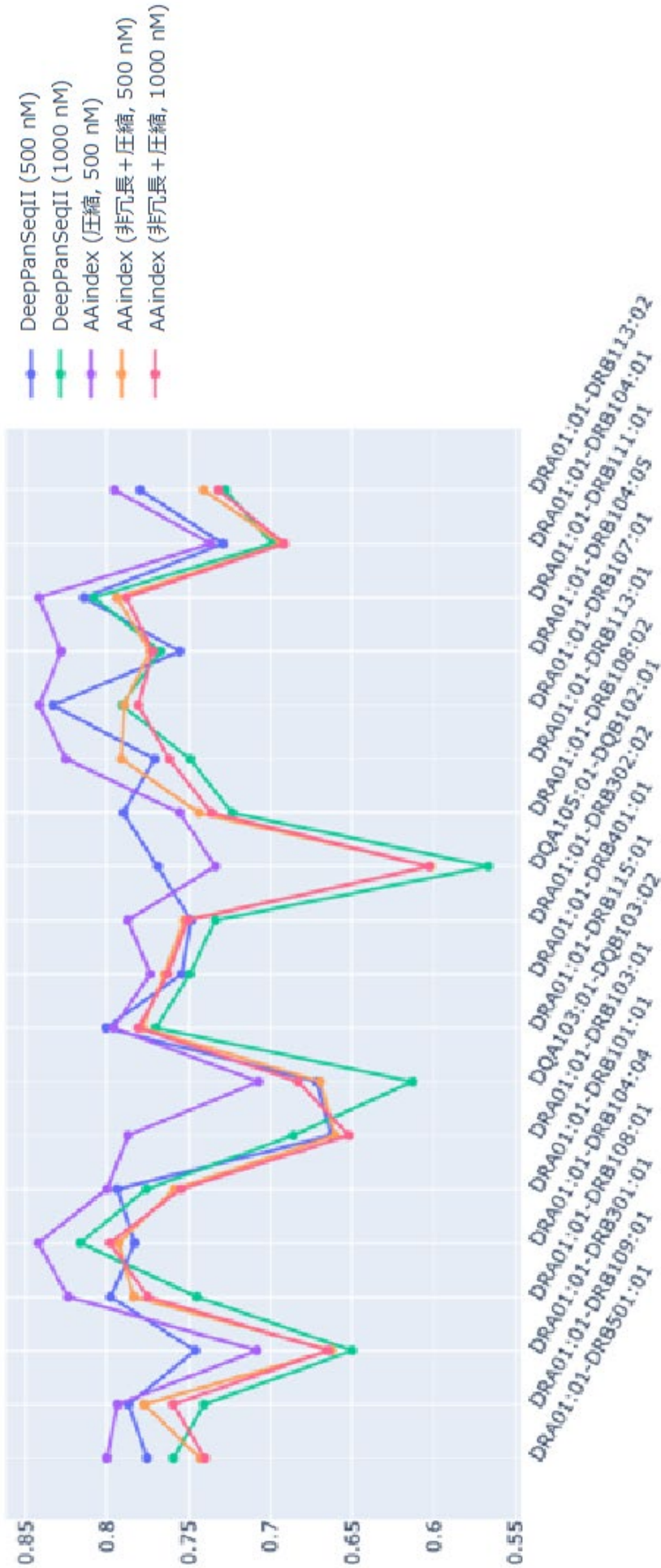
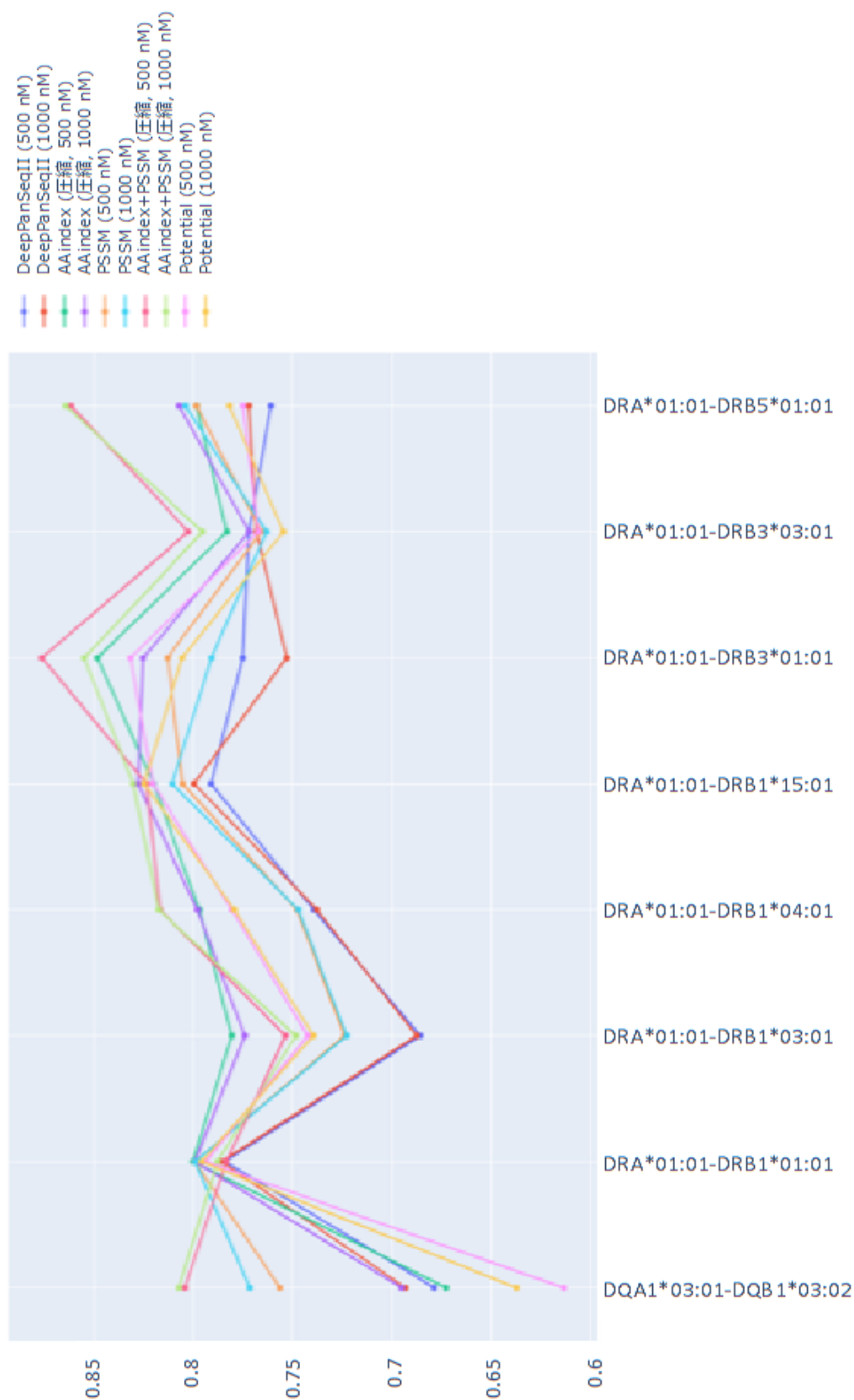


図6. 「PDB-HBデータセット」に対するLOAO交差検証テストの結果（結合ペプチド数1000以上のアレル：8）



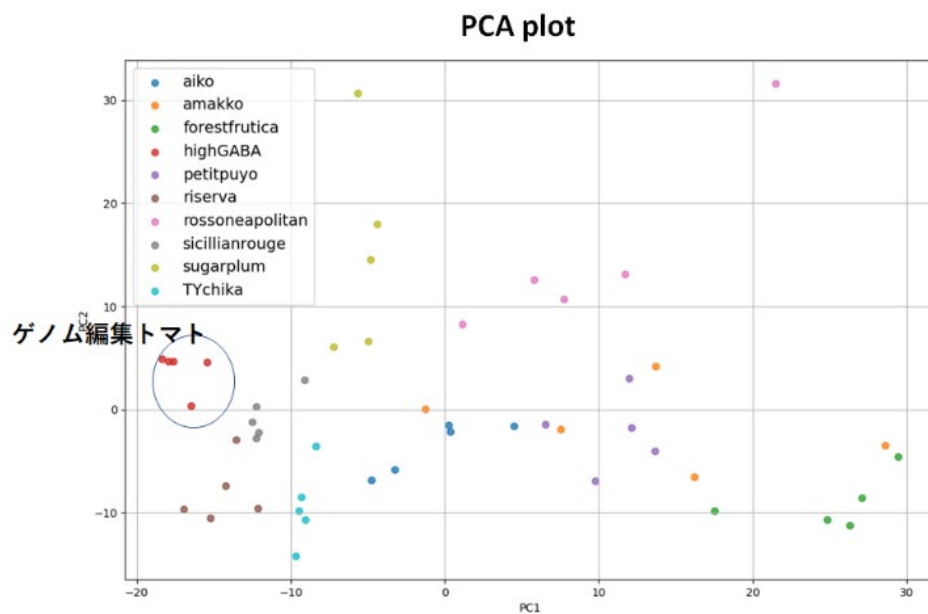
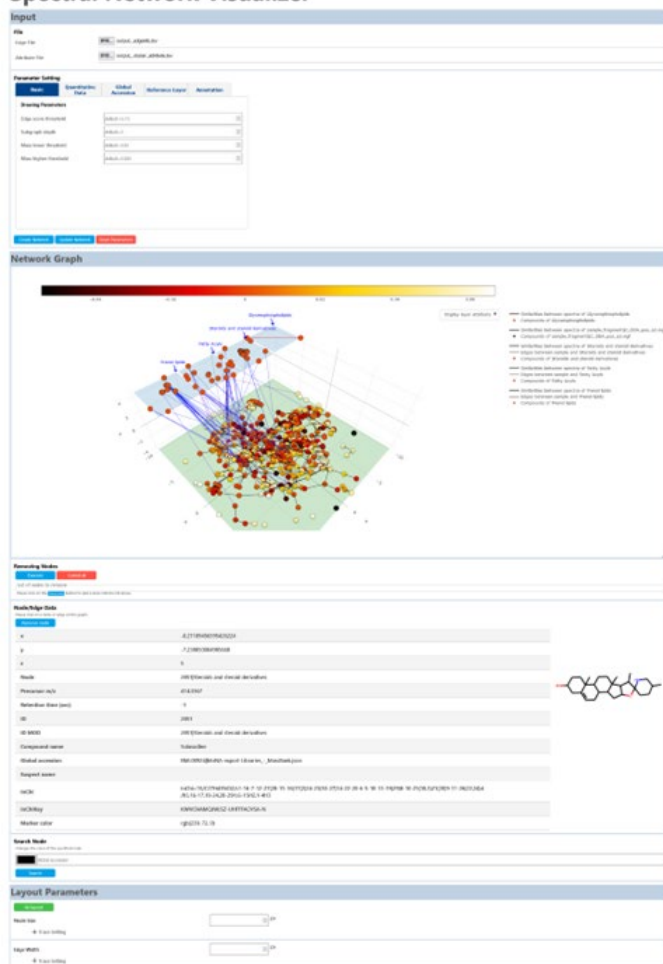


図1 トマトLC-MSデータの主成分解析



ネットワーク描写設定

ネットワーク可視化

ノード情報

グラフィックアウト変更

図2 Webアプリ版解析ツール

研究成果の刊行に関する一覧表
(令和4年度)

書籍

著者氏名	論文タイトル名	書籍全体の編集者名	書籍名	出版社名	出版地	出版年	ページ
小泉 望	リスクコミュニケーションのために求められること		ゲノム編集技術～実験上のポイント／産業利用に向けた研究開発動向と安全性周知	情報機構	東京	2023	p295 -p301
小泉 望、 山口 タ、 標葉隆馬			OECD 加盟国におけるゲノム編集作物に関するパブリック・エンゲージメント事例集	科学コミュニケーション研究所	神奈川	2022	

雑誌

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
Takeda, K.F., Yazawa, A., Yamaguchi, Y., Koizumi, N., & Shineha, R.	Comparison of public attitudes toward five alternative proteins in Japan.	<i>Food Quality and Preference.</i>	105	104787	2023
Takabatake R, Egi T, Soga K, Narushima J, Yoshida S, Shibata N, Nakamura K, Kondo K, Kishine M, Mano J, Kitta K	Development and interlaboratory validation of a novel reproducible qualitative method for GM soybeans using comparative Cq-based analysis for the revised non-GMO labeling system in Japan.	<i>Anal. Chem.</i>	94	13447-13454	2022

Soga K, Nakamura K, Egi T, Narushima J, Yoshiba S, Kishine M, Mano J, Kitta K, Takabatake R, Shibata N, Kondo K	Development and validation of a new robust detection method for low-content DNA using DDCq-based real-time PCR with optimized standard plasmids as a control sample.	<i>Anal. Chem.</i>	94	14475-14483	2022
Shibata N, Soga K, Sugino M, Nakamura K, Narushima J, Yoshiba S, Egi T, Takabatake R, Kondo K	Evaluation of conversion factor for rapid quantification of authorized genetically modified maize and soybean in Japan.	<i>BPB Reports.</i>	5	115-120	2022
Narushima J, Kimata S, Shiwa Y, Gondo T, Akimoto S, Soga K, Yoshiba S, Nakamura K, Shibata N, Kondo K	Unbiased prediction of off-target sites in genome-edited rice using SITE-Seq analysis on a web-based platform.	<i>Genes to Cells.</i>	27	706-718	2022
Goto K, Tamehiro N, Yoshida T, Hanada H, Sakuma T, Adachi R, Kondo K, Takeuchi I	AllerStat: Finding Statistically Significant Allergen-Specific Patterns in Protein Sequences by Machine Learning.	<i>J Biol Chem.</i>	submitted.		

厚生労働大臣 殿

機関名 国立医薬品食品衛生研究所
所属研究機関長 職 名 所長
氏 名 合田 幸広

次の職員の令和4年度厚生労働科学研究費の調査研究における、倫理審査状況及び利益相反等の管理については以下のとおりです。

1. 研究事業名 食品の安全確保推進研究事業
2. 研究課題名 新たなバイオテクノロジーを用いて得られた食品の安全性確保と
リスクコミュニケーションのための研究
3. 研究者名 (所属部署・職名) 生化学部・部長
(氏名・フリガナ) 近藤 一成・コンドウ カズナリ

4. 倫理審査の状況

	該当性の有無 有 無	左記で該当がある場合のみ記入 (※1)		
		審査済み	審査した機関	未審査 (※2)
人を対象とする生命科学・医学系研究に関する倫理指針 (※3)	☐ 有 ☒ 無	☐		☐
遺伝子治療等臨床研究に関する指針	☐ 有 ☒ 無	☐		☐
厚生労働省の所管する実施機関における動物実験等の実施に関する基本指針	☐ 有 ☒ 無	☐		☐
その他、該当する倫理指針があれば記入すること (指針の名称：)	☐ 有 ☒ 無	☐		☐

(※1) 当該研究者が当該研究を実施するに当たり遵守すべき倫理指針に関する倫理委員会の審査が済んでいる場合は、「審査済み」にチェックし一部若しくは全部の審査が完了していない場合は、「未審査」にチェックすること。

その他 (特記事項)

(※2) 未審査に場合は、その理由を記載すること。
(※3) 廃止前の「疫学研究に関する倫理指針」、「臨床研究に関する倫理指針」、「ヒトゲノム・遺伝子解析研究に関する倫理指針」、「人を対象とする医学系研究に関する倫理指針」に準拠する場合は、当該項目に記入すること。

5. 厚生労働分野の研究活動における不正行為への対応について

研究倫理教育の受講状況	受講 ☒ 未受講 ☐
-------------	---

6. 利益相反の管理

当研究機関におけるCOIの管理に関する規定の策定	有 ☒ 無 ☐ (無の場合はその理由：)
当研究機関におけるCOI委員会設置の有無	有 ☒ 無 ☐ (無の場合は委託先機関：)
当研究に係るCOIについての報告・審査の有無	有 ☒ 無 ☐ (無の場合はその理由：)
当研究に係るCOIについての指導・管理の有無	有 ☐ 無 ☒ (有の場合はその内容：)

厚生労働大臣 殿

機関名 国立医薬品食品衛生研究所
所属研究機関長 職 名 所長
氏 名 合田 幸広

次の職員の令和4年度厚生労働科学研究費の調査研究における、倫理審査状況及び利益相反等の管理については以下のとおりです。

1. 研究事業名 食品の安全確保推進研究事業
2. 研究課題名 新たなバイオテクノロジーを用いて得られた食品の安全性確保と
リスクコミュニケーションのための研究
3. 研究者名 (所属部署・職名) 生化学部・室長
(氏名・フリガナ) 柴田 識人・シバタ ノリヒト

4. 倫理審査の状況

	該当性の有無 有 無	左記で該当がある場合のみ記入 (※1)		
		審査済み	審査した機関	未審査 (※2)
人を対象とする生命科学・医学系研究に関する倫理指針 (※3)	☐ 有 ☒ 無	☐		☐
遺伝子治療等臨床研究に関する指針	☐ 有 ☒ 無	☐		☐
厚生労働省の所管する実施機関における動物実験等の実施に関する基本指針	☐ 有 ☒ 無	☐		☐
その他、該当する倫理指針があれば記入すること (指針の名称：)	☐ 有 ☒ 無	☐		☐

(※1) 当該研究者が当該研究を実施するに当たり遵守すべき倫理指針に関する倫理委員会の審査が済んでいる場合は、「審査済み」にチェックし一部若しくは全部の審査が完了していない場合は、「未審査」にチェックすること。

その他 (特記事項)

(※2) 未審査に場合は、その理由を記載すること。
(※3) 廃止前の「疫学研究に関する倫理指針」、「臨床研究に関する倫理指針」、「ヒトゲノム・遺伝子解析研究に関する倫理指針」、「人を対象とする医学系研究に関する倫理指針」に準拠する場合は、当該項目に記入すること。

5. 厚生労働分野の研究活動における不正行為への対応について

研究倫理教育の受講状況	受講 ☒ 未受講 ☐
-------------	---

6. 利益相反の管理

当研究機関におけるCOIの管理に関する規定の策定	有 ☒ 無 ☐ (無の場合はその理由：)
当研究機関におけるCOI委員会設置の有無	有 ☒ 無 ☐ (無の場合は委託先機関：)
当研究に係るCOIについての報告・審査の有無	有 ☒ 無 ☐ (無の場合はその理由：)
当研究に係るCOIについての指導・管理の有無	有 ☐ 無 ☒ (有の場合はその内容：)

厚生労働大臣 殿

機関名 国立医薬品食品衛生研究所
所属研究機関長 職 名 所長
氏 名 合田 幸広

次の職員の令和4年度厚生労働科学研究費の調査研究における、倫理審査状況及び利益相反等の管理については以下のとおりです。

1. 研究事業名 食品の安全確保推進研究事業
2. 研究課題名 新たなバイオテクノロジーを用いて得られた食品の安全性確保とリスクコミュニケーションのための研究
3. 研究者名 (所属部署・職名) 生化学部・室長
(氏名・フリガナ) 安達 玲子・アダチ レイコ

4. 倫理審査の状況

	該当性の有無 有 無	左記で該当がある場合のみ記入 (※1)		
		審査済み	審査した機関	未審査 (※2)
人を対象とする生命科学・医学系研究に関する倫理指針 (※3)	☐ 有 ☒ 無	☐		☐
遺伝子治療等臨床研究に関する指針	☐ 有 ☒ 無	☐		☐
厚生労働省の所管する実施機関における動物実験等の実施に関する基本指針	☐ 有 ☒ 無	☐		☐
その他、該当する倫理指針があれば記入すること (指針の名称：)	☐ 有 ☒ 無	☐		☐

(※1) 当該研究者が当該研究を実施するに当たり遵守すべき倫理指針に関する倫理委員会の審査が済んでいる場合は、「審査済み」にチェックし一部若しくは全部の審査が完了していない場合は、「未審査」にチェックすること。

その他 (特記事項)

(※2) 未審査に場合は、その理由を記載すること。
(※3) 廃止前の「疫学研究に関する倫理指針」、「臨床研究に関する倫理指針」、「ヒトゲノム・遺伝子解析研究に関する倫理指針」、「人を対象とする医学系研究に関する倫理指針」に準拠する場合は、当該項目に記入すること。

5. 厚生労働分野の研究活動における不正行為への対応について

研究倫理教育の受講状況	受講 ☒ 未受講 ☐
-------------	---

6. 利益相反の管理

当研究機関におけるCOIの管理に関する規定の策定	有 ☒ 無 ☐ (無の場合はその理由：)
当研究機関におけるCOI委員会設置の有無	有 ☒ 無 ☐ (無の場合は委託先機関：)
当研究に係るCOIについての報告・審査の有無	有 ☒ 無 ☐ (無の場合はその理由：)
当研究に係るCOIについての指導・管理の有無	有 ☐ 無 ☒ (有の場合はその内容：)

(留意事項) ・該当する□にチェックを入れること。
・分担研究者の所属する機関の長も作成すること。

厚生労働大臣 殿

機関名 国立医薬品食品衛生研究所
所属研究機関長 職 名 所長
氏 名 合田 幸広

次の職員の令和4年度厚生労働科学研究費の調査研究における、倫理審査状況及び利益相反等の管理については以下のとおりです。

1. 研究事業名 食品の安全確保推進研究事業
2. 研究課題名 新たなバイオテクノロジーを用いて得られた食品の安全性確保と
リスクコミュニケーションのための研究
3. 研究者名 (所属部署・職名) 生化学部・主任研究官
(氏名・フリガナ) 吉場 聡子・ヨシバ サトコ

4. 倫理審査の状況

	該当性の有無 有 無	左記で該当がある場合のみ記入 (※1)		
		審査済み	審査した機関	未審査 (※2)
人を対象とする生命科学・医学系研究に関する倫理指針 (※3)	☐ 有 ☒ 無	☐		☐
遺伝子治療等臨床研究に関する指針	☐ 有 ☒ 無	☐		☐
厚生労働省の所管する実施機関における動物実験等の実施に関する基本指針	☐ 有 ☒ 無	☐		☐
その他、該当する倫理指針があれば記入すること (指針の名称：)	☐ 有 ☒ 無	☐		☐

(※1) 当該研究者が当該研究を実施するに当たり遵守すべき倫理指針に関する倫理委員会の審査が済んでいる場合は、「審査済み」にチェックし一部若しくは全部の審査が完了していない場合は、「未審査」にチェックすること。

その他 (特記事項)

(※2) 未審査に場合は、その理由を記載すること。
(※3) 廃止前の「疫学研究に関する倫理指針」、「臨床研究に関する倫理指針」、「ヒトゲノム・遺伝子解析研究に関する倫理指針」、「人を対象とする医学系研究に関する倫理指針」に準拠する場合は、当該項目に記入すること。

5. 厚生労働分野の研究活動における不正行為への対応について

研究倫理教育の受講状況	受講 ☒ 未受講 ☐
-------------	---

6. 利益相反の管理

当研究機関におけるCOIの管理に関する規定の策定	有 ☒ 無 ☐ (無の場合はその理由：)
当研究機関におけるCOI委員会設置の有無	有 ☒ 無 ☐ (無の場合は委託先機関：)
当研究に係るCOIについての報告・審査の有無	有 ☒ 無 ☐ (無の場合はその理由：)
当研究に係るCOIについての指導・管理の有無	有 ☐ 無 ☒ (有の場合はその内容：)

2023年5月17日

厚生労働大臣
~~(国立医薬品食品衛生研究所長)~~ 殿
~~(国立保健医療科学院長)~~

機関名 大阪公立大学

所属研究機関長 職 名 学長

氏 名 辰巳砂 昌弘

次の職員の（令和）4年度厚生労働科学研究費の調査研究における、倫理審査状況及び利益相反等の管理については以下のとおりです。

1. 研究事業名 食品の安全確保推進研究事業

2. 研究課題名 新たなバイオテクノロジーを用いて得られた食品の安全性確保とリスクコミュニケーションのための研究

3. 研究者名 （所属部署・職名） 農学研究科 教授

（氏名・フリガナ） 小泉 望・コイズミ ノゾム

4. 倫理審査の状況

	該当性の有無		左記で該当がある場合のみ記入（※1）		
	有	無	審査済み	審査した機関	未審査（※2）
人を対象とする生命科学・医学系研究に関する倫理指針（※3）	☐	☒	☐		☐
遺伝子治療等臨床研究に関する指針	☐	☒	☐		☐
厚生労働省の所管する実施機関における動物実験等の実施に関する基本指針	☐	☒	☐		☐
その他、該当する倫理指針があれば記入すること（指針の名称：）	☐	☒	☐		☐

（※1）当該研究者が当該研究を実施するに当たり遵守すべき倫理指針に関する倫理委員会の審査が済んでいる場合は、「審査済み」にチェックし一部若しくは全部の審査が完了していない場合は、「未審査」にチェックすること。

その他（特記事項）

（※2）未審査に場合は、その理由を記載すること。

（※3）廃止前の「疫学研究に関する倫理指針」、「臨床研究に関する倫理指針」、「ヒトゲノム・遺伝子解析研究に関する倫理指針」、「人を対象とする医学系研究に関する倫理指針」に準拠する場合は、当該項目に記入すること。

5. 厚生労働分野の研究活動における不正行為への対応について

研究倫理教育の受講状況	受講 ☒ 未受講 ☐
-------------	---

6. 利益相反の管理

当研究機関におけるCOIの管理に関する規定の策定	有 ☒ 無 ☐ （無の場合はその理由：）
当研究機関におけるCOI委員会設置の有無	有 ☒ 無 ☐ （無の場合は委託先機関：）
当研究に係るCOIについての報告・審査の有無	有 ☒ 無 ☐ （無の場合はその理由：）
当研究に係るCOIについての指導・管理の有無	有 ☐ 無 ☒ （有の場合はその内容：）

- (留意事項)
- ・該当する□にチェックを入れること。
 - ・分担研究者の所属する機関の長も作成すること。

令和5年 4月 17日
(公 印 省 略)

厚生労働大臣 殿

機関名 国立研究開発法人産業技術総合研究所

所属研究機関長 職 名 理事長

氏 名 石村 和彦

次の職員の令和4年度厚生労働科学研究費の調査研究における、倫理審査状況及び利益相反等の管理については以下のとおりです。

1. 研究事業名 食品の安全確保推進研究事業
2. 研究課題名 新たなバイオテクノロジーを用いて得られた食品の安全性確保とリスクコミュニケーションのための研究
3. 研究者名 (所属部署・職名) 人工知能研究センター・研究チーム長
- (氏名・フリガナ) 富井 健太郎 (トミイ ケンタロウ)

4. 倫理審査の状況

	該当性の有無		左記で該当がある場合のみ記入 (※1)		
	有	無	審査済み	審査した機関	未審査 (※2)
人を対象とする生命科学・医学系研究に関する倫理指針 (※3)	☐	☒	☐		☐
遺伝子治療等臨床研究に関する指針	☐	☒	☐		☐
厚生労働省の所管する実施機関における動物実験等の実施に関する基本指針	☐	☒	☐		☐
その他、該当する倫理指針があれば記入すること (指針の名称：)	☐	☒	☐		☐

(※1) 当該研究者が当該研究を実施するに当たり遵守すべき倫理指針に関する倫理委員会の審査が済んでいる場合は、「審査済み」にチェックし一部若しくは全部の審査が完了していない場合は、「未審査」にチェックすること。

その他 (特記事項)

(※2) 未審査に場合は、その理由を記載すること。
(※3) 廃止前の「疫学研究に関する倫理指針」、「臨床研究に関する倫理指針」、「ヒトゲノム・遺伝子解析研究に関する倫理指針」、「人を対象とする医学系研究に関する倫理指針」に準拠する場合は、当該項目に記入すること。

5. 厚生労働分野の研究活動における不正行為への対応について

研究倫理教育の受講状況	受講 ☒ 未受講 ☐
-------------	---

6. 利益相反の管理

当研究機関におけるCOIの管理に関する規定の策定	有 ☒ 無 ☐ (無の場合はその理由：)
当研究機関におけるCOI委員会設置の有無	有 ☒ 無 ☐ (無の場合は委託先機関：)
当研究に係るCOIについての報告・審査の有無	有 ☒ 無 ☐ (無の場合はその理由：)
当研究に係るCOIについての指導・管理の有無	有 ☐ 無 ☒ (有の場合はその内容：)

(留意事項) ・該当する□にチェックを入れること。
・分担研究者の所属する機関の長も作成すること。

機関名 学校法人沖縄科学技術大学院大学学園

所属研究機関長 職 名 理事長

氏 名 ピーター・グルース

次の職員の令和4年度厚生労働科学研究費の調査研究における、倫理審査状況及び利益相反等の管理については以下のとおりです。

1. 研究事業名 食品の安全確保推進研究事業
2. 研究課題名 新たなバイオテクノロジーを用いて得られた食品の安全性確保とリスクコミュニケーションのための研究
3. 研究者名 (所属部署・職名) 進化神経生物学ユニット・グループリーダー
(氏名・フリガナ) 早川 英介 (ハヤカワ エイスケ)

	該当性の有無		左記で該当がある場合のみ記入 (※1)		
	有	無	審査済み	審査した機関	未審査 (※2)
人を対象とする生命科学・医学系研究に関する倫理指針 (※3)	☐	☒	☐		☐
遺伝子治療等臨床研究に関する指針	☐	☒	☐		☐
厚生労働省の所管する実施機関における動物実験等の実施に関する基本指針	☐	☒	☐		☐
その他、該当する倫理指針があれば記入すること (指針の名称：)	☐	☒	☐		☐

(※1) 当該研究者が当該研究を実施するに当たり遵守すべき倫理指針に関する倫理委員会の審査が済んでいる場合は、「審査済み」にチェックし一部若しくは全部の審査が完了していない場合は、「未審査」にチェックすること。

その他 (特記事項)

(※2) 未審査に場合は、その理由を記載すること。

(※3) 廃止前の「疫学研究に関する倫理指針」、「臨床研究に関する倫理指針」、「ヒトゲノム・遺伝子解析研究に関する倫理指針」、「人を対象とする医学系研究に関する倫理指針」に準拠する場合は、当該項目に記入すること。

5. 厚生労働分野の研究活動における不正行為への対応について	
研究倫理教育の受講状況	受講 ☒ 未受講 ☐

6. 利益相反の管理	
当研究機関におけるCOIの管理に関する規定の策定	有 ☒ 無 ☐ (無の場合はその理由：)
当研究機関におけるCOI委員会設置の有無	有 ☒ 無 ☐ (無の場合は委託先機関：)
当研究に係るCOIについての報告・審査の有無	有 ☒ 無 ☐ (無の場合はその理由：)
当研究に係るCOIについての指導・管理の有無	有 ☐ 無 ☒ (有の場合はその内容：)

(留意事項) ・該当する□にチェックを入れること。
・分担研究者の所属する機関の長も作成すること。