

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たら、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, 7.5 mM Tris-HCl pH8.0, 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  $\mu$ 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  $\mu$ L および RNase A (ニッポンジーン) を 1.5  $\mu$ 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  $\mu$ L の HE Buffer (5 mM 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  $\mu$ L、各対象プライマー溶液 (50  $\mu$ mol/L) 各 0.25  $\mu$ L、対象プローブ溶液 (10  $\mu$ mol/L) 0.5  $\mu$ L を混合し、純水で全量 20  $\mu$ L に調製後、各 DNA 試料液 (10 ng/ $\mu$ L) 5  $\mu$ L を添加して全量 25  $\mu$ 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  $\mu$

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

P6-25-R5: GCTCTGCCTATTGTCCCATATAT  
AACC

### 5. ゲノム編集 GABA トマトにおける標的部 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  $\mu$ L、2.5mM dNTPs 溶液 4  $\mu$ L、各対象プライマー溶液 (50  $\mu$ mol/L) 各 0.25  $\mu$ L、PrimerSTAR GXL ポリメラーゼ 1  $\mu$ L、各トマトゲノム DNA 100 ng を混合し、純水で全量 50  $\mu$ 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。

SIGAD3\_Fwd1: CCAAGGTTTGAGATGCTAG  
AGAAGTC

SIGAD3\_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。

```
SIGAD3_seq_Fwd: GAGACCCTCCGTAGGTT  
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ズ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、特に入力はしていない。

ロングリードシーケンズ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. シークエンスリードのランダムサンプリング

シークエ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ズ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ズ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ズデータの統計値を示す。なお推定シーケンズ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ズデータでは Velvet、ロングリードシーケンズデータでは Flye でアセンブリを行った。

ショートリードシーケンズ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ズ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ケンスタデータのランダムサンプリング

ショートリードシーケンサーによる WGS データを用いて残存した外来遺伝子を検出する際に、どの程度のシーケ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 コマンドでシーケンスカバレッジを算出する。今回の場合シーケンスタにより得られた総塩基数 (66,929,724,803 b) をサイズのゲノムサイズ (1.1 Gb) で割ると、シーケ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 (目標のシーケンスカバレッジ) / 60.8 (実際のシーケンスカバレッジ) = 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 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ケンスタデータを用いた各シーケンスカバレッジでの標的配列のアセンブリ (Velvet)

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 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×のシーケ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×のシーケ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ケンスタデータを用いた各シーケ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ない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ケ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×と比較的低いシーケンスカバレッジで正確に再現できることが示唆された。

#### ロングリードシーケンスタデータを用いた各シーケ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ケンスタ同様に、各シーケンスカバレッジにおけるロングリードシーケン

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 SSIb 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 SSIb 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ズデータの場合でも、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ズ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ズデータ、ロングリードシーケンズ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ズデータのランダムダウンサンプリング

シーケンズ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ズ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。

#### 各シーケンズカバレッジにおけるアセンブリ結果

各シーケンズ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×以上のシーケンズ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×のシーケンズ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×以上のシーケンズ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のリードデータをとマトリファレンスゲノムSL4.0 ([https://solgenomics.net/organism/Solanum\\_lycopersicum/genome/](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×

のシークエ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クエ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クエンサーおよびロングリードシークエンサーによる WGS を実施した。各々 79.2× および 55.1× というシークエ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クエ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、シークエ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クエ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クエ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クエンサーによる WGS データについて得られた「残存配列の検出には 25×以上のシークエ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×程度のシークエ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することが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サ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、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, Yoshihara 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, Yoshihara 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、橋田和美：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