

修士論文

遺伝子発現解析による

Angelica acutiloba の防御応答機構の解明

原田 将光

2019 年 3 月 8 日

奈良先端科学技術大学院大学

情報科学研究科

本論文は奈良先端科学技術大学院大学情報科学研究科に
修士(理学) 授与の要件として提出した修士論文である。

原田将光

審査委員：

金谷 重彦	教授	(主指導教員)
笠原 正治	教授	(副指導教員)
小野 直亮	准教授	(副指導教員)
黄 銘	助教授	(副指導教員)
MD. ALTAF-UL-AMIN	准教授	(副指導教員)

遺伝子発現解析による

Angelica acutiloba の防御応答機構の解明*

原田将光

内容梗概

Angelica acutiloba (和名: トウキ) は古くから生薬として用いられてきたセリ科シシウド属の多年草である。その薬効は鎮痛・鎮痙・抗菌作用と多岐にわたり当帰芍薬散や潤腸湯をはじめとする様々な漢方薬に用いられている。しかしながら、*A. acutiloba* は国内での生産量は決して多いとは言えず、その一因として *Papilio machaon* (和名: キアゲハ) による食害がある。通常、植物は外部より食害によるストレスを受けた時、防御応答機構を活性化させ生存性を高めようとする。この際に代謝される化合物はストレスを排除するのみでなく、人に対して薬効成分を示す物質も多い。よって食害の有無により活性化される代謝経路を解明することにより、人為的に食害に強いのみでなく品質が高い *A. acutiloba* を作り出すことが期待できる。

本研究では mRNA-seq を用いて遺伝子発現解析を行うことにより、食害に対する防御応答機構の解明を行った。これにより、二次代謝経路を中心に食害により変動する遺伝子・代謝経路を明らかにした。

キーワード

バイオインフォマティクス、mRNA-seq、遺伝子発現解析、*Angelica acutiloba*

*奈良先端科学技術大学院大学 情報科学研究科 修士論文,
2019 年 3 月 8 日

*Insight into defense mechanism of *Angelica acutiloba* based on gene expression analysis **

Masamitsu Harada

Abstract

Angelica acutiloba is a perennial herb known as toki in Japanese. Its roots are one of the popular varieties used in the Kampo medicine and Traditional Chinese medicine. Those medicine including *A. acutiloba* are mainly prescribed for suppression of women diseases. Because of its hard cultivation methods due to some insect attack, only around 200t are made in Japan. Recently there is the report that the insect attack increases amount and variation of both medical compounds and the compounds used for molecular defense. We examined transcriptome to clarify molecular mechanism between *A. acutiloba* and herbivorous insects using 9 samples (3 condition, 3 replicate) of *A. acutiloba* exposed to *Papilio machaon* for 0 (control plants), 6 and 12h. The experiment revealed the underlying molecular defense mechanism of *A. acutiloba* against *P. machaon*.

Keywords:

bioinformatics, mRNA-seq, gene expression analysis, *Angelica acutiloba*

*Master's Thesis, Graduate School of Information Science, Nara Institute of Science and Technology, March 8, 2019.

—目次—

1 序論	1
1.1 背景と目的	1
1.2 本論文の構成	1
2 手法	2
2.1 解析の流れ	2
2.2 シーケンシング	3
2.2.1 次世代シーケンサー	3
2.2.2 ブリッジ PCR	4
2.2.2 Sequencing By Synthesis 法	4
2.3 <i>De novo</i> アセンブリ	5
2.4 マッピングと発現量推定	7
2.4.1 マッピング	7
2.4.2 発現量推定	7
2.5 機能アノテーション	8
2.5.1 アノテーション	8
2.5.2 DIAMOND	8
2.6 発現変動遺伝子の検出	9
2.6.1 TMM 正規化	9
2.6.2 二群間比較	11
2.7 パスウェイ解析	12
2.7.1 代謝経路	12
2.7.1 クラスタリング	12
2.8 代謝経路解析	13
3 結果と考察	14
3.1 データセット	14
3.2 遺伝子発現量推定	14

3.3 発現変動遺伝子の検出.....	15
3.4 パスウェイ解析	16
3.4.2 植物ホルモン及び MAPK 伝達経路.....	17
3.4.2 フェニルプロパノイド生合成経路 (Phenylpropanoid biosynthesis pathway) .	19
3.4.3 フラボノイド類生合成経路 (Flavonoid biosynthesis).....	21
3.5 代謝経路解析	23
3.5.1 テルペノイド生合成経路及びその周辺における遺伝子発現.....	24
4 結論	27
謝辞	28
参考文献	29
付表	31

-図表目次-

図 1 本研究における遺伝子発現解析の概要	2
図 2 Illumina 社製 NGS HiSeq 3000	3
図 3 <i>De novo</i> アセンブリの概略図	6
図 4 M-A プロット	10
図 5 発現変動遺伝子の内訳	15
図 6 ジャスモン酸とエチレンによる MAPK 伝達経路の発現変動	18
図 7 フェニルプロパノイド生合成経路における遺伝子発現変動	20
図 8 フェニルプロパノイド生合成経路における遺伝子発現変動	22
図 9 対数尤度の各値におけるセグメントの出現頻度	23
図 10 テルペノイド生合成経路における遺伝子発現	25
図 11 ステロイド・モノテルペノイド生合成経路における遺伝子発現	26
表 1 解析に用いるデータセット	14
表 2 クラスタリングにより検出されたパスウェイ	16

1 序論

1.1 背景と目的

近年、次世代シーケンサーや質量分析器などをはじめとする分析技術の急速な発達により、遺伝子やタンパク質の発現情報をはじめとする膨大な量の生物学的データを一度に取得することが可能となった。また、これら膨大な量の生物学的データを効率的に処理するために生物学と情報科学が融合したバイオインフォマティクスと呼ばれる分野が誕生した。バイオインフォマティクスの発達により生物のゲノムやたんぱく質をはじめとする様々な情報を網羅的に解析することが可能となり、生命を遺伝子やたんぱく質のネットワークとしてとらえ総体をシステムとして理解しようとする試みが行われるようになった。これら技術の発展により特に農作業分野では、遺伝情報の発現・代謝メカニズムが明らかにされることにより、農業上有用な形質に関与する遺伝子を調べ、効率よく品種改良を行うことが可能となるなど様々な恩恵を受けている。

本研究の対象である *Angelica acutiloba* (和名：トウキ)は古くから生薬として用いられるセリ科シシウド属の多年草である。その根は当帰芍薬散をはじめとする多くの漢方薬に用いられており、その薬効は抗炎症作用・鎮痛作用・抗菌作用など様々であるが、国内での栽培量は約 185t と決して多くない[1]。これは、*A. acutiloba* は収穫まで 2 年の歳月を要し、*Papilio machaon* (和名：キアゲハ)や *Epiblema leucantha* (和名：クロモンシロハマキ)などの害虫が存在するため、栽培に多くの手間がかかるためである。

そこで、本研究では次世代シーケンサーにより産出された mRNA-seq を解析することにより、二次代謝経路を中心に薬効や防御機構に関わる遺伝子・代謝経路を明らかにすることにより良質な *A. acutiloba* の生産に必要な因子の解明を目的とする。

1.2 本論文の構成

本論文は全 4 章で構成されている。2 章では、本研究のトランスクリプトーム解析で用いた手法を解説、3 章で *A. acutiloba* の解析の結果及び考察を述べ、4 章にて本論文のまとめを述べる。

2 手法

生物はセントラルドグマと呼ばれる分子生物学上の概念に従い DNA から mRNA に転写、mRNA からタンパク質に翻訳を行うことにより遺伝子情報を伝達し、機能を発現させることにより成長促進や生命活動維持などをおこなっている。この遺伝子情報を機能に変換する過程を遺伝子発現と呼び、遺伝子発現調整の大部分を担っているのが DNA から mRNA への転写である。そこで、トランスクリプトーム解析では mRNA の変動量から各遺伝子の発現量を推定可能である点に着目して解析を行う。本章ではトランスクリプトーム解析に用いた手法および周辺技術について解説する。

2.1 解析の流れ

本研究では第 2.2 節から第 2.6 節で述べる通り、mRNA-seq を用いたトランスクリプトーム解析を行った後に統計解析により発現変動遺伝子の検出を行った。また、第 2.7 節および第 2.8 節では発現変動遺伝子に対しパスウェイ解析を行った。

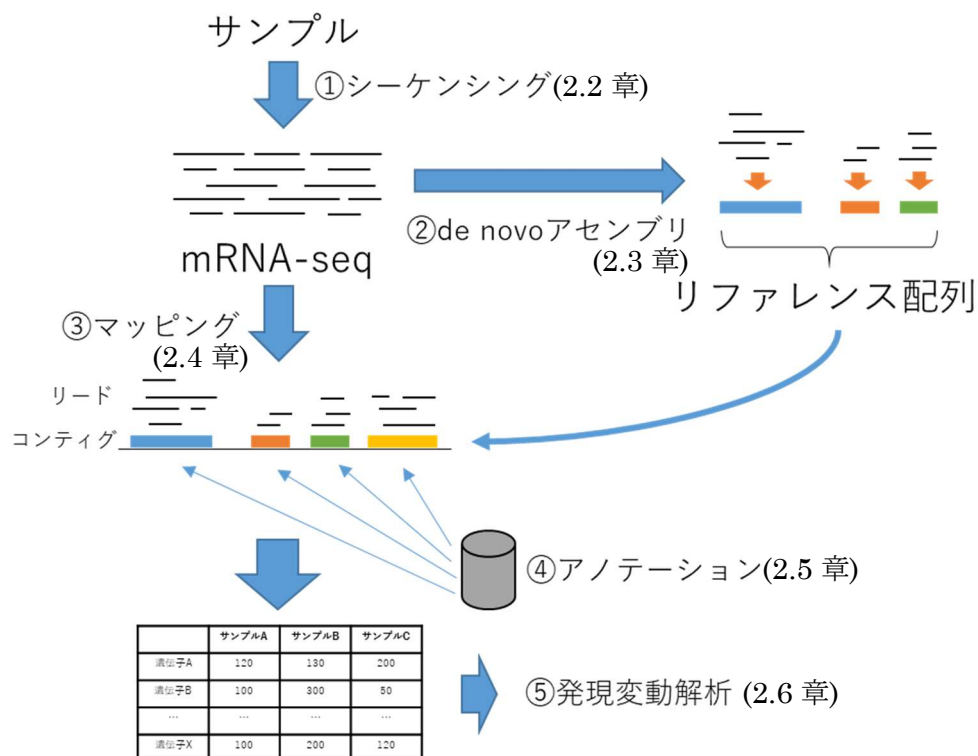


図 1 本研究における遺伝子発現解析の概要

2.2 シーケンシング

2.2.1 次世代シーケンサー

生物の遺伝子情報は A (アデニン), T (チミン), G (グアニン), C (シトシン) の四つの塩基から構成される DNA により保持されており、その配列を決定することをシーケンシング、作業を行う機器をシーケンサーと呼ぶ。塩基配列の決定は 1960 年代より、低分子 RNA を対象に塩基特異的切断性を有する RNA 分解酵素と電気泳動やクロマトグラフィーを用いた分子分離同定技術を組み合わせることにより始まり、1976 年には初めて全ゲノム配列が 3569bp からなる RNA ウイルスのバクテリオファージ MS2 にて決定された[2]。また、1977 年には Maxam-Gilbert 法が発表され、DNA のシーケンシングを行うことも容易となった[3]。しかしながら、Maxam-Gilbert 法は試薬に毒性や不安定性があるなどの問題を抱えており、同年発表された Sanger 法が後にシーケンシングの主流となっていった[4]。この Sanger 法の電気泳動および蛍光検出を自動化したものがキャピラリーシーケンサーであり、この機器の出現により 2000 年には約 130Mbp のシロイヌナズナが、2003 年には約 3Gbp のヒトの遺伝子が決定されるなど、長大な遺伝子長を持った生物種の全ゲノム配列を決定することが可能となった。また、2005 年以降、Smart Molecule Real Time (SMRT) 法や Sequencing By Synthesis (SBS) 法をはじめとする新たな手法を活用した次世代シーケンサー (NGS) が登場した。従来のシーケンサーが数百から千数百塩基という比較的長い断片を、一度に十数本から百本程度ずつ読み取るのに対し、NGS では数重から百塩基程度の短い塩基断片を数千万本から数億本、大量に並列処理するため、従来のシーケンサーに比べ高速かつ多量にリードと呼ばれる短い塩基情報を得られる。本研究ではブリッジ PCR 法および SBS 法を用いた次世代シーケンサーである Illumina HiSeq 3000 (図 2) を用いて mRNA-seq の解析を行った。



図 2 Illumina 社製 NGS HiSeq 3000

2.2.2 ブリッジ PCR

HiSeq を用いた NGS のトランスクリプトーム解析では、まず前処理として解析対象の mRNA を断片化後、ランダムプライマを用い mRNA 断片から cDNA を合成、cDNA の 5' 末端と 3' 末端にそれぞれ別のアダプタを付加することにより、ストランド特異的な mRNA-Seq ライブラリを調整する。調整された mRNA-Seq ライブラリは量が少なくこの状態では NGS により検出することができないため、クラスタ形成を行いシーケンスの際に検出可能な十分量に増やす必要がある。このクラスタ形成に用いる手法がブリッジ PCR 法であり、まず調整されたライブラリを一本鎖 DNA に変性させフローセルと呼ばれるガラス基板上にアダプタと相補的なオリゴヌクレオチドを合成したデバイスに流し、一本鎖 DNA 断片をフローセル上のオリゴヌクレオチドに結合させる。その状態で DNA ポリメラーゼを用い相補鎖の DNA を複製すると、伸長した DNA 断片の両端のアダプタ配列がフローセルと結合するため、橋渡しをしたような状態が形成される。この状態で二本鎖を変性させ結合を外すと、複製された DNA のそれぞれがフローセルに結合した二本の一本鎖 DNA 断片を得ることができる。このブリッジ結合と伸長、変性を繰り返すことにより多数の一本鎖 DNA を局所的に固定したまま増幅することができ、この密集した DNA 断片のコピーをクラスタと呼ぶ。

2.2.2 Sequencing By Synthesis 法

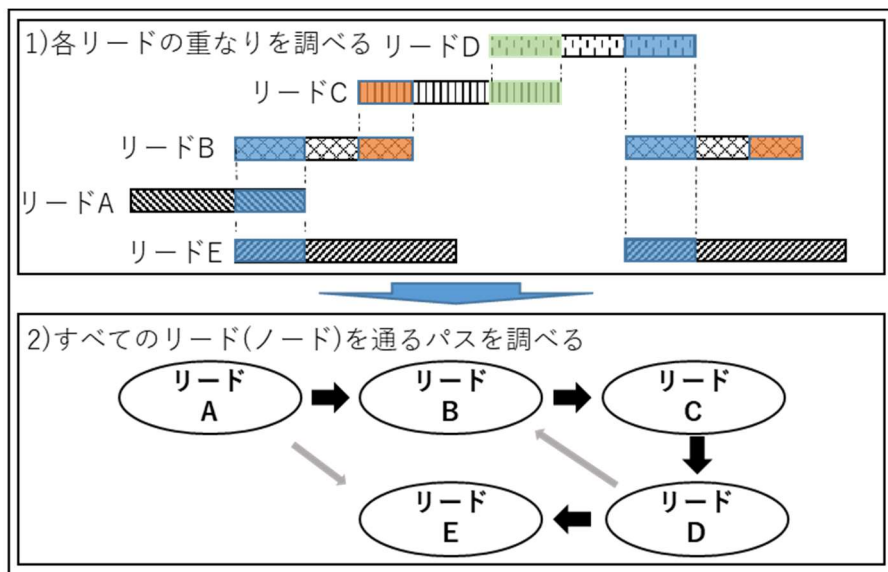
ブリッジ法により増幅固定された DNA 断片に対し可逆的ターミネーターを用いることにより一塩基ずつシーケンスする手法を Sequencing By Synthesis (SBS)法と呼ぶ。HiSeq ではフローセルの表面上に形成されたクラスタに対し、4 種類に蛍光標識されたヌクレオチドを用いてシーケンスを行う。蛍光標識とキャップ付加されたデオキシリボ核酸 (dNTP) を DNA 鎖に追加し、傾向により追加された塩基を読み取る。このキャップされたヌクレオチドはポリメラーゼ反応を阻害するターミネーターとなるため、一塩基以上の伸長反応が起こらないように調整されている。その後、キャップと蛍光色素を取り除き、次の一塩基を伸長させる。これを繰り返すことにより、(1)一塩基伸長、(2)蛍光による塩基の読み取り、(3)蛍光分子とターミネーターの 1 サイクル毎に一塩基の読み取りをおこない、塩基配列を決定している。

2.3 *De novo* アセンブリ

RNA-seq によるトランスクリプトーム解析において、全ゲノムが解読されており既にリファレンスとなる遺伝子配列が存在するモデル生物以外の生物種を扱うには、次世代シーケンサーより読み取られたリードを用いてその生物種がどのような遺伝子配列を所持しているかをあらかじめ推定し、リファレンス配列を構築する必要がある。mRNA の断片から得られたリードをつなぎ合わせてリファレンスとして用いるコンティグと呼ばれる塩基配列を構築する作業を *De novo* アセンブルと呼び、次世代シーケンサーの登場以前は **Overlap-Layout-Consensus(OLC)**法と呼ばれるリード同士の相同性に基づきアライメントを行う手法が頻用された[5]。OLC 法は図 3a のように、各リードを頂点として共通連続塩基があるノード同士を辺で結んだグラフを作成後、全てのノードを通るパスを検索する手法であり、各リードの塩基長が比較的長い旧世代シーケンサーにより読み取られたリードのアセンブルに対しては有効なアセンブル手法であった。しかしながら、この手法では NP 完全であるハミルトンパス問題を解く必要があり、リード数に対して多項式時間以下で解くアルゴリズムは知られていないため、短い配列長のリードを多量に読み取る NGS を用いた解析では計算量が爆発してしまう。

これらの問題に対応し効率的にコンティグを生成する手法として **de Bruijn Graph**を使用したアルゴリズムが提案されており、本研究ではその実装の一つである **Trinity**と呼ばれるソフトウェアを使用している[6]。de Bruijn Graph は図 3b のように初めに個々のリードを k 個の塩基配列 (**k-mer**) の連鎖したものとして表現し、それぞれの **k-mer** を頂点、配列の連鎖を頂点間の辺として結ぶことでグラフとして表したものである。その後、各リードから作成されたグラフの和集合をとることによって全リードを合成したグラフを作成し、出来るだけ多くの辺を通るパスを検索する。与えられたグラフの全ての辺を一度ずつ通るような経路はオイラーパスと呼ばれるが、この検索は辺の数 E のオーダー計算量 $O(E)$ で可能なことが知られている。独立したグラフごとの最大数の辺を通るパスを検索し、そのパスに対応する配列をコンティグとして保存する。Trinity では計算量が非常に大きいハミルトンパス問題である OLC 法の代わりに de Bruijn グラフを用いたオイラーパス問題に変換することにより効率的なアセンブリを可能としている。また、トランスクリプトームにおける *de novo* アセンブルでは通常のゲノム解析でも起こりうるシーケンスエラーなどの実験手法上の問題のほかにも、遺伝子の発現量の差による断片数のバイアス、選択的スプライシングによるスプライシングバリエانتの存在、重複遺伝子、偽遺伝子の存在などに注意する必要がある。

a)



b)

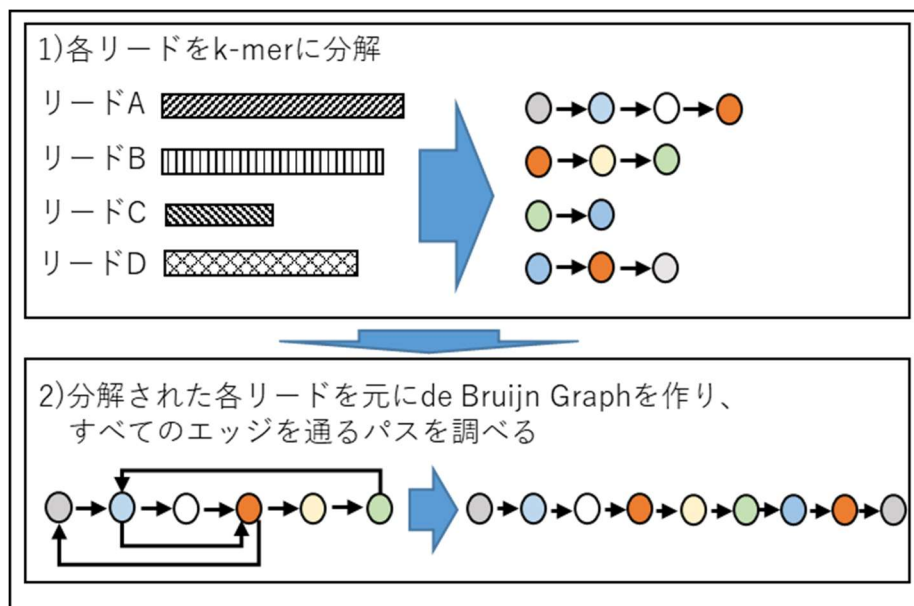


図 3 *De novo* アセンブリの概略図

(a)OLC 法による *de novo* アセンブリ、

(b)de Bruijn Graph を用いた *de novo* アセンブリ、

OLC 法による *de novo* アセンブリがハミルトンパス問題としてアセンブリを行うのに対し、de Bruijn Graph は k-mer 分解を行いオイラーパス問題としてアセンブリを扱っている。

2.4 マッピングと発現量推定

2.4.1 マッピング

モデル生物の既知のリファレンス配列や *de novo* アセンブルで生成されたライブラリをリファレンス配列として、これに対して mRNA から得た個々のリードを対応付ける操作をマッピングと呼ぶ。マッピングを行う代表的なソフトとして Bowtie ,Bowtie2 , TopHat2 ,HISAT2 ,BWA ,STAR などが存在し、それぞれ扱えるリード長やエクソンの対応の有無などの違いがある。本研究では、短いリードを効率よくマッピングすることができる Bowtie2 を用いてリードをリファレンスへマッピングした[7]。

2.4.2 発現量推定

発現量推定には Alignment-free transcript quantification と Alignment-based transcript quantification と呼ばれる二つの方法が存在する [8]。Alignment-free transcript quantification はマッピングを行わずに発現量推定を行う手法であり、マッピングを省いているため精度は落ちる可能性があるものの非常に高速で有ることが特徴である。一方、Alignment-based transcript quantification はマッピングの結果をもとに発現量推定を行う手法であり高精度であることが多い。前者の手法を用いている代表的ソフトは Sailfish, Salmon, kallisto などがあげられ、後者は Cufflinks や RSEM などである。

本研究では、Alignment-based transcript quantification を採用した RSEM (RNA-Seq by Expectation Maximization) と呼ばれるソフトを用いて発現量推定を行った[9]。Alignment-based transcript quantification によるマッピングにおける問題点として、トランスクリプトーム解析において転写産物が同じエクソンを持つ場合や、一つの遺伝子が複数の転写物に転写される際のリードの配分があげられる。例として、ある遺伝子が 3 つのエクソン領域 (E_1, E_2, E_3) を持ち、そこから 3 つの転写産物を ($T_1:E_1-E_2-E_3$, $T_2:E_1-E_2$, $T_3:E_1-E_3$) を作るとする。この時、あるリードが E_1 へマップされた場合、転写産物にリードが分配されるかにより後の解析結果に差異が生じる可能性がある。この問題に対し RSEM では複数のマップ候補が一つのリードにある場合、転写産物の配列に基づき尤度計算を行うことで複数箇所に適切に配分を行っている。

2.5 機能アノテーション

2.5.1 アノテーション

De novo アセンブリにより得られたコンティグは単なる配列情報であるため、個々のコンティグの機能は未知である。個々のコンティグの機能を明らかにする作業をアノテーションと呼び、実験的手法を用いて一つ一つの機能を同定する他、ホモロジーサーチと呼ばれる手法が存在している。ホモロジーサーチとは対象となる配列を既に機能が同定されている遺伝子やタンパク質をまとめたデータベースの配列と比較を行うことで類似性を持つ配列を検索する手法である。これは、共通の先祖を持ち機能が似ている遺伝子である相同遺伝子（ホモログ）は類縁度が高いと推測できることに基づいており、BLAST や FASTA などがホモロジーサーチ用のソフトウェアとして一般的に用いられている。

2.5.2 DIAMOND

本研究ではホモロジーサーチのソフトウェアに DIAMOND を用いた[10]。ホモロジーサーチを行う上で異なる配列同士の類似度を求めるために、下式で表される Smith-Waterman アルゴリズムが一般的に用いられている。

$$F(i, j) = \max \left\{ \begin{array}{l} 0 \\ F(i-1, j-1) + s(x_i, y_j) \\ F(i-1, j) - d \\ F(i, j-1) - d \end{array} \right\} \quad (1)$$

ここで、 $F(i, j)$ は塩基 x_i と y_j のスコア、 s は塩基 x_i と y_j が一致・不一致の時に与えられるスコア、 d はギャップペナルティ($d>0$)とする。

しかしながら、上記アルゴリズムのみを用いた手法ではデータ量が多い場合、非常に多くの計算時間が必要となる。そこで、DIAMOND では、Seed-And-Extend strategy と呼ばれる手法を用い、まず検索対象の配列とデータベース上の配列で類似度が高い領域を検索行い、類似度が高い領域のみで Smith-Waterman ベースのスコア計算を行うことで高速な検索を実現している。

2.6 発現変動遺伝子の検出

2.6.1 TMM 正規化

マッピングされた遺伝子の発現量は分析機器の性能の問題等によりサンプルごとに偏りが発生する。この偏りを補正し、より高い精度で発現変動遺伝子を検出するために TMM (Trimmed mean of M values) 正規化を行う [11]。任意の二群間の発現量を比較した場合、発現変動をしていない遺伝子が大部分を占め、二群間の発現量変動していない遺伝子の $\log_2$ fold-change (M 値) は 0 で有ることが期待されるがサンプル間の偏りにより多くの場合差異が発生する。この状態では、発現変動遺伝子の検出において発現変動していない遺伝子を誤検出する可能性があるため、M 値が 0 になるように正規化を行うことで誤検出を減らす必要がある。TMM 正規化は下の計算により行われる。

(1) すべての遺伝子の M 値および A 値を算出する。

遺伝子 $g = 1, 2, 3, \dots, G$ において、任意のライブラリ l, k に対し M 値及び A 値は以下のよう求める

$$M_g = \log_2 \frac{Y_{gk}}{N_k} - \log_2 \frac{Y_{gl}}{N_l} \quad (2)$$

$$A_g = \frac{1}{2} \left(\log_2 \frac{Y_{gk}}{N_k} + \log_2 \frac{Y_{gl}}{N_l} \right) \quad (3)$$

(2) M 値の上位 30% および下位 30% を除外する

(3) A 値の上位 5% および下位 5% を除外する

(4) 残った遺伝子の発現量を用いて TMM 係数を算出する。

$$\log_2(TMM_k^r) = \frac{\sum_{g=1}^G w_{gk}^r M_{gk}^r}{\sum_{g=1}^G w_{gk}^r} \quad (4)$$

$$w_{gk}^l = \frac{N_k - Y_{gk}}{N_k Y_{gk}} + \frac{N_r - Y_{gr}}{N_r Y_{gr}} \quad (5)$$

ここで、 Y_{gk} はライブラリ k における遺伝子 g の発現量、 N_k はライブラリ k のサイズ ($N_k = \sum_{g=1}^G Y_{gk}$)、 r は k 以外のすべてのライブラリを表している。

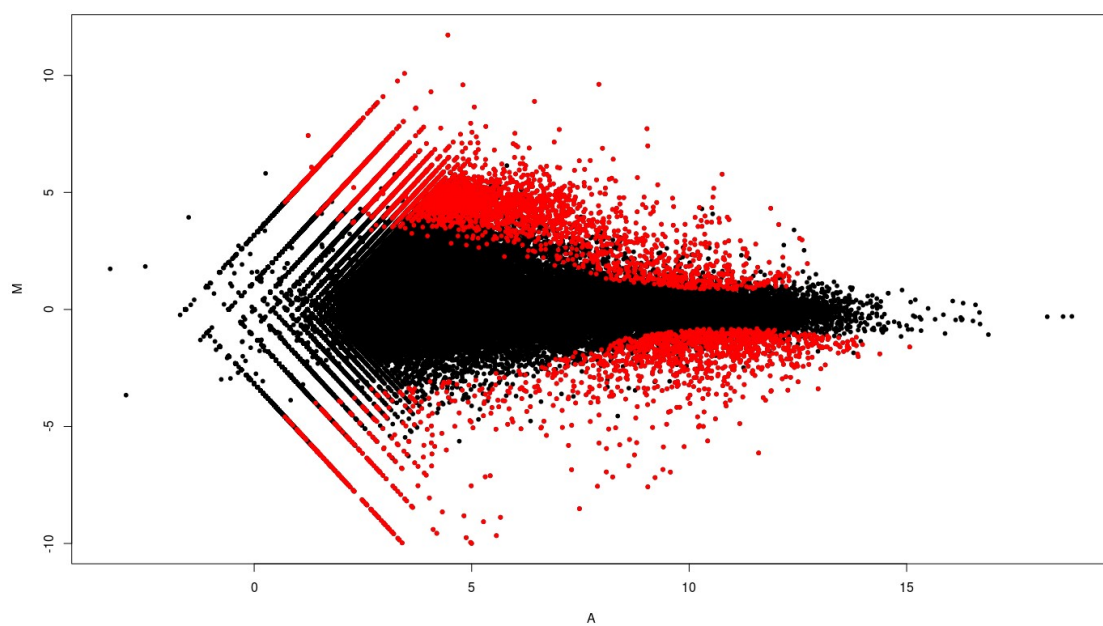


図 4 M-A プロット

縦軸は M 値、横軸は A 値である。赤で描画された点は遺伝子発現変動している遺伝子を黒で描画された点は発現変動なしの遺伝子を表している。

2.6.2 二群間比較

二群間の遺伝子発現量に対し有意差検定を行い、発現変動している遺伝子の検出を行う。mRNA は一般的に負の二項分布に従うことが報告されているため、今回は負の二項分布を用いた有意差検定を行った[12]。

負の二項分布は確率 p で成功するベルヌーイ試行を行った際に r 回成功するまでに必要な試行回数の確率分布であり、式6から式8で与えられる。

$$P(Y = y) = \binom{y + r - 1}{y} p^r (1 - p)^y \quad (6)$$

$$E(Y) = \frac{pr}{1 - p} \quad (7)$$

$$V(Y) = \frac{pr}{(1 - p)^2} \quad (8)$$

負の二項分布を用いた検定では、発現量の類似している遺伝子は分散と平均が類似しているとの考えに基づき、上式から近似する分布を求める。その後、発現量の合計値を推定された分布に当てはめ、発現量の合計値よりも大きい値が観測される確率を p 値として検定を行っている。また、本検定は多重比較検定であるため補正を行うために FDR (False Discovery Rate) の算出を行っている。

2.7 パスウェイ解析

2.7.1 代謝経路

代謝とは、生命が成長や生殖、生命維持のために生体内で行う酵素による化学反応のことであり、代謝のうち生命維持に不可欠な化合物を生産する代謝を一次代謝、生命維持に直接関与しない副次的な化合物を生産する代謝を二次代謝とよぶ。また、代謝によって産生される化合物を代謝物と呼ぶ。代謝にかかわる多数の化学反応は生体内で複雑に繋がっているが、その中でも化学反応の反応物と生産物が連鎖的につながっている一連の経路のことを代謝経路と呼ぶ。一般に、生物が目的となる代謝物を効率よく生産するためには関連する一連の代謝経路を活性化させる必要があり、同じ代謝経路上の反応を触媒する複数の酵素は相関して発現することが多いと考えられている。

2.7.1 クラスタリング

活性化された酵素に関連する酵素は同様の傾向を持って発現している可能性が高い。よってクラスタリングを行いパスウェイごとに各クラスタの有意差検定を行うことにより、同様の傾向をもつ遺伝子が多く含まれる代謝経路を検出することができる。

本研究では非階層的クラスタリングの一種である **x-means** によりクラスタリングを行った。**x-means** は **k-means** を元にクラスタ数を自動決定する手法であり、下の手順により行われる。

- (1) 少数の任意のクラスタ数で **k-means** を行う。
- (2) 作成されたクラスタに対し **2-means** を行い、クラスタを二分割する。
- (3) 新たに作成された 2 クラスタの **BIC** を計算する。

$$BIC = I(D) + \frac{p}{2\log(|D|)} \quad (9)$$

$$I(D) = \sum_{D_i} \sum_{x \in D_i} -\log(\sqrt{2\pi}\sigma^p) - \frac{1}{2\sigma^2} |x - \mu_i|^2 + \log \frac{|D_i|}{|D|} \quad (10)$$

ここで、 p を次点数、 D を p 次点数の集合、 μ_i を D_i の重心とする。

- (4) 2 クラスタの **BIC** が分割前のクラスタより小さくなるまで手順 2 を繰り返す。

上述した **x-means** によるクラスタリング結果をもとに、各パスウェイにマッピングを行い、各パスウェイに属する各クラスタの遺伝子発現数を用いてフィッシャーの正確確率検定により p 値を算出している。また、本検定は多重比較検定のため **FDR** を産出している。

2.8 代謝経路解析

クラスタリングによるパスウェイ解析ではパスウェイごとの大まかな変動傾向の把握することは可能であるが、同一パスウェイ内に目的となる複数の代謝産物が存在しかつ複数の遺伝子が発現している場合などに、どの代謝経路が一連の連動した化学反応として変動しているかを推定することが困難となる。そこで、より精密に発現変動している代謝経路の推定を行うために、発現変動遺伝子および既知のパスウェイを用い下記のように尤度推定を行う。

- (1)KEGG パスウェイデータベースを元に発現変動遺伝子のみで構成されたネットワークを作成する。
- (2)作成したネットワークを元に連続する 3 つの遺伝子をセグメントとして切り出す。
- (3)発現変動解析の結果をもとに各時点の遺伝子量を増加、減少、変動なしの 3 グループに分ける。
- (4)各時点の結果を時系列順に並べ、変動パターンを各時点 3 グループに分類する。
- (5)下式で尤度推定を行い、同様の傾向で変動している遺伝子の抽出を行う。

$$P(D|N) = \prod_{i=1}^n \prod_{j=1}^{q_{ij}} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} (N_{ijk})! \quad (11)$$

ここで r_i は i 番目のセグメントに存在する遺伝子の取り得る状態数、 N_{ijk} は i 番目のセグメントにおける分類後の変動パターンの総数($N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$)を表している。

3 結果と考察

3.1 データセット

A. acutiloba の葉 *P. machaon* の幼虫を設置後、0, 6, 12 時間後に葉をサンプルとして各時点 3 サンプルずつ採取を行い、次世代シーケンサーを用いて解析により mRNA-seq を表 1 の通りに得た。

表 1 解析に用いるデータセット

条件	サンプル名	総配列長(Mbp)
食害なし	Aa-0h-1-a	9546
	Aa-0h-1-b	9061
	Aa-0h-3	9481
食害あり (6 時間後に採取)	Aa-6h-1	8553
	Aa-6h-2	8101
	Aa-6h-3	8276
食害あり (12 時間後に採取)	Aa-12h-1-a	9117
	Aa-12h-1-b	8092
	Aa-12h-2	8194

3.2 遺伝子発現量推定

手法で述べた通りにすべてのサンプルを用いて Trinity で *de novo* アセンブリを行い計 512104 コンティグ得た。512104 コンティグをリファレンス配列とし、RSEM を用いてマッピング及び重複配列の削除を行った結果少なくとも一本のリードが 284945 コンティグにマッピングされた。De novo アセンブルにより得られたコンティグは配列のみであり遺伝子機能に関する情報を有していないため、DIAMOND で各コンティグを KEGG Orthology データベースに対して相動性検索を行い、E-value<10⁻⁸ となった配列はアノテーションができたものとした。結果として 284945 コンティグ中 83925 コンティグがアノテーションされた。

3.3 発現変動遺伝子の検出

食害なし・食害後 6 時間の 2 群間および食害なし・食害後 12 時間の 2 群間で有意差検定を行い、食害を受けた際に発現変動した遺伝子を検出した。284945 コンティグに対し有意差検定を行ったため、多重検定の補正を行うために FDR を算出し、 $FDR < 0.05$ を満たしているものを遺伝子発現変動しているとみなした。有意差検定の結果、284945 コンティグ中 11670 コンティグが発現変動しているとみなされ、内 3645 遺伝子が KEGG Orthology データベースによりアノテーションされている。これら遺伝子発現変動をしていると見なされた遺伝子の内、8118 コンティグは 0h vs 6h で発現しており、5728 コンティグの発現量が増加、2390 コンティグが減少していた。また、0h vs 12h では 2060 コンティグが増加、2632 コンティグが減少している。0h vs 6h および 0h vs 12h 共に発現変動を起こしていた遺伝子は 2387 コンティグである。

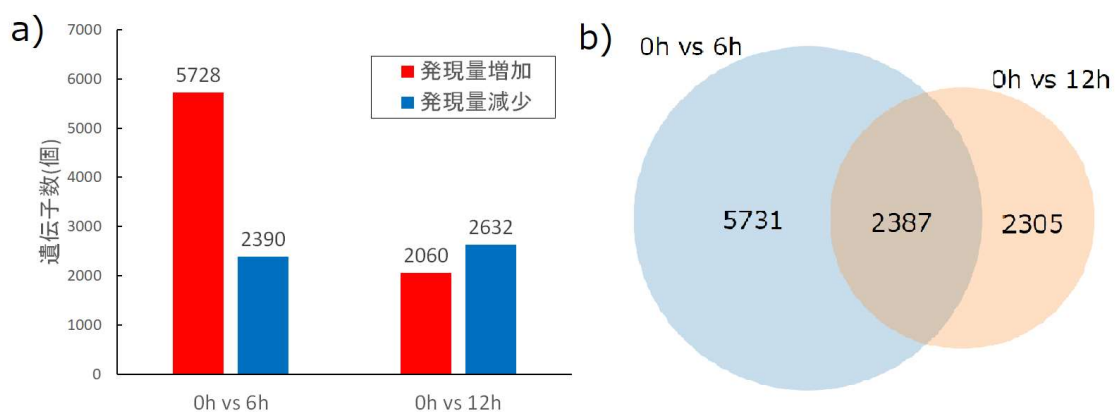


図 5 発現変動遺伝子の内訳

3.4 パスウェイ解析

食害を受けたことにより発現変動した遺伝子の内、KEGG ORTHOLOGY によりアノテーションされた 3645 コンティグの遺伝子を用い KEGG データベース上に存在する既知のパスウェイ情報にマップを行い、どのパスウェイにどの程度存在するかを調べた(附表 1)。また、遺伝子の発現傾向ごとに 3645 コンティグにたいし x-means 法を用いたクラスタリングを行い、結果 45 クラスタを得た。この結果を用いて同様の変動傾向の遺伝子が有意に存在するパスウェイの検出するために、各クラスタの遺伝子を KEGG パスウェイにマップを行い、パスウェイ内の各クラスタに含まれる遺伝子数を用いてフィッシャーの正確確率検定を行った。FDR<0.05 以下を有効に変動しているパスウェイであるとした結果、9 パスウェイが検出された(表 2)。これら代謝経路の多くは食害の有無にかかわらず生命活動維持のために恒常的に変動しているパスウェイであり、食害の有無に直接関係していると考えられるパスウェイは Phenylpropanoid biosynthesis, Flavonoid biosynthesis, MAPK signaling pathway – plant, Plant hormone signal transduction の 4 つである。

表 2 クラスタリングにより検出されたパスウェイ

KO_ID	パスウェイ名	クラスタ番号	p-value	FDR
map00190	Oxidative phosphorylation	45	8.76×10^{-22}	1.59×10^{-18}
map00460	Cyanoamino acid metabolism	16	0.000262	0.036606
map00510	N-Glycan biosynthesis	7	9.22×10^{-5}	0.020974
map00592	alpha-Linolenic acid metabolism	4	0.000413	0.044194
map00940	Phenylpropanoid biosynthesis	18	0.000312	0.039835
map00940	Phenylpropanoid biosynthesis	7	2.34×10^{-6}	0.000608
map00941	Flavonoid biosynthesis	5	0.000127	0.025756
map03018	RNA degradation	7	0.000328	0.039835
map04016	MAPK signaling pathway - plant	7	1.21×10^{-7}	3.67×10^{-5}
map04075	Plant hormone signal transduction	6	0.00054	0.049912
map04075	Plant hormone signal transduction	7	3.99×10^{-8}	1.45×10^{-5}

3.4.2 植物ホルモン及び MAPK 伝達経路

植物ホルモンは植物自身が生産し、形成・成長・代謝を始めとする植物全体の機能制御に関与する有機化合物の総称である。この植物ホルモンのうちジャスモン酸は食害や障害ストレスを受けた時に応答することにより合成される植物ホルモンであり、成長抑制や形態形成をはじめとする多義にわたる成長調整作用により、ストレスに対する体制誘導を担っている[13]。また、エチレンも同様に植物の防御応答や成長を制御することによって知られている植物ホルモンの一種である[14]。これら植物ホルモンは MAPK と呼ばれる Ser/Thr キナーゼを介して細胞外から細胞内にシグナル伝達を行うことにより細胞の代謝・増殖・分裂・アポトーシスをはじめとする様々な機能に関与する。また、ジャスモン及びエチレンは MAPK 伝達経路を介して協奏的に作用することによりストレス応答をおこなうことが知られている[15]。

本解析においては、図 6 のとおり MAPK 伝達経路においてジャスモン酸とエチレンが協奏的に作用するように遺伝子発現していることが見て取られる。ジャスモン酸により MKK3-MPK6 を介して拮抗関係にある MYC2 の発現量を抑えることにより、MYC2 と拮抗関係にあり防御応答機構を活性化させる ChiB の発現量を増加させている。また、エチレン側の遺伝子であり、防御応答を抑制する CTR1 が 6 時間で高発現が増加している。これは RAN1 の発現量が同様に 6 時間目で増加していることに起因すると考えられる。

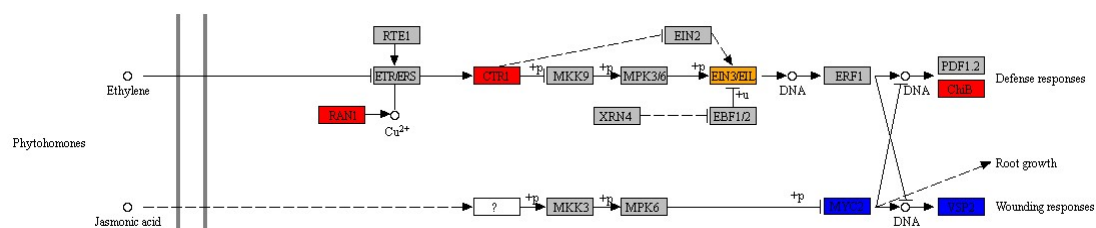


図 6 ジャスモン酸とエチレンによる MAPK 伝達経路の発現変動

赤は 6 時間において発現量が増加した遺伝子、オレンジは 12 時間において発現量が増加した遺伝子、青は 6 時間目以降発現量が減少した遺伝子、灰色はアノテーションされたが発現変動していない遺伝子を表す。食害により生産されるジャスモン酸により発現が抑制される MYC2 及び VSP2 の発現量減少が確認できる。一方エチレンにより抑制される CTR1 は発現量が増加しており、RAN1 の発現量が増加していることに起因していると考えられる。

3.4.2 フェニルプロパノイド生合成経路 (Phenylpropanoid biosynthesis pathway)

フェニルプロパノイド類はフェニルアラニンから合成され、シグナル伝達物質および分子骨格として多種多様な機能を有する植物二次代謝産物群である。初めにフェニルアラニンが脱アミノ化によりケイ皮酸に変換されることにより代謝が始まり、ヒドロキシル化やメチル化を得ることによりクマル酸をはじめとする様々なフェニルプロパン(C6-C3)単位を有する酸が生成される。これらの酸が還元されるとモノリグノールと呼ばれるアルコールになり、モノリグノールがリグニン生合成の出発物質となる。また、フェニルプロパノイド類がフラボノイド類やヒドロキシケイ酸と共に食害により誘導されることが報告されている[16]。

本解析で発現量が変動した遺伝子は図 7 のとおりである。これら遺伝子の中で特筆すべきは lignin の合成にかかわる katG、Coumarine の中間化合物である Coumarinate の合成にかかわる酵素 (EC:3.2.1.21)、フラボノイド類の中間経路に存在する 4CL の発現量が上昇していることである。Coumarine は *Leptinotarsa decemlineata*(和名: コロラドハムシ)に対する殺卵子が報告されているほか *Phaedon cochleariae* をはじめとする多くの虫に対する毒性が報告されている[17]。また、lignin は食害時にセルロースとヘミセルロースの消化率を低下させることが報告されている[18]。これらは害虫に対する直接的な防御作用を有する化合物であり、*A. acutiloba* においても食害防御機構の一端としてこれら代謝物が生産されていることが推測される。

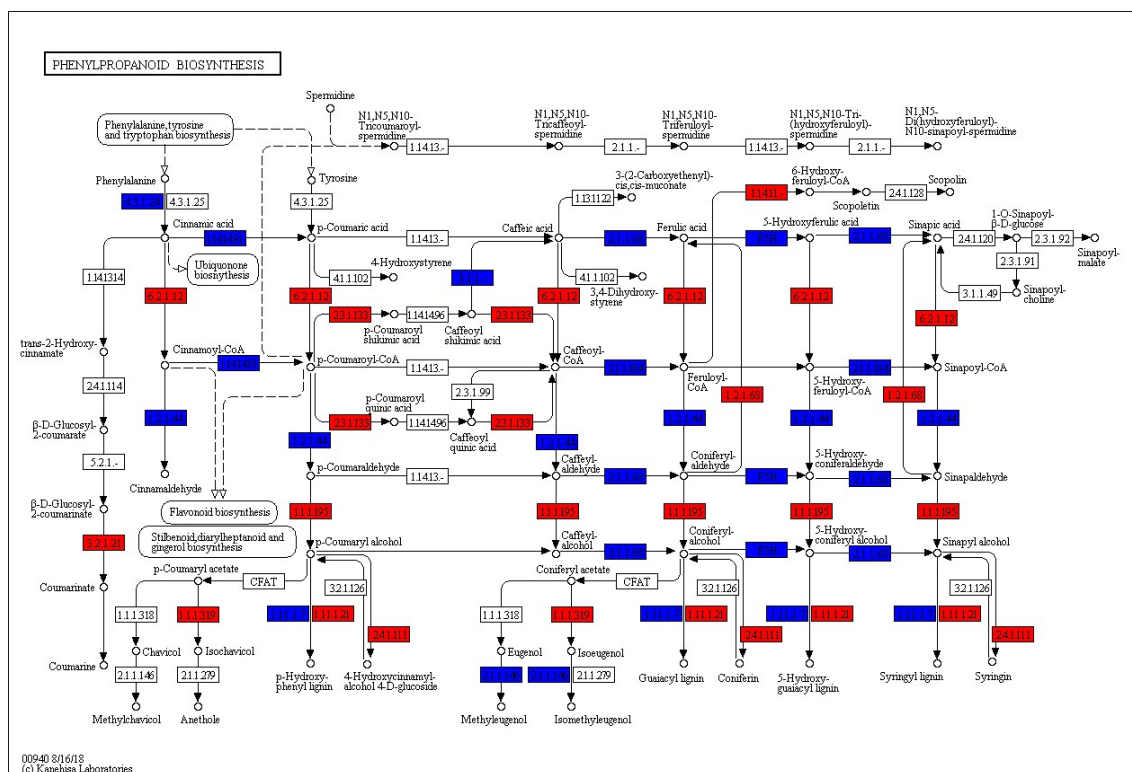


図 7 フェニルプロパノイド生合成経路における遺伝子発現変動

赤は 6 時間において発現量が増加した遺伝子、青は 6 時間目以降発現量が減少した遺伝子を表している。また、白はアノテーション時にアノテーションされなかった遺伝子である。

3.4.3 フラボノイド類生合成経路 (Flavonoid biosynthesis)

フラボノイド類はフェニルプロパノイド誘導体とマロニル・CoA の縮合により合成される化合物群であり、色素や抗酸化活性を含む多様な機能を持つ植物二次代謝産物である。フェニルプロパノイドと同様に食害に対する防御に用いられる化合物も多く存在し、Naringenin, Dihydroquercetin, Catechin, Fisetin の4化合物は *Isoptera* (和名：シロアリ) に対する、Phloretin は *Schizaphis graminum* に対する摂食障害作用が報告されている[19] [20]。

本解析の結果、フラボノイド類生合成経路では6時間時点で発現量変動により高発現量になったものと、12時間時点で発現量が減少する即ち0時間、6時間時点では発現量が多い遺伝子が多くみられ、本代謝経路が活発に動いていることが推測される。6時間時点では摂食障害作用が報告されている Dihydroquercetin, Naringenin, Phloretin の生合成にかかわる, CHS や Chalcone isomerase が高発現しているほか、抗菌作用を有する Luteoforol の生合成する DHR の発現量も増加している。これら抗菌・摂食障害作用を持つ代謝産物を生合成する遺伝子が6時間時点で高発現であることから、*A. acutiloba* はフェニルプロパノイド類のようにフラボノイド類に関しても他植物と同様に防御応答機構の一部としてこれら代謝産物を活用していると推測される。

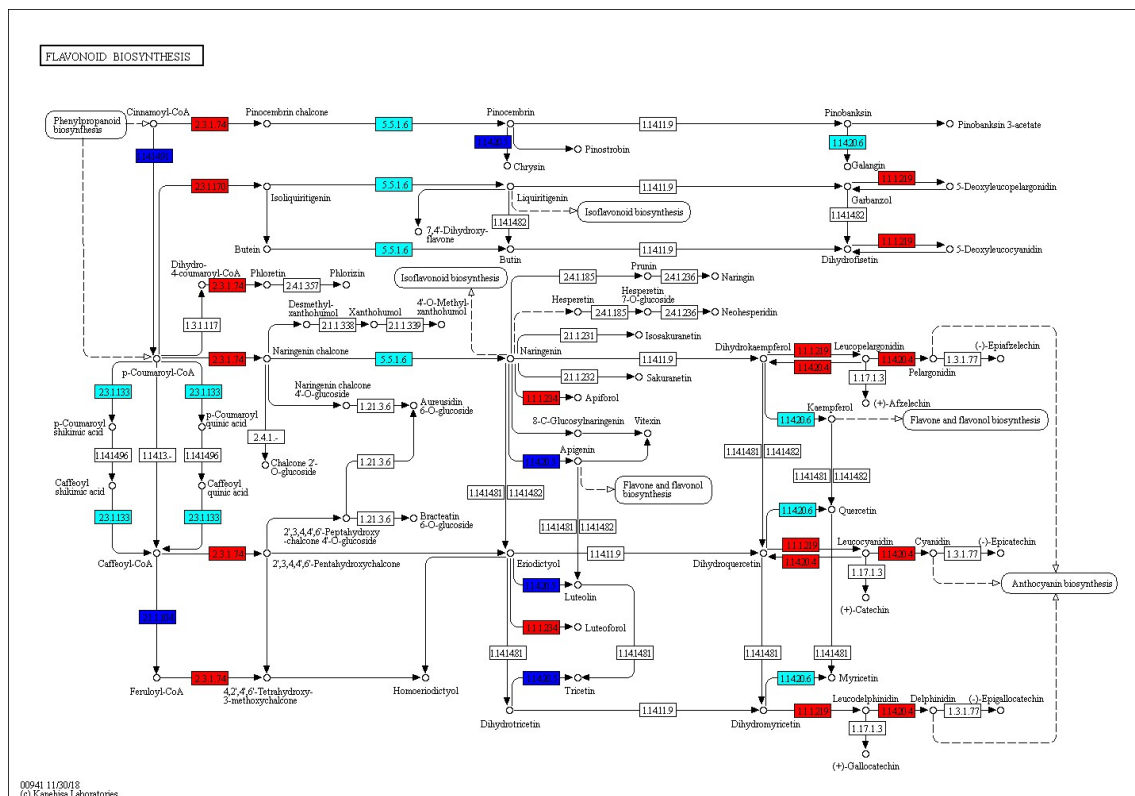


図 8 フェニルプロパノイド生合成経路における遺伝子発現変動

赤は 6 時間後において発現量が増加していた遺伝子、青は 6 時間以降に減少した遺伝子、水色は 12 時間後に減少した遺伝子を表している。また、白はアノテーション時にアノテーションされなかった遺伝子である。

3.5 代謝経路解析

クラスタリングでは検証が困難であったこの遺伝子の変動を確認するために前述した手法を用いて個々の遺伝子を対象としたネットワーク解析を行う。発現変動遺伝子の中、KEGG ORTHOLOGY によりアノテーションされた 3645 個の遺伝子を KEGG パスウェイにマップし、発現変動している遺伝子のみで構成されたネットワークを作成した。作成されたネットワークより連続している 3 つの遺伝子のリストを作成した結果 74330 セグメントを得た。これら全セグメントに対し、尤度計算を行い、特に尤度の高い上位 5% にあたる 3700 セグメントに関する検討を行った。

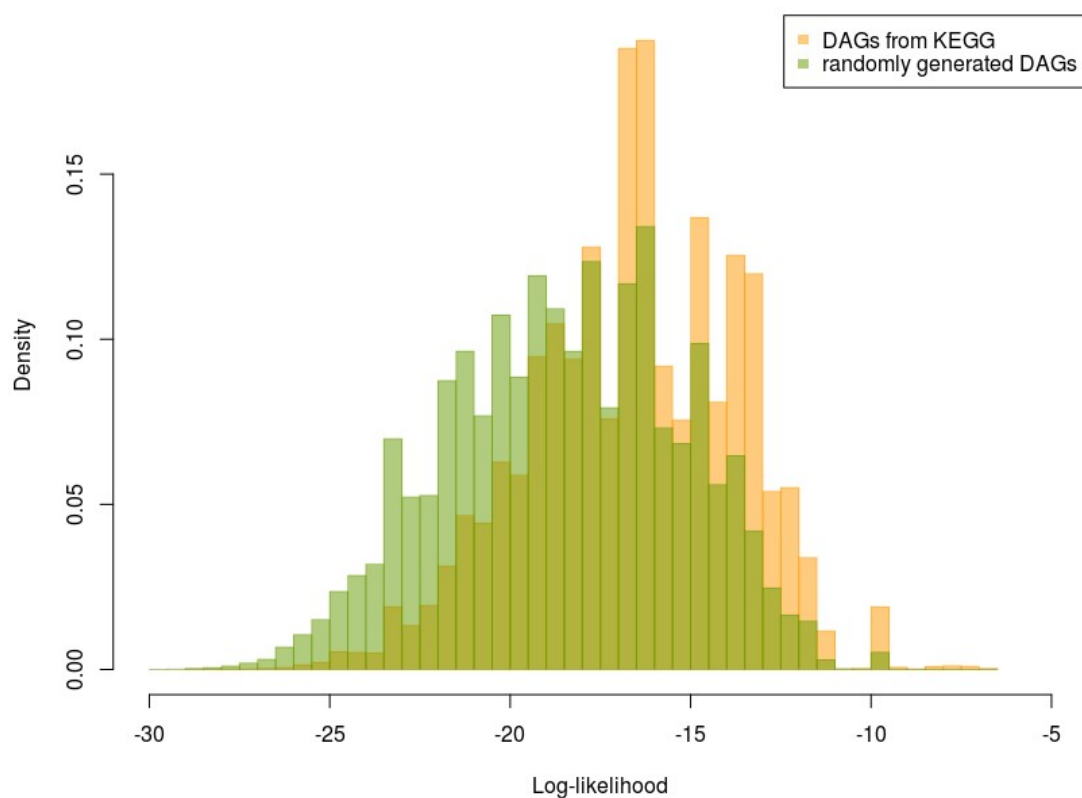


図 9 対数尤度の各値におけるセグメントの出現頻度

3.5.1 テルペノイド生合成経路及びその周辺における遺伝子発現

A. acutiloba において食害を受けた際に生産量の増加が報告されている代謝物質として、テルペノイド生合成経路を通じ代謝されるモノテルペノイドが報告されている。モノテルペノイドの生合成にかかわる代謝経路はアセチル CoA を出発物質とするメバロン酸経路とピルビン酸を出発物質とする非メバロン酸経路の二つがある。これらの経路はイソペンテニル二リン酸(IPP)と呼ばれる中間生産物を生成し、IPP よりモノテルペン、セスキテルペン、ステロイドなど様々なテルペノイドへと代謝される。このテルペノイド生合成経路は β -カリオフィレンをはじめとする *A. acutiloba* の様々な薬効成分に関わっているため、本代謝経路の発現行うことは防御機構および薬効成分を明らかにする上で重要である[21] [22]。

代謝経路解析を通じて *A. acutiloba* がメバロン酸経路を選択的に活性化し IPP を生産していることが明らかとなった。また、メバロン酸経路に始まりフォトステロールの一種である Stigmasterol, Brassicasterol, Campesterol 及びモノテルペノイドの Limonen, Alpha-Terpineol までの代謝経路を一連の反応として同様の傾向で発現している。これら活性化された遺伝子はすべて 6 時間に特異的に発現量が上昇している。ステロイド合成経路における最終代謝物である Stigmasterol は細胞膜を構成する植物ステロールの一種であり人に対する薬効としてはコレステロール抑制が報告されている[23]。また、モノテルペノイドの Limonen に関しても同様に *A. acutiloba* の薬効成分でありこれら化合物が食害により上昇していることは食害が品質の向上につながる可能性があると考えられる。前述したように生物が必要とする代謝産物を効率よく生産するためには代謝経路を一連の反応として効率よく活性化する必要がある。よって、*A. acutiloba* はステロイドおよびモノテルペノイド生合成経路を活性化させるために上流ではメバロン酸経路を選択的に活性化させていると考えられる。

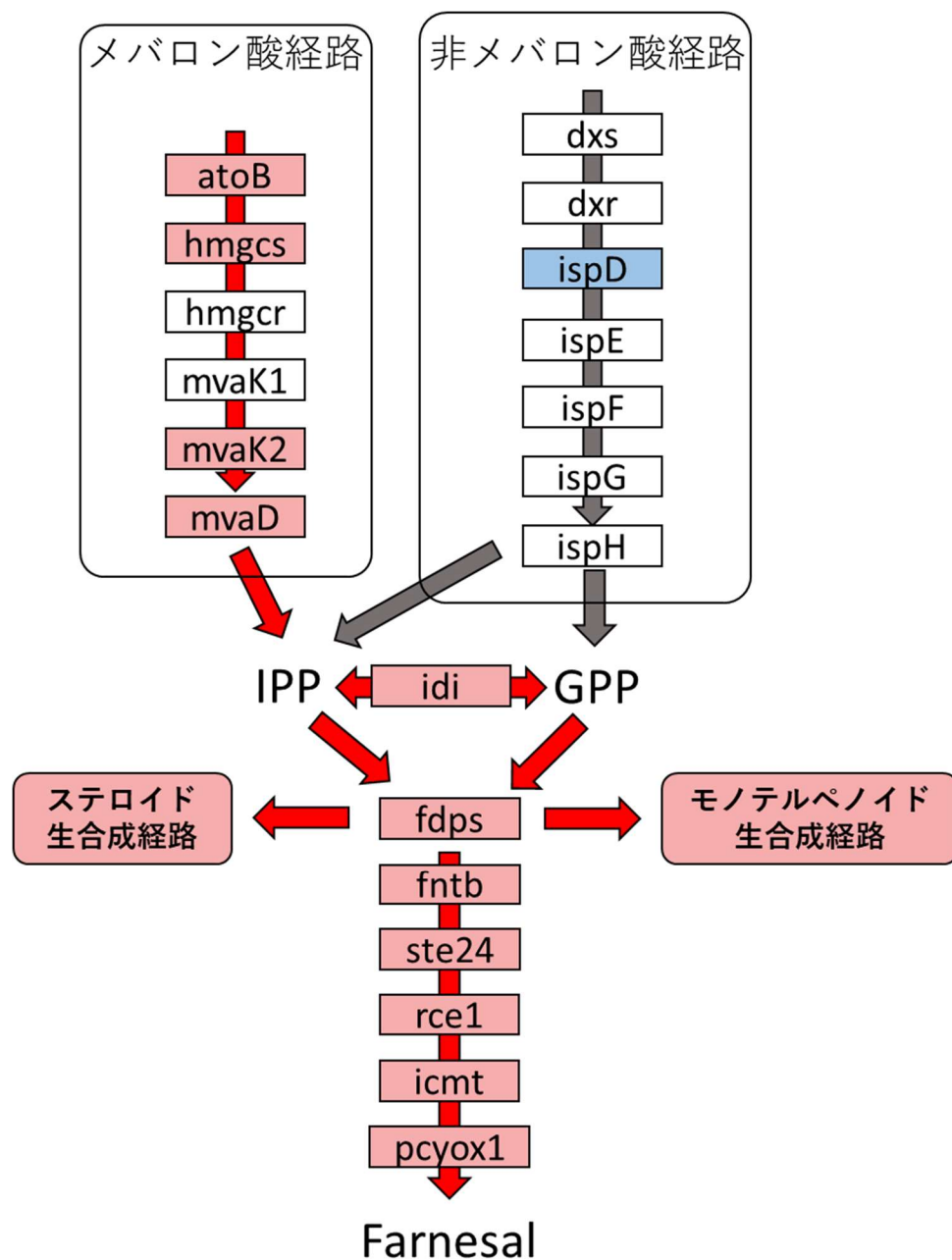


図 10 テルペノイド生合成経路における遺伝子発現

赤は代謝経路解析により同様の傾向で遺伝子発現している一連の遺伝子群を表す。青は発現変動解析で変動は検出されたものの代謝経路解析で除外された遺伝子である。また、白はアノテーションされていない遺伝子である。メバロン酸経路を選択的に活性化させステロイド、モノテルペノイド生合成経路へと繋がっている。

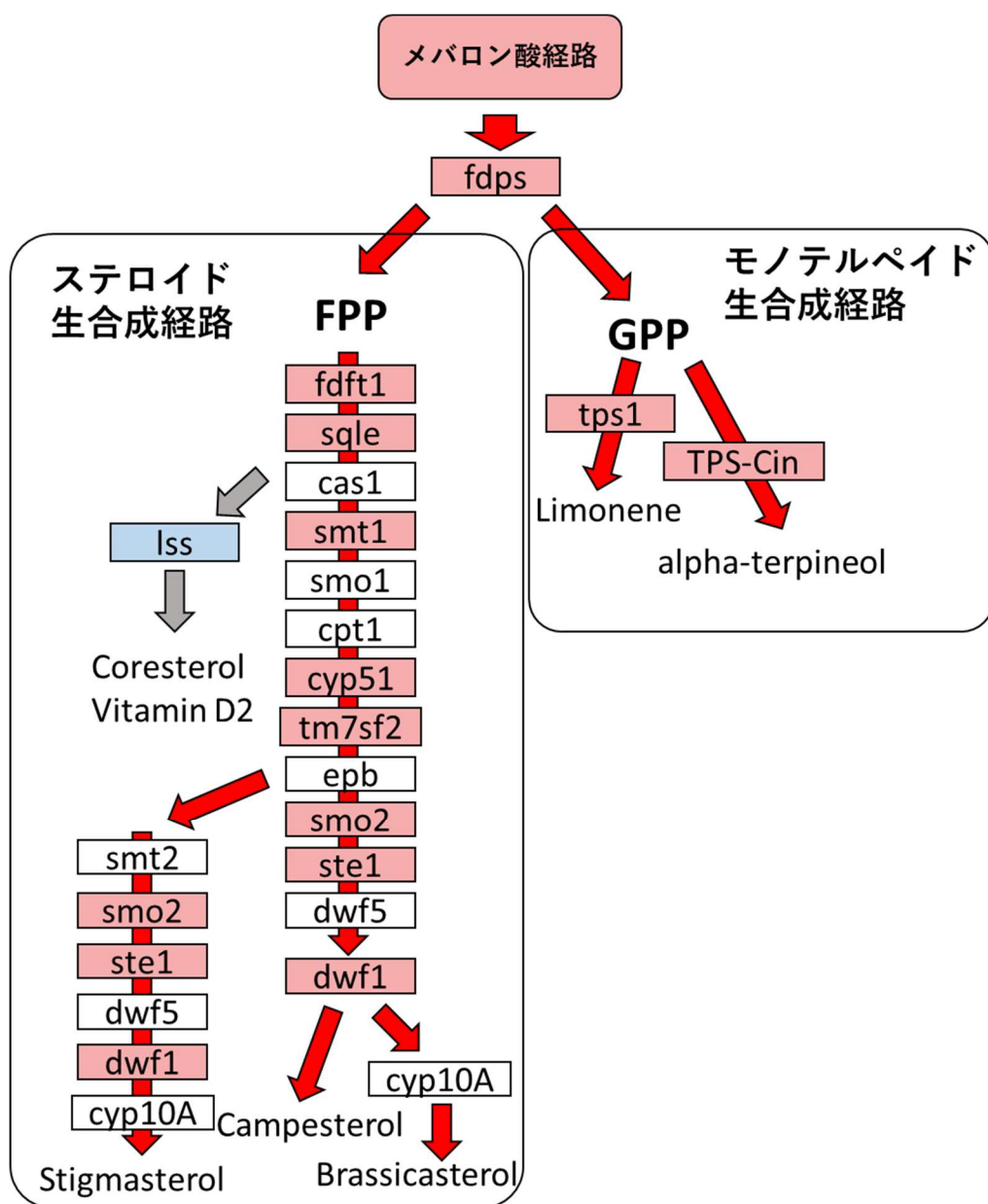


図 11 ステロイド・モノテルペノイド生合成経路における遺伝子発現

赤は代謝経路解析により同様の傾向で遺伝子発現している一連の遺伝子群を表す。青は発現変動解析で変動は検出されたものの代謝経路解析で除外された遺伝子である。また、白はアノテーションされていない遺伝子である。代謝経路解析によりメバロン酸経路と同様の傾向をもった発現遺伝子がステロイドおよびモノテルペノイド生合成経路にて発現していることが明らかとなった。

4 結論

本研究では、*A. acutiloba* の食害に対する防御応答機構の解明を行うために mRNA-seq を用いた遺伝子発現解析を行った。パスウェイ解析を通して、食害により活性化される植物ホルモンであるジャスモン酸を介して MAPK 伝達経路が変動していることが確認され、*A. acutiloba* が一般的な植物の食害応答機構を有することが明らかとなった。また、二次代謝であるフラボノイド類生合成経路及びフェニルプロパノイドを活性化していることも判明した。これら合成経路により産出される代謝産物は多くの植物で食害を含むストレスに対する防御反応に用いられており、これらの代謝産物は他植物で防御応答に用いられる化合物が多く含まれる。特にフラボノイド類は植物の色に関連する化合物を多く含み訪花動物の誘因などに用いられている。色と間接防御のために用いられる害虫の天敵の誘引効果は未だ明らかにされていないが、スズメバチが特定の色の元により害虫を捕食することが報告されているなど[24]、一部の研究ではその影響が明らかとなっている。

代謝経路解析では先行研究により生化学的に変動が報告されているメバロン酸経路から始まるテルペノイド生合成経路を中心とする二次代謝経路に関する発現変動が明らかとなり、テルペノイド生合成経路からステロイド生合成経路およびモノテルペノイド生合成経路までの代謝経路が一連の反応として活性化していることが判明した。本解析に用いた手法は従来のパスウェイ解析では関連性や連続性が見つけにくかった個々の遺伝子をネットワークとして解析することにより効率よく抽出することが可能であることを示した。また、テルペノイド合成経路により合成される化合物の中には揮発性テルペン類と呼ばれる害虫に直接作用するのではなく、害虫の天敵を誘引することにより害虫を駆除する為に用いられる化合物等が含まれている。よって、テルペノイド生合成経路やフラボノイド生合成経路が活性化していることは、*A. acutiloba* が直接的な防御機構の他に色や匂いなどで害虫の天敵を引き寄せる間接的な防御機構を有することが推測される。

本研究では *A. acutiloba* の防御応答機構の十分を解析したとはいえ、今後の生化学的実験により更なる知見が明らかになることが望まれる。しかしながら、本研究により明らかとなった代謝経路の変動は、今後解析を進めるうえで重要な知識であると言え、*A. acutiloba* の安定生産に寄与する足掛かりとなると考えられる。

謝辞

本研究は多くの方々のご指導・ご協力により行うことができました。本研究を行うにあたり、終始懇切丁寧なご指導と研究環境を与えてくださった奈良先端科学技術大学院大学 情報科学研究科 金谷 重彦 教授に深く感謝いたします。本研究を進めるにあたり、日ごろの研究に加え学会発表及び修士論文作成などで多くの助言を賜りました副指導教官である情報科学研究科 小野 直亮 准教授に深く御礼を申し上げます。また、研究室ゼミで研究に関する的確なご意見を頂きました情報科学研究科 黄 銘 助教授と情報科学研究科 Md. Altaf-Ul-Amin 准教授に深く御礼を申し上げます。最後に計算システムズ生物學研究室の皆様には研究生活全般を支えていただいたことをここに感謝申し上げます。

参考文献

1. Yamaoka, D., et al., *The Current Situation and Problems of Domestic Crude Drug Production*. Kampo Medicine, 2017. **68**(3): p. 270-280.
2. Min Jou, W., et al., *Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein*. Nature, 1972. **237**(5350): p. 82-8.
3. Maxam, A.M. and W. Gilbert, *A new method for sequencing DNA*. Proc Natl Acad Sci U S A, 1977. **74**(2): p. 560-4.
4. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors*. Proc Natl Acad Sci U S A, 1977. **74**(12): p. 5463-7.
5. Kececioğlu, J. and J. Ju, *Separating repeats in DNA sequence assembly*, in *Proceedings of the fifth annual international conference on Computational biology*. 2001, ACM: Montreal, Quebec, Canada. p. 176-183.
6. Grabherr, M.G., et al., *Full-length transcriptome assembly from RNA-Seq data without a reference genome*. Nat Biotechnol, 2011. **29**(7): p. 644-52.
7. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2*. Nat Methods, 2012. **9**(4): p. 357-9.
8. Jin, H., Y.W. Wan, and Z. Liu, *Comprehensive evaluation of RNA-seq quantification methods for linearity*. BMC Bioinformatics, 2017. **18**(Suppl 4): p. 117.
9. Li, B. and C.N. Dewey, *RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome*. BMC Bioinformatics, 2011. **12**(1): p. 323.
10. Buchfink, B., C. Xie, and D.H. Huson, *Fast and sensitive protein alignment using DIAMOND*. Nature Methods, 2014. **12**: p. 59.
11. Robinson, M.D. and A. Oshlack, *A scaling normalization method for differential expression analysis of RNA-seq data*. Genome Biol, 2010. **11**(3): p. R25.
12. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data*. Bioinformatics, 2010. **26**(1): p. 139-40.
13. Galis, I., et al., *Molecular mechanisms underlying plant memory in JA-mediated defence responses*. Plant Cell Environ, 2009. **32**(6): p. 617-27.
14. Yoo, S.D. and J. Sheen, *MAPK signaling in plant hormone ethylene signal transduction*. Plant Signal Behav, 2008. **3**(10): p. 848-9.
15. Hettenhausen, C., M.C. Schuman, and J. Wu, *MAPK signaling: a key element*

- in plant defense response to insects*. Insect Sci, 2015. **22**(2): p. 157-64.
16. AlJabr, A.M., et al., *Toxicity of Plant Secondary Metabolites Modulating Detoxification Genes Expression for Natural Red Palm Weevil Pesticide Development*. Molecules, 2017. **22**(1).
 17. S. Sadasivam and B. Thayumanayan, *Molecular Host Plant Resistance to Pests*. 2003: CRC Press. 244.
 18. Muslima Khanam, L.A., D. Talukder, and M.A. Hye, *Toxic and repellent action of sugarcane bagasse-based lignin against some stored grain insect pests*. Vol. 25. 2007.
 19. S, M., et al., *Phenolic Extractives and Natural Resistance of Wood*, in *Biodegradation - Life of Science*. 2013.
 20. Mierziak, J., K. Kostyn, and A. Kulma, *Flavonoids as important molecules of plant interactions with the environment*. Molecules, 2014. **19**(10): p. 16240-65.
 21. Kato, R., et al., *Metabolomic and proteomic analysis of the response of angelica acutiloba after herbivore attack*. Vol. 12. 2017. 1547-1551.
 22. Eguchi, R., *Differential Expression Analysis of RNA-seq data in Angelica acutiloba*. 奈良先端科学技術大学院大学情報科学研究科修士論文 ; 2016, 2016.
 23. Batta, A.K., et al., *Stigmasterol reduces plasma cholesterol levels and inhibits hepatic synthesis and intestinal absorption in the rat*. Metabolism, 2006. **55**(3): p. 292-299.
 24. Masayoshi, U., et al., *Starvation and herbivore-induced plant volatiles affect the color preferences of parasitic wasps*. BioControl, 2013. **58**(2): p. 187-193.

付表

附表 1 各パスウェイの遺伝子発現数

KO_ID	パスウェイ名	DEG (コンティグ 数)
ko01100	Metabolic pathways	713
ko01110	Biosynthesis of secondary metabolites	318
ko01130	Biosynthesis of antibiotics	188
ko01120	Microbial metabolism in diverse environments	169
ko03010	Ribosome	97
ko01200	Carbon metabolism	88
ko04714	Thermogenesis	86
ko01230	Biosynthesis of amino acids	82
ko00190	Oxidative phosphorylation	80
ko04141	Protein processing in endoplasmic reticulum	60
ko00230	Purine metabolism	59
ko03013	RNA transport	56
ko04932	Non-alcoholic fatty liver disease (NAFLD)	54
ko03040	Spliceosome	49
ko04144	Endocytosis	46
ko00240	Pyrimidine metabolism	44
ko00500	Starch and sucrose metabolism	38
ko04120	Ubiquitin mediated proteolysis	38
ko04142	Lysosome	37
ko04146	Peroxisome	37
ko00270	Cysteine and methionine metabolism	35
ko04723	Retrograde endocannabinoid signaling	35
ko04145	Phagosome	34
ko04151	PI3K-Akt signaling pathway	34
ko00260	Glycine, serine and threonine metabolism	33
ko00630	Glyoxylate and dicarboxylate metabolism	33
ko03008	Ribosome biogenesis in eukaryotes	32
ko03018	RNA degradation	32
ko04111	Cell cycle – yeast	32

ko01212	Fatty acid metabolism	31
ko00520	Amino sugar and nucleotide sugar metabolism	30
ko00620	Pyruvate metabolism	30
ko04810	Regulation of actin cytoskeleton	30
ko04150	mTOR signaling pathway	29
ko03015	mRNA surveillance pathway	28
ko03050	Proteasome	27
ko00010	Glycolysis / Gluconeogenesis	26
ko01210	2-Oxocarboxylic acid metabolism	26
ko04910	Insulin signaling pathway	26
ko00040	Pentose and glucuronate interconversions	25
ko00250	Alanine, aspartate and glutamate metabolism	25
ko00562	Inositol phosphate metabolism	25
ko00564	Glycerophospholipid metabolism	25
ko04075	Plant hormone signal transduction	25
ko04110	Cell cycle	25
ko04510	Focal adhesion	25
ko00561	Glycerolipid metabolism	24
ko04016	MAPK signaling pathway – plant	24
ko04260	Cardiac muscle contraction	24
ko00280	Valine, leucine and isoleucine degradation	23
ko00330	Arginine and proline metabolism	23
ko00860	Porphyrin and chlorophyll metabolism	23
ko00940	Phenylpropanoid biosynthesis	22
ko04530	Tight junction	22
ko04919	Thyroid hormone signaling pathway	22
ko00020	Citrate cycle (TCA cycle)	21
ko00030	Pentose phosphate pathway	21
ko00051	Fructose and mannose metabolism	21
ko00380	Tryptophan metabolism	21
ko00480	Glutathione metabolism	21
ko04068	FoxO signaling pathway	21
ko04071	Sphingolipid signaling pathway	21
ko04113	Meiosis – yeast	21
ko04152	AMPK signaling pathway	21

ko04626	Plant-pathogen interaction	21
ko04712	Circadian rhythm – plant	21
ko00900	Terpenoid backbone biosynthesis	20
ko04114	Oocyte meiosis	20
ko04921	Oxytocin signaling pathway	20
ko00680	Methane metabolism	19
ko04010	MAPK signaling pathway	19
ko04014	Ras signaling pathway	19
ko04210	Apoptosis	19
ko04218	Cellular senescence	19
ko04261	Adrenergic signaling in cardiomyocytes	19
ko00071	Fatty acid degradation	18
ko00710	Carbon fixation in photosynthetic organisms	18
ko00770	Pantothenate and CoA biosynthesis	18
ko00983	Drug metabolism – other enzymes	18
ko02010	ABC transporters	18
ko04015	Rap1 signaling pathway	18
ko04922	Glucagon signaling pathway	18
ko00052	Galactose metabolism	17
ko00220	Arginine biosynthesis	17
ko00310	Lysine degradation	17
ko00640	Propanoate metabolism	17
ko00760	Nicotinate and nicotinamide metabolism	17
ko00970	Aminoacyl-tRNA biosynthesis	17
ko04024	cAMP signaling pathway	17
ko04070	Phosphatidylinositol signaling system	17
ko04217	Necroptosis	17
ko04360	Axon guidance	17
ko04371	Apelin signaling pathway	17
ko04666	Fc gamma R-mediated phagocytosis	17
ko04931	Insulin resistance	17
ko00350	Tyrosine metabolism	16
ko00920	Sulfur metabolism	16
ko03320	PPAR signaling pathway	16
ko04066	HIF-1 signaling pathway	16

ko04213	Longevity regulating pathway – multiple species	16
ko04721	Synaptic vesicle cycle	16
ko04722	Neurotrophin signaling pathway	16
ko00410	beta-Alanine metabolism	15
ko00510	N-Glycan biosynthesis	15
ko00790	Folate biosynthesis	15
ko00950	Isoquinoline alkaloid biosynthesis	15
ko04022	cGMP-PKG signaling pathway	15
ko04310	Wnt signaling pathway	15
ko00360	Phenylalanine metabolism	14
ko00513	Various types of N-glycan biosynthesis	14
ko00600	Sphingolipid metabolism	14
ko00650	Butanoate metabolism	14
ko00980	Metabolism of xenobiotics by cytochrome P450	14
ko01040	Biosynthesis of unsaturated fatty acids	14
ko02020	Two-component system	14
ko03020	RNA polymerase	14
ko03420	Nucleotide excision repair	14
ko04020	Calcium signaling pathway	14
ko04728	Dopaminergic synapse	14
ko04914	Progesterone-mediated oocyte maturation	14
ko04972	Pancreatic secretion	14
ko00100	Steroid biosynthesis	13
ko00400	Phenylalanine, tyrosine and tryptophan biosynthesis	13
ko00830	Retinol metabolism	13
ko04214	Apoptosis – fly	13
ko04621	NOD-like receptor signaling pathway	13
ko04670	Leukocyte transendothelial migration	13
ko04912	GnRH signaling pathway	13
ko00053	Ascorbate and aldarate metabolism	12
ko00590	Arachidonic acid metabolism	12
ko00910	Nitrogen metabolism	12
ko03060	Protein export	12
ko04072	Phospholipase D signaling pathway	12
ko04520	Adherens junction	12

ko04612	Antigen processing and presentation	12
ko04915	Estrogen signaling pathway	12
ko04966	Collecting duct acid secretion	12
ko04976	Bile secretion	12
ko00061	Fatty acid biosynthesis	11
ko00062	Fatty acid elongation	11
ko00563	Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	11
ko00720	Carbon fixation pathways in prokaryotes	11
ko01220	Degradation of aromatic compounds	11
ko04062	Chemokine signaling pathway	11
ko04211	Longevity regulating pathway	11
ko04390	Hippo signaling pathway	11
ko04540	Gap junction	11
ko04916	Melanogenesis	11
ko04928	Parathyroid hormone synthesis, secretion and action	11
ko04934	Cushing syndrome	11
ko04975	Fat digestion and absorption	11
ko04979	Cholesterol metabolism	11
ko00130	Ubiquinone and other terpenoid-quinone biosynthesis	10
ko00300	Lysine biosynthesis	10
ko00362	Benzoate degradation	10
ko00450	Selenocompound metabolism	10
ko00460	Cyanoamino acid metabolism	10
ko00670	One carbon pool by folate	10
ko01521	EGFR tyrosine kinase inhibitor resistance	10
ko01523	Antifolate resistance	10
ko02024	Quorum sensing	10
ko03022	Basal transcription factors	10
ko03030	DNA replication	10
ko04115	p53 signaling pathway	10
ko04136	Autophagy – other	10
ko04270	Vascular smooth muscle contraction	10
ko04320	Dorso-ventral axis formation	10
ko04611	Platelet activation	10
ko04660	T cell receptor signaling pathway	10

ko04727	GABAergic synapse	10
ko04918	Thyroid hormone synthesis	10
ko04933	AGE-RAGE signaling pathway in diabetic complications	10
ko00140	Steroid hormone biosynthesis	9
ko00195	Photosynthesis	9
ko00340	Histidine metabolism	9
ko00511	Other glycan degradation	9
ko00592	alpha-Linolenic acid metabolism	9
ko00627	Aminobenzoate degradation	9
ko00740	Riboflavin metabolism	9
ko00941	Flavonoid biosynthesis	9
ko00982	Drug metabolism – cytochrome P450	9
ko01522	Endocrine resistance	9
ko01524	Platinum drug resistance	9
ko04370	VEGF signaling pathway	9
ko04720	Long-term potentiation	9
ko04726	Serotonergic synapse	9
ko04925	Aldosterone synthesis and secretion	9
ko04926	Relaxin signaling pathway	9
ko04974	Protein digestion and absorption	9
ko00073	Cutin, suberine and wax biosynthesis	8
ko00254	Aflatoxin biosynthesis	8
ko00290	Valine, leucine and isoleucine biosynthesis	8
ko00565	Ether lipid metabolism	8
ko00906	Carotenoid biosynthesis	8
ko03410	Base excision repair	8
ko03430	Mismatch repair	8
ko03460	Fanconi anemia pathway	8
ko04012	ErbB signaling pathway	8
ko04216	Ferroptosis	8
ko04350	TGF-beta signaling pathway	8
ko04380	Osteoclast differentiation	8
ko04512	ECM-receptor interaction	8
ko04550	Signaling pathways regulating pluripotency of stem cells	8
ko04623	Cytosolic DNA-sensing pathway	8

ko04724	Glutamatergic synapse	8
ko04920	Adipocytokine signaling pathway	8
ko04962	Vasopressin-regulated water reabsorption	8
ko04971	Gastric acid secretion	8
ko04978	Mineral absorption	8
ko00196	Photosynthesis – antenna proteins	7
ko00591	Linoleic acid metabolism	7
ko00904	Diterpenoid biosynthesis	7
ko04122	Sulfur relay system	7
ko04130	SNARE interactions in vesicular transport	7
ko04625	C-type lectin receptor signaling pathway	7
ko04650	Natural killer cell mediated cytotoxicity	7
ko04657	IL-17 signaling pathway	7
ko04664	Fc epsilon RI signaling pathway	7
ko04668	TNF signaling pathway	7
ko04725	Cholinergic synapse	7
ko04911	Insulin secretion	7
ko04917	Prolactin signaling pathway	7
ko00514	Other types of O-glycan biosynthesis	6
ko00901	Indole alkaloid biosynthesis	6
ko00909	Sesquiterpenoid and triterpenoid biosynthesis	6
ko00960	Tropane, piperidine and pyridine alkaloid biosynthesis	6
ko04340	Hedgehog signaling pathway	6
ko04662	B cell receptor signaling pathway	6
ko04710	Circadian rhythm	6
ko04713	Circadian entrainment	6
ko04730	Long-term depression	6
ko04750	Inflammatory mediator regulation of TRP channels	6
ko04960	Aldosterone-regulated sodium reabsorption	6
ko04964	Proximal tubule bicarbonate reclamation	6
ko04970	Salivary secretion	6
ko00072	Synthesis and degradation of ketone bodies	5
ko00120	Primary bile acid biosynthesis	5
ko00430	Taurine and hypotaurine metabolism	5
ko00643	Styrene degradation	5

ko00730	Thiamine metabolism	5
ko00902	Monoterpenoid biosynthesis	5
ko04330	Notch signaling pathway	5
ko04630	JAK–STAT signaling pathway	5
ko04659	Th17 cell differentiation	5
ko04924	Renin secretion	5
ko04930	Type II diabetes mellitus	5
ko04961	Endocrine and other factor–regulated calcium reabsorption	5
ko04973	Carbohydrate digestion and absorption	5