

201522051A

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

発芽前後におけるGMダイズの  
遺伝子発現プロファイリングに関する基盤研究

平成27年度 総括研究報告書  
(H27-食品-若手-023)

研究代表者 中村公亮

平成 28 年 5 月

# 目 次

## I. 総括研究報告書

発芽前後におけるGMダイズの遺伝子発現プロファイリングに関する基盤研究

研究要旨-----	1
A. 研究目的-----	1
B. 研究方法-----	1
C. 研究結果-----	5
D. 考察-----	9
E. 結論-----	11
F. 健康危険情報-----	11
G. 研究発表-----	12
H. 知的財産権の出願・登録状況-----	14
図表一覧-----	15
II. 研究成果の刊行に関する一覧表-----	37

## 発芽前後における GM ダイズの遺伝子発現プロファイリングに関する基盤研究

研究代表者 中村公亮 国立医薬品食品衛生研究所・生化学部

### 研究要旨

昨今、発芽ダイズは発芽前を上回る栄養価が注目されており、菓子や健康食品などの加工食品に利用されている。種子は、発芽に伴って休眠状態にある遺伝子の発現を活発化させるため、発芽前後ではタンパク質をはじめとする代謝産物の構成が大きく異なることが予想される。食品加工への発芽ステップの導入といった食品科学技術の著しい進展に伴い、遺伝子組換え（GM）ダイズの構成成分の変化等に関する検討を進める必要が出てきた。そこで本研究では、初年度（H27）は、RNA-Seq 解析によりダイズの全ゲノム上の遺伝子を網羅した発芽により誘導される遺伝子の解析を実施する。最終年度（H28）は、発芽時の外的環境因子や組換え技術による遺伝子発現プロファイルへの影響について解析する。こうして、ダイズが発芽する際に発現誘導される遺伝子（発芽遺伝子）について、転写及び翻訳レベルで全ゲノム上の遺伝子を網羅し解析する新しい手法を開発し、非 GM 型と GM 型ダイズを比較する際の有用性について検証を行う。

### 協力研究者

石垣拓実（国立医薬品食品衛生研究所）明  
石良、権藤崇裕、田中秀典、橋口正嗣（宮  
崎大学）菅野陽平（北海道立衛生研究所）

### A. 研究目的

本研究では、ダイズが発芽する際に発現誘導される遺伝子（発芽遺伝子）について、転写及び翻訳レベルで全ゲノム上の遺伝子を網羅し解析する新しい手法を開発し、遺伝子組換え（GM）型と非 GM 型ダイズの構成成分の変化等を比較する際の有用性について検証を行うことを目的とする。

### B. 研究方法

#### 1. 試料と発芽条件

試験には、非 GM 型ダイズ 2 品種（国内で流通しているアメリカ産 GL3494 品種とカナダ産 OAC Kent 品種）、及び、Williams 品種の種子を供した。種子の発芽は、以下の 2 種類の方法を検討した。発芽条件①では、発芽ダイズ食品生産の際に使用される発芽条件「発芽ダイズ加工食品及び前記食品の製造方法」（特開 2006-345708）を基に、48 時間 27℃又は 40℃の滅菌水中に浸漬した。発芽条件②では、滅菌処理を

行った種子をそれぞれ上下2枚のキムタオルで挟み込み滅菌水で湿らせた後、48時間27°C又は40°Cで発芽させた。種子の滅菌処理は以下のように行った。シャーレー一枚あたり、20粒の種子を準備し、それをデシケーター内にフタをあけて入れて、次亜塩素酸ナトリウム（有効塩素6%）100 mLに濃塩酸3.5 mLを入れ、デシケーターの蓋を閉じて、発生する塩素ガスでガス滅菌を行った。

## 2. 次世代シーケンシング解析用試料の調製

発芽ダイズ生産の発芽条件（40°C）と常温発芽条件（27°C）下で48時間培養し発芽させた1のダイズ3品種より、粒単位でトータルRNAの抽出・精製を行った。1粒又は5粒を試料に、乳鉢・乳棒を用いて液体窒素を加えながら粉状になるまで粉碎し、Qiagen RNeasy Plant Mini Kitの2カラム分を1試料に使用してトータルRNAを精製した。ゲノムDNAは、RNase-free DNaseを使用して、完全に分解させた。得られたRNAの品質は、Bioanalyzer（アジレントテクノロジー社）を使用しRNA Integrity Number（RIN）値を測定することにより評価した。RNA4 µgを使用し、oligo dTによりmRNAを濃縮した。TrueSeq Stranded mRNA Sample Prep Kit（イルミナ社）を使用して、RNA-Seqライブラリー（5

plex）を作成した。イルミナHiSeq2500によるシーケンス解析は、フローセルの1レーンを用いてhigh output mode 100-base paired-endシーケンシングを外部委託した。

## 3. 次世代シーケンシング解析データの解析

得られたデータは、ダイズゲノム解析（Nature, 463, 178-183, 2010）より得られた配列データベース（V1.0.29, [ftp://ftp.ensemblgenomes.org/pub/release-29/plants/fasta/glycine\\_max/dna/](ftp://ftp.ensemblgenomes.org/pub/release-29/plants/fasta/glycine_max/dna/)）をリファレンス配列としてマッピングを行った。発現差解析は、Empirical Analysis of Digital Gene Expression解析手法（*Biostatistics*, 9, 321-332, 2008, *Bioinformatics*, 26, 139-140, 2010）を使用し、得られたリード数から各遺伝子の発現を定量した。また、発現した遺伝子については既知のアレルゲンとの比較（相同性検索）をデータベース上で行った。すなわち、80以上のアミノ酸残基で35%以上の相同性、及び、6~8アミノ酸残基の完全一致の条件（FAO/WHOガイドライン）でアレルゲン性を米国ネブラスカ大Food Allergen Research and Resource Program（FARRP）の開発したAllergenOnlineデータベース上で検索した。

#### 4. リアルタイム PCR による遺伝子発現の定量

3. の RNA-Seq 解析より得られたデータを基に、個々の遺伝子発現については ABI7900HT リアルタイム PCR システムを使用し定量した。2. の方法で発芽ダイズより調製した 500 ng の精製 RNA を、逆転写酵素 SuperScript II reverse transcriptase (Invitrogen) と oligo dT20 のプライマーを使用して 20  $\mu$ L の反応液中で逆転写反応を行い、cDNA を作成した。2  $\mu$ l の cDNA を鋳型に、exon-intron 間でスプライシング標的的境界領域に設定したプライマー対による QuantiTect SYBR® Green PCR (QIAGEN) を用いたリアルタイム PCR により遺伝子発現の定量を行った。PCR 反応液は、20  $\mu$ L/well として、以下のとおり調製した。2 $\times$ QuantiTect SYBR® Green PCR master mix 10  $\mu$ L、対象プライマー対溶液（各プライマー、50  $\mu$ mol/L）各 0.2  $\mu$ L を混合し、cDNA 試料液 0.5  $\mu$ L を添加し滅菌蒸留水で全量 20  $\mu$ L に調製した。反応条件は、50°C で 2 分間、95°C で 10 分間加温し、その後、95°C で 15 秒間、60°C で 1 分間を 1 サイクルとして、50 サイクルの増幅反応を行った。リファレンス遺伝子の発現解析には、ダイズ由来内在性遺伝子 *Actin* の発現を検知するプライマー対を使用した。選定したターゲットフラグメントの参照元配列を参考にプライマー設計を行った。Feature ID を、

The Samuel Roberts Noble Foundation の HP 内のマメ科ゲノムデータベース LegumeIP で検索し、ヒットした参照配列より CDS とゲノム中の遺伝子構造の情報を得た（URL <http://plantgrn.noble.org/LegumeIPv2/index.jsp>）。次に、RT-PCR プライマーの設計元の配列を得るため、イントロンを含む 500~3000 bp をターゲットとしたプライマーを作成した。プライマーの設計は、primer3 プログラムを用いた。なお、遺伝子構造に関してはデータベースのものと次世代シーケンスより得られたコンティグをもとにした遺伝子構造と食い違っているものもみられるため、両方で構造が一致している部分を検知した。

各種遺伝子（温度により発現量に差のある遺伝子 no.2~6、品種により発現量に差のある遺伝子 no.1~3 及び発現差定量のコントロール遺伝子 *Actin*）の発現量を定量するために使用したプライマー配列は以下の通りである。

*Actin* (*BMC Mol Biol*, 10, 93, 2009) :

5'-ATCTTGACTGAGCGTGGTTATTCC-3'

5'-GCTGGTCCTGGCTGTCTCC-3'

温度 no.2:

5'-CGGAAAGGAGGAGATTGTGG-3',

5'-CACCTGCCGCTTCTATCCTC-3'

温度 no.3:

5'-ACATTGCCTTGCACCCTTG-3',

5'-GATCTTTGTCCCCTTTTCCAC-3'

温度 no.4:

5'-GGCTTCCAATCCAAGAACGA-3',

5'-TCCCAATGTTGATCCCTTTGA-3'

温度 no.5:

5'-ACCGACGCCGACTGTTTTAC-3',

5'-TGCTTCGCTCTCGTCAGAAC-3'

温度 no.6:

5'-CTCGATGCTGCTGGAAACAC-3',

5'-TGTTTTGATGCCTGCTCTGC-3'

品種 no.1:

5'-GAAAAGGTGATTGGAGGGGTA-3',

5'-GCTAATTGGATTTGTGATGTCG-3'

品種 no.2:

5'-TGACTTCAAGAACCCCAACC-3',

5'-GACAAGGTCACATCATCCTCAA-3'

品種 no.3:

5'-TGACTTCAAGAACCCCAACC-3',

5'-GGTCCAGCTTCACAATATTTATCA-3'

温度 no.2~6 は、27°C発芽サンプルの発現量を 1 とした場合の 40°C発芽サンプルの相対発現量を比較するために使用し、品種 no.1~3 は、アメリカ産 GL3494 品種における発現量をリファレンスとした場合のカナダ産 OAC Kent 品種の相対発現量を比較するために使用した。

#### 5. PCR を使用したゲノム DNA 配列の検出

ダイズのゲノム DNA は、アメリカ産 GL3494 品種とカナダ産 OAC Kent 品種の発芽ダイズ (種子) 10 粒程度をボールミル

(MM200 Retch) で粉砕し、粉末状試料 20 mg をとり、QIAGEN DNeasy plant mini kit を用いて DNA の抽出精製を行った。標的配列の増幅断片長によって 2 種類の PCR 反応を使用して行った。標的アンプリコンの長さが 500~1500 bp の場合、PCR 反応液は、25  $\mu$ L/well として、以下の組成のとおり調製した。10xExTaqBuffer ( $Mg^{2+}$ ) 10  $\mu$ L、対象プライマー対溶液 (各プライマー、50  $\mu$ mol/L) 各 3  $\mu$ L、2.5 mM dNTP 5  $\mu$ L、DNA 試料液 50 ng を添加し、ExTaq 1U を加え、滅菌蒸留水で全量 25  $\mu$ L となるよう調製した。反応条件は、50°C で 2 分間、95°C で 10 分間加温し、その後、95°C で 15 秒間、60°C で 1 分間を 1 サイクルとして、50 サイクルの増幅反応を行った。標的アンプリコンの長さが 2000~3000 bp の場合、PCR 反応液は、25  $\mu$ L/well として、以下の組成のとおり調製した。10xLATAqBuffer ( $Mg^{2+}$ ) 2.5  $\mu$ L、対象プライマー対溶液 (各プライマー、50  $\mu$ mol/L) 各 3  $\mu$ L、2.5 mM dNTP 4  $\mu$ L、DNA 試料液 50 ng を添加し、LATAq HS 1U を加え、滅菌蒸留水で全量 25  $\mu$ L となるよう調製した。反応条件は、94°C で 1 分間加温後、98°C で 10 秒間、50°C で 30 秒間、72°C で 1 分間/kb を 1 サイクルとして、40 サイクルの増幅反応を行った。

## C. 研究結果

### 1. 発芽ダイズと RNA-Seq 解析用サンプルの調製

アメリカ産 GL3494 品種、カナダ産 OAC Kent 品種、及び、Williams 品種は、B.の1.で示した2種類の方法の発芽の条件(発芽条件①、発芽条件②)を検討した(Figure 1)。発芽条件①で培養したダイズは、27°C又は40°Cともに腐敗が進行し(Figure 1A)、発芽条件②で培養したダイズは、腐敗することなく発芽した(Figure 1B,C)。発芽ダイズ1粒又は5粒よりトータルRNAを抽出し、精製されたRNAの評価を行った(Figure 2)。27°C又は40°Cで発芽させたアメリカ産 GL3494 品種、カナダ産 OAC Kent 品種および40°Cで発芽させたWilliams 品種より精製したトータルRNAを吸光度及びバイオアナライザーより、収量、精製度、並びに、品質を評価した。その結果、発芽ダイズ1粒より10~75 µg、5粒より140~411 µgが精製された。精製度は、いずれのサンプルも  $A_{260\text{ nm}}/A_{280\text{ nm}}$  は2.10~2.18、 $A_{260\text{ nm}}/A_{230\text{ nm}}$  は1.32~2.26であり、高い精製度であることが示唆された。また、バイオアナライザーのエレクトロフェログラム(Figure 2A)より、RIN値を算出した。その結果、いずれのサンプルも $\geq 7$ であったため、RNAの分解は進んでおらず高い品質であることが示唆された(Figure 2B)。得られたトータルRNAか

ら mRNA を oligo dT により精製し、RNA-Seq ライブラリー (5 plex) のサンプルを作成後、RNA-Seq 解析を行った。まず、次世代シーケンサーより得られたリードのトリミングを行った。次世代シーケンシングデータ解析用ソフト Genomic Workbench ver.8.5.1 を使用し、得られたデータ (ペアエンド・リード) のトリミングを行った (Figure 3)。その結果、アメリカ産 GL3494 品種では、78,843,604 リード (27°C発芽条件) と 72,260,274 リード (40°C発芽条件)、カナダ産 OAC Kent 品種では、76,550,425 リード (27°C発芽条件) と 71,928,417 リード (40°C発芽条件)、Williams 品種では、71,209,645 リード (40°C発芽条件) のデータセットが得られた (表 1)。平均 101 base の全リード数の 99.6%以上 (7千万以上のリード数) の配列がトリミング後に残り、リード長も 99.7 base が得られた。この結果から、RNA-Seq 解析を行うための十分な品質を有したリード数を得ることができたと判断された。

### 2. ダイズゲノム塩基配列へのマッピング

得られたリードは、ダイズゲノム塩基配列 (V1.0.28) をリファレンスにマッピングを行った。5種類のサンプル (アメリカ産 GL3494 品種 27°C/40°C、カナダ産 OAC Kent 品種 27°C/40°C、Williams 品種 40°C) より得られたリードをゲノムに最も高いアラ

インメントスコアを示す場所にマッピングすると同時に、マッピング後のデータを部分修正 (local realignment) した。その結果、全ゲノムを網羅して品種及び温度の違いに関わらず発現の相同性が見られた遺伝子、及び、品種及び温度の違いによって発現差のある遺伝子を同定することができた (Figure 4)。以上の結果から、発芽サイズの遺伝子発現を RNA-Seq より解析する場合、品種、及び、発芽温度の条件を一定に揃えることが重要であることが示唆された。

### 3. RNA-Seq 解析

RNA-Seq 解析より、発現した遺伝子のゲノム上において存在する位置 (染色体別) と発現量を示すデータを得た。また、得られたリードは、データベース上に登録されている遺伝子情報を付加することでアノテーションを行った。Figure 5A は、遺伝子別にマッピングされたリード数やゲノム上において存在する位置などの情報別にリスト化したデータの一部を示す。Figure 5B は、16 番染色体上にマッピングされたリード数の分布図をサンプル別に表示させた図の例である。ゲノム配列にマップされたリード数もしくは発現量の違いが顕著に検出された位置を、発芽温度については四角で、品種間については丸で示す。このようなデータ情報をまとめた結果、

アメリカ産 GL3494 品種 27°C 発芽条件下では 2,115,935 リード、アメリカ産 GL3494 品種 40°C 発芽条件下では 2,123,999 リード、カナダ産 OAC Kent 品種 27°C 発芽条件下では 1,998,598 リード、カナダ産 OAC Kent 品種 40°C 発芽条件下では 2,057,249 リード、Williams 産 40°C 発芽条件下では 2,146,500 リードがダイズゲノム塩基配列へマッピングされた。マッピング結果より、発芽温度及び品種の違いによって、発現パターンの違う遺伝子群が確認された。例として、16 番染色体上の GLYMA16G04770.3 遺伝子は、27°C で発芽させた際にアメリカ産 GL3494 品種とカナダ産 OAC Kent 品種で発現するが、40°C では発現しなかった。一方、同じく 16 番染色体上の GLYMA16G08570.1 遺伝子は、27°C 発芽条件下では発現しないが、40°C 発芽条件下では品種にかかわらず発現が検出された。また、品種及び発芽温度の違いに関係なく、一定の発現量を検出した遺伝子群も特定した。この結果をデータベース中に登録されアノテーションされている全遺伝子について、アメリカ産 GL3494 品種及びカナダ産 OAC Kent 品種を 27°C もしくは 40°C で発芽させた際の遺伝子発現情報をヒートマップ上に表示させた結果を Figure 6 に示した。

### 4. サンプル間の発現差解析

サンプル間の遺伝子発現差解析では、得



られたリードのカウントデータを使用し、データが負の二項分布に従うと仮定して平均値と dispersion を推定し、検定を行った。Genomic Workbench ソフトに搭載されている Empirical Analysis of DGE (proportional test) プログラムを使用して 2 群の比較検定を行った。まず、発芽温度による遺伝子発現差を解析するため、27°C で発芽させたアメリカ産 GL3494 品種とカナダ産 OAC Kent 品種の 2 品種のデータが同等のものと仮定し、同様に 40°C で発芽させた上記 2 品種のデータを同等のものと仮定して解析を行った。27°C で発芽させたサンプルをリファレンスに、40°C で発芽させたサンプルにおいて、2 倍量以上の差 ( $p < 0.05$ ) のある遺伝子をリスト化し、その結果、12,722 遺伝子を特定した (Figure 7)。最大の差は、染色体 7 番 (5,415,828 ~ 5,421,767) に存在する遺伝子 (GLYMA07G06750.2) で 1,174 倍の差 ( $p$  値  $9.42 \times 10^{-44}$ ) であった (Figure 8)。遺伝子配列を BLASTn 検索したところ、BAG family molecular chaperone regulator 6 (BAG6A), mRNA (GenBank accession no. XM\_006583196.2) と 100% 相同性を示した。当該遺伝子の解析に関する報告は、植物では皆無であることが判った。次に、品種による遺伝子発現差を解析するため、27°C と 40°C で発芽させたアメリカ産 GL3494 品種を同等のものと仮定し、同様に 27°C と 40°C

で発芽させたとカナダ産 OAC Kent 品種を同等のものと仮定して、解析を行った。アメリカ産 GL3494 品種の発芽サンプルをリファレンスにカナダ産 OAC Kent 品種の発芽サンプルにおいて、2 倍量以上の差 ( $p < 0.05$ ) のある遺伝子をリスト化し、8,418 遺伝子を特定した (Figure 9)。最大の差は、染色体 9 番 (19,201,064 ~ 19,203,101) に存在する遺伝子 (GLYMA09G16090.1) で 6,852 倍の差 ( $p$  値  $2.41 \times 10^{-228}$ ) であった (Figure 10)。遺伝子配列を BLASTn 検索したところ、F-box/LRR-repeat protein 3-like (LOC100820469) (GenBank accession no. XR\_415935) と 100% 相同性を示した。当該遺伝子の解析に関する報告は、植物では皆無であることが判った。

##### 5. RT-リアルタイム PCR による遺伝子発現差解析の確認

RNA-Seq 解析の結果より得られた、発芽温度の違い (27°C と 40°C) により遺伝子発現の差が見られた遺伝子リストから、遺伝子をランダムに選択し、数十倍の差の遺伝子 (GLYMA01G43460.2 [温度 No.2] と GLYMA15G08420.1 [温度 No.3]) と数百倍の差の遺伝子 (GLYMA19G34031.1 [温度 No.5]) の発現量の定量を RT-リアルタイム PCR より行った。まず、アメリカ産 GL3494 品種ゲノム中の No.2、No.3、No.5 遺伝子のゲノム配列をシーケンシングした結

果、すべての遺伝子において配列はリファレンス配列と完全に一致した (Figure 11~13)。シーケンシングした配列を基に、リアルタイム PCR 定量用のプライマー配列を設計した (Figure 14~16)。Figure 17A~C は、RNA-Seq データの出力リード配列をダイズゲノム配列にマッピングを行った結果を示す。Figure 17D は、ダイズ由来 Actin 遺伝子をリファレンスに 27°C で発芽させた際と比較して、40°C で発芽させた際の遺伝子発現量比を算出した結果を示す。RNA-Seq 解析の結果、温度 No.2、No.3、No.5 遺伝子の発現量はそれぞれ 75.16 倍、92.98 倍、363.66 倍の差が検出された。RT-リアルタイム PCR により、発現量比を定量したところ、温度 No.2、No.3、No.5 遺伝子の発現量はそれぞれ  $26.21 \pm 5.35$  倍、 $556.42 \pm 364.59$  倍、 $666.37 \pm 880.91$  倍の差が検出された (Figure 17D)。RT-PCR より算出された発現量の差は RNA-Seq 解析の No.2 遺伝子で 35%、No.3 遺伝子で 598%、No.5 遺伝子で 183% であった。

RNA-Seq 解析の結果より、アメリカ産 GL3494 品種と比較し、カナダ産 OAC Kent 品種において発現量に 637.96 倍の違いのある遺伝子 GLYMA17G18930.1 の RT-リアルタイム PCR 定量を行った。まず、アメリカ産 GL3494 品種及びカナダ産 OAC Kent 品種においてゲノム配列をシーケンシングしたところ、データベース上のリ

ファレンス配列と完全一致した (Figure 18)。Figure 19 は、RT-リアルタイム PCR 用のプライマーの認識配列位置を示す。遺伝子発現定量を行った結果、アメリカ産 GL3494 品種の遺伝子発現量は、カナダ産 OAC Kent 品種では 55327.40 倍、Williams 品種では 28452.20 倍の発現量の差が検出された (Figure 20)。

#### 6. アレルゲンデータベースなどの既存のデータベースとバッチ解析を行う方法の検討

遺伝子発現量に差のある遺伝子の配列リストを基に、その遺伝子がコードするタンパク質のアミノ酸配列リストを得た。ダイズ cDNA 配列 (V1.0.30) の FASTA フォーマットデータは、「EnsemblPlants」

([http://plants.ensembl.org/Glycine\\_max/Info/Index](http://plants.ensembl.org/Glycine_max/Info/Index)) よりダイズ cds 配列

([ftp://ftp.ensemblgenomes.org/pub/plants/release-30/fasta/glycine\\_max](ftp://ftp.ensemblgenomes.org/pub/plants/release-30/fasta/glycine_max)) の取得を行った。cds 配列は、タンパク質への翻訳をソフト上で行い、配列を得た。(ファイル名: Glycine\_max.V1.0.30.cds.all translations) 得られたファイルは、「Sequence CSV」フォーマットで export した。保存したアミノ酸配列リストは、エクセル上で遺伝子発現量の変化のあった遺伝子リストと照合を行った。データ照合の無い遺伝子に関しては、「EnsemblPlants」のデータベースを参照し

て、アミノ酸配列を取得した。得られたすべての遺伝子翻訳後のアミノ酸配列は、データベース上に登録されているアレルゲン配列との照合を行った。発現量差のある遺伝子については既知のアレルゲンとの比較（相同性検索）をデータベース上で行った。アレルゲンの検索サイトには、University of Nebraska-Lincoln が運営する「AllergenOnline」を使用した。「AllergenOnline」では、FASTA format よりアレルゲンデータベースを検索可能であることが確認された。**Figure 21** は、遺伝子発現量の変化のあった遺伝子発現差リストと遺伝子番号を参照して照合した結果を示した。このリストには、遺伝子発現量の差のあったデータベース中の遺伝子番号、ゲノム配置場所、及び遺伝子がコードするアミノ酸配列が含まれる。このリストを、FASTA format に置換して、アレルゲンデータベースを使用したアレルゲン検索を行った結果の例を **Figure 22** に示す。

## 7. アノテーションされていない新規遺伝子のリスト化

RNA-Seq で遺伝子領域にマッピングされなかったリードのリスト化を行った。まず、ダイズゲノムにマッピングされたリードの除去を行った。アメリカ産 40°C 発芽種子より得られたデータ (CTTGTA タグ配列) を例に用いて、RNA-Seq Analysis ツー

ルで、遺伝子の領域のみマッピングを行い、その際に作成される Un-mapped reads データを作成した。Unmapped リードの中から再度ゲノムへマッピングされた配列を得た。RNA-Seq Analysis ツールで「Also map to inter-genic regions」を選択して、再度、マッピングを行った。遺伝子アノテーションされていない領域の配列を FASTA フォーマットで得ることができた。

次に得られた配列の reading frame (+1 ~+3 または-1~-3) を選択して、ペプチドのアミノ酸配列を得た。その配列を、Genomic Workbench より Sequence CSV フォーマットで Export して、テキストもしくはエクセル上で表示させた (**Figure 23**)。その結果、リーディングフレームを 5' 及び 3' 側から両方に翻訳される可能性のあるアミノ酸を全て推測したペプチド配列のリストが得られた。

## D. 考察

初年度は、本研究の最終年度に行う組換え技術により開発した GM 型ダイズと非 GM 型ダイズとの発芽遺伝子の発現に関する比較実験を行うための発芽条件の検討とリファレンスデータの構築を目的に、非 GM 型ダイズの RNA-Seq 解析を行った。

サンプル間の遺伝子発現量を統計学的に解析するためには、RNA-Seq による出力データ量を増やす必要がある。本研究では、

1 サンプルあたり 7,000 万リードを出力データ量の目標とした。出力データ中の非特異的なリード数を極力抑える解析プロトコルを作成する必要があるがあった。発芽ダイズ食品生産の際に使用される発芽条件では、種子を発芽させる際に腐敗し、細菌が増殖する (LB-agar プレートにて白金耳を使用して植菌後 37°Cにて培養、細菌のコロニー形成を確認した[データ提示せず]) ことによって、細菌由来のゲノム DNA 又は RNA は、ダイズ由来の RNA を抽出・精製する際にサンプル中に濃縮されてしまうことが考えられた。そのため本研究では、ガス滅菌した種子と滅菌水を使用して発芽させ、発芽ダイズサンプル中の細菌の増殖を抑え、細菌由来のゲノム DNA 又は RNA の混入を防ぐ発芽プロトコルを構築した。

RNA-Seq より出力されるリードの数は、各遺伝子の発現量に比例すると仮定して、サンプル間の発現量を定量する際のデータとして使用するため、mRNA の分解によりシーケンスリードの重複やゲノム配列へのマッピングができない様な問題を避けられるよう、発芽ダイズの粉碎は、液体窒素を利用して種子と器具を冷却した上で、乳鉢・乳棒を使用して均一に粉碎することで、精製 RNA の分解を防ぐプロトコルを構築することができた。また、RNA 精製時に、DNaseI を使用してゲノム DNA を完全に分解し、ゲノム DNA の RNA サン

プルへの混入を防ぐプロトコルとした。以上の試料調製法にて、品種及び発芽温度を変えた環境下で発現する遺伝子の特定及び発現定量を行った。その結果、本年度は、発芽ダイズを試料とした RNA-Seq 解析のプロトコルを立ち上げることができた。

RNA-Seq 解析の結果、27°Cと 40°Cで発芽させた際の温度差によって 2 倍以上 ( $p < 0.05$ ) 発現量に差のある遺伝子 12,722 遺伝子が検出された。最大の発現量の差のある遺伝子 (GLYMA07G06750.2) で 1,174 倍 ( $p$  値  $9.42 \times 10^{-44}$ ) の差が検出された。ランダムに選んだ、75~364 倍の差のあった 3 種類の遺伝子を RT-リアルタイム PCR より定量した結果、RNA-Seq 解析で得られた発現の差が確認された。また、発芽温度も含め一定の環境下で発芽させたダイズの品種間においても 2 倍以上の差 ( $p < 0.05$ ) のある 8,418 遺伝子が検出され、RT-リアルタイム PCR において発現の差が確認された。最終年度では、発現を検出した遺伝子群のリード数が、真の発現量を反映しているかを検証するため、リアルタイム PCR を使用して解析を継続し発現量の定量確認を試みる予定である。

本研究結果より、構築した発現量の差のある遺伝子リストに関しては、発現するタンパク質のアミノ酸配列を既存のアレルゲンデータベースとバッチ比較する方法を検討した。「AllergenOnline」を使用する

ことで、RNA-Seq解析から得られる FASTA format のデータよりアレルゲンデータベースで相同性のある既知アレルゲンを検索可能であることが示唆された。アノテーションされていないリード配列から、新規遺伝子の発現に関する新しい知見が得られた。また、サンプル間の遺伝子発現量の比較情報を基に、リアルタイム PCR 定量用プライマー対の設計が可能であることが示された。

今後は、実際に GM ダイズ中に発現している遺伝子の非 GM ダイズと比較して発現差のある遺伝子のリスト化及び既存のアレルゲンデータベースなどの検索サイトへの適用性に関して検証を行う予定である。

#### E. 結論

初年度の結果より、発芽ダイズの RNA-Seq を行う場合は、①同じ品種間で発芽時の環境を揃える、②発現量の再現性及びリード数からの発現量算出の統計学的処理のため、1 サンプルにつき 2 試験以上の結果からトータル RNA を抽出精製その後 RNA-Seq を行う必要性が示唆された。RNA-Seq で得られたデータは、研究期間 2 年間における解析が終了した後、発芽時における遺伝子発現量及び質、並びに、コーディング領域の情報として DDBJ BioProject Database へ登録し、また、次世

代シーケンスからの出力データは DDBJ Sequence Read Archive へ登録しデータベース化後、公開する。そのデータベースより、発芽遺伝子から発現すると予想されるタンパク質のアミノ酸配列は、アレルゲンデータベースなどのデータと照合し、既知アレルゲンとの相同性検索に使用する事が可能であった。また、ダイズの未知の発芽遺伝子を複数特定し、発現するタンパク質のアミノ酸配列予測を行うことができた。今後、RNA-Seq より得られたデータは、転写レベルではリアルタイム PCR、翻訳レベルでは LC-MS-MS 及び 2D-DIGE を使用したプロテオーム解析で定量・定性的に確認を行う予定である。

#### F. 健康危険情報

なし

## G. 研究発表

### 1. 論文発表 :

- 1) Nakamura, K., Kondo, K., Akiyama, H., Ishigaki, T., Noguchi, A., Katsumata, H., Takasaki, K., Futo, S., Sakata, K., Fukuda, N., Mano, J., Kitta, K., Tanaka, H., Akashi, R., Nishimaki-Mogami, T. Interlaboratory study on unauthorized genetically modified papaya PRSV-YK real-time PCR detection method. *Data in Brief*, 7, 1165-1170, 2016.
- 2) Nakamura, K., Kondo, K., Akiyama, H., Ishigaki, T., Noguchi, A., Katsumata, H., Takasaki, K., Futo, S., Sakata, K., Fukuda, N., Mano, J., Kitta, K., Tanaka, H., Akashi, R., Nishimaki-Mogami, T. Whole genome sequence analysis of unidentified genetically modified papaya for development of a specific detection method. *Food Chemistry*, 205, 272-279, 2016.
- 3) Nakamura, K., Matsuoka, H., Nakashima, S., Kanda, T., Nishimaki-Mogami, T., Akiyama, H. Oral administration of apple condensed tannins delays rheumatoid arthritis development in mice via down-regulation of T helper 17 (Th17) cell responses. *Molecular Nutrition & Food Research*, 59, 1406-1410, 2015.
- 4) Nakamura, K., Ishigaki, T., Hanada, K., Akimoto, S., Kondo, K., Nishimaki-Mogami, T. DNA methylation pattern analysis of common plant virus promoter used to develop genetically modified crops, PacifiChem2015, Hawaii, USA, 2015年12月
- 5) Kondo, K., Sakata, K., Noguchi, A., Nakamura, K., Fukuda, N., Ishigaki, T., Nishimaki-Mogami, Tomoko. A new analytical methodology for unknown genetically modified organisms using linear-amplified mediated PCR (LAM-PCR), 7th International Symposium on Recent Advances in Food Analysis, Prague, Czech Republic, 2015年11月
- 6) Kondo, K., Sakata, K., Nakamura, K., Noguchi, A., Fukuda, N., Ishigaki, T., Nishimaki-Mogami, Tomoko. Rapid identification of poisonous *Entoloma rhodopolium* and edible *E. sarcopum* mushrooms using PCR-restriction fragment length polymorphism and real-time PCR, 7th International Symposium on Recent Advances in Food Analysis, Prague, Czech Republic, 2015年11月
- 7) 中村公亮、石垣拓実、近藤一成、最上(西巻)知子：汎用性ウィルスプロモーター導入によるクロマチンループ内の内在性遺伝子発現への影響、日本薬学会 第136年会、横浜、2016年3月

### 2. 学会発表

- 1) Nakamura, K., Ishigaki, T., Hanada, K., Akimoto, S., Kondo, K.,

- 5) 中村公亮、近藤一成、穂山浩、石垣拓実、野口秋雄、坂田こずえ、福田のぞみ、大森清美、布施谷実聡、川上浩、田中秀典、明石良、真野潤一、橘田和美、最上（西巻）知子：我が国における未承認遺伝子組換えパパイヤの食品への混入に関する事例と検知法開発の現状、第 52 回全国衛生化学技術協議会年会、静岡、2015 年 12 月
- 6) 野口秋雄、中村公亮、真野潤一、高畠令王奈、橘田和美、近藤一成、最上（西巻）知子：遺伝子組換えトウモロコシの新規スク 52 回全国衛生化学技術協議会年会、静岡、2015 年 12 月
- 7) 福田（佐藤）のぞみ、近藤一成、坂田こずえ、中村公亮、野口秋雄、最上（西巻）知子：遺伝毒性試験および全ゲノム解析を用いた CRISPR/Cas9 の DNA2 本鎖切断ポテンシャル、第 38 回日本分子生物学会年会、神戸、2015 年 12 月
- 8) 坂田こずえ、近藤一成、野口秋雄、中村公亮、福田のぞみ、石垣拓実、最上（西巻）知子、LAM-PCR を用いた組換え作物中の未知領域解析法の検討、第 110 回 日本食品衛生学会学術講演会、京都、2015 年 10 月
- 9) 野口秋雄、町井香苗、中村公亮、真野潤一、高畠令王奈、橘田和美、川上浩、近藤一成、最上（西巻）知子：遺伝子組換えトウモロコシの簡易粒検査法の開発、第 110 回 日本食品衛生学会学術講演会、京都、2015 年 10 月
- 10) 菅野陽平、坂田こずえ、野口秋雄、中村公亮、小林友子、福田のぞみ、佐藤正幸、青塚圭二、鈴木智宏、最上知子、手島玲子、近藤一成：ツキヨタケの PCR-RFLP を用いた迅速同定法の検討（第 2 法）：加熱、消化処理サンプルへの適用、第 110 回 日本食品衛生学会学術講演会、京都、2015 年 10 月
- 11) 中村公亮、近藤一成、石垣拓実、野口秋雄、坂田こずえ、福田のぞみ、大森清美、真野潤一、橘田和美、最上（西巻）知子：安全性未承認遺伝子組換えパパイヤ（PRSV-HN 系統）の検出と検知法開発、第 110 回 日本食品衛生学会学術講演会、京都、2015 年 10 月
- 12) 中村公亮、石垣拓実、坂田こずえ、福田のぞみ、野口秋雄、穂山浩、近藤一成、真野潤一、高畠令王奈、橘田和美、最上（西巻）知子：未承認遺伝子組換え食品検知法の開発：未承認遺伝子組換えジャガイモ検知を例に、第 1 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、千葉、2015 年 9 月
- 13) 真野潤一、波田野修子、布藤聡、峯岸恭孝、二宮健二、中村公亮、近藤一成、手島玲子、高畠令王奈、橘田和美：食

品遺伝子検査を簡易化するダイレク  
トリアルタイム PCR、2015 年度 AOAC  
International 日本セクション年次大会、  
東京、2015 年 6 月

modified papaya PRSV-YK genomic sequence  
data, DDBJ Sequenced Read Archive  
accession no. PRJDB3976 (Registration date:  
8-Mar-2016).

- 14) 中村公亮、石垣拓実、近藤一成、最上  
（西巻）知子：次世代ゲノム編集技術  
による遺伝子組換え食品の内在性遺  
伝子発現への影響、日本食品化学学会  
第 21 回総会・学術大会、東京、2015  
年 5 月
- 15) 石垣拓実、中村公亮、近藤一成、最上  
（西巻）知子：遺伝子組換えヒヨコマ  
メ検査法確立に向けたヒヨコマメ内  
在性遺伝子（CaNCED）特異的検知法  
の開発、日本食品化学学会 第 21 回  
総会・学術大会、東京、2015 年 5 月

## H. 知的財産権の出願・登録状況

### 1. 特許

中村公亮、小林友子、近藤一成：遺伝子組  
換え植物の判定法、特開 2016-019480、公  
開日：2016 年 2 月 4 日

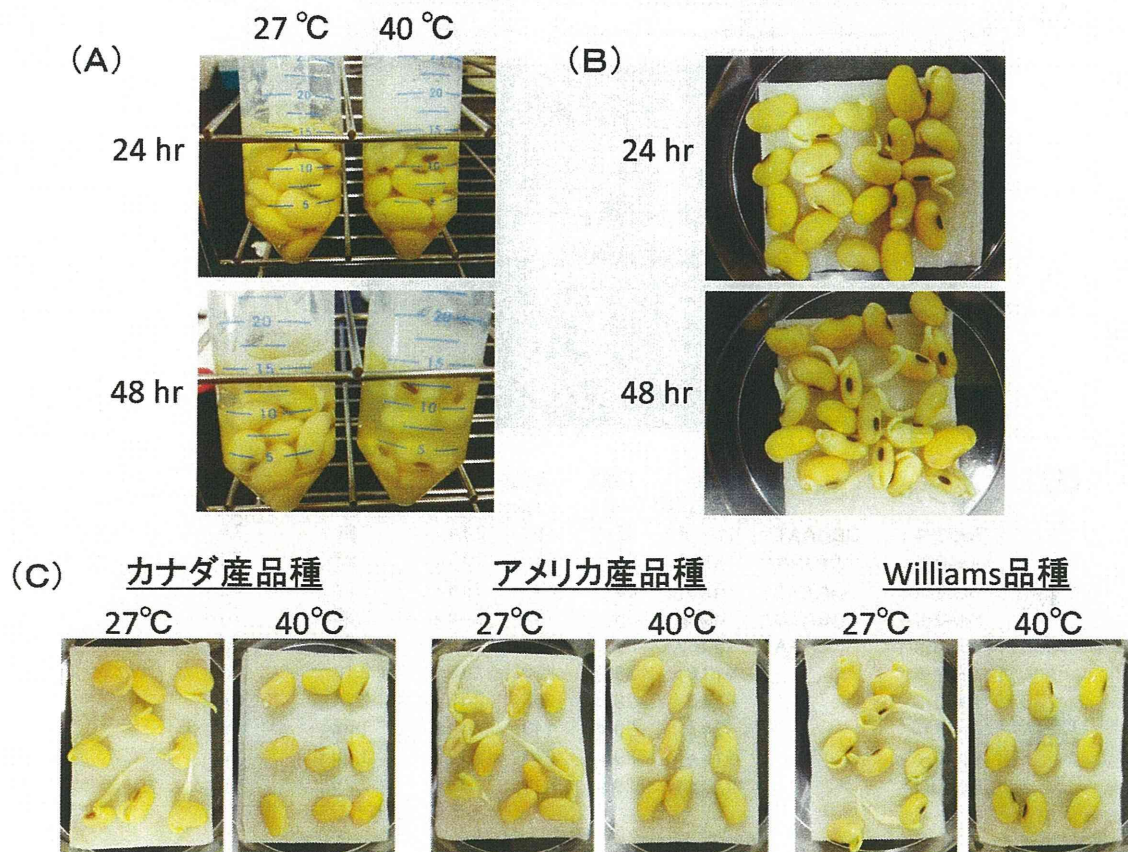
### 2. 実用新案登録

なし

### 3. その他

次世代シーケンサーを使用したゲノム解  
析データのデータベースへの登録：  
Nakamura, K., Takasaki, K. Genetically





**Figure 1.** ダイズ播種条件の検討

(A) 27°C又は40°C水浸漬下で24時間及び48時間培養した後のダイズ (B) 滅菌発芽条件下で27°C 24時間または48時間培養したダイズ (C) アメリカ産品種、カナダ産品種およびWilliams品種を、27°Cと40°Cで48時間発芽

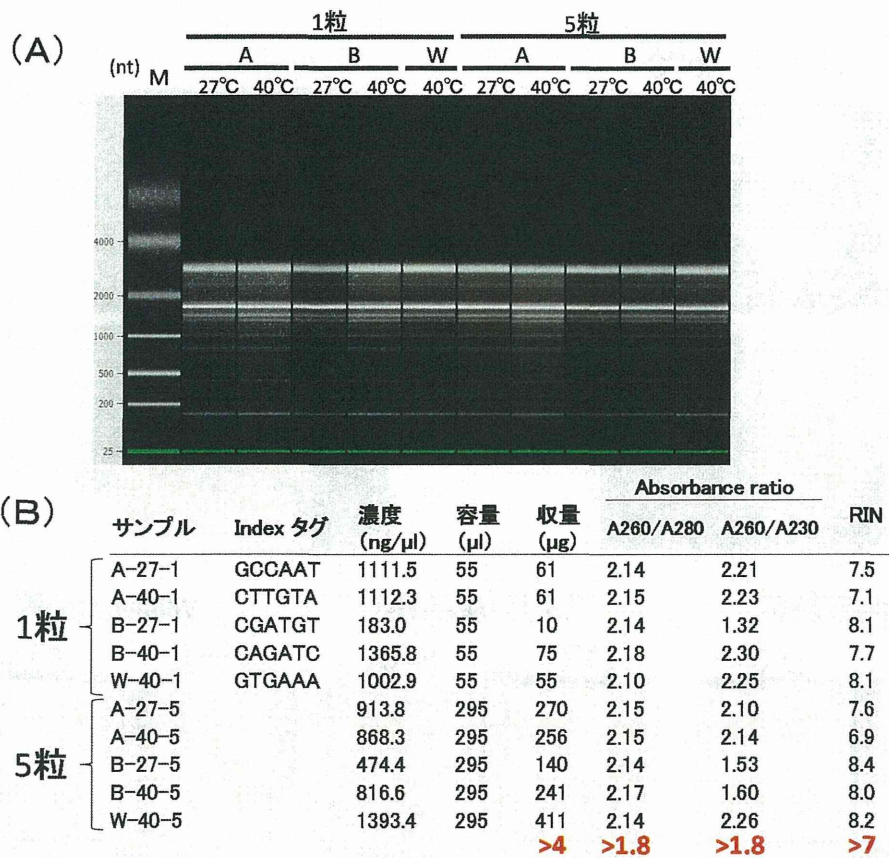


Figure 2. ダイブトータル RNA の精製結果

(A) Bioanalyzer を使用した電気泳動図 (B) 精製 RNA の評価結果



表1 リード配列のトリミング結果一覧

サンプル	アダプター配列	リード数	平均長 (base)	リード数(トリミング後)	トリミング率 (%)	トリミング後の平均リード長 (base)
アメリカ産27°C	GCCAAT	79,160,348	101	78,843,604	99.60	99.7
アメリカ産40°C	CTTGTA	72,546,664	101	72,260,274	99.61	99.7
カナダ産27°C	CGATGT	76,819,176	101	76,550,425	99.65	99.7
カナダ産40°C	CAGATC	72,208,242	101	71,928,417	99.61	99.7
Williams品種40°C	GTGAAA	71,448,556	101	71,209,645	99.67	99.7

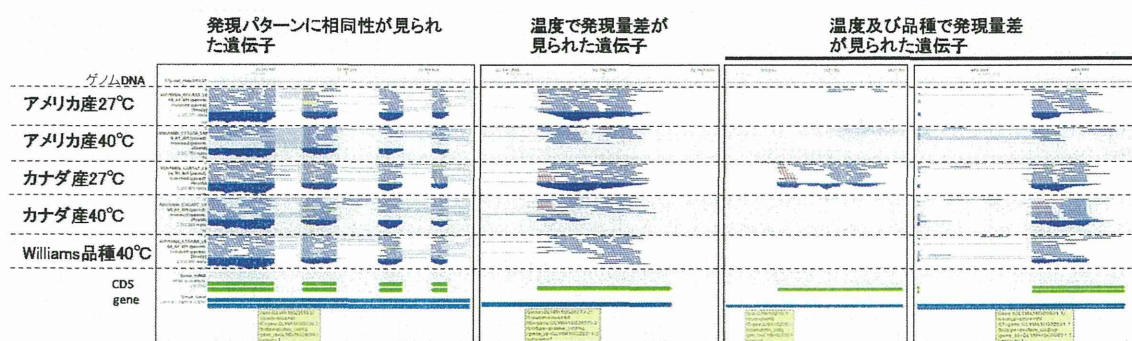


Figure 4. ドライズゲノムへのマッピング結果

リファレンスゲノムとしてドライズゲノム V1.0.28 を使用した。サンプルより得られたリードをゲノムに最も高いアラインメントスコアを示す場所にマッピングすると同時に、マッピング後のデータを local realignment した。GCCAAT (アメリカ産 27°C) は 2,115,935 リード、CTTGTA (アメリカ産 40°C) は 2,123,999 リード、CGATGT (カナダ産 27°C) は 1,998,598 リード、CAGATC (カナダ産 40°C) は 2,057,249 リード、GTGAAA (Williams 産 40°C) は 2,146,500 リードがゲノムデータベース上にマッピングされた。結果は、ドライズ成体由来の CDS 及び遺伝子リストを使用し、アノテーション付加を行ったトラックリスト上に示す