

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28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 29 (2019) 年 5 月

目 次

I. 総括研究報告書

研究要旨-----	2
A. 研究目的-----	2
B. 研究方法-----	2
C. 研究結果-----	9
D. 考察-----	13
E. 結論-----	14
F. 健康危険情報-----	14
G. 研究発表-----	14
H. 知的財産権の出願・登録状況-----	17
図表一覧-----	18 ~ 38

II. 研究成果の刊行に関する一覧表-----	39
-------------------------	----

厚生労働科学研究費補助金（食の安全確保推進研究事業）
研究報告書（平成 28 年度）

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

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、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（発芽遺伝子）の RNA-Seq 解析を実施した。また、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した。こうして、ダイズの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方法を開発し、GM と非 GM ダイズを比較する際の有用性について検証を行った。

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）

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。

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手法を開発し、遺伝子組換え（GM）型と非 GM ダイズの

B. 研究方法

1. 試料と発芽条件

試験には、非 GM 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）、Jack 品種、Williams 品種、及び、Green Fluorescent Protein（GFP）とダイズ由来 SYNC1 を発現するよう組換えた Williams 品種（GM ダイズ）の 5 品種を供した。種子は、発芽前に滅菌処理を行った。10 cm シャーレー枚

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、次亜塩素酸ナトリウム（有効塩素 6%）100 mL に濃塩酸 3.5 mL を加え、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した。種子の発芽は、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 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. RNA-Seq 解析

2.1 試料の調製

発芽ダイズ生産の発芽条件（40 °C）下で 48 時間培養し発芽させた 1. のダイズ 5 品種より、粒単位でトータル RNA の抽出・精製を行った。1 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、Agilent Bioanalyzer 2100 system（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。RNA の濃度と精製度は、NanoDrop 2100 spectrophotometer（サーモサイエンティフィック社）を使用して推定した。得られたトータル RNA 1.5 μg を試料に次世代シーケンシング用のライブラリの調製に供した。ライブラリの調製には、NEBNext® Ultra™ RNA Library Prep Kit for Illumina（NEB 社）を使用し、各試料にはタグ配列を付加した。以下にその概要を記す。mRNA 精製は、poly-T oligo を付加した磁石ビーズで行った。得られた mRNA は、NEBNext First Strand Synthesis Reaction Buffer（5X）中で加熱し、二価カチオン存在下で断片化させた。断片化させた mRNA は、ランダムヘキサマープライマーを使用し、M-MuLV Reverse Transcriptase（RNaseH-）により逆転写させた。cDNA の相補鎖は、dTTP の代わりに dUTP を含む dNTP を使用して DNA Polymerase I により合成し RNase H を使用して mRNA を分解させて行った。3' 末をアデニル化した後、NEBNext Adaptor を付加した。合成した cDNA は、150~200 bp の鎖長を AMPure XP system（ベックマンコルター社）を使用して単離した。USER Enzyme（NEB 社）を使用して、ウラシルを含む DNA 鎖を断片化した。次に、Phusion High-Fidelity DNA polymerase、Universal PCR プライマー、Index タグプライマーを使用して PCR を行った。得られ

た PCR 産物は、AMPure XP system を使用して精製を行い、Agilent Bioanalyzer 2100 system を使用してクオリティチェックを行った。Index タグを付加したサンプルは、cBot Cluster Generation System (イルミナ社) を使用して、TruSeq PE Cluster Kit v3-cBot-HS キット (イルミナ社) によるフローセルへのクラスター化を行った。シーケンシングは、100-base paired-end でフローセルの 5 plex /1 レーンを用いてイルミナ HiSeq2500 により行った。

2.2. データの解析

シーケンサーより得られた Fastq ファイルは、Genomic Workbench ver.9.0.1 を使用して、リード配列のトリミングを行った。トリミングは、アダプター配列の除去すること、10%以上の未解読塩基配列を含むリードであること、50%以上の塩基配列でクオリティスコア (Q 値 5) を有するリードの除去することを条件に行った。本試験に使用したアダプター配列は、以下の通りである。

5'アダプター :

5'-AATGATACGGCGACCACCGAGATCTCACTCTTTCCCTACACGACGCTCTTCCGATCT-3'、

3'アダプター :

5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCACATCACGATCTCGTATGCCGTCTTCTGCTTG-3' (アンダーラインした 6 塩基は、タグ配列)

トリミングを行ったリードは、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/) をリファレンス配列に使用し、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 (リマッピング) した。各サンプルに関して、RNA-Seq を行いサンプル間の遺伝子発現差解析を行った。発現差解析の条件は、各品種のデータを Two-group comparison (paired) で解析した。発現差解析では、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。Empirical Analysis of Digital Gene Expression (edgeR ソフトウェア , *Biostatistics*, 9, 321-332, 2008; *Bioinformatics*, 26, 139-140, 2010) を使用して 2 群の比較検定を行った。条件の設定は、発現量があるとするための最初のカウント数を 5 リード数とした。リファレンス配列と比較し、

遺伝子発現量が 2 倍量以上の差 ($p < 0.05$) のある遺伝子を選抜した。

3. リアルタイム PCR による遺伝子発現の定量

2. 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. の方法で発芽サイズより調製した 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 μ L の反応液中で逆転写反応を行い、cDNA を合成した。2 μ 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と QuantiTect SYBR® Green PCR (QIAGEN) を使用したリアルタイム PCR により遺伝子発現の定量を行った。PCR 反応液は、20 μ L/well として、以下のとおり調製した。2 \times QuantiTect SYBR® Green PCR master mix 10 μ L、対象プライマー対溶液 (各プライマー、50 μ mol/L) 各 0.2 μ L を混合し、cDNA 試料液 0.5 μ L を添加し滅菌蒸留水で全量 20 μ L に調製した。反応条件は、50 で 2 分間、95 で 10 分間加温し、その後、95 で 15 秒間、60 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、サ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 (Table 12)。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 ~ 3,000 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。種子の品種別において、発現量の異なる遺伝子の発現量を検証するために使用した RT-PCR 用プライマー対の配列は、Table 12 に示した。各プライマーセットの特異性は、PCR 後、2% (w/v) アガロースを使用した電気泳動により検証した。

4. PCR を使用したゲノム DNA 配列の検出

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

(MM200 Retch) で粉碎し、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 反応を使用した。標的アンプリコンの長さが 500 ~ 1,500 bp の場合、PCR 反応液は、25 μ L/well として、以下の組成のとおり調製した。10xExTaqBuffer (Mg^{2+}) 10 μ L、対象プライマー対溶液 (各プライマー、50 μ mol/L) 各 3 μ L、2.5 mM dNTP 5 μ 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 μ L となるよう調製した。反応条件は、50 $^{\circ}C$ で 2 分間、95 $^{\circ}C$ で 10 分間加温し、その後、95 $^{\circ}C$ で 15 秒間、60 $^{\circ}C$ 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 2,000 ~ 3,000 bp の場合、PCR 反応液は、25 μ L/well として、以下の組成のとおり調製した。10xLATAqBuffer (Mg^{2+}) 2.5 μ L、対象プライマー対溶液 (各プライマー、50 μ mol/L) 各 3 μ L、2.5 mM dNTP 4 μ 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 μ L となるよう調製した。反応条件は、94 $^{\circ}C$ で 1 分間加温後、98 $^{\circ}C$ で 10 秒間、50 $^{\circ}C$ で 30 秒間、72 $^{\circ}C$ 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。

5. プロテオーム解析用の試料調製

5.1. 分析試料

発芽ダイズ(Williams 品種、Jack 品種、及び、GM Williams 品種) は、解析するまでの間、-80 $^{\circ}C$ で保存した。

5.2. 試料粉碎と沈殿処理

ダイズの各品種より 1 粒ずつを液体窒素中で凍結し、乳鉢を用いて別々に粉碎した。粉碎物を、10% トリクロロ酢酸と 0.07% 2-メルカプトエタノールを含むアセトン溶液 8 mL に懸濁した。懸濁液を 10-mL 遠沈管に回収し、-20 $^{\circ}C$ で 45 分間静置した。遠心分離 (35,000 \times g、0 $^{\circ}C$ 、15 分間) してから上清を除いた。沈殿は 0.07% 2-メルカプトエタノールを含むアセトン溶液で 3 回洗浄した。洗浄後の沈殿を凍結乾燥して粉末試料を得た。

5.3. 粉末試料からのタンパク質抽出

粉末試料から 3 mg を 1.5-mL マイクロ容器に分取した。分取した粉末に抽出溶液 [6 M 尿素、2 M チオ尿素、60 mM ジチオトレイトール (DTT)、100 mM 重炭酸アンモニウム、pH 8.8] を 100 μ L 加え、攪拌しながら 37 $^{\circ}C$ で 1 時間保温した。その後、遠心分離 (15,000 \times g、20 $^{\circ}C$ 、15 分間) し、タンパク質を含む上清を回収した。

5.4. タンパク質濃度測定

回収した上清の一部を、Bradford 法による総タンパク質定量に供した。定量用の

検量線は、ウシ血清アルブミンの溶液を段階希釈して作成した。

5.5. トリプシンによる試料タンパク質の加水分解

試料溶解液から総タンパク質 50 µg 分を 1.5-mL マイクロ容器に分取した。分取液に DTT を加え 37 °C で 30 分間保温し、続いてヨードアセトアミドを加えてから室温で 1 時間静置した(還元アルキル化処理)。処理後の溶液に 100 mM 重炭酸アンモニウムを加え、溶液の尿素濃度を 2 M まで下げた。最後にトリプシン 2.5 µg を加え、37 °C で 16 時間保温して加水分解反応を行った。反応後のペプチド溶液には C18 STAGE Tip による脱塩処理を施した (Anal. Chem., 75, 663-70, 2003)。脱塩後の試料を減圧下で乾燥した。

5.6. LC-MS/MS

乾燥状態のペプチド試料を、水、アセトニトリル、及び、トリフルオロ酢酸からなる溶媒 (体積比 98:2:0.1) に溶解した。出発総タンパク質量に換算して 200 ng 相当量を LC-MS/MS に供した。LC-MS/MS システムの仕様と設定条件は、下記に示す。

LC : Ultimate3000 液体クロマトグラフ (ダイオネクス社)

・分析用 C18 カラム (Tip column) : Nano HPLC Capillary Column (粒径 3 µm、内径 75 µm、長さ 15 cm、日京テクノス株式会社)

・移動相 A の組成 : [水] : [アセトニトリル] : [ギ酸] = 98:2:0.1 (体積比)

・移動相 B の組成 : [水] : [アセトニトリル] : [ギ酸] = 5:95:0.1 (体積比)

・アセトニトリル送液勾配 (分, %B, %アセトニトリル) : (0, 2, 3.86) → (5, 2, 3.86) → (120, 33, 32.69) → (120.01, 95, 90.35) → (130, 95, 90.35) → (130.01, 2, 3.86) → (145, 2, 3.86)

・流速 : 毎分 350 nL

MS/MS : Q Exactive 質量分析計 (サーモフィッシャーサイエンティフィック)

・イオンモード : 陽イオンモード

・イオントランスファーキャピラリーの設定温度 : 250

・FullScan の m/z 走査範囲 (Scan range) : 300 ~ 1,500

・質量分解能 (Resolution) : 70000 (MS), 17500 (MS/MS)

・Lock Mass : Or(Reference m/z = 391.28429, 445.12003)

・測定条件ファイル : 「Top 10 Method」を用いた。すなわち、FullScan (m/z 300 ~ 1500) の質量スペクトルの上で、検出強度の高いピークから順に 10 個の MS/MS デー

タを取得した。この FullScan と MS/MS データの取得を交互に実施するよう測定変数を設定した。前の試料由来のペプチドの検出 (キャリーオーバー) を抑えるため、各試料の測定の間それぞれ 3 回分の空測定を挿入した。

5.7. 配列データベース検索によるペプチド/タンパク質の同定

MS/MS データを配列データベース検索に供した。検索ソフトウェアとして Matrix Science 社の Mascot (ver. 2.5; <http://www.matrixscience.com/>) を用いた。検索用配列データセットは下記のとおり 4 種類作成し、アノテーションに用いた。

CDS (*Glycine max*) : ダイズ CDS の配列データセット (計 73,319 件) に、3 種類の配列 [Bialaphos resistant gene (bar) , Enhanced green-fluorescent protein (eGFP) , 及び、lysyl-tRNA synthetase (SYNC1)] を加えて構築した。

Uniprot (*Glycine max*) : Uniprot (<http://www.uniprot.org/>) から出力したダイズ (*Glycine max*) のアミノ酸配列データベース (reference proteome set) 計 66,206 件 (<http://www.uniprot.org/proteomes/UP000008827>) に、上記と同様 3 種類の配列 (bar、

eGFP、および SYNC1) を加えて構築した。

Uniprot/SwissProt (Green Plants) : Uniprot/SwissProt (http://ftp.ebi.ac.uk/pub/databases/uniprot/current_release/knowledgebase/) 2016_01 版 (計 550,299 件) から Green Plants に分類されるタンパク質配列を抜粋した (計 37,228 件)

NCBI/Genome (*Glycine max*) : NCBI Genome データベース (<https://www.ncbi.nlm.nih.gov/genome/>) から出力したダイズ (*Glycine max*) のアミノ酸配列データベース (計 71,526 件) に、上記と同様 3 種類の配列 (bar、eGFP、及び、SYNC1) を加えて構築した。

データベース検索の条件は次の通り :
Enzyme, Semi Trypsin
Maximum missed cleavage, 2
Peptide tolerance, ± 5 ppm
MS/MS tolerance, ± 0.02 Da
Mass, monoisotopic mass
Fixed Modification, Carbamidomethyl (C, +57.021)
Variable Modification, Oxidation (M, +15.995)

有意なペプチド同定は、False discovery rate (FDR) を指標にして選定した。すな

わち、FDR が 1% になるようにペプチド同定のスコア閾値を調整した。

5.8. ペプチドの同定情報と検出強度値の連結

各試料から得られた LC-MS/MS のデータを Nonlinear Dynamics 社の Progenesis QI for proteomics (ver. 2.0; <http://www.nonlinear.com>) に入力し、各検出ピークの強度値を取得した。続いて、ペプチドの同定情報を各検出ピークに連結し、連結された検出ピークの強度を当該ペプチドの検出強度とした。また、タンパク質の計量値は、各タンパク質に帰属するユニークペプチドの検出強度の積算値とした。

C. 研究結果

1. 発芽サイズと RNA-Seq 用サンプルの調製

本研究で使用した発芽サイズを **Table 1** に示す。**Scheme 1** に従って、5 品種 (カナダ産、アメリカ産、Jack 品種、Williams 品種、GM Williams 品種) を発芽させ、発芽させた各品種 2 粒ずつを選択し、それぞれ 1 粒からトータル RNA を抽出精製した。その結果、2.29 ~ 37.69 μg のトータル RNA を得た (**Table 2**)。得られたトータル RNA の吸光度比は、 A_{260}/A_{280} 2.06 ~ 2.24 と A_{260}/A_{230} 0.37 ~ 2.06、及び、RNA Integrity

Number (RIN 値) は、7.1 ~ 8.4 であった。この結果より、トータル RNA は、非常に高い回収量と精製度で、得られたことが判った。

2. RNA-Seq 解析

得られたトータル RNA を用いて、RNA-Seq 用サンプルの調製を行った (**Scheme 2**)。Oligo (dT) を使用して、トータル RNA より mRNA を単離後、逆転写させ cDNA を合成し、両末端側にアダプター配列とインデックスタグを付加した後、Illumina HiSeq2500 を使用してシーケンシング解析を行った。その結果、Raw read のクオリティー値の閾値を Q20 として設定した結果、全リード中の塩基数として 97.64 ~ 98.38%、又は、全リード中のリード数として 98.62 ~ 98.89% (44,034,849 ~ 60,574,989 リード) 使用することが可能であった (**Table 3**)。この結果から、**Scheme 1**、及び、**Scheme 2** に従って調製したサンプルは、RNA-Seq へ十分に実用可能であることが示唆された。

3. サンプル間における遺伝子発現差の解析

3.1. 非 GM と GM の発現差解析

GM サイズは、サイズ由来 Lys-tRNA 遺伝子 (*Sync1*)、ビアラホス耐性遺伝子

(BAR) 及び、ever green fluorescent protein 遺伝子 (EGFP) を高発現させるよう構築したトランスジェニックベクターを非 GM Williams 品種 (非 GM ダイズ) へトランスフェクションしたものである (Figure 1 A, B)。GM ダイズのトランスジェニック構造配列は、ホモ型でゲノム中に 1 か所のみ導入されたことを確認した (データ示さず)。このように開発した GM ダイズと非 GM ダイズに発現する発芽遺伝子を比較するため、RNA-Seq は、それぞれ 2 粒ずつを試料に供した。その結果、トリミング後に得られたリード数は、非 GM ダイズで (種子 G, 43,843,369 リード、種子 H, 53,372,808)、GM ダイズで (種子 I, 56,974,554 リード、種子 J, 49,153,651 リード) であった。それぞれのリードは、ダイズリファレンスゲノム v.1.01 にマッピングし、リアライメント後のデータを得た。そのデータより、品種間で 2 倍以上の差 ($p < 0.05$) のある発芽遺伝子を選抜した (Figure 1 C, D)。Table 4 と Table 5 は、非 GM ダイズの種子間で発現量の差の高い上位 20 位の遺伝子と、非 GM ダイズと GM ダイズの種子間で発現量差の高い上位 20 位の遺伝子をそれぞれ示す。その結果、染色体 6 番 (配列番号 47071378 ~ 47075313) に存在する遺伝子

GLYMA06G44091.1 は、非 GM ダイズ間で最も発現量の差が高く 613.48 倍 (p 値 $= 1.76 \times 10^{-46}$) の差を検出した。また、非 GM ダイズと GM ダイズを比較して最も発現量の差が高かった染色体 11 番 (10078670 ~ 10082298) に存在する遺伝子 GLYMA11G14040.2 は、519.19 倍 (p 値 $= 9.72 \times 10^{-12}$) の差を検出した。Figure 2 は、発現量差上位 3 位までの遺伝子をダイズゲノムリファレンス配列にマッピングした結果を示す。これらの遺伝子は、種子間の発現量差として最も高い (最大 613.48 倍) のものであった。以上の結果より、種子間で変動が見られる発芽遺伝子の発現量差は、edgeR ソフトウェアで算出した場合、約 600 倍以下で検出されることが判った。

3.2. 非 GM 品種間における発芽遺伝子の発現量差の解析

非 GM 品種間における発芽遺伝子の発現量の差を解析するため、アメリカ品種、カナダ品種、Williams 品種、及び、Jack 品種の非 GM4 品種を試料に供した (Figure 3)。その結果、アメリカ産をリファレンスにした場合、発現量の差 600 倍以上 ($p < 0.05$) ある遺伝子は、アメリカ産 vs カナダ産の場合、32 遺伝子 (発現量差の最高値

=15,220.93 倍、p 値 9.58×10^{-23} , **Table 6, Figure 4A**) アメリカ産 vs Williams 品種の場合、14 遺伝子 (発現量差の最高値=6,920.15 倍、p 値 3.72×10^{-27} , **Table 7, Figure 4B**) アメリカ産 vs Jack 品種の場合、24 遺伝子 (発現量差の最高値=8,765.89 倍、p 値 5.24×10^{-25} , **Table 8, Figure 4C**) 検出された。また、Williams 品種をリファレンスに、発現量差 600 倍以上 ($p < 0.05$) ある遺伝子は、Williams 品種 vs アメリカ産の場合、11 遺伝子 (発現量差の最高値=9,818.73 倍、p 値 2.93×10^{-30} , **Table 9, Figure 4D**) Williams 品種 vs カナダ産の場合、15 遺伝子 (発現量差の最高値=12,060.30 倍、p 値 1.68×10^{-32} , **Table 10, Figure 4E**) Williams 品種 vs Jack 品種の場合、16 遺伝子 (発現量差の最高値=21,151.66 倍、p 値 7.7210^{-32} , **Table 11, Figure 4F**) 検出された。**Figure 5,6** は、品種間で発現量の差のある遺伝子を例に、RNA-Seq 解析により得られたリード配列のマッピング結果を示す。**Figure 5** は、アメリカ産以外の品種では、発現している遺伝子 GLYMA17G18930.1 (品種 No5 遺伝子) **Figure 6** は、Williams 品種以外の品種では、発現している遺伝子 GLYMA12G09400.1 の結果を示す。以上の結果から、転写レベルにおける発芽遺伝子

発現パターンは、品種間で異なることが示唆された。

4 . RT-PCR を用いた発現量の差の確認

発芽遺伝子は、RT-PCR を用いて確認した (**Table 12, Figure 7A**)。RT-PCR 用のプライマー対は、アノテーション情報を基にスプライシングサイトを同定し、イントロン配列を挟んだエキソン領域の塩基配列を標的とした。発芽ダイズより調製した cDNA を鋳型に、PCR を行いダイレクトシーケンシングしたところ、アノテーション情報通りのスプライシング位置と成熟 mRNA 配列を確認した (データ示さず)。各品種より調製した cDNA サンプルは、設計したプライマー対を使用して RT-PCR 試験に供した。すべての cDNA サンプルは、リファレンス陽性コントロールとして Act1 遺伝子の発現を検出した (**Figure 7C**)。遺伝子 GLYMA17G18930.1 (品種 No5) の発現量は、アメリカ産のみ定性的に確認されなかった (**Figure 7B**)。また、この結果は、RNA-Seq のリード配列をゲノム上にマッピングし、トラッキングで示した **Figure 5** の結果と合致した。

また、RNA-Seq の結果より、遺伝子 GLYMA12G09400.1 は、Williams 品種をリファレンスとした場合、アメリカ産は 1403.12 倍、カナダ産は 3565.85 倍、Jack 品種は 2070.67 倍の差を検出した (**Table 9**

~11)。遺伝子 GLYMA12G09400.1 の発現を特異的に検知する RT-PCR 用プライマー対を使用して、RT-PCR 試験に供したところ、Williams 品種のみ発現が確認されなかった (Figure 7C)。この結果は、RNA-Seq で得られた結果 (Figure 6) と合致した。

5. LC-MS/MS を使用したプロテオームデータとの比較

5.1. LC-MS/MS

各試料の LC-MS/MS データから抜き出した Total ion current (TIC) クロマトグラムを Figure 8 に示す。各試料ともに類似の TIC クロマトグラムが得られた。この結果から、試料調製から LC-MS/MS までの各工程が一様に行われたと判断した。

5.2. ペプチドおよびタンパク質の同定

「B. 研究方法 5.6. LC-MS/MS」の手順に従って、各試料から取得した MS/MS データを 4 種類の配列データベースと照合した。照合以降の解析は、配列データセットごとにまとめて行った。Table 13 には、ペプチドの同定情報に連結されたピークの数、同定タンパク質の数および計量値を算出したタンパク質の数をデータベース毎に列挙した。同定数は、CDS (*Glycine max*)、Uniprot (*Glycine max*)、及び、NCBI (*Glycine max*) の間に大差はなかったが、Uniprot (Green plants) では他の 2 つよりも少なか

った。そこで、以降の解析には、各タンパク質名の情報が含まれる Uniprot (*Glycine max*) のデータベースセットを用いた。また、GM サイズに過剰発現していることがわかっている 3 種類のタンパク質 [Bialaphos resistant gene (Bar), Enhanced green-fluorescent protein (eGFP) および lysyl-tRNA synthetase (SYNC1)] のうち、今回の分析では eGFP、及び、SYNC1 が同定された。また、Table 11 に列挙した、RNAseq 解析で変動を示した 16 種類の遺伝子 (GLYMA_02G03250.1, GLYMA_0892S50.1, GLYMA_03G07121.1, GLYMA_02_G10320.1, GLYMA_12G13090.1, GLYMA_01G30060.1, GLYMA_11G15310.1, GLYMA_12G09400.1, GLYMA_02G03230.1, GLYMA_18G08015.1, GLYMA_03G05730.2, GLYMA_13G37101.1, GLYMA_16G24590.1, GLYMA_08G03381.2, GLYMA_20G27410.1, GLYMA_03G06976.1) のうち、GLYMA_02G10320.1 のみが同定された。

5.3. タンパク質の計量値の比較

「B. 研究方法 5.8. ペプチドの同定情報と検出強度値の連結」の手順に従って、

帰属ユニークペプチドの検出強度を足し合わせて各タンパク質の計量値を算出した。同定されたタンパク質のうち、ユニークペプチドが同定されなかったタンパク質については、計量値を算出しなかった。この計量値情報をもとにして、Jack/Williams 間、及び、Jack/GM 間におけるタンパク質倍数値 (Fold)-計量値の散布図を作成した (Figure 9)。両グラフを比較すると Jack/Williams 間の変動は、より小さい傾向にあった。

Jack/Williams と GM/Williams の各々について、倍数値が 4 倍以上、1/4 以下のタンパク質をそれぞれ Table 14~17 に抜粋した。このうち、GM/Williams における SYNK1 の倍数値は 300.11 倍、eGFP の倍数値は 54.61 倍であった (Table 16)。RNA-Seq 解析で変動を示した 16 種類の遺伝子のひとつ、GLYMA02G10320.1 にタンパク質レベルで有意な差は認められなかった (倍数値 1.06 倍)。

D. 考察

本研究では、GM サイズの発芽遺伝子を解析する方法の確立を目的に、RNA-Seq、並びに、LC-MS/MS 解析を行った。その過程で、非 GM サイズと比較する際の基盤となる、非 GM サイズの発芽遺伝子に関するデータを構築した。まず、発芽条件の検討を行った。発芽条件を一定に保つよう設定

したが、サイズ品種間、又は、粒間における遺伝子発現量の差 (edgeR 算出値 600 倍程度) を検出した。これは、発芽サイズの解析には、品種を揃え、複数粒の発芽サイズを試料に供する必要性を示唆した。edgeR 算出値 600 倍以下の発現差のある遺伝子については、非 GM と GM 種子間の遺伝子発現差を解析した際にも検出された。よって、非 GM サイズと GM サイズの RNA-Seq 解析による比較には、同一品種を用いる必要が示された。

本研究で得られた発芽遺伝子に関する RNA-Seq データと RT-PCR のデータは、合致した。GM サイズと非 GM サイズの発芽遺伝子を定量的に比較する場合、遺伝子発現量差の閾値 (edgeR 算出値 > 600 倍) を設定し、それ以上の差のある遺伝子については、個々に RT-PCR 法等を用いて精査する必要があった。

RNA-Seq については、スプライシングバリエーションや、サイズのアノテーション情報の不完全性から、RNA-Seq を使用し全ゲノムを網羅した遺伝子の発現量の定量は、現時点では不可能であると考えられた。よって、RNA-Seq は、発芽遺伝子の「スクリーニング」目的の用途として適している方法と考えられた。また、RNA-Seq より得られるデータは、遺伝子のスプライシング位置を、ゲノム配列上で特定することが可能であったことから、RNA-Seq は、発芽遺伝

子を定量的に解析する際の RT-PCR 用プライマーを設計する有効なツールであることが示唆された。

本研究では、LC-MS/MS 解析用の発芽ダイズサンプルを計 2 回調製し、LC-MS/MS 分析を実施した。eGFP は GM ダイズ品種のみ組換えタンパク質として発現させているが、他の非 GM ダイズ 2 品種でも GM の 100 分の 1 程度の検出強度が得られていた。これは、タンパク質分離に使用した HPLC カラムに吸着したタンパクの洗浄が不足し「キャリーオーバー」として検出されていたと考えられた。よって、LC-MS/MS 分析を行う場合、測定試料の「キャリーオーバー」が無視できる程度であることを確かめる必要があった(データ非開示)。また、1 回目の試料は各系統とも複数の発芽ダイズ粒を一つの容器内で凍結保存したため、保存・輸送中に根または芽がダイズ本体から脱落する可能性があった。そのため、発芽ダイズは、1 粒ずつ容器に入れて、試料調製の際には脱落した芽と根をともに破碎できるようにした。LC-MS/MS を使用したプロテオーム解析の結果、GM ダイズは、非 GM ダイズの Williams 品種の遺伝子を改変して作成された系統なので、両者のタンパク質発現が類似していることが予想されたが、倍数值-検出強度の分布図 (Figure 9) は、むしろ Jack と Williams との類似を示していた。

また、RNA-Seq と LC-MS/MS にて検出された発現量差のある発芽遺伝子群の相関性は確認されなかった(データ示さず)。これは、転写レベルと翻訳レベルでの発芽遺伝子のプロセッシングの違い、又は、発芽ダイズ中に存在するタンパク質を検出する機器等の性能の違いを示唆した。

E. 結論

RNA-Seq は、各発芽遺伝子の発現量や発芽遺伝子の特定をスクリーニングするための有効なツールであることが示唆された。しかし、発芽ダイズの RNA-Seq を行う場合は、同じ品種間で発芽時の環境を揃える必要があり、発現量の再現性比較のため、1 サンプルにつき 2 粒以上からトータル RNA を抽出精製し分析する必要があった。Scheme 3 は、発芽ダイズの RNA-Seq 解析を行う際の工程を示す。発現量の確実な定量値の算出に関しては、各遺伝子に特異的な RT-PCR を行う必要があった。RNA-Seq を用いたトランスクリプトームデータと LC-MS/MS を用いたプロテオームデータの、定性・定量的な相関性は確認されなかった。これは、転写レベルと翻訳レベルの発芽遺伝子のプロセッシングの違い、又は、発芽ダイズ中に存在するタンパク質を検出する機器等の性能の違いを示唆した。

F. 健康危険情報

なし

G. 研究発表

1. 論文発表 :

1. 論文発表

- 1) Miyahara, T., Miyake, N., Sawahuji, K., Kitta, K., Nakamura, K., Kondo, K., Ozeki, Y. Wheat DNA fragmentation of commercial processed foods. *Japanese Journal of Food Chemistry and Safety*, 23, 141-148, 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. Interlaboratory validation data on real-time polymerase chain reaction detection for unauthorized genetically modified papaya line PRSV-YK. *Data in Brief*, 7, 1165-1170, 2016.
- 3) 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.

- 4) Noguchi, A., Nakamura, K., Sakata, K., Sato-Fukuda, N., Ishigaki, T., Mano, J., Takabatake, R., Kitta, K., Teshima, R., Kondo, K., Nishimaki-Mogami, T. Development and interlaboratory validation of a simple screening method for genetically modified maize using $\Delta\Delta Cq$ -based multiplex real-time PCR. *Analytical Chemistry*, 88, 4285-4293, 2016.
- 5) 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.

2. 学会発表

- 1) 中村公亮、石垣拓実、野口秋雄、坂田こずえ、加藤怜子、高畠令王奈、岸根、真野潤一、橘田和美、最上(西巻)知子、近藤一成：アクリルアミド産生低減並びに打撲黒班低減を目的に開発

- された遺伝子組換えジャガイモ (J3, F10, E12 系統) の検知法開発(第 1 報)、第 53 回全国衛生化学技術協議会年会、青森、2016 年 11 月
- 2) 坂田こずえ、中村公亮、野口秋雄、石垣拓実、加藤怜子、近藤一成: ITS-RPB2 領域を用いたクサウラベニタケ系統分類と中毒事例検体の分析、第 53 回全国衛生化学技術協議会年会、青森、2016 年 11 月
 - 3) 野口秋雄、中村公亮、坂田こずえ、石垣拓実、加藤怜子、真野潤一、高畠令王奈、橘田和美、最上(西巻)知子、近藤一成: 遺伝子組換えトウモロコシ粒検査法の簡易化、第 53 回全国衛生化学技術協議会年会、青森、2016 年 11 月
 - 4) 野口秋雄、中村公亮、坂田こずえ、石垣拓実、加藤怜子、真野潤一、高畠令王奈、橘田和美、最上(西巻)知子、近藤一成: 遺伝子組換えトウモロコシの簡易粒検査法の開発(続報)、第 112 回 日本食品衛生学会学術講演会、函館、2016 年 10 月
 - 5) 真野潤一、西辻泰之、野間聡、菊池洋介、福留真一、川上裕之、佐藤恵美、新畑智也、栗本洋一、布藤聡、野口秋雄、中村公亮、近藤一成、高畠令王奈、橘田和美: リアルタイム PCR を用いた DNA 断片化測定法の開発と性能評価、AOAC International 日本セクション年次大会、東京、2016 年 7 月
 - 6) 中村公亮、近藤一成、穠山浩、石垣拓実、野口秋雄、勝又啓史、高崎一人、布藤聡、坂田こずえ、福田のぞみ、真野潤一、橘田和美、田中秀典、明石良、最上(西巻)知子: 安全性未審査の遺伝子組換えパパイア検知に向けた全ゲノムシーケンシング技術応用の検討について、日本食品化学学会 第 22 回 総会・学術大会、高知、2016 年 6 月
 - 7) 石垣拓実、中村公亮、布施谷実聡、川上浩、近藤一成: ドライフルーツ食品を例とした、標的遺伝子コピー数の差に伴う内在性遺伝子の検出感度の違いについて、日本食品化学学会 第 22 回 総会・学術大会、高知、2016 年 6 月
 - 8) 三宅奈穂、宮原平、沢藤ことは、中村公亮、近藤一成、小関良宏: 小麦加工食品におけるゲノム DNA 断片化の評価、日本食品化学学会 第 22 回 総会・学術大会、高知、2016 年 6 月

- 9) 真野潤一、野間聡、菊池洋介、福留真一、川上裕之、栗本洋一、布藤聡、野口秋雄、中村公亮、近藤一成、最上(西巻)知子、高畠令王奈、橘田和美：デジタル PCR を用いた遺伝子組換えトウモロコシ簡易定量スクリーニング法の開発、第 111 回日本食品衛生学会学術講演会、東京、2016 年 5 月

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

3. その他

なし

平成28年度 研究成果の刊行に関する一覧表

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
なし							

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
Miyahara, T., Miyake, N., Sawahuji, K., Kitta, K., Nakamura, K., Kondo, K., Ozeki, Y.	Wheat DNA fragmentation of commercial processed foods	Japanese Journal of Food Chemistry and Safety	23	141-148	2016
Noguchi, A., Nakamura, K., Sakata, K., Sato-Fukuda, N., Ishigaki, T., Mano, J., Takabatake, R., Kitta, K., Teshima, R., Kondo, K., Nishimaki-Mogami, T.	Development and interlaboratory validation of a simple screening method for genetically modified maize using $\Delta\Delta Cq$ -based multiplex real-time PCR	Analytical Chemistry	88	4285-4293	2016

Table 1. Samples used in this study

Sample no.	Name of cultivar
A	Product of Canada
B	(OAC Kent [Ippanhakumokudaizu])
C	Product of USA
D	(GL3494 [Tamabijin])
E	Jack
F	
G	Williams (Wild-type)
H	
I	Williams (GM)
J	

Table 2. Evaluation of total RNA purified

Sample no.	Total volume (μL)	Concentration ($\text{ng}/\mu\text{L}$)	Total amount obtained (μg)	Absorbance ratio		RIN
				A_{260}/A_{280}	A_{260}/A_{230}	
A	47	802	37.69	2.06	1.65	8
B	40	764	30.56	2.19	2.06	7.9
C	39	794	30.97	2.20	1.37	7.7
D	45	396	17.82	2.17	1.91	7.2
E	41	176	7.22	2.24	0.37	7.1
F	44	210	9.24	2.17	1.60	8.4
G	44	52	2.29	2.21	1.45	7.1
H	42	108	4.54	2.20	0.49	7.6
I	45	404	18.18	2.17	1.18	7.5
J	45	280	12.60	2.17	0.82	7.5

Table 3. Evaluation of RNA-Seq reads

Sample no.	Raw paired reads	Clean paired reads	Effective rate (%)	Q20 (%)	Total single reads	Single reads mapped in pairs	Single reads mapped in broken pairs	Single reads not mapped
A	51,441,409	50,729,766	98.62	97.79	101,459,532	61,672,590 (60.79%)	28,691,993 (28.28%)	11,094,949 (10.94%)
B	49,683,359	49,016,783	98.66	97.68	98,033,566	56,209,740 (57.34%)	29,043,653 (29.63%)	12,780,173 (13.04%)
C	44,531,292	44,034,849	98.89	97.73	88,069,698	46,011,594 (52.24%)	30,690,052 (34.85%)	11,368,052 (12.91%)
D	47,882,082	47,317,833	98.82	97.69	94,635,666	52,819,398 (55.81%)	29,460,833 (31.13%)	12,355,435 (13.06%)
E	61,319,942	60,574,989	98.79	98.12	121,149,978	69,789,842 (57.61%)	38,011,810 (31.38%)	13,348,326 (11.02%)
F	48,122,897	47,577,266	98.87	98.38	95,154,532	55,738,078 (58.58%)	28,605,777 (30.06%)	10,810,677 (11.36%)
G	45,459,810	43,843,369	96.44	98.12	87,686,738	51,498,634 (58.73%)	25,660,013 (29.26%)	10,528,091 (12.01%)
H	55,188,240	53,372,808	96.71	98.14	106,745,616	61,030,950 (57.17%)	33,180,743 (31.08%)	12,533,923 (11.74%)
I	58,862,750	56,974,554	96.79	98.19	113,949,108	65,403,516 (57.40%)	35,492,083 (31.15%)	13,053,509 (11.46%)
J	49,725,464	49,153,651	98.85	97.64	98,307,302	58,565,980 (59.57%)	29,295,701 (29.80%)	10,445,621 (10.63%)

Table 4. List of differential genes between Non-GM-type and Williams cultivar*

Feature ID	Chromosome	Region	P-value	Fold change
GLYMA06G44091.1	6	47071378..47075313	1.76E-46	613.48
GLYMA12G34075.2	12	37241711..37243166	7.72E-34	444.24
GLYMA17G31900.1	17	35078089..35079829	1.98E-31	412.01
GLYMA17G20450.1	17	19471430..19474526	3.16E-30	395.89
GLYMA20G01430.1	20	962695..963815	1.26E-29	387.83
GLYMA12G00991.1	12	542484..543426	6.47E-27	355.59
GLYMA17G33075.1	17	36682979..36684646	2.65E-23	307.24
GLYMA10G32820.1	10	41214830..41215683	4.24E-22	291.12
GLYMA16G22474.1	16	25971208..25973609	1.70E-21	283.06
GLYMA03G19460.1	3	24727108..24729564	1.70E-21	283.06
GLYMA09G08330.1	9	7472274..7473626	2.71E-20	266.95
GLYMA04G05040.1	4	3784240..3786870	2.71E-20	266.95
GLYMA10G22390.2	10	28500279..28506009	2.71E-20	266.95
GLYMA05G33481.1	5	38116443..38119643	1.09E-19	258.89
GLYMA09G35840.1	9	41727496..41729798	4.34E-19	250.83
GLYMA16G34840.1	16	37366383..37367407	1.74E-18	242.77
GLYMA16G33750.1	16	36551930..36553977	2.78E-17	226.65
GLYMA11G22373.1	11	19578803..19579699	2.78E-17	226.65
GLYMA17G35105.1	17	39104206..39105398	2.78E-17	226.65
GLYMA20G05250.1	20	5992299..5994285	2.78E-17	226.65

*Twentieth most different expression level genes are listed

Table 5. List of differential genes between non-GM- and GM-type Williams cultivar*

Feature ID	Chromo	Region	P-value	Fold change
GLYMA11G14040.2	11	10078670..10082298	9.72E-12	519.19
GLYMA17G02260.1	17	1453405..1456485	7.18E-14	418.87
GLYMA07G32100.2	7	36998904..36999990	6.50E-09	333.37
GLYMA06G01391.1	6	866877..869760	1.62E-06	276.46
GLYMA09G03090.1	9	2178668..2180818	2.79E-07	231.93
GLYMA11G12770.1	11	9120544..9124159	7.15E-07	201.67
GLYMA17G08051.1	17	5950558..5951790	1.98E-07	194.5
GLYMA01G44640.2	1	55233128..55234930	8.31E-07	188.1
GLYMA13G42100.1	13	42191818..42193149	3.87E-05	185.59
GLYMA17G03480.2	17	2316582..2318978	2.90E-07	184.55
GLYMA09G28575.1	9	35505909..35507116	1.16E-04	177.78
GLYMA13G09707.1	13	11308589..11311684	8.47E-07	172.61
GLYMA14G17863.1	14	19931432..19937024	1.63E-04	171.31
GLYMA12G15951.1	12	14972263..14973737	4.60E-07	164.74
GLYMA01G40460.1	1	52219822..52222031	7.03E-06	156.28
GLYMA05G10075.1	5	10080145..10085029	1.75E-05	148.59
GLYMA02G13800.2	2	12123744..12127608	8.76E-06	140.3
GLYMA02G13800.1	2	12123744..12127608	5.71E-05	134.32
GLYMA12G33711.1	12	36913314..36916524	9.98E-06	134.16
GLYMA08G06330.1	8	4487558..4489934	8.60E-06	133.2

*Twentieth most different expression level genes are listed

Table 6. Gene lists having >600 times differential gene expression level between US and Canada cultivar using US cultivar as a reference

Feature ID	Chromosome	Region	P-value	Fold change
GLYMA20G01945.1	20	1467724..1469637	9.58E-23	15,220.93
GLYMA09G16090.1	9	19201065..19203101	2.47E-25	14,006.49
GLYMA06G10840.1	6	8215351..8217052	4.27E-08	8,690.69
GLYMA14G33161.2	14	40620309..40624931	3.34E-07	6,197.73
GLYMA03G03451.1	3	3197245..3198319	1.38E-07	5,411.74
GLYMA14G23895.1	14	28387656..28392096	2.11E-06	3,855.04
GLYMA08G10460.1	8	7589047..7589812	6.03E-07	2,422.29
GLYMA09G04500.2	9	3316706..3319908	2.65E-15	1,730.21
GLYMA19G31674.1	19	39473386..39476650	1.83E-06	1,516.51
GLYMA06G02170.1	6	1433698..1435146	5.77E-13	1,358.87
GLYMA08G22640.2	8	17188397..17189549	7.18E-14	1,339.68
GLYMA10G38440.1	10	46239325..46240490	3.31E-08	1,311.02
GLYMA09G05660.1	9	4446091..4450837	2.06E-10	1,176.59
GLYMA17G26556.1	17	27707046..27712214	2.34E-05	1,104.14
GLYMA14G33161.1	14	40620309..40624930	9.32E-07	1,094.60
GLYMA17G19007.1	17	16773936..16774706	2.30E-12	1,026.82
GLYMA01G06250.2	1	6303784..6304910	6.98E-09	958.13
GLYMA03G08170.2	3	9017226..9018873	2.69E-05	934.17
GLYMA12G14980.1	12	13759744..13760377	2.00E-11	922.42
GLYMA20G27410.1	20	36585232..36588760	2.18E-22	834.97
GLYMA10G31991.1	10	40469872..40471094	6.28E-11	790.3
GLYMA19G34210.1	19	41826839..41830086	3.08E-23	774.34
GLYMA19G26435.1	19	33159771..33163336	2.13E-14	757.64
GLYMA07G33511.1	7	38428874..38432132	7.89E-15	721.56
GLYMA15G05391.1	15	3808741..3815631	4.47E-08	720.97
GLYMA17G13820.1	17	10616633..10617551	1.47E-10	716.37
GLYMA10G13450.1	10	15138921..15139984	2.32E-10	696.05
GLYMA14G15605.1	14	16729271..16738914	4.37E-08	684.1
GLYMA15G05391.2	15	3808944..3815631	5.09E-08	682.81
GLYMA18G19100.2	18	20858053..20865070	3.46E-05	653.36
GLYMA01G11640.1	1	14972819..14973806	8.89E-10	628.88
GLYMA01G42500.2	1	53760884..53761948	1.08E-08	613.82

Table 7. Gene lists having >600 times differential gene expression level between US and Williams cultivar using US cultivar as a reference

Feature ID	Chromosome	Region	P-value	Fold change
GLYMA09G16090.1	9	19201065..19203101	3.72E-27	6,920.15
GLYMA20G01945.1	20	1467724..1469637	5.78E-20	1,884.69
GLYMA03G03451.1	3	3197245..3198319	7.01E-14	1,818.09
GLYMA18G19100.2	18	20858053..20865070	7.99E-19	1,577.05
GLYMA05G36750.2	5	40529554..40531831	2.13E-17	1,447.52
GLYMA04G10621.1	4	8833901..8835822	1.02E-06	1,446.89
GLYMA17G26556.1	17	27707046..27712214	1.11E-17	1,376.39
GLYMA14G12170.1	14	10900868..10903684	3.29E-13	1,288.85
GLYMA14G33161.2	14	40620309..40624931	2.84E-16	1,179.78
GLYMA05G25340.1	5	31481126..31482429	7.31E-13	945.94
GLYMA19G31674.1	19	39473386..39476650	5.52E-14	829.77
GLYMA15G39520.1	15	46174815..46178992	8.98E-14	804.45
GLYMA15G39510.1	15	46171119..46171865	3.86E-13	735.45
GLYMA17G19007.1	17	16773936..16774706	2.07E-12	695.49

Table 8. Gene lists having >600 times differential gene expression level between US and Jack cultivar using US cultivar as a reference

Feature ID	Chromosome	Region	P-value	Fold change
GLYMA09G16090.1	9	19201065..19203101	5.24E-25	8,765.89
GLYMA20G01945.1	20	1467724..1469637	4.33E-19	4,306.51
GLYMA08G10460.1	8	7589047..7589812	2.54E-20	2,366.75
GLYMA17G26556.1	17	27707046..27712214	3.34E-19	2,088.06
GLYMA12G14980.1	12	13759744..13760377	1.01E-08	1,895.40
GLYMA05G36750.2	5	40529554..40531831	5.05E-17	1,799.88
GLYMA18G08015.1	18	6758624..6760903	6.17E-17	1,636.07
GLYMA18G03640.3	18	2502550..2503445	9.57E-10	1,427.06
GLYMA18G19100.2	18	20858053..20865070	6.08E-13	1,395.99
GLYMA08G22640.2	8	17188397..17189549	1.63E-08	1,390.54
GLYMA16G24590.1	16	28569677..28572015	1.34E-16	1,360.38
GLYMA19G31674.1	19	39473386..39476650	3.75E-15	1,218.72
GLYMA05G25340.1	5	31481126..31482429	2.22E-15	1,148.69
GLYMA07G25536.1	7	27944328..27951602	2.45E-13	1,026.02
GLYMA07G21018.1	7	21452281..21452961	1.63E-14	1,011.88
GLYMA17G19007.1	17	16773936..16774706	7.77E-13	1,010.06
GLYMA10G38440.1	10	46239325..46240490	6.13E-08	1,006.76
GLYMA07G17000.2	7	16666326..16668062	5.77E-14	929.67
GLYMA09G04500.2	9	3316706..3319908	7.25E-08	925.72
GLYMA02G03250.1	2	2518997..2520592	4.23E-19	882.39
GLYMA19G26435.1	19	33159771..33163336	1.27E-21	844.81
GLYMA18G50595.1	18	59676735..59681611	4.08E-12	700.38
GLYMA20G27410.1	20	36585232..36588760	7.99E-25	657.15
GLYMA09G29250.1	9	36191240..36192018	5.22E-10	619.38

Table 9. Gene lists having >600 times differential gene expression level between Williams and US cultivars using Williams cultivar as a reference

Feature ID	Chromosome	Region	P-value	Fold change
GLYMA03G07121.1	3	7458627..7491220	2.93E-30	9,818.73
GLYMA15G15140.1	15	11581588..11582282	3.16E-17	1,472.81
GLYMA12G09400.1	12	7168483..7172381	8.33E-18	1,403.12
GLYMA03G06976.1	3	7275123..7298068	2.94E-16	1,145.18
GLYMA01G30060.1	1	40769687..40771162	1.67E-15	1,108.67
GLYMA14G17863.1	14	19931432..19937024	6.86E-15	1,091.65
GLYMA17G20770.1	17	19844171..19845334	8.68E-14	1,081.14
GLYMA03G05650.1	3	5984323..5988698	7.75E-15	997.9
GLYMA08G03381.2	8	2382841..2385344	5.42E-15	989.8
GLYMA05G36950.1	5	40688684..40690756	2.11E-10	661.24
GLYMA20G03670.1	20	3517516..3519869	3.47E-11	637.15

Table 10. Gene lists having >600 times differential gene expression level between Williams and Canada cultivars using Williams cultivar as a reference

Feature ID	Chromosome	Region	P-value	Fold change
GLYMA12G13090.1	12	11449237..11449791	1.68E-32	12,060.30
GLYMA06G10840.1	6	8215351..8217052	1.77E-08	8,091.06
GLYMA12G09400.1	12	7168483..7172381	3.50E-23	3,565.85
GLYMA03G07121.1	3	7458627..7491220	2.43E-20	3,367.50
GLYMA01G30060.1	1	40769687..40771162	5.66E-24	3,340.16
GLYMA11G34613.1	11	36389844..36393874	2.43E-17	1,394.44
GLYMA14G23895.1	14	28387656..28392096	3.75E-06	1,338.77
GLYMA13G37101.1	13	38277892..38278797	6.96E-17	1,317.92
GLYMA01G31890.1	1	43130698..43137550	4.61E-05	927.18
GLYMA20G27410.1	20	36585232..36588760	7.11E-31	925.68
GLYMA01G06250.2	1	6303784..6304910	9.75E-10	854.93
GLYMA03G08170.2	3	9017226..9018873	1.59E-05	826.91
GLYMA01G18990.1	1	23424892..23425203	1.05E-26	805.92
GLYMA03G05730.2	3	6051893..6063501	1.09E-04	732.43
GLYMA10G13450.1	10	15138921..15139984	1.41E-11	633.67

Table 11. Gene lists having >600 times differential gene expression level between Williams and Jack cultivars using Williams cultivar as a reference

Feature ID	Chromosome	Region	P-value	Fold change
GLYMA02G03250.1	2	2518997..2520592	7.72E-32	21,151.66
GLYMA0892S50.1	scaffold_892	2250..3229	6.22E-28	18,414.07
GLYMA03G07121.1	3	7458627..7491220	4.34E-34	7,673.40
GLYMA02G10320.1	2	8186068..8188789	1.74E-20	5,717.26
GLYMA12G13090.1	12	11449237..11449791	4.04E-22	3,568.04
GLYMA01G30060.1	1	40769687..40771162	4.43E-27	2,907.23
GLYMA11G15310.1	11	10952617..10954051	4.80E-24	2,846.41
GLYMA12G09400.1	12	7168483..7172381	2.49E-22	2,070.67
GLYMA02G03230.1	2	2494913..2496466	3.60E-22	1,546.78
GLYMA18G08015.1	18	6758624..6760903	4.96E-20	1,541.70
GLYMA03G05730.2	3	6051893..6063501	2.72E-08	1,372.22
GLYMA13G37101.1	13	38277892..38278797	7.53E-12	1,305.76
GLYMA16G24590.1	16	28569677..28572015	1.10E-19	1,278.84
GLYMA08G03381.2	8	2382841..2385344	2.16E-15	842.83
GLYMA20G27410.1	20	36585232..36588760	1.78E-35	746.13
GLYMA03G06976.1	3	7275123..7298068	2.18E-13	660.36

Table 12. RT PCR primer sets used to analyze gene GLYMA17G18930.1 and gene GLYMA12G09400.1

FeatureID	Primer sequence (5'→3')	Amplicon (bp)
GLYMA17G18930.1	sense TTCCCTGTTGGATTTGAGAAGCTAGGAAAAG anti- CCAAAGCACAACTTCATGTTTCAGCACCAGC	349
GLYMA12G09400.1	sense GAGAGAAGGTTTTGGATTCCAACCTCCTGCAC anti- GACTGCAAATGTTTTCCCACATTTTCTGCAC	342
<i>Actin1</i> (XM_003552652)	sense ATCTTGACTGAGCGTGGTTATTCC anti- GCTGGTCCTGGCTGTCTCC	126

Table 13. Number of identified peptides and proteins

	Sequence database			
	CDS (<i>Glycine max</i>)	Uniprot (<i>Glycine max</i>)	Uniprot/SwissProt (Green Plants)	NCBI/Genome (<i>Glycine max</i>)
Number of peaks with identified peptide	3,179	3,274	1,844	3,205
Number of identified proteins	451	457	289	450
Number of proteins that are quantitatively analyzable	424	429	244	420

Table 14. Protein fold change of more than 4 between Jack and Williams

No.	Protein name	Uniprot AccessionNumber	Gene Name	Confidence Score	Unique Peptide Counts	Normalized abundance		Fold Change Jack/Williams
						Williams	Jack	
1	Uncharacterized protein	I1NDJ4	GLYMA_20G027200	42.3	1	0	55,840	infinity
2	Beta-conglycinin, beta chain	P25974	CG-4	4975.0	4	5,093,644	1,087,833,393	213.57
3	Uncharacterized protein	I1MG41	GLYMA_15G119800	38.5	1	28,809	2,906,869	100.90
4	Uncharacterized protein	I1L3Q2;K7LE73	GLYMA_09G158600	219.9	3	4,098,609	283,449,042	69.16
5	Uncharacterized protein	I1LY75	GLYMA_13G111700	32.7	1	8,134	303,090	37.26
6	Uncharacterized protein	I1LRP2	GLYMA_12G097400	39.4	1	424,319	5,112,728	12.05
7	Uncharacterized protein	I1JCB6	GLYMA_02G042500	41.9	1	416,136	3,630,876	8.73
8	Uncharacterized protein	K7KLI9	GLYMA_04G216900	24.6	1	393,577	3,236,300	8.22
9	Uncharacterized protein	I1MT39	GLYMA_17G075800	25.3	1	36,737	175,677	4.78
10	Uncharacterized protein	I1N3G2;I1L6T4	GLYMA_18G222200	57.7	2	615,022	2,930,016	4.76
11	Uncharacterized protein	C6SWF6;I1L3J0	GLYMA_16G202500	99.3	2	350,671	1,593,327	4.54
12	Uncharacterized protein	C6SYG6	GLYMA_19G236100	74.4	1	1,283,949	5,683,054	4.43
13	Uncharacterized protein	I1L5X5	GLYMA_09G238000	39.6	1	75,873	328,339	4.33
14	Alpha-1,4 glucan phosphorylase	I1KYU6	GLYMA_08G334000	28.5	1	31,889	135,800	4.26
15	Uncharacterized protein	K7KVC7	GLYMA_06G160300	42.0	1	178,246	757,851	4.25
16	Alpha-1,4 glucan phosphorylase	I1LWR4	GLYMA_13G057800	48.4	2	63,829	270,360	4.24

Table 15. Protein fold change of less than 0.25 between Jack and Williams

No.	Protein name	Uniprot AccessionNumber	Gene Name	Confidence Score	Unique Peptide Counts	Normalized abundance		Fold Change Jack/Williams
						Williams	Jack	
1	Uncharacterized protein	K7N4I7		314.4	1	10,655	0	0.00
2	Uncharacterized protein	I1N2I4	GLYMA_18G182600	45.2	1	53,895	0	0.00
3	Glycinin G3	P11828	GY3	3452.8	25	2,839,742,376	331,645,220	0.12
4	G.max mRNA from stress-induced gene (H4)	Q43453	GLYMA_17G030200	116.2	1	686,801	145,678	0.21
5	Uncharacterized protein	I1JPV1	GLYMA_03G189300	24.1	1	146,880	31,709	0.22

Table16. Protein fold change of more than 4 between GM and WT Williams

No.	Protein name	Uniprot AccessionNumber	Gene Name	Confidence Score	Unique Peptide Counts	Normalized abundance		Fold Change GM/Williams
						Williams	GM	
1	Uncharacterized protein	I1NDJ4	GLYMA_20G027200	1	42.32	0	59,882	Infinity
2	SYNC1 Soybean	I1K000;I1KPU2;I1K0Q4	Sync1	5	237.86	49,680	14,909,520	300.11
3	Uncharacterized protein	I1LY75	GLYMA_13G111700	1	32.68	8,134	481,321	59.17
4	Enhanced green fluorescent protein(eGFP)	X1XX11	eGFP	8	583.25	8,473,101	462,686,460	54.61
5	Uncharacterized protein	K7N4I7		1	314.43	10,655	189,604	17.79
6	Uncharacterized protein	C6SWF6;I1L3J0	GLYMA_16G202500	2	99.26	350,671	3,558,002	10.15
7	Uncharacterized protein	I1MG41	GLYMA_15G119800	1	38.51	28,809	287,122	9.97
8	Uncharacterized protein	I1MT39	GLYMA_17G075800	1	25.27	36,737	275,245	7.49
9	Uncharacterized protein	K7KLI9	GLYMA_04G216900	1	24.58	393,577	2,753,076	7.00
10	Uncharacterized protein	I1N3G2;I1L6T4	GLYMA_18G222200	2	57.69	615,022	4,233,763	6.88
11	Uncharacterized protein	I1L3Q2;K7LE73	GLYMA_09G158600	3	219.91	4,098,609	26,178,334	6.39
12	Uncharacterized protein	A0A0R0GZ14;A0A0R4J3J4;A0A0R0I189	GLYMA_13G176500	1	89.69	116,136	630,523	5.43
13	Uncharacterized protein	I1L0N6		1	52.61	291,650	1,497,587	5.13
14	Uncharacterized protein	I1JBN4;C6TFW8	GLYMA_02G019600	2	86.54	578,708	2,910,093	5.03
15	Uncharacterized protein	C6SYG6	GLYMA_19G236100	1	74.35	1,283,949	6,432,535	5.01
16	Alpha-mannosidase	I1ML97;I1JXQ5		2	77.18	420,057	1,930,058	4.59
17	Superoxide dismutase	I1JYA9	GLYMA_04G221300	1	40.22	14,174	64,430	4.55
18	Uncharacterized protein	I1LRP2	GLYMA_12G097400	1	39.41	424,319	1,909,809	4.50
19	G.max mRNA from stress-induced gene (H4)	Q43453	GLYMA_17G030200	1	116.21	686,801	3,008,123	4.38
20	Uncharacterized protein	A0A0R0GA14;A0A0R0J5K1	GLYMA_14G063800	1	100.61	588,700	2,516,664	4.27
21	Uncharacterized protein	I1NH02	GLYMA_20G164700	3	106.52	4,954,399	20,068,875	4.05

Table 17. Protein fold change of less than 0.25 between GM and WT Williams

No.	Protein name	Uniprot AccessionNumber	Gene Name	Confidence Score	Unique Peptide Counts	Normalized abundance		Fold Change Jack/Williams
						Williams	GM	
1	DHAR class glutathione S-transferase	C6TGU0	DHAR3	1	80.86	161,915	0	0.00
2	Uncharacterized protein	C6SVF6	GLYMA_16G072400	1	34.23	656,925	0	0.00
3	Acyl-[acyl-carrier-protein] desaturase	B8XJY3	SAD1	1	79.4	208,272	3,366	0.02
4	Uncharacterized protein	I1JXJ3	GLYMA_04G195100	1	250.28	1,307,307	46,826	0.04
5	Uncharacterized protein	I1KAK6	GLYMA_06G123700	1	36.1	703,867	28,449	0.04
6	Glycinin G3	P11828	GY3	25	3452.76	2,839,742,376	195,798,455	0.07
7	Uncharacterized protein	I1KVT5	GLYMA_08G224100	1	46.39	202,364	14,872	0.07
8	Uncharacterized protein	I1MGM0	GLYMA_15G148900	1	58.63	1,338,846	122,214	0.09
9	Polyadenylate-binding protein	I1JE12	GLYMA_02G103900	1	68.43	568,208	53,096	0.09
10	Uncharacterized protein	K7LKW2		1	53.15	1,522,954	184,174	0.12
11	Uncharacterized protein	I1JE09;C6T5U0	GLYMA_02G103600	2	71.88	798,138	139,381	0.17
12	60S ribosomal protein L27	I1KRI7	GLYMA_08G087200	1	25.18	1,070,867	193,508	0.18
13	Uncharacterized protein	A0A0R4J5A5	GLYMA_14G176900	1	82.46	304,399	55,076	0.18
14	Annexin	G3E7M9	GLYMA_13G088700	1	84.06	1,853,384	343,251	0.19
15	Tau class glutathione S-transferase	I1MJ34	GSTU48	1	73.69	373,220	70,822	0.19
16	Uncharacterized protein	I1MAC4	GLYMA_14G117700	1	172.26	2,912,510	648,510	0.22
17	Uncharacterized protein	A0A0R0FE22	GLYMA_17G166400	1	19.69	165,078	40,880	0.25
18	Ferritin	C6TCN5;C6TI81;I1JL80; I1MYZ8;K7MTN2;I1JL81;K7KA92	GLYMA_18G205800	2	107.2	7,262,792	1,822,452	0.25
19	Uncharacterized protein	I1JVB9;A0A0R0J4X0;Q96450	GLYMA_04G099900	1	113.69	1,179,122	299,188	0.25

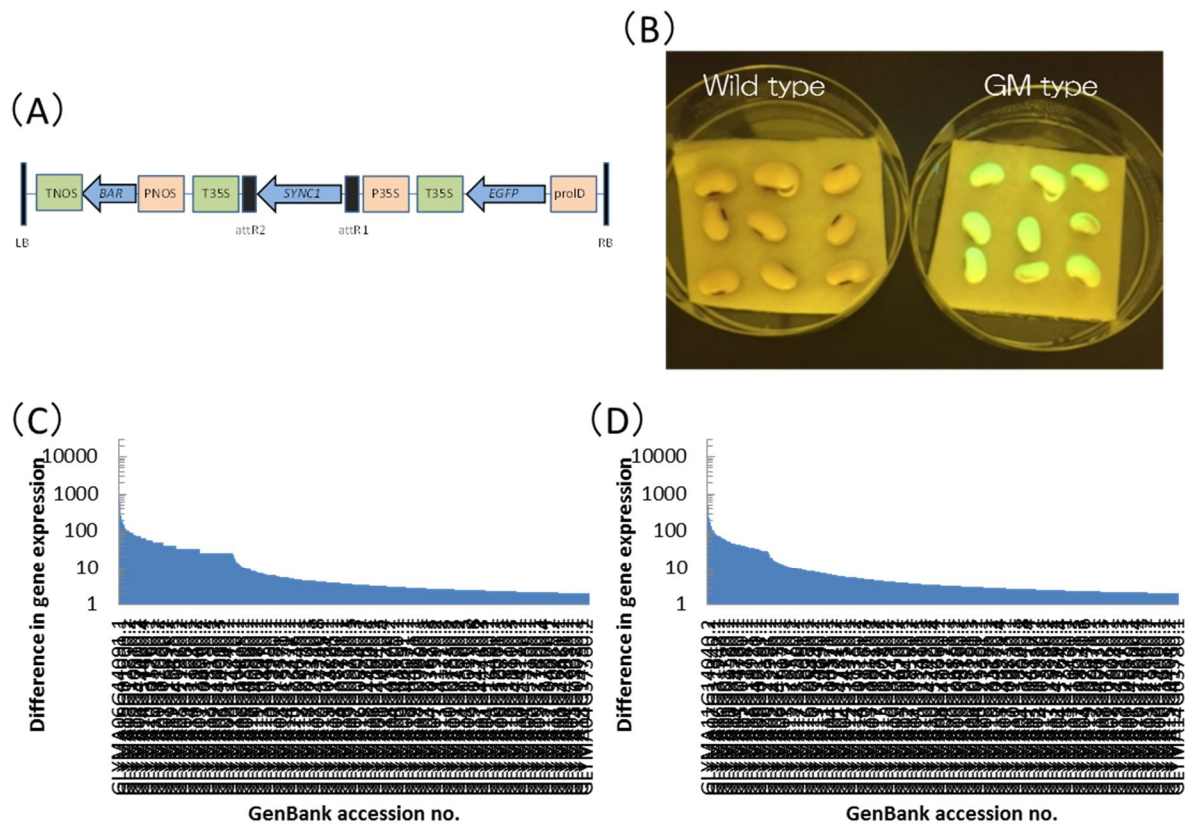


Figure 1. Identification of genes having differential gene expression between Non-GM and GM type soybean seedlings

(A) Transgenic construct in GM type (LB, left-border sequences; TNOS, nos terminator; BAR, bialaphos resistance; PNOS, nos promoter; T35S, 35S RNA terminator; attR2, recombination site; SYNC1, G. max SYNC1; attR1, recombination site, P35S, 35S RNA promoter; EGFP, ever green fluorescent protein; proID, roID promoter; RB, right-border sequences)

(B) Non-GM and GM type soybean seedlings under the UV light

Differential gene expression level between (C) Non-GM vs Non-GM type, (D) Non-GM vs GM type

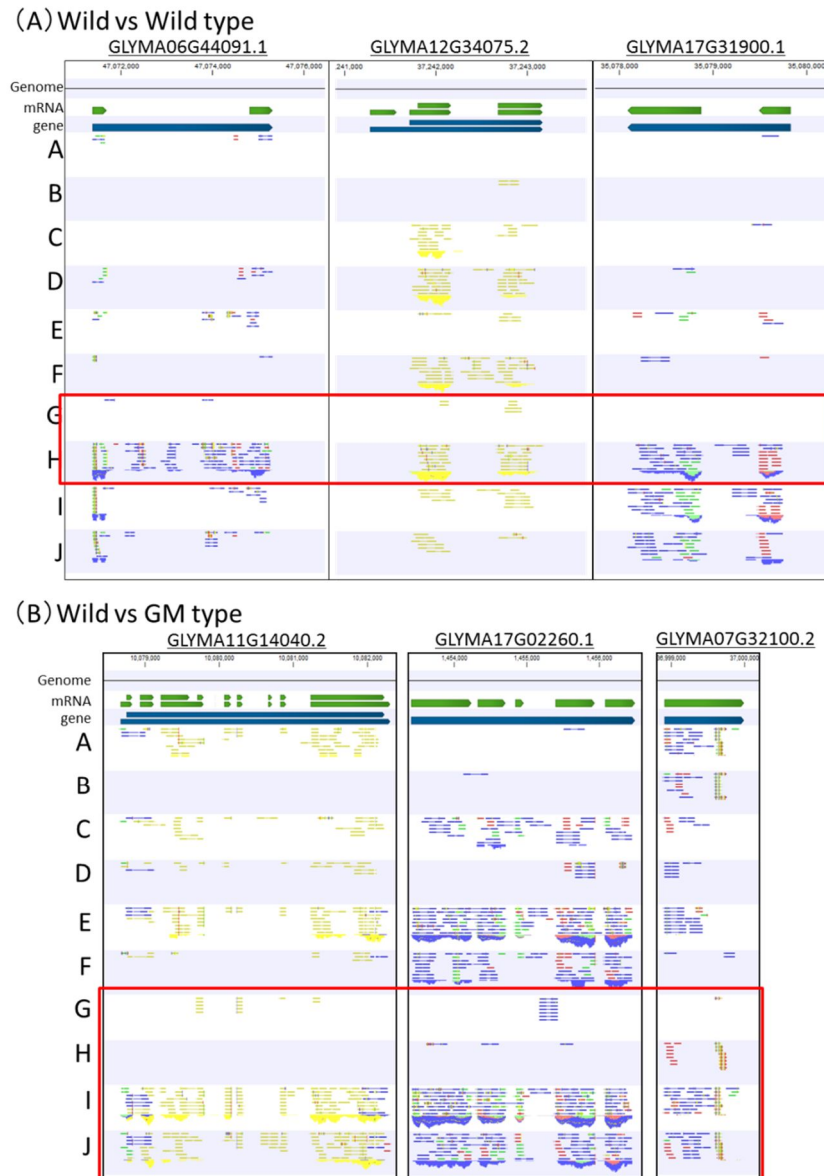


Figure 2. Mapped reads to the reference genome
 Top three genes identified having gene expression difference among (A) non-GM-type Williams cultivars and (B) non-GM- and GM-type Williams cultivars

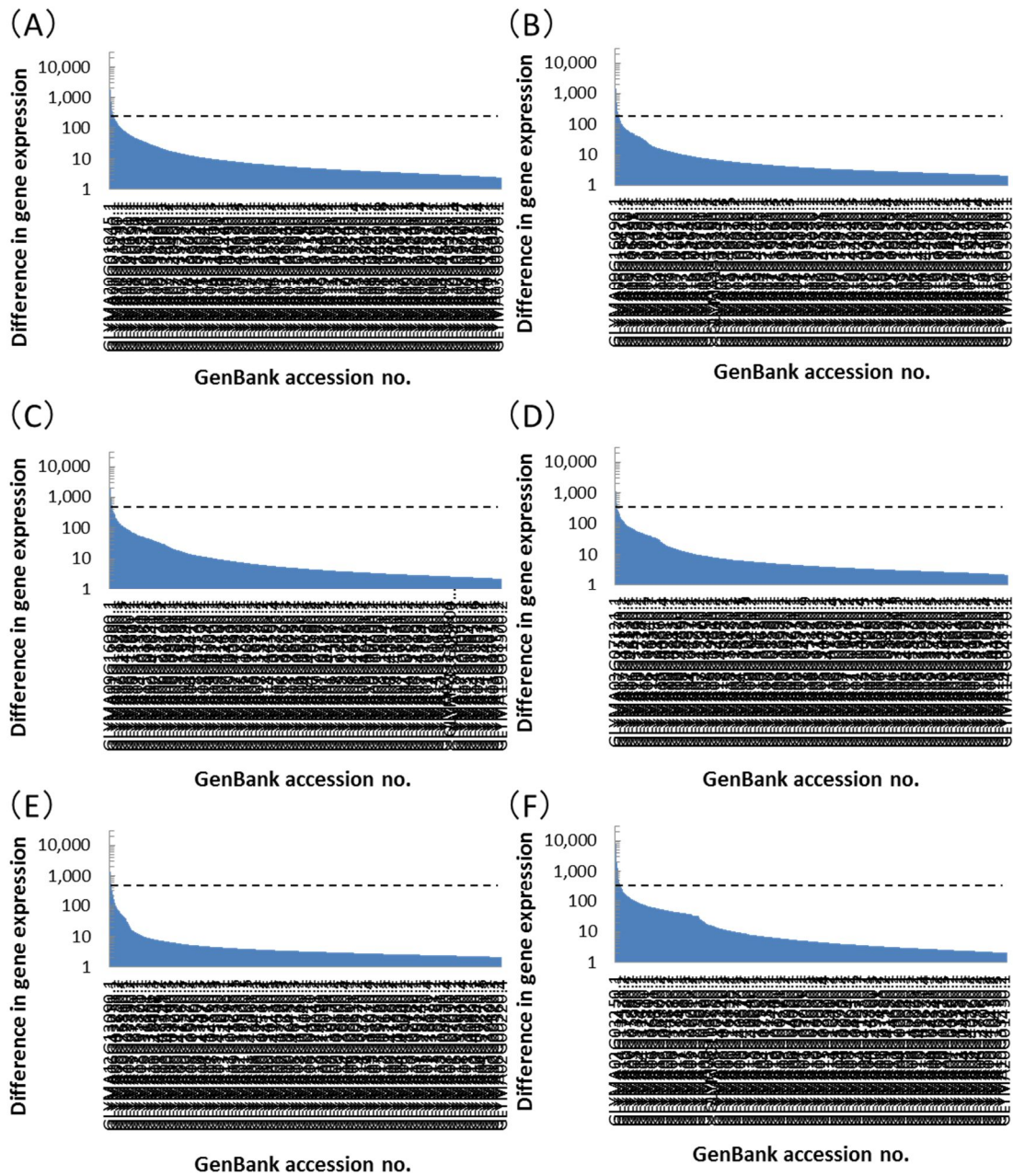


Figure 3. Differential genes expression among different varieties

Using US variety as a reference (A-C), (A) US vs Canada cultivar, (B) US vs Williams cultivar, (C) US vs Jack cultivar

Using Williams cultivar as a reference (D-F), (D) Williams vs US cultivar, (E) Williams vs Canada cultivar, (F) Williams vs Jack cultivar

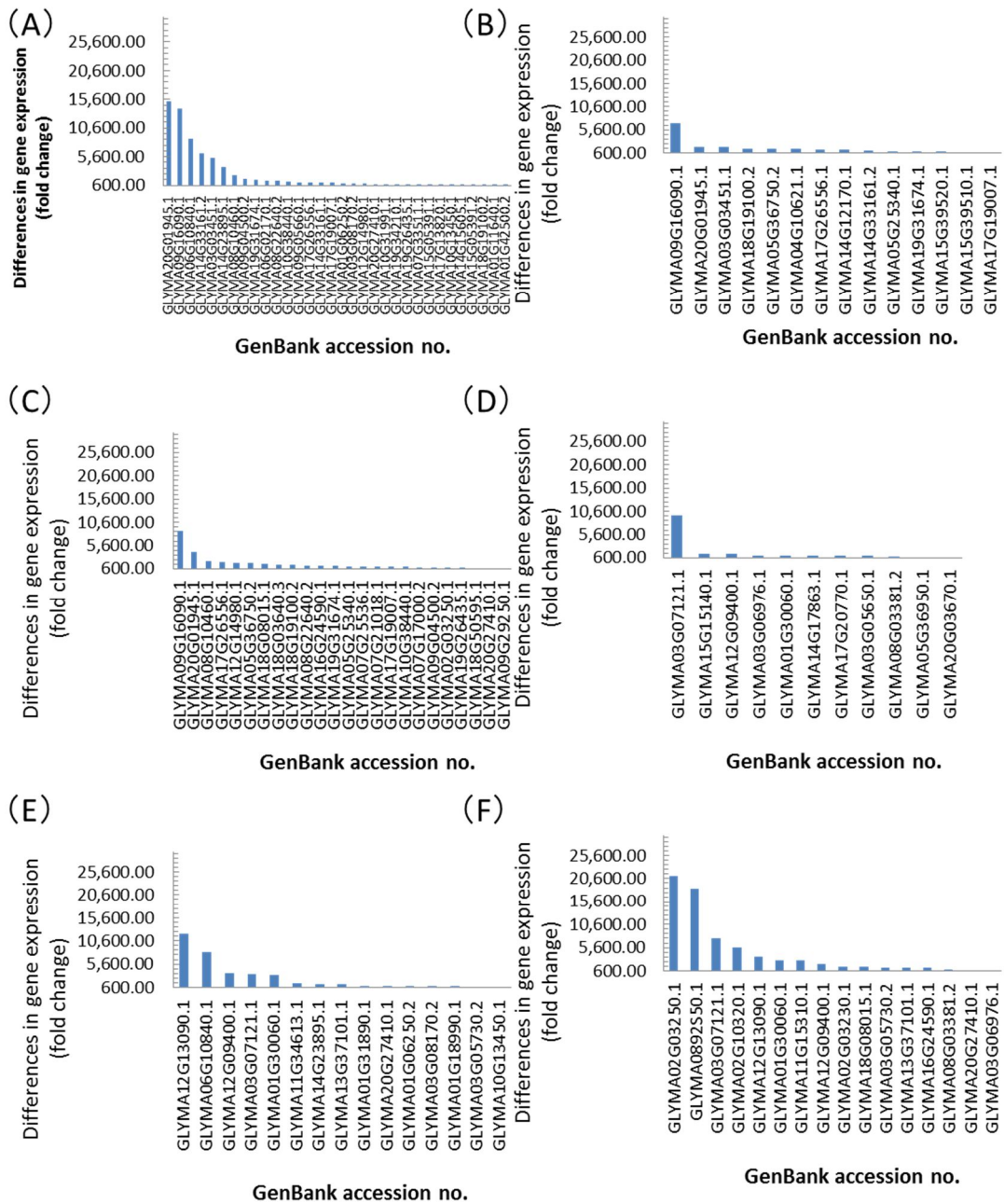


Figure 4. Identification of genes having differential gene expression >600 times using Empirical Analysis of DGE

Using US cultivar as a reference (A-C), (A) US vs Canada cultivar, (B) US vs Williams cultivar, (C) US vs Jack cultivar

Using Williams cultivar as a reference (D-F), (D) Williams vs US cultivar, (E) Williams vs Canada cultivar, (F) Williams vs Jack cultivar

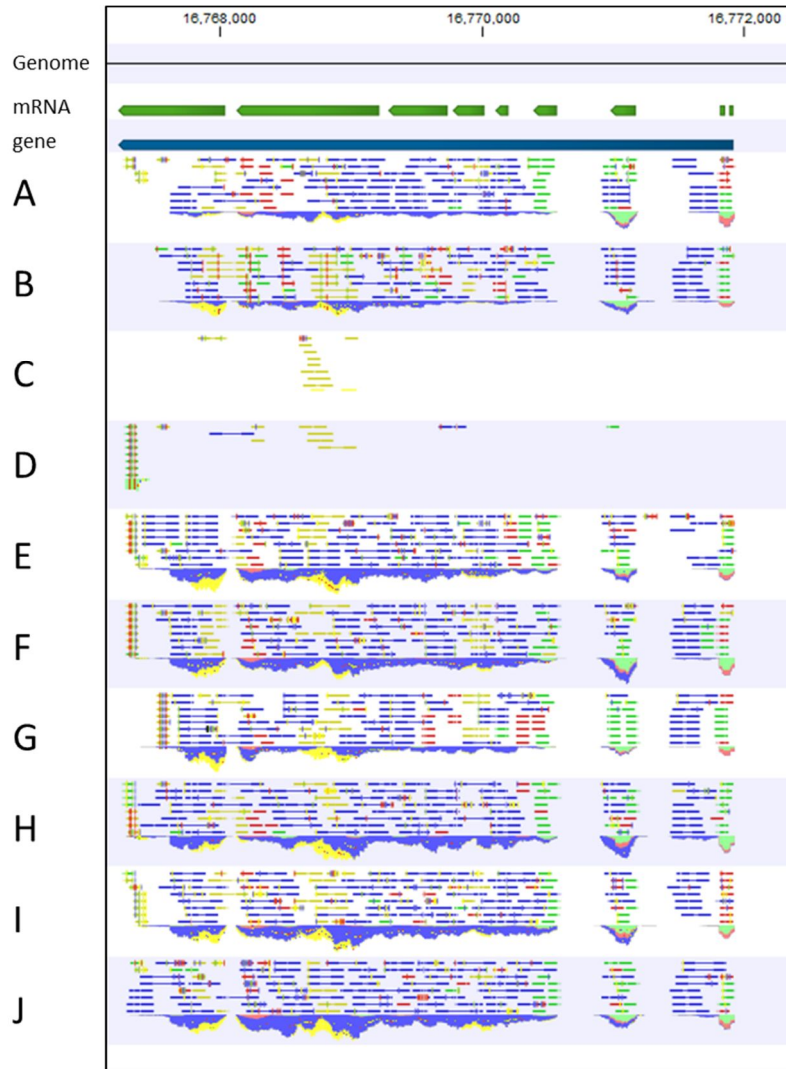


Figure 5. Gene GLYMA17G18930.1 expression among soybean varieties

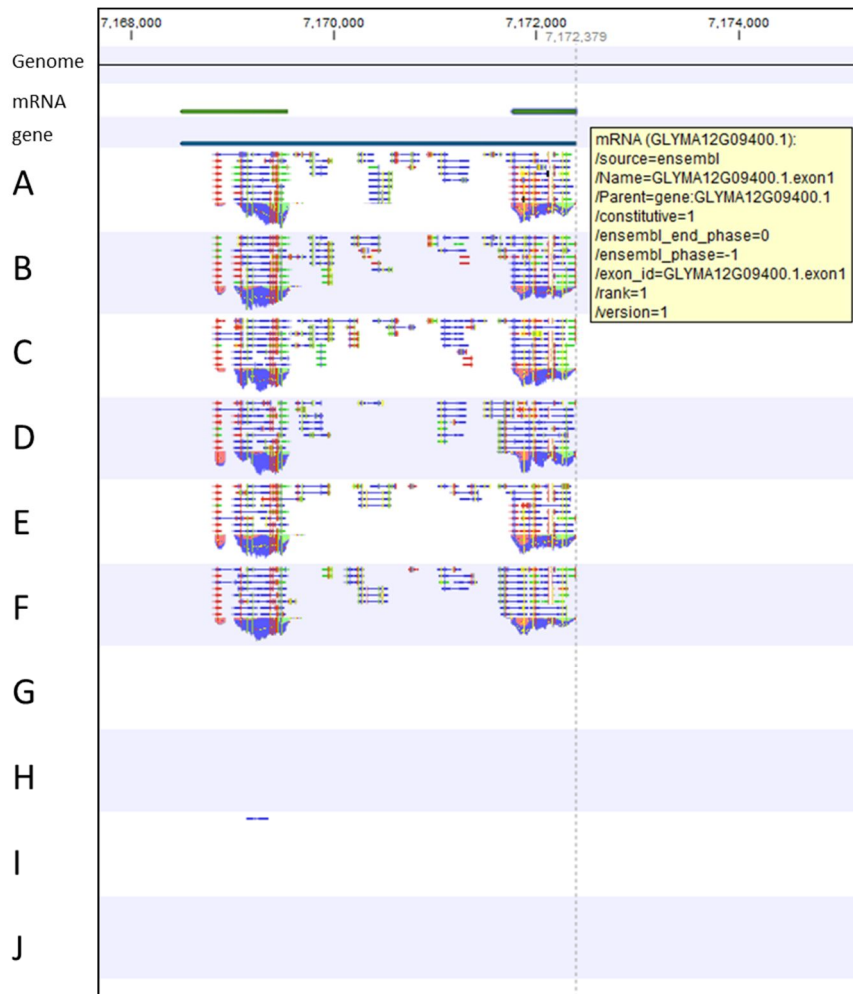


Figure 6. Gene GLYMA12G09400.1 expression among soybean varieties

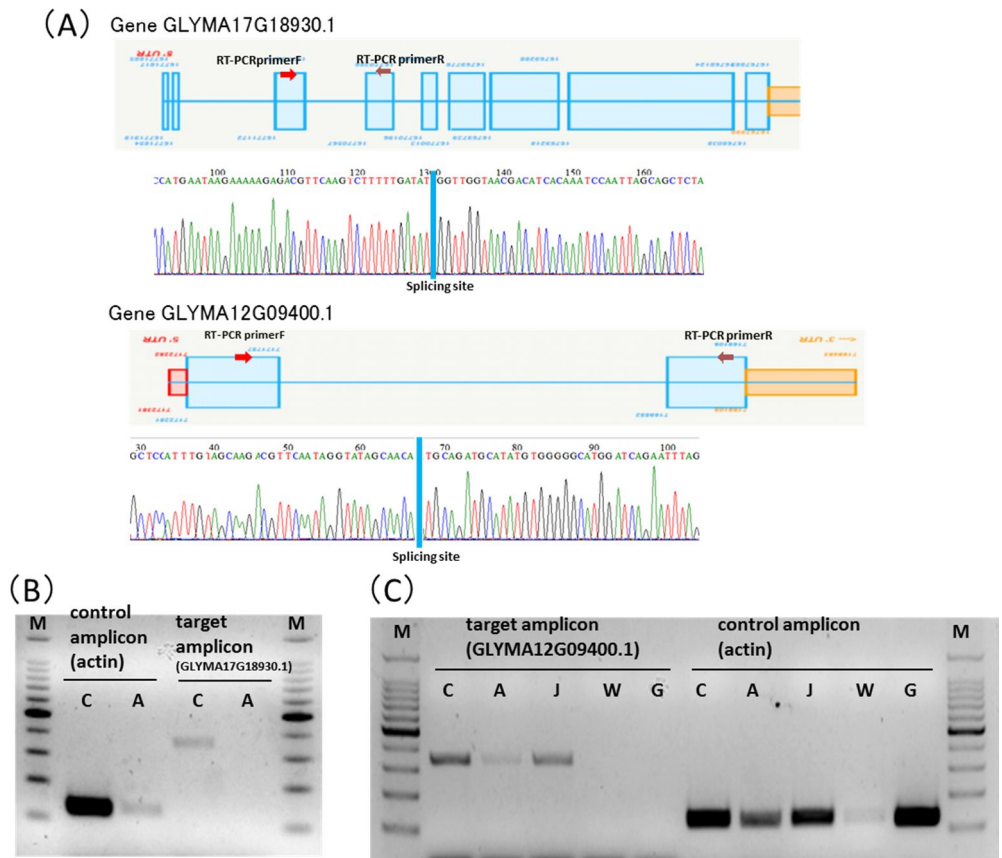


Figure 7. 1% (w/v) agarose gel electrophoresis of the RT-PCR samples

(A) Gene structure of gene GLYMA17G18930.1 and gene GLYMA12G09400.1 in the soybean genome, (B) gene GLYMA17G18930.1 expression analysis, (C) gene GLYMA12G09400.1 expression analysis

C, Canada; A, US; J, Jack; W, Non-GM-type Williams; G, GM-type Williams; M, 100 bp marker

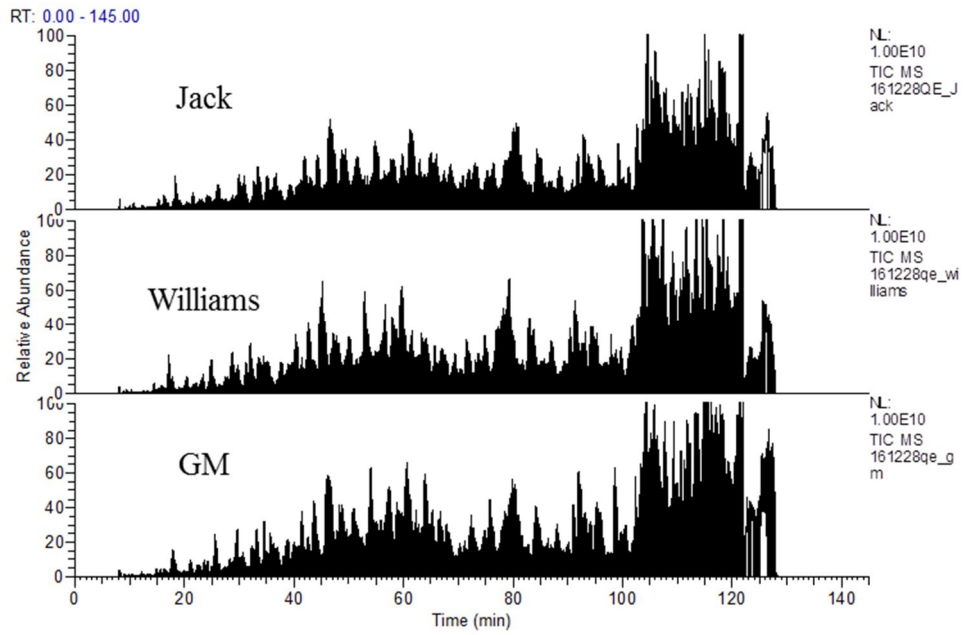


Figure 8. Comparison of TIC chromatography (0-130 min)

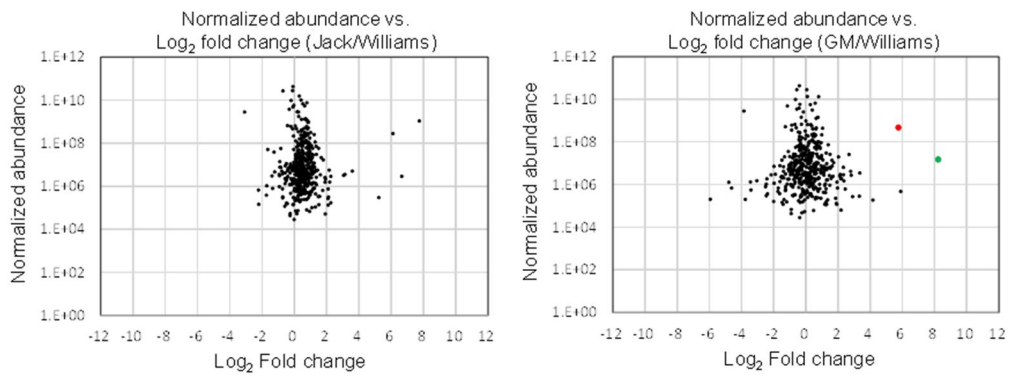
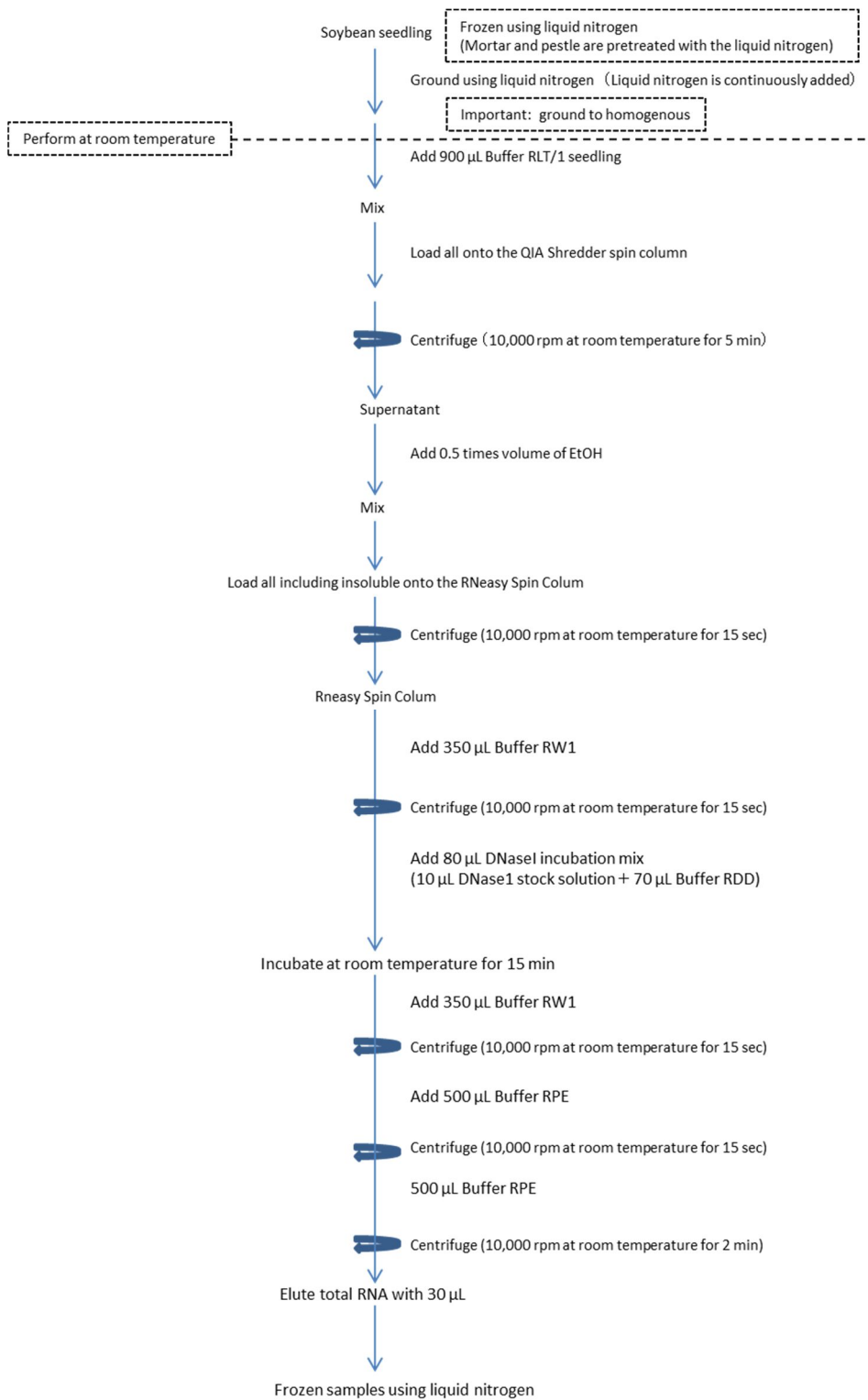
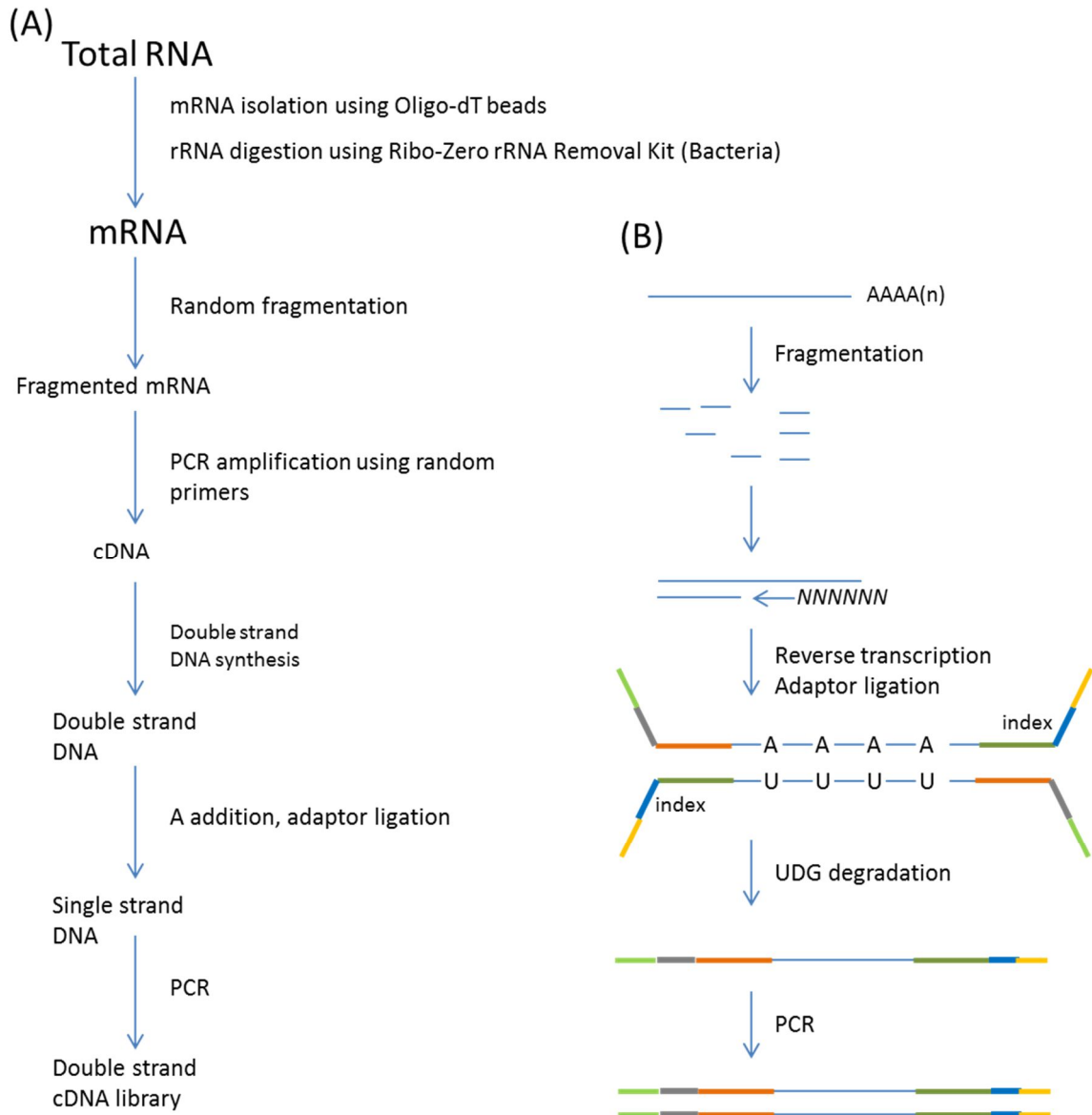


Figure 9 . Scatter plots of normalized abundance and fold change of the proteins eGFP and SYNC1 are marked with red and green points, respectively.



Scheme1. Purification scheme for the total RNAs



Scheme 2. RNA-Seq sample preparation

(A) Experimental steps for double strand cDNA library preparation

(B) Schematic diagram for the double strand DNA preparation

- ・野生型品種、及び、種子粒間の発現差解析を行う
- ↓
- ・得られる最大上限fold値を閾値に設定する
- ↓
- ・閾値を超えた遺伝子については、RNA-Seq(リファレンスゲノムへのマッピング)データを参照し、差を定性的に確認する
- ↓
- ・検出エラーでないものは、RT-PCRにて、定性的・定量的に確認する

Scheme 3. The flow of the gene expression analysis using RNA-Seq data for germinated non-GM and GM soybeans