

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）
研究分担者 木下 政人 （京都大学）
研究分担者 小泉 望 （大阪府立大学）
研究分担者 竹内 一郎 （名古屋工業大学）
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 為広 紀正 （国立医薬品食品衛生研究所）
研究分担者 中村 公亮 （国立医薬品食品衛生研究所）

研究要旨：

ゲノム編集技術を利用した作物（ゲノム編集作物）から作られる新たな食品の研究開発が国内外で活発に行なわれている。しかし、安全性審査が必要な従来の遺伝子組換え食品とは異なり、導入遺伝子は存在しない。ゲノム編集食品の届出制度が平成31年度（令和元年）10月に開始となり食品衛生法上の取扱いも明確化された。今後制度実施にあたり、そこに至った科学的背景等を整理しておくことが必要である。また、安全性確認を科学的なエビデンスをもって行うための手法開発整備も重要な課題である。本研究では、種々の手法による遺伝子改変の影響、ゲノム編集作物の開発状況や規制状況の情報収集を行い施策に反映するとともに、安全性確認に必要な項目や問題点を明らかにした。また、ゲノム編集技術や合成生物学など新たなバイオテクノロジー技術を用いた新開発食品の安全性を確認するために必要な新たな手法の開発検討を行った。ゲノム編集ではオフターゲットが課題になっていることから、配列類似性によらないバイアスのないゲノム全体のDNA切断部位を検出する方法の開発、非アレルゲンタンパクのアミノ酸情報も加味し、既知アレルゲンタンパクとの相同性に依存しない人工知能を用いた全く新たなタンパクアレルゲン性予測アルゴリズムの開発、新開発食品試料中に出現する未知成分の質量分析インフォマティクスを用いた同定手法の開発、の検討を行った。その結果、ゲノム解析では、SITE-seq法を出発点にしたオフターゲット検出法を確立するとともに、webツールを開発した。新規アレルゲン性予測では、アレルゲンタンパクにのみ出現するアミノ酸配列パターンを抽出、データセットの改良を行いながらアレルゲン性予測が従来よりも精度が高いことが確認できた。アレルゲン性とも関連するタンパクの分解性試験について、pH、酵素濃度について細かく設定して検討した。その結果、分解されやすいタンパクにおいてもペプシン濃度よりもpH変化が分解性に大きく影響することが分かった。質量分析インフォマティクスでは、基になる高品質な質量分析スペクトル情報が必要なため公共データベース、および標品測定からスペクトル情報を取得することでデータベース化するとともに、Pythonプログラム等を用いて結果をネットワーク化して可視化できた。さらに、ゲノム編集マダイ、トラフグ開発において、ゲノム編集食品の事前相談・届出制度にある必要項目に沿った十分なデータを取得してその安全性を確認した。

A. 研究目的

ゲノム編集技術を応用した新たな食品（ゲノム編集食品）の研究開発が国内外で活発に行なわれている。ゲノム編集食品では、従来の遺伝子組換え食品のような外来遺伝子を導入することはなく、もともとの性打つが有する内在性遺伝子の配列を数塩基欠失により機能欠失させることで新た

な形質（もち性向上、筋肉量増加、GABA量増加など）を付与することが期待されている。そのため、国民受容の改善の点でも大きく期待されている。また、合成生物学を利用した物質生産も米国を中心に活発に研究されている。酵母などの微生物に、新たな物質生産に必要な多数の遺伝子を導入することで、その生物が元来合成できない化合物の生

産が可能になっている。

ゲノム編集食品では、安全性評価の対象は内在性遺伝子改変に伴う塩基配列変化とゲノム編集時の意図しない変化（いわゆる、オフターゲット）となると考えられる。一方で、合成生物学利用作物では、生合成経路に関わる多数の遺伝子を導入するため、安全性評価対象は導入した遺伝子群とその影響であるが、組換え範囲が大きい従来からの遺伝子組換え前後の比較による実質的同等性の考え方が適用できないことも想定される。

従来の遺伝子組換え食品における安全性確認の基本的な考え方は、組換えをする前後の作物を用いた比較解析からの実質的同等性（リスクが組換え前と比較して同等かそれ以下）で判断している。すなわち、導入遺伝子に関する分子生物学的特性、ヒトによる長期間にわたる安全な食経験、構成成分変化、使用方法等について同等性を失っていないかである。しかし、改変されるのは内在性遺伝子上における塩基の挿入・欠失であり、標的部位（オンターゲット）での変化が十分解析されることが重要で、その上で潜在的なリスクは意図しない改変であるオフターゲットの影響である。オン・オフターゲット部位での変化によって生じるリスクは、新たな毒性・アレルギータンパクの生成である。ゲノム解析が進んだ現在においても、ゲノム配列のみから毒性タンパクやアレルギータンパクが生成しないことを明らかにするのは容易ではない。また、意図しない有害成分産生の可能性があったとしても、現在の質量分析を用いた解析では未知ピークの同定や推定は困難である。さらに、タンパクアレルギー性の確認も、現在実行可能な *in silico* 解析は既知のアレルギータンパク質との相同性比較のみであり、相同性がない新規アレルギータンパク質の予測や非天然型アミノ酸から構成されるタンパクのアレルギー性を予測することは極めて難しい。このような状況を鑑みて、ゲノム編集食品の開発状況情報収集をもとにしたケーススタディーや開発者との連携で申請側の問題点を明らかにするとともに、上記のゲノム編集食品や合成生物学利用食品の安全性確認のために必要な評価手法の新たな開発が急務と考えられた。

本研究では、手法開発において、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、全ゲノム解析をすることなく潜在的な DNA 2 本鎖切断部位を網羅的に検出する手法、

新たな成分が産生した場合の質量分析インフォマティクスを用いた成分同定あるいは基本構造推定手法、人工知能を活用して相同性がないアレルギータンパクの予測や非天然型アミノ酸から構成されるタンパクのアレルギー性を予測する手法、の開発検討を行う。また、諸外国の規制・ゲノム編集・合成生物学に関する情報収集を行い、その結果から仮想的モデル生物を用いたケーススタディーを行い、安全性確認に必要なデータや問題点を明らかにすることとした。また、平成 31 年度（令和元年）10 月に、ゲノム編集食品の届出・事前相談制度が開始されたことから、それに伴う科学的知見の整理に必要な文献情報を整理した。リスクコミュニケーションにおいては、ゲノム編集技術に関する知識がほとんど無い層をターゲットにした、チラシ、パンフレットの作成を行う。

B. 研究方法

(1) ゲノム編集に関する情報収集解析、ケーススタディーおよびアレルギー分解性の検討

植物・動物（細胞）を主な対象に、自然変異、放射線による突然変異誘導、ゲノム編集による変異誘導について、定量的な解析がされた科学論文を中心に調査整理した。

(2) リスクコミュニケーションに関する研究

専門的な知識を持たない一般の人の、遺伝子組換え食品やゲノム編集食品の疑問や不安が大きいため、疑問点の整理、専門家と一般の人での認識の違いを調査しながらパンフレットや小冊子の作成を行った。

(3) ゲノム網羅的に DNA 2 本鎖切断部位を検出する手法とツールの開発検討

現在、ゲノム編集技術を用いた時のオフターゲットについては、標的部位（オンターゲット）と類似したゲノム上の場所を *in silico* に検索することしかできない。そのため、オンターゲットと配列類似性がない部位でのオフターゲットやその影響は検出把握できない。これを解決するために、既報である SITE-Seq 法を出発点に動物や植物に適用可能で簡便かつ再現性の高い、ゲノムワイドな DNA 2 本鎖切断部位解析手法とそれを利用者が使用する環境ツールの開発検討を行った。

(4) 質量分析インフォマティクスによる化合物

同定

化合物の質量スペクトル（フラグメントスペクトル）類似度をもとにして、食品中の未知化合物の検出と構造推定を行うシステムの構築を行った。データ解析のワークフローとして、安全と考えられる食品試料（非ゲノム編集体）と分析対象の試料（ゲノム編集体等）の液体クロマトグラフィー質量分析データから得られる化合物イオンの比較定量値およびフラグメントスペクトルを解析データとして用いた。化合物の構造・クラス推定に不可欠なスペクトルデータの取得拡充に努めて、独自の代謝物の標準品 300 種類を質量分析装置 Q-exactive HF (Thermo Fischer Science) で分析したライブラリを含めた統合フラグメントスペクトルデータベースを構築した。データ解析には、プログラミング言語:Python を使い、ケモインフォマティクスライブラリ:RDKit、データ可視化ライブラリ:Plotly 等と連携させることでデータ解析環境の構築を行った。

(5) アレルギーデータベース ADFS のアップデート、および新規タンパクアレルギー性予測に必要な情報の作製

現アレルギーデータベース ADFS の情報更新のため、アレルギー情報の追加と 2018 年 6 月から 2019 年 5 月までの 1 年間に NCBI PubMed に収載された論文からのエピトープ情報を追加した。また、ADFS サイトの脆弱性対策のために、JAVA、MySQL などのバージョンアップの他スクリプトを改訂した。

新規アレルギー予測手法の検討のために、アレルギーデータベース Compare からアレルギータンパク情報を、非アレルギータンパク情報は、Uniprot からアレルギー情報を除くことによって構築した。

(6) 機械学習を用いた新規タンパクアレルギー性予測手法とツールの開発

既知のアレルギータンパクのアミノ酸配列のみに依存しない、新たなアレルギー予測システム構築のために、既知アレルギータンパクのほか非アレルギータンパクのアミノ酸配列情報を加えたデータセットをもとに検討を行うこととした。

使用するデータセットについて、食品および非食品タンパク質を追加して検討した。アレルギータンパクに特徴的なパターンの抽出について、デ

ータマイニング分野の技術（系列マイニング）を利用して、アレルギーにのみ出現する部分アミノ酸配列を検索した。また、アレルギーには食品・非食品タンパクを含むため、非アレルギーデータにも食品・非食品タンパクを含むデータを構築することを検討した。予測システムの構築は上述の Leave-Food-Out クロスバリデーションを利用した教師あり学習によって行った。さらに、2019 年度は、さらに、抽出されたパターンの生物学的な考察として、既存のエピトープとの一致度の確認や、結合性の確認なども行った。

(7) ゲノム編集生物作製における現象解析と規制の進め方

ミオスタチン遺伝子破壊マダイおよびトラフグ各 3 系統、レプチン受容体遺伝子破壊トラフグ 2 系統、メラノコルチン 4 型受容体遺伝子トラフグ 1 系統について、アレルギー性、オフターゲット、外来遺伝子残存性および継代安定性について検討した。改変部位での予想される全アミノ酸配列、新生ペプチドとその直上 10 アミノ酸部分、および、塩基欠失部位を挟んだ両側の終止コドン内で予想されるペプチドを用い、web 上のアレルギー検索サイトによりアレルギー性の有無を検討した。オフターゲットおよび外来遺伝子残存性は、レプチン受容体遺伝子破壊トラフグの全ゲノム解析をもとに解析した。

(8) 人材育成（統計学、バイオインフォマティクス、AI 分野）

分担研究者および協力研究者と共同で行うことで、インフォマティクス関連技術の取得に努めた。

C. 研究結果および考察

(1) ゲノム編集に関する情報収集解析、ケーススタディーおよびアレルギー分解性の検討

2018 年途中までの新規育種技術 (NBT) を用いた動物および植物について調査した結果、動物では食品用途（食用）は全 39 報中 20 報であった。使用技術はほとんど CRISPR/Cas9 およびその改変型であり、主な獲得形質はブタの筋肉量増大やウイルス抵抗性である。植物でも使用技術はほとんど CRISPR/Cas9 で、食用は全 122 報中 42 報、研究用は 76 報であった。食用では、トマトの保存性向上や種子がなくても果実ができるもの、コムギの光合成能向上やうどんこ病抵抗性、イネの除草剤耐

性、イネでは収量の増加のほかウイルス抵抗性キュウリなどがある。詳細は分担報告書を参照のこと。ケーススタディーでは、開発直近の筋肉量増大マダイやフグ、もち性向上トウモロコシの実際の事例からとゲノム編集技術で仮想の農作物等を設定して、確認すべき事項や問題点を明らかにした。詳細は分担報告書に記載しているので参照のこと。

(2) リスクコミュニケーションに関する研究

一般人 4,000 人を対象にした Web 調査の結果から、ゲノム編集技術に関して知らないかあまり知らないが 8 割に上り、ゲノム編集食品を食べたい人は 1 割程度、またゲノム編集食品の安全性についての専門家の意見を信頼できるとしたのは 2 割に満たないことが分かった。社会受容についてリスクの程度や対応、発生確率などが重要な事項と考えていることが伺える結果となった。

詳細は分担報告書及び別添資料に記載している。

(3) ゲノム網羅的に DNA 2 本鎖切断部位を検出する手法とツールの開発検討

標的部位と類似していない箇所のオフターゲットの検出が可能な unbiased な手法を、SITE-Seq 法をもとに、イネもとに詳細に検討した。ALS 遺伝子を標的にした実験結果から、標的部位が切断された実験条件 (Cas9 濃度が 64、256 nM) で検出されたオフターゲット部位は 9 か所あった。このうち 4 つは PAM 配列も存在することからオフターゲットの候補と考えられた。一方で、ミスマッチ数は 4 塩基以上で確率的にはかなり小さいと推察された。このうち、PAM を含む 2 つのオフターゲット候補について切断効率を算出すると、オンターゲット 74% に対して、13% (4 塩基ミスマッチあり) および 6% (7 塩基ミスマッチあり) であった。これらは通常のオンラインツール (CRISPRdirect や Cas-OFFinder など) の通常の検索条件では検査されない。同様に、ほかの 6 か所についても同様の結果であった。以上の結果は、オフターゲットのオンライン検索ではすべてのオフターゲットを検出することはできないことを示している。今回の実験結果は、主に抽出 DNA を用いた結果であり *in vivo* においては DNA 修復されて検出されない可能性も考えられるが、ゲノム編集技術の一つ CRISPR/Cas9 は技術が持つ本質的性質として、ミスマッチが多

くても切断されることを示している (この場合でも 3' 側 10 塩基で見れば 1、2 塩基ミスマッチが主である)。最終的に検出されるかどうかは、用いる生物の DNA 修復能力に大きく依存するため、評価においてはその点も考慮に入れて行うべきと考えられた。

詳細は分担報告書に記載している。

(4) 質量分析インフォマティクスによる化合物同定

本研究では、化合物の質量スペクトル類似度をもとに試料中の未知化合物の検出と構造推定を行う解析システムの構築を行うための検討を行った。既存のスペクトルライブラリに加え標準品・標準試料から大量の質量スペクトルライブラリを取得すること統合スペクトルライブラリの拡充を行うとともに、比較定量情報を反映した可視化機能の実装など解析ツールの高機能化を行った。

代表的な食品・モデル植物 30 種 (大豆、トマト、ジャガイモ等) に関しては低分子化合物の抽出と実際の測定によるスペクトルデータ取得を行った。生物種一代謝物関係データベース (KNpSACK) から試料ごとの代謝物情報を抽出し、スペクトルデータと照合することで標準試料由来のスペクトルデータベースを構築した。今後、実際の試料 (遺伝子組換え前後の試料) などを使用して、変化のある成分の同定、推定がどの程度可能か検討する。

(5) アレルゲンデータベース ADFS のアップデート、および新規タンパクアレルゲン性予測に必要な情報の作製

アレルゲン情報は、AllergenOnline の登録アレルゲンと統合するためアップデートを行った。エピトープ配列は、キーワード検索により抽出した 20 報について、アレルゲン・エピトープ情報が記載されている 10 報についてピアレビューを行った。その結果、7 報の論文から 7 種のアレルゲンについて、総数 22 のエピトープ情報を新たに追加した。

新規予測法のために、アレルゲンデータベース Compare からアレルゲンタンパク情報 2038 種入手し、アレルギー表示が義務付けされている特定原材料 7 品目並びに推奨されている原材料のうち 4 品目の非アレルゲンタンパク配列情報について UniProt から 10577 種を取得した。非アレルゲン学習データの種類を増やして解析ができるよう

に、アレルゲンとして登録された全ての種について情報を取得し解析できるよう調整したが、Uniprot から取得したデータからすべてのアレルゲン情報を削除して、非アレルゲン情報を作成するのは困難と考えられた。

(6) 機械学習を用いた新規タンパクアレルゲン性予測手法とツールの開発

特定の食物に頻出するアミノ酸部分配列を誤ってアレルゲン特異的パターンとして抽出してしまうリスクを避けるために、順に一つの作物を除き残りの食物で訓練する Leave-Food-Out クロスバリデーションを行った。抽出したアレルゲンパターンについて、アレルゲン性タンパク質が多くのアレルゲン特異的パターンを含んでいることが確認できた。実際、これらのアレルゲン特異的パターンの生物学的特徴を調べたところ、既知のエピトープと類似していることが確認されている。従来の予測方法に比べて、本研究で構築した方法ではおおむねすべての場合において最もよい判定・予測性能を示していることが確認できた。次年度はさらにほかのアプローチとの比較も行うことで本システムの有効性の実証を行う。

(7) ゲノム編集生物作製における現象解析と規制の進め方

ミオスタチン遺伝子破壊したマダイおよびトラフグでは、E-value<0.05 以下においていずれの系統もアレルゲン性が疑われるアミノ酸配列は検出されなかった。レプチン受容体遺伝子破壊トラフグにおいても、E-value<1 ではイネの α -アミラーゼと相同性が認められたが、E-value<0.05 以下においていずれの系統もアレルゲン性が疑われるアミノ酸配列は検出されなかった。

レプチン受容体遺伝子破壊トラフグにおいてオフターゲットおよび外来遺伝子残存性について解析した結果、オフターゲットによる欠失は観察されなかった。また、残存性については NGS による全ゲノム解析データからのリードを用いたベクターにマッピングさせることで検索した結果、ベクターのバックボーン配列の一部 (ColE1 ori) が検出されたがゲノム編集魚および野生型魚の両方で観察されていることから、トラフグゲノム内に一般的に侵入した細菌断片であると思われる。それ以外の外来遺伝子とその断片配列は認められなかった。その他に可食部のメタボローム解析を実施

した。詳細は分担報告書に記載している。

D. 結論

自然変異、放射線による突然変異育種、ゲノム編集技術の各技術について放射線による突然変異育種は、後代交配を経ることで最終的な変異は、考えられていたよりも小さく数塩基の変異が中心であるのに対して、ゲノム編集により変異は同様に小さいものの、その頻度は非常に高いため、相対的に意図しない変異頻度も高いと考えられた。ゲノム編集作物では、イネの研究が非常に活発であるほかに多様な植物で研究開発が進んでいることが分かった。

リスクコミュニケーションでは、専門的知識がない人に向けての新しいパンフレット(厚労省用)を作成した。各種の説明会や web 形式アンケートから、ゲノム編集食品を知っているのは 2 割で十分に知られていない。また、専門家の意見について信頼できるとした割合も 2 割以下など、一般の人への理解や信頼性が低く、これを改善することが不可欠である。対象別にきめ細かいコミュニケーションが必要である。

網羅的なオフターゲット検出法 SITE-Seq は、イネでも有効であることを確認した。SITE-Seq 法は、オンラインオフターゲット予測ツールでは予測不可能であったオフターゲットの予測が可能であり、ゲノム編集食品の安全性評価時に有用な手法と考えられる。

意図しない新たな代謝物を同定予測するための手法について、質量分析インフォマティクスを用いて試料間比較による代謝物の変動と連携した化合物のクラス推定・可視化という解析フレームワークを用いて確立した。

既存の ADFS データベースの情報更新を行うとともに、新規アレルゲン性予測に活用するためにタンパク情報を整備した。これらのデータセットを用いて、アレルゲンに特徴的なアミノ酸モチーフを検索した。これまでに構築したアレルゲン性判定・予測システムのプロトタイプの問題点を抽出し、改良を加えた結果、訓練データベースの大規模化と高精度化、アレルゲン特異的パターンの信頼性向上、判定・予測システムの精度向上が可能となった。

ミオスタチンゲノム編集マダイ、トラフグ、レプチン受容体ゲノム編集トラフグ、メラノコルチン4型受容体ゲノム編集トラフグにおいて、予想

されるタンパク質およびペプチドはアレルギー性を示さないこと、オフターゲット・外来遺伝子残存性もないことが示され、届出・事前相談における必要項目をカバーするデータの取得ができた。

E. 業績

論文、学会発表、説明会、リスコミ開催などの業績詳細は、各分担報告書に記載。