

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3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性 DNA の残存の有無を網羅的に調べる方法として、全ゲノムシーケンズ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だ。本研究成果より、挿入が想定される外来性 DNA 配列が予め分かっている場合、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って得られた全ゲノムシーケンズデータを SV 解析やアセンブリ解析に供することで、外来性 DNA の残存の有無、挿入箇所、挿入された配列の内容を明らかにすることが可能であると示唆された。この一連の解析が、高い再現性を持つ標準的な手法として確立できれば、開発者および評価者双方が同一の解析環境で外来性 DNA の残存を評価できることになり、ゲノム編集食品の安全性評価の精緻化に役立つと共に、当該食品に対する国民の理解促進にも貢献できると期待される。

研究協力者

曾我 慶介（国立医薬品食品衛生研究所）
吉場 聡子（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）

A. 研究目的

2019 年より開始されたゲノム編集食品の事前相談・届出制度では、外来遺伝子およびその一部が残存しないことを調べる手法の一つとして次世代シーケンサー（Next generation sequencer; NGS）を挙げている。また外来性 DNA の残存ばかりでなく、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網羅的解析において、NGS データによる解析は有効であると考えられていることから、こうした遺伝子変化に起因すると思われる、新たなアレルゲンや毒性タンパク質・代謝物が生成される可能性を論じる上でも、今後これが重要な解析手法になると予想される。しかしながら NGS 解析を公定試験法とするには、

解析の必要要件や標準的手順が明確になっておらず、これらの点が課題となっている。また NGS のように大規模に塩基配列を解読する技術は、PCR 増幅された短い塩基配列を決定するショートリードシーケンサーのような従来からある手法の他に、長い単一分子の塩基配列を決定するロングリードシーケンサーが開発されるなど技術革新も進んでいることから、これらシーケンサーの特性を把握した上で、ゲノム編集食品における遺伝子変化の網羅的な解析にこれらがどう活用できるか検討する必要もある。一方でこれらシーケンサーによって行われてきたこれまでの変異解析研究では、50 bp 以上の比較的大きな変異である構造変異 (Structural Variant, SV) の検出は、ヒト以外ではあまり検討されていないのが実状であり、食用生物を研究対象とする際にはこの点も解決すべき課題となる。

本研究では、ゲノム編集食品の意図しないゲノム上の変化を対象に、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全

ゲノムシーケンスデータを用いた解析手法の開発と標準化を目的として、外来性 DNA を有する試料をモデルとした SV 検出手法の検討を行った。

B. 研究方法

1. サンプル

外来 DNA 残存モデルサンプルとして除草剤耐性遺伝子組換えダイズ RRS2 系統の粉砕試料 Monsanto MON89788 Soybean Powder 994.0 g 超/kg (AOCS; 0906-B、以後 RRS2 と表記)、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、以後 nonGM と表記) を購入した。

2. ゲノム DNA 抽出

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

- 1) 1.5 mL の CTAB 抽出 buffer (1.5% CTAB, 75 mM Tris-HCl pH 8.0, 15 mM 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, 50 mM Tris-HCl pH8.0, 10 mM 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。

3. イルミナライブラリー調製およびショートリードシーケンス

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ーを調製。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 ライブラリーとした。シーケンスは HiSeq X Ten (Illumina) による 150 bp \times 2 のペアエンドシーケンスを受託企業へ外注した。

4. ナノポアライブラリー調製およびロングリードシーケンス

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。続いて、1 μ g の DNA を用いて Ligation Sequencing Kit (Oxford Nanopore Technologies; SQK-LSK109) によりシーケンスライブラリーを調製し、Qubit Fluorometer により二本鎖 DNA 濃度を測定。調製したライブラリーを 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。

5. ショートリードシーケンスデータを用いた SV 解析

得られたリードデータをクオリティコントロールツール「FastQC (ver. 0.11.9)」で確認すると、概ね良好なデータであったが、Phred スコアが 20 を下回るリードや、ライブラリー調製に用いたアダプター配列が含まれていたことから、リードデータのトリミングツール「Trim Galore! (ver.

0.6.7)」を用いて $Q < 20$ リードの除去、シーケンスアダプター配列の除去および 40 bp 以下のリードの除去を行った。ダイズの参照ゲノム配列「Glycine_max_v2.1.dna.toplevel.fa」に対して、上記処理済みのリードシーケンスデータをマッピングツール「BWA-MEM (ver. 0.7.17-r1188)」によってマッピングした。ダイズの参照ゲノム配列は EnsemblPlants (<http://plants.ensembl.org/>) より取得した。「Samtools (ver. 1.13)」によってマッピングファイルを圧縮し、ソート、インデックスファイルを作成。また「Picard (ver. 2.26.0)」の MarkDuplicates を用いて PCR duplicate リードを除去。SV caller には「DELLY (ver. 0.8.7)」「GRIDSS (ver. 2.12.1)」「Manta (ver. 1.6.0)」「SvABA (ver. 1.1.0)」を用いて、参照ゲノム配列に対する SV 解析を実施し、全て Normal/Tumor 解析を実施し、RRS2 にのみ特異的に見られる SV 変異を検出した。

パラメーターは全てデフォルトで行い、コールされた SV のうち、Filter が PASS のみを抽出。

「DELLY」で生成された binary variant call format (bcf) ファイルは「bcftools (ver. 1.13)」で variant call format (vcf) ファイルに変換。「GRIDSS」のフィルタリングには「GRIPSS (ver. 1.11)」を使用。

6. ロングリードシーケンスデータを用いた SV 解析

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、base caller「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。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。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。ダイズの参照ゲノム配列「Glycine_max_v2.1.dna.toplevel.fa」に対して、上記シーケンスデータをマッピングツール「Minimap2 (ver. 2.21-r1071)」により -ax map-ont オプションを指定してマッピングした。「Samtools」によってマッピングファイルを圧縮し、ソート、インデックスファイルを作成。SV caller には「Sniffles (ver. 1.0.12)」または「cuteSV (ver. 1.0.11)」を用いた。

デフォルトパラメータでは SV を検知する最小リードは両ツールともに 10 であったが、有効なアライメントリードの最小値が「Sniffles」では 2,000 bp、「cuteSV」では 500 bp と異なっている。また、ナノポアシーケンサーで得られたデータの SV 解析では、「cuteSV」のオプション --max_cluster_bias_DEL を 200 から 100 に、--diff_ratio_merging_DEL を 0.5 から 0.3 変更することが推奨されているが、本検討ではパラメーターはデフォルトとして、参照ゲノム配列に対する SV 解析を実施した。得られた vcf ファイルを用いて、「bcftools」の isec オプションを用いて、nonGM と RRS2 で共通する変異を差し引き、RRS2 にのみ特異的に見られる SV 変異を抽出した。

7. 全ゲノムシーケンスデータを用いた RRS2 T-DNA 配列の再構築

RRS2 の T-DNA 配列情報は米国特許 (US76232985 B2) から取得した (Figure 1)。この RRS2 T-DNA 配列に対し、「FASTX Toolkit (ver. 0.0.14)」の fasta_formatter コマンドの -w オプションを 60 に指定して、一行当たり 60 文字の FASTA ファイルとした。作成した T-DNA 配列 (4385 bp) を参照ゲノム配列として、ショートリードシーケンスデータは「BWA-MEM」にて、ロングリードシーケンスデータは「Minimap2」により -ax map-ont を指定してマッピングし、各全ゲノムシーケンスデータから RRS2 T-DNA

を含むリードを抽出。「Samtools」 view コマンドの -F オプションを 4 に指定してマッピングされなかったデータを除き、「Samtools」で FASTQ ファイルに変換した。ショートリードシーケンズデータは、「KmerGenie (ver. 1.7051)」を用いてアセンブリに最適な k-mer 値を推定、「SeqKit(ver. 2.0.0)」を使って作成した FASTQ の平均リード長を調べ、Picard CollectInsertSizeMetrics を行ってライブラリーの平均インサート長を調べた。さらに、抽出した T-DNA リードを「Velvet (ver. 1.2.10)」を用いてアセンブリした。ロングリードシーケンズデータは「Flye (ver. 2.9)」を用いて --nano-hq および -i 5 オプションを指定してアセンブリした。各アセンブリ結果は RRS2 の T-DNA 配列とのマルチプルアライメントに供したが、GenomeNet (<https://www.genome.jp/>) の CLUSTALW にて実施した。

C. 研究結果

1. 全ゲノムシーケンズデータ解析の標準的手順に向けて検討すべき課題

1) データ解析の概要

ショートリードシーケンサーやロングリードシーケンズによって得られたデータの一般的な解析の流れは以下の通りとなる。

- ① Base calling: データを塩基配列情報に変換
- ② Quality check: 塩基配列情報の質を確認
- ③ Mapping: 参照ゲノム配列との相同性を基に、解読された塩基配列情報の位置を決定
- ④ SV calling: SV 変異の探索
- ⑤ Filtering: SV 変異候補の選抜

これら各段階の技術や手法は日進月歩で進化しているが、公定試験法としての利用という観点で考えると、その性能比較や手順の整備などの面で、例えば以下のような検討すべき課題がある。

- ① シーケンサーの性能比較
- ② 参照ゲノム配列の有無および整備

③ マッピングツール

④ SV caller の選択

上記課題のうち、本研究では①, ②, ④について検討を行なった。マッピングツールについても種々の手法が提唱されている。例えばロングリードシーケンズデータのマッピングツールとしては数年前まで「BWA-MEM」が主流であったが、より高速で動く「Minimap2」が開発されてからはこちらに置き換わっている。その他にも SV caller の「Sniffles」の開発者が変異解析用に開発したマッピングツールである「NGMLR」を用いた解析報告も多い。本研究では、これらマッピングツールを使用した解析報告や評判などを考慮して、ショートリードシーケンズデータには「BWA-MEM」、ロングリードシーケンズデータには「Minimap2」にて実施した。

2) シーケンサーについて

大規模に塩基配列を解読するシーケンサーは、ショートリードシーケンサーとロングリードシーケンズの2種類に大別できる。代表的な機種についてその性能を Table 1 にまとめた。概して、ショートリードシーケンサーは正確性が高い反面、PCR による増幅を行うためライブラリー中に配列依存的な PCR バイアスが生じる可能性があることおよび、長いリピート配列の解読が弱い。一方ロングリードシーケンサーは正確性の面では劣るものの、PCR 等の増幅を行わず、1分子の DNA を直接シーケンズするためショートリードと比較して配列依存性が低く、リピート配列の解読に強いといった特徴があるとされている。またその変異解析ツールについても各々のシーケンサーの特徴を踏まえた専用のツールが開発されつつある。本研究では、ショートリードシーケンサーとロングリードシーケンズとして各々1機種ずつ選び、解析手順、変異判定にかかる性能を比較検討することにした。なおロングリードシーケンサーについては、Oxford Nanopore

Technologies 社 (ONT) のナノポアシーケンサーおよび Pacific Bioscience 社 (PacBio) の Sequel システムの 2 種類が有力である (Table 1)。PacBio の Sequel は正確性も高いが、ランニングコストが高く、必要ゲノム量も多く要求されることが特徴である。一方で、ナノポアシーケンサーの正確性は 90% ほど悪いが、データ量選択の自由度が高く、ランニングコストが比較的低いことと、ゲノム必要量が Sequel より少ないことが特徴である。また、近年の base caller の修正により、一番正確なモードでは約 98% に上ることが報告された。以上より、ロングリードシーケンサーはナノポアシーケンサーを検討することとした。

3) 様々な生物種の参照ゲノム配列について

ゲノム編集技術が報告されて以降、様々な動物・植物においてこうした技術の応用研究がなされてきた。例えば、令和元年度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」報告書によると、2018 年から 2019 年にかけて文献報告された食品となる植物へのゲノム編集技術の応用例では、コメ (130 件)、トマト (45 件)、コムギ (21 件)、トウモロコシ (20 件) が上位 4 品目であった。すなわちこれら品目は、今後ゲノム編集食品として事前相談・届出される可能性が高い品目であると考えられ、全ゲノムシーケンスで得られたデータを解析ができるよう、参照ゲノム配列を優先的に整備しておくべきであると考えられる。今回のモデルサンプルとして用いたサイズも含め、5 品目の参照ゲノム配列を調査し、Table 2 にまとめた。各品目とも系統毎に登録されており、また参照ゲノム配列としては未完成のものもあるなど、整備度合いについても違いがあることが分かった。

4) 変異解析ツールについて

変異解析で使用するバイオインフォマティクスツールについて、これまでに様々なものが開発されて GitHub 等に公開されている。しかし、開発されて間もないツールも数多くあることから、それらのツールを利用して解析を進める前に各ツールの長所・短所を文献等で予め整理することが重要である。また、多くのツールは Linux 環境で開発されており、これらのツールを Windows や Mac の OS で動作させることが困難なこともしばしばみられる。しかし、Mac ではターミナルで Unix コマンドが利用できること、および Windows も Windows Subsystem for Linux (WSL) の開発により、Windows OS 上で Linux を利用できるようになったことで、ツール利用への敷居は徐々に下がっている。さらに近年は Anaconda という Python ディストリビューションにより、同一環境における多くのツールの共存による互換性の問題を、各環境を隔離してパッケージ管理を行うことで回避できるようになった。また、Docker というコンテナ仮想化を用いて、アプリケーションを配置するオープンプラットフォームにもバイオインフォマティクスのツールがライブラリーとしてアップされるようになり、純正 Linux 機でなくても、手軽にツールを導入し利用できる環境が整いつつある。このような環境を鑑みて、汎用性が高い方法として適用可能な方法の情報収集および整理を行った (Table 3, 4)。さらに SV caller の性能比較を行った論文 (Kosugi et al. Genome Biology (2019) 20, 117, Mahmoud et al. Genome Biology (2019) 20, 246) などを参考にしつつ、ショートリードシーケンスデータには「Manta」「DELLY」「GRIDSS」「SvABA」、ロングリードシーケンスデータには「Sniffles」「cuteSV」にて SV 解析を行うことにした。

2. シークエンス結果の概要

1) ショートリードシーケンスについて

ショートリードシーケンサーとして、HiSeq (Illumina) を検討した。ゲノム DNA の状態はアガロースゲル電気泳動を実施し、概ね高分子量であることを確認したのち、ライブラリー調製に供した (Figure 2A)。調製した DNA ライブラリーの品質は Bioanalyzer 2100 (Agilent) で確認した (Figure 2B)。Bioanalyzer のエレクトロフェログラムより 150-1500 bp の平均分子量は、RRS2 サンプルで 568 bp、コントロール (non-GM) サンプルで 589 bp であった。また各サイズの DNA が概ね正規分布し、プライマーダイマーや分解された DNA 等の低分子 DNA のピークが見られないことから、正常な DNA ライブラリーであることを確認した。ショートリードシーケンスを実施したところ、ダイズゲノムサイズを約 1.1 Gb として見積ると、その 60 倍以上のカバレッジを満たすデータを取得した。生データ量およびその統計値は Table 5 に示した。

2) ロングリードシーケンスについて

ロングリードシーケンサーは ONT 社のナノポアシーケンサーである PromethION を検討した。長鎖リードデータを大量に得るために低分子塩基を除去し、シーケンサーにアプライする前の DNA の状態はアガロースゲル電気泳動で確認した (Figure 3)。ロングリードシーケンスを実施し、得られた生データ量およびその統計値を Table 6, 7 に示した。RRS2 サンプルでデータ量は 79 Gb、N50 は 25.1 kb、nonGM サンプルでデータ量は 48.6 Gb、N50 は 5.5 kb であった。ダイズゲノムサイズを約 1.1 Gb として見積ると、取得データでは、RRS2 サンプルで 71.8 倍、nonGM サンプルで 44.2 倍のカバレッジ、ベースコールおよびフィルタリング後のデータでは、RRS2 サンプルで 47.2 倍、nonGM サンプルで 23.5 倍のカバレッジであった。

3. SV 解析の結果

各シーケンズデータについて実施した SV 解析で検出された変異の結果をまとめた (Table 8)。また RRS2 サンプルにおける T-DNA 配列の挿入について、各 SV caller ツールの結果をまとめた (Figure 4)。偽陽性を除くため、RRS2 サンプルで検出された SV から nonGM サンプルで検出された SV を差し引いた結果 (RRS2/nonGM) を載せている。但し、「Sniffles」で解析したデータでは差し引くことができなかった。また「GRIDSS」および「SvABA」での解析については、変異の種類を同定することができなかったため、BND (breakend) としてまとめて表記している。変異の種類毎の総数で比較すると、「Manta」および「DELLY」と比較して、「Sniffles」および「cuteSV」では挿入と欠損変異の数が多い傾向にあった。「cuteSV」による解析について差し引き前後の SV 総数を比較すると、RRS2→RRS2/nonGM で、INS (26891→26793)、DEL (24721→23252)、DUP (587→459)、INV (156→148)、BND (3090→2597) と、各 SV の 1~22% の減少にとどまった。

ショートリードシーケンズデータでは、RRS2 の T-DNA 配列挿入部を変異として検出できたのは GRIDSS のみであったが、アルゴリズムの問題により非ヒト生物を対象にした解析では「挿入」変異とは判定されなかった。一方ロングリードシーケンズデータでは、「Sniffles」および「cuteSV」とともに RRS2 の T-DNA 配列挿入部を「挿入」変異として判定できた。挿入配列として「Sniffles」は 3509 bp、「cuteSV」は 4254 bp と予測したが、RRS2 サンプルの実際の T-DNA 挿入長は 4324 bp であり、「cuteSV」の方がより近い値であった。

4. RRS2 T-DNA 配列の再構築

RRS2 の T-DNA 配列に対して、ショートリードシーケンズデータを「BWA-MEM」によりマッピングしたところ、全リード中、RRS2 T-DNA 配列にマッピングしたのは 1481 リードのみであり、全て RRS2 サンプル由来で、nonGM サンプ

ルは1リードもマッピングされなかった。抽出した1481リードをFASTQ形式に戻し、「SeqKit」を使って平均リード長を調べたところ、147.7 bpであった。これらのことから、RRS2 T-DNA 領域のシークエンスカバレッジは、1481 (リード数) × 147.7 (平均リード長) / 4385 (T-DNA 配列長) = 49.88…となることから、およそ50倍とした。またこのライブラリーのMEDIAN_INSERT_SIZEは337 bpで、アセンブリに最適なk-mer値はk=91であった。よって、「velveth」のhash値は91とし、「velvetg」の-ins_lengthは330、-exp_covは50としてアセンブリを実行したところ、4463 bpのコンティグが1本のみ生成された。

RRS2のT-DNA配列に対して、ロングリードシークエンスデータを「Minimap2」によりマッピングしたところ、全リード中、RRS2 T-DNA配列にマッピングしたのは58リードのみであり、全てRRS2サンプル由来で、nonGMサンプルは1リードもマッピングされなかった。抽出した58リードをロングリードアセンブラー「Flye」を用いてアセンブリしたところ、長さ91213 bpの1本のコンティグが得られた。平均カバレッジは13倍であった。

生成された各コンティグ配列の内部配列にはRRS2 T-DNA配列に相当する配列が確認された。そこで、この各コンティグ配列を参照ゲノム配列のRRS2 T-DNA配列と比較した結果、ショートリードシークエンスデータにより得られたコンティグ配列はT-DNA配列(4385 bp)と完全に一致していたが、ロングリードシークエンスデータにより得られたコンティグ配列は4384 / 4385 bp (99.9%)と一塩基の不一致が見られた(Figure 6)。不一致が見られた箇所はTが8塩基連続する配列であり、8番目のTが欠損していた。

D. 考察

本研究では、ゲノム編集食品の意図しないゲノム上の変化を検出するために、全ゲノムシークエン

スデータを用いた解析手法の確立と標準化を目的としている。今年度は、モデル試料として遺伝子組換えダイズRRS2を用い、外来性DNAの残存を確認する手法について検討した。

一般的なショートリードシークエンスおよびロングリードシークエンスによって得られたデータの解析スキームをもとに、測定機種、参照ゲノム配列、SV callerなどによる解析結果への影響をまず検討すべきであると考えた。そこで、測定機種についてはショートリードシークエンサーとロングリードシークエンサー各々の代表的な機種を用いてモデル試料の全ゲノムシークエンスデータを取得し、各々の機種でよく使われているSV callerによりその性能を比較した。その結果、ロングリードシークエンサーによって得られた全ゲノムシークエンスデータの解析により、遺伝子組換えダイズRRS2において外来性遺伝子挿入が検出されたこと、さらにどちらの機種で得られた全ゲノムシークエンスデータであっても遺伝子組換えダイズRRS2ゲノムサンプルにおいて挿入されたDNA配列が再構成できることが分かった。すなわち、当該ゲノム編集食品に残存している可能性が疑われる外来性DNA (Cas9やgRNAを細胞導入する際に使用したベクターなど)が予め分かっていたら、その残存の有無や位置、そして挿入されたDNA配列などを調べることが可能であることを示唆している。したがって今後はこうした解析スキームの必要要件を考える必要がある。今回は例えばショートリードシークエンサーのデータでは60倍以上のカバレッジがあり、対象としたRRS2サンプルには4000 bp以上の遺伝子挿入があった。シークエンスデータの解析における必要要件を考える上では、以下の点などについて検討する必要があると考えている。

- ・どの程度のカバレッジが必要か？
- ・長さの異なる遺伝子挿入でも検出できるのか？
- ・挿入長と必要なカバレッジの関係

ゲノム編集食品における外来性 DNA の残存を調べる方法として、ショートリードシーケンスおよびロングリードシーケンスで得られた全ゲノムシーケンスデータを用いて、SV 解析による挿入位置の同定やアセンブリによる残存配列の同定が可能であることが今回の検討より示唆された。残存性を調べる既存の方法として他にサザンブロット法や PCR 法などがあるが、挿入位置と配列をほぼ正確に決定できることは全ゲノムシーケンスデータを用いた今回の解析手法の利点であると言える。一方こうした全ゲノムシーケンスデータを用いた外来性 DNA 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. (2020) Sci. Rep.)。これは、挿入が想定される外来性 DNA 配列を断片化し、対象生物個体の全ゲノムシーケンスデータと網羅的に比較照合することで、外来性 DNA 配列の断片を検出する手法である。20 塩基以上の断片であればその挿入箇所を同定できると言われている。簡便なスクリーニング方法としては優れた方法と言えるが、20 塩基より短い挿入では判定の正確性に欠ける、挿入配列の全容はこの手法だけでは分からないなどの課題もある。今回我々が試みた SV 解析とアセンブリ解析を組み合わせた手法との比較や両手法の併用の可能性なども今後検討する必要がある。

ゲノム編集食品の意図しないゲノム上の変化という観点で考えると、欠損・逆位・重複・これらの複合など挿入以外の遺伝子変異も十分に考えられる。挿入変異同様、これら他の変異は当該食品の安全性を考える上でハザードとなり得ることから、これらの同定も重要な課題となる。今回の検討でも様々な SV caller で他種類の遺伝子変異についても調べたが、Table 8 に示す通り、SV caller 毎にその種類や数が大きく異なっていた。従ってこの Table 8 に示す SV の総数には偽陽性を含んでいる可能性が高い。また挿入変異についても、RRS2 の T-DNA 配列のような挿入される可能性

がある外来性 DNA が予め分かっていたら、挿入の有無やその位置を特定することも可能であるが、そのような可能性のある外来性 DNA が不明であった場合、ゲノム編集技術に起因する意図しない挿入変異を正しく評価できる保証がない。従ってこうした様々な変異を正しく評価できる系の構築が望まれる。現状検出される SV 総数の違いは、SV caller の性能や判定アルゴリズムの違いにも起因するが、参照ゲノム配列の不備による偽陽的な変異も、この要因の一つと考えられる。今後こうした点も考慮しながら、評価系の構築を行う必要がある。また変異検出の性能評価という観点から、各変異の種類毎に標準的サンプルを用意する必要もあると考えられる。合わせて検討したい。

E. 結論

本年度の検討より、残存する可能性がある外来性 DNA 配列が予め分かっていたら、ショートリードシーケンスおよびロングリードシーケンスによって得られた全ゲノムシーケンスデータを用いた SV 解析とアセンブリ解析を組み合わせることにより、その残存した DNA 配列のゲノム挿入の有無や位置、そしてその挿入配列を明らかにできることが分かった。これは、ゲノム編集食品の事前相談・届出制度で求められている「外来遺伝子およびその一部が残存しないことを調べる手法」として全ゲノムシーケンスデータを利用した評価手順を提起したことになり、開発者および評価者にとって信頼性の高い安全性評価を可能にするものである。今後手順の標準化やその解析に必要な要件について検討する必要がある。

F. 研究発表・業績

1. 論文発表

該当なし

2. 学会発表

- 1) 曾我慶介、中村公亮、成島純平、吉場聡子、木俣真弥、江木智宏、岸根雅宏、真野潤一、橘田和美、高島令王奈、柴田識人、近藤一成：改正食品表示基準に対応した遺伝子組換えとうもろこし混入の判定に係る定性 PCR 検査法の開発、日本食品衛生学会第 117 会学術講演会、東京、2021 年 10 月 26 日
- 2) 高島令王奈、江木智宏、曾我慶介、峯岸恭孝、成島純平、吉場聡子、柴田識人、中村公亮、近藤一成、岸根雅宏、真野潤一、橘田和美：改正食品表示基準に対応した遺伝子組換えダイズ混入の判定に係る定性 PCR 検査法の開発、日本食品衛生学会第 117 会学術講演会、東京、2021 年 10 月 26 日
- 3) 吉場聡子、成島純平、曾我慶介、柴田識人、近藤一成：安全性未承認の遺伝子組換えナタネの試験法開発、第 58 回全国衛生化学技術協議会年会、名古屋、2021 年 11 月 25-26 日
- 4) 柴田識人、曾我慶介、成島純平、吉場聡子、江木智宏、高島令王奈、近藤一成：遺伝子組換えダイズ・トウモロコシ定量のための内標比の算出、第 58 回全国衛生化学技術協議会年会、名古屋、2021 年 11 月 25-26 日
- 5) 曾我慶介、成島純平、吉場聡子、江木智宏、真野潤一、高島令王奈、橘田和美、中村公亮、柴田識人、近藤一成：「遺伝子組換え出ない」表示確認に係る新定性検査法の試験室間共同試験による妥当性評価、第 58 回全国衛生化学技術協議会年会、名古屋、2021 年 11 月 25-26 日
- 6) 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：ゲノム編集作物におけるオフターゲット予測法の検討、第 44 回日本分子生物学会年会、横浜、2021 年 12 月 2 日
- 7) 曾我慶介、吉田光範、吉場聡子、成島純平、柴田識人、近藤一成：ナノポアシーケンス技術を用いたスギヒラタケの全ゲノムアセンブリの検討、第 44 回日本分子生物学会年会、横浜、2021 年 12 月 3 日
- 8) 近藤一成、成島純平、曾我慶介、吉場聡子、柴田識人、坂田こずえ、田口千恵、加藤怜子：ヒト細胞 TK6 を用いた CRISPR/Cas による構造変異(SV)解析の検討、第 44 回日本分子生物学会年会、横浜、2021 年 12 月 3 日

G. 知的財産権の出願・登録状況

該当なし