

アブラナ科植物 Fast Plants における全ゲノム転写解析

坂場美咲・川邊隆大

Genome-wide transcriptome analysis in Rapid cycling *Brassica rapa* (Fast Plants type)

Misaki SAKABA and Takahiro KAWANABE

(Received 6 December 2018; accepted 30 January 2019)

Abstract

The species *Brassica rapa* is an important vegetable and oilseed crop, such as Chinese cabbage (var. *pekinensis*), turnip (var. *rapa*), pak choi (var. *chinensis*), and Komatsuna (var. *perviridis*). Rapid cycling *Brassica rapa*, known as Wisconsin Fast Plants, are widely used organism in science education and genetic research. However there are few studies on the transcriptome analysis in Fast Plants. In this study, we investigated the transcript profiles of five different tissues (bud, leaf, stem, root, seed) in Fast Plants using RNA-seq. 87.7% of the total annotated genes were detected as expressed and 569 genes were constitutively expressed across all tissues. We further defined 1,641, 875, 428 and 965 tissue-specific expressed genes in bud, leaf, stem and root, respectively. A comprehensive expression analysis revealed that many auxin-inducible genes were a tissue-specific manner. Notably, some *SAUR* genes were highly expressed in 10, 13 day after pollination seed, indicating their putative role in seed development. The comprehensive analysis of RNA-seq data would provide a foundation for future comprehensive analysis in *B. rapa*.

緒言

アブラナ科に属する植物には、モデル植物であるシロイヌナズナ (*Arabidopsis thaliana*) や、ハクサイ、キャベツ、ダイコン等の野菜類、植物油の原料となるナタネなどが含まれる。特に、*Brassica* 属作物には多くの主要な野菜が含まれており、ハクサイやカブなどの *Brassica rapa* 種、キャベツやブロッコリーなどの *B. oleracea* 種などがある。これらは、日本の農業生産額において上位になる品目であることから重要な野菜である。*B. rapa* は農業上重要な種であるが、先にも述べたように、ハクサイ、ミズナ、コマツナ、カブなどのように形態学的に多様な変種を含み、さらにはイエローサルソンなどの油糧作物も含むことからゲノム、進化学的研究のモデル植物として非常に重要である。また、*Brassica* 属は、U の三角形として知られているように、二倍体種の *B. rapa* (2n=20, AA), *B. nigra* (2n=16, BB), *B. oleracea* (2n=18,

CC) および異質倍数体の *B. juncea* (AABB), *B. napus* (AACC), *B. carinata* (BBCC) のゲノムの関係性が明らかとなっている (1)。*B. napus* は世界的に重要な油糧作物でもあり、そのゲノム構成種である *B. rapa* は *Brassica* 属の比較研究には欠かせないものである。

2011年に *B. rapa* のゲノム配列が公開され、ゲノムワイドスケールでの解析が可能となった (2)。その後、*B. oleracea* (3)、複二倍体種である *B. napus* (4, 5)、*B. juncea* (6) のゲノム情報も公開されており、*Brassica* 属間のゲノム比較も可能となってきている。RNA-sequencing (RNA-seq) 法は、次世代シーケンサーを用いて mRNA の配列を解読し、転写産物の発現量定量化を行う手法である。配列を読むことから新規転写産物や新規スプライシングバリエーションを網羅的に解析することができ、その発展に伴い、全ゲノム転写解析の手法として近年広く用いられるようになってきている。

Wisconsin Fast Plants として知られている Rapid cycling *Brassica rapa* (RCBr) は、科学教育における植物としてよく用いられている。この植物は、ウィスコンシン大学の Paul Williams 博士によってアブラナ科植物の遺伝学

研究を効率よく行う目的で開発された *B. rapa* の変種であり (7), その生活環はとても短く, 連続した光, 水および栄養が好適な条件下であれば, 種を植えてから約14日で開花し, 約40日で成熟種子を採ることができるという特徴がある.

オーキシンは胚発生や様々な器官の分化や伸長, 非生物のストレスへの応答など, 植物の生活環全体で発生および生長を調節する主要なホルモンである. その調節は, オーキシン応答遺伝子である *auxin/indole-3-acetic acid* (*Aux/IAA*) 遺伝子ファミリー, *auxin-response factor* (*ARF*) 遺伝子ファミリー, *small auxin-up RNAs* (*SAUR*) 遺伝子ファミリー, *gretchen hagen 3* (*GH3*) 遺伝子ファミリーらにより制御されている (8). シロイヌナズナでは *Aux/IAA* 遺伝子29個, *ARF* 遺伝子23個, *SAUR* 遺伝子82個, *GH3* 遺伝子35個が存在しており, ダイズやキュウリ, タバコなどのその他様々な植物種においてもオーキシン応答遺伝子の同定が行われてきている (8). *B. rapa* でもオーキシン応答遺伝子の探索は行われており, その系統の一つである 'Chiifu' では *ARF* 遺伝子ファミリーが31個, *Aux/IAA* 遺伝子ファミリーが55個同定されている (9, 10).

本研究では, Fast Plants の各組織の転写産物の発現定量化を目的として, 5つの組織 (蕾, 葉, 茎, 根および種子) で RNA-seq を行った. また, オーキシン応答遺伝子である *ARF*, *Aux/IAA*, *SAUR*, *GH3* の各遺伝子ファミリーについてそれぞれの組織における遺伝子発現について解析を行った.

材料および方法

植物材料

Brassica rapa Fast Plants は25℃の温室で育成し, 明期16時間・暗期8時間で育成した. Fast Plants における蕾 (2サンプル; FP-bud), 葉 (2サンプル; FP-leaf), 茎 (2サンプル; FP-stem), 根 (FP-root) および交配10 (FP-10-seed), 13 (FP-13-seed), 17日後 (FP-17-seed) の種子を含む5つの組織から RNeasy Plant Mini kit (Qiagen, USA) を用いて total RNA を抽出した. RNA のクオリティは2100 bioanalyzer (Agilent Technologies, USA) を用いて確認し, 高品質であった RNA は NeoPrep ライブラリー調整システム (illumina, USA) により cDNA ライブラリーを作成した. 作成した10個のライブラリーは NextSeq 500 (illumina, USA) を用いてシーケンズを行い, 75bp シングルエンドの配列を取得した.

RNA-seq データの解析

取得した RNA-seq のデータ配列は, FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) を用いてクオリティチェックを行った. トリミングには, Trim galore! v0.4.5 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) を用いて低クオリティのリードを除去した. リードの5'末端から14bpの除去, クオリティまたはアダプターのトリミングのため33bpよりも短いリードの除去, および配列のトリミングを必要とするアダプター配列のオーバーラップを3に設定して行った. その後, hisat2 v2.1.0 (11) を用いて, BrapaFPsc_277_v1.3にマッピングを行った. レファレンスゲノムとして用いた BrapaFPsc_277_v1.3及び, そのアノテーション情報は phytozome (<https://phytozome.jgi.doe.gov/>) より入手した. FPKM (Fragment Per Kilo-base of exon per Million mapped fragments) 値は, cufflinks v2.2.1 (12) を用いて算出した.

発現の組織幅

組織特異的発現遺伝子の同定は Yanai らの方法により行った (13).

$$Ti = \frac{\sum_{j=1}^n (1-S(i, j) / S(i, \max))}{n-1}$$

式における指数 T は組織特異性を表し, n は, 調査対象である *B. rapa* 組織の数を示し, S (i, j) は, j 組織における i 遺伝子の FPKM 値, S (i, max) は, n 組織における遺伝子 i の FPKM の最高値を示している. 各組織において, 対象組織が遺伝子 i における FPKM の最大値であること, 対象組織の FPKM が5以上, Ti が0.99以上, そして, 他の組織の1-S (i, j) が0.99以上であるものを組織特異的に発現している遺伝子であると定義した.

ヒートマップの作製

ヒートマップは, R studio (<https://www.rstudio.com/>) の genefilter (<https://bioconductor.org/packages/release/bioc/html/genefilter.html>) および ggplot2 (<https://bioconductor.org/packages/release/bioc/html/ggplot2.html>) のパッケージを用い, FPKM 値を Z-score 化して作成した.

結果および考察

Fast Plants の5つの組織 (蕾, 葉, 茎, 根および種子) から10個の cDNA ライブラリーを作成し, 75bp 長のシングルエンドリードを取得した. 低品質のリードを

Table.1 Statistic of mapping reads

Sample name	Raw reads	Filtering reads	Alignment rate
FP-bud-1	47,552,981	47,515,481	95.71%
FP-bud-2	52,182,362	53,455,898	95.72%
FP-leaf-1	44,070,669	43,751,839	96.05%
FP-leaf-2	43,162,073	42,776,871	96.24%
FP-stem-1	53,115,878	53,058,521	96.05%
FP-stem-2	36,951,002	36,911,186	96.52%
FP-root	43,503,849	43,435,063	96.21%
FP-10DAP-seed	56,595,184	55,472,658	95.53%
FP-13DAP-seed	50,284,931	49,211,737	95.13%
FP-17DAP-seed	39,933,743	39,414,928	93.59%

除去したところ、高品質 ($Q>25$) の4.6億リードが得られ、それらをその後の解析に用いた。公開されているFast Plantsのゲノム配列にマッピングを行なったところ、それぞれのライブラリーは約95%のマッピング率であった (Table. 1)。

Fast Plantsの蕾、葉、茎、根および種子の発現データをもとに、転写産物の発現はFPKM値を用いて定量化し、各組織における発現レベルに関して比較した。同じ組織が同じクラスターに存在し、組織ごとにクラスタリングされていることから、各組織において遺伝子発現が異なっていることが示された (Fig. 1)。35,528個の遺伝子 (全40,492個のアノテーションされている遺伝子の87.7%) は、少なくとも一つの組織において発現が確認され、1,216個の遺伝子は全ての組織において発現が確認された (Fig. 2)。また、交配10、13日後の種子では他

の組織に比べ、多くの遺伝子が発現していた (Fig. 2)。全ての組織において遺伝子発現が見られた1,216個の遺

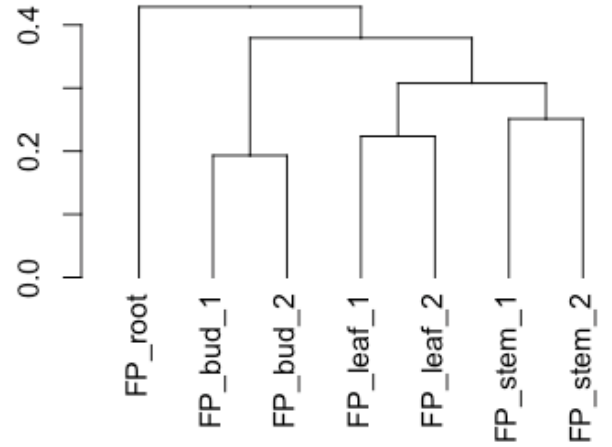


Fig.1. Dendrogram of four tissue transcriptomes based on clustering of log2-transformed FPKM values of constitutive expressed genes.

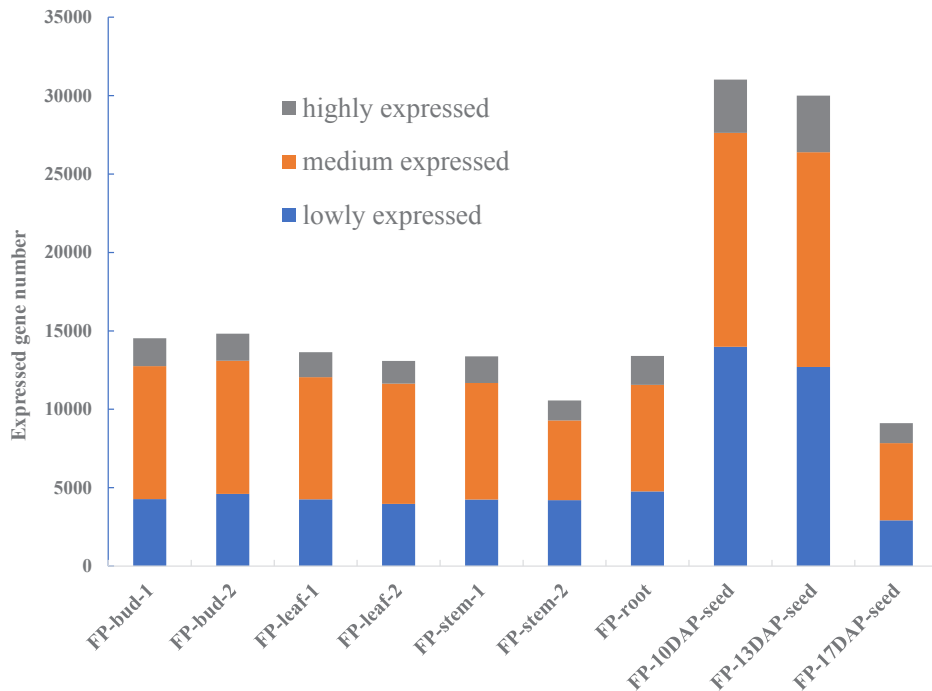
Fig.2. The number of highly (FPKM > 50), medium ($5 < \text{FPKM} \leq 50$), and lowly ($\text{FPKM} \leq 5$) expressed genes in each tissues.

Table. 2 The constitutively expressed gene whose FPKM value were greater than 5 in each tissues

Gene ID	FPKM							<i>A.thaliana</i> ID	<i>Arabidopsis</i> gene name	Gene description
	bud	leaf	stem	root	FP10	FP13	FP17			
Brara.A00999	34.61	20.32	33.38	36.09	41.61	37.51	43.69	AT4G17920		RING/U-box superfamily protein
Brara.A01511	1014.40	1203.17	1334.75	330.75	68.22	289.95	1277.80	AT4G23800		HMG (high mobility group) box protein
Brara.A02070	55.07	23.80	49.42	34.84	48.39	25.03	22.51	AT1G66910		protein kinase superfamily protein
Brara.B01172	44.83	41.07	49.83	56.73	68.75	70.56	66.82	AT5G18450	<i>XTH32</i>	integrase-type DNA-binding superfamily protein
Brara.C02570	37.21	23.00	21.85	29.37	41.36	32.48	26.28	AT2G36870	<i>HTA8</i>	xyloglucan endotransglucosylase/hydrolase 32
Brara.C02712	56.62	25.54	25.64	29.08	30.80	28.70	22.99	AT2G38810		histone H2A 8
Brara.C03057	21.84	22.99	20.95	14.64	29.69	24.52	17.66	AT2G45220		plant invertase/pectin methylesterase inhibitor superfamily
Brara.E01410	17.72	20.03	23.41	17.36	27.74	18.86	15.73	AT2G42350	<i>ACT11</i>	RING/U-box superfamily protein
Brara.F00547	11.65	11.75	9.86	15.50	14.24	13.90	10.73	AT3G12110		actin-11
Brara.F02033	38.44	40.63	22.36	43.44	33.28	24.78	28.21	AT1G09440	<i>ADH1</i>	protein kinase superfamily protein
Brara.H01691	63.06	32.91	15.55	63.82	33.09	41.86	7.45	AT1G77120		alcohol dehydrogenase 1
Brara.K00094	52.27	39.62	27.79	53.77	39.40	29.93	38.86	AT1G05600	<i>MAPKKK13</i>	tetratricopeptide repeat (TPR)-like superfamily protein
Brara.K01036	2257.47	1202.22	1414.23	2520.25	3027.83	4135.06	2213.38	AT1G07150		mitogen-activated protein kinase kinase kinase 13
Brara.K01475	911.66	429.65	585.14	816.81	1037.30	1948.50	2143.62	AT5G19390		rho GTPase activation protein (RhoGAP) with PH domain
Brara.K01516	62.96	64.92	90.43	49.36	67.30	50.48	58.19	AT5G18950		tetratricopeptide repeat (TPR)-like superfamily protein

伝子から FPKM が5以上のものを選抜したところ、非常に安定して発現している遺伝子が569個得られた。これらの遺伝子は、*protein kinase protein*, *tetratricopeptide repeat protein*, *RING/U-box protein* や *integrase-type DNA binding protein* などをコードする遺伝子が多く含まれており、ハウスキーピング遺伝子であると考えられることから、RT-qPCR などのコントロールとして用いることができ、非常に有用な情報であると考えられる (Table.2)。

蕾、葉、茎および根のそれぞれ4つの組織を比較し、特異的な発現を示す遺伝子群の同定を行なった。蕾では、1,641個の遺伝子が特異的な発現を示し、*plant invertase/pectin methylesterase inhibitor*, *RING/U-box protein*, *WRKY DNA-binding protein*, *serine carboxypeptidase-like*, *ribosomal protein*, *protein phosphatase 2C protein*, *actin12* および *protein kinase protein* などをコードする遺伝子群が含まれていた。葉特異的な発現遺伝子は、875個の遺伝子が確認され、*pentatricopeptide repeat protein*, *protein kinase protein*, *NAD(P)-binding rosmann-fold protein*, *alpha/beta-hydrolases protein* および *major facilitator protein* などをコードする遺伝子群が含まれていた。茎特異的な発現遺伝子は、428個の遺伝子が特異的であり、*basic helix-loop-helix (bHLH) DNA-binding protein*, *p-loop containing nucleoside triphosphate hydrolases protein*, *leucine-rich repeat protein kinase protein* および *protein kinase protein* などが含まれていた。根特異的な発現遺伝子に関しては、965個の遺伝子が特異的であり、*major facilitator protein*, *pectin lyase-like protein*, *zinc finger protein*, *ribosomal protein* および *leucine-rich repeat protein kinase protein* などをコードする遺伝子群が含まれていた。

種子において、各サンプルを他の4つの組織と比較し、

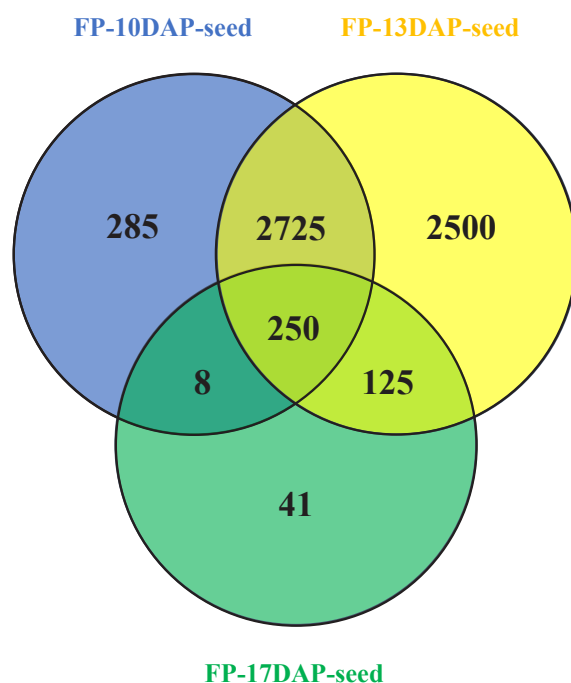


Fig. 3. Venn diagram of showing the overlap of specific genes at 10, 13, and 17 day after pollinated (DAP) seed.

それぞれにおいて特異的な発現を示す遺伝子の探索を行った。その結果、交配10日後で3,268個、交配13日後で5,600個、交配17日後で424個の特異的な発現を示す遺伝子がえられた。このうち250個の遺伝子は全てのステージにおいて発現がみられた (Fig. 3)。交配10日後においては、*pentatricopeptide repeat protein* や *major facilitator with SPX (SYG1/Pho81/XPR1) domain-containing protein*, *PPR protein*, *SCR-like 1* をコードする遺伝子が確認された。交配13日後においては、*galactose oxidase/kelch repeat protein* や、*hydroxyproline-rich glycoprotein protein* をコードする遺伝子が特異的に発現していた。交配17日後の種

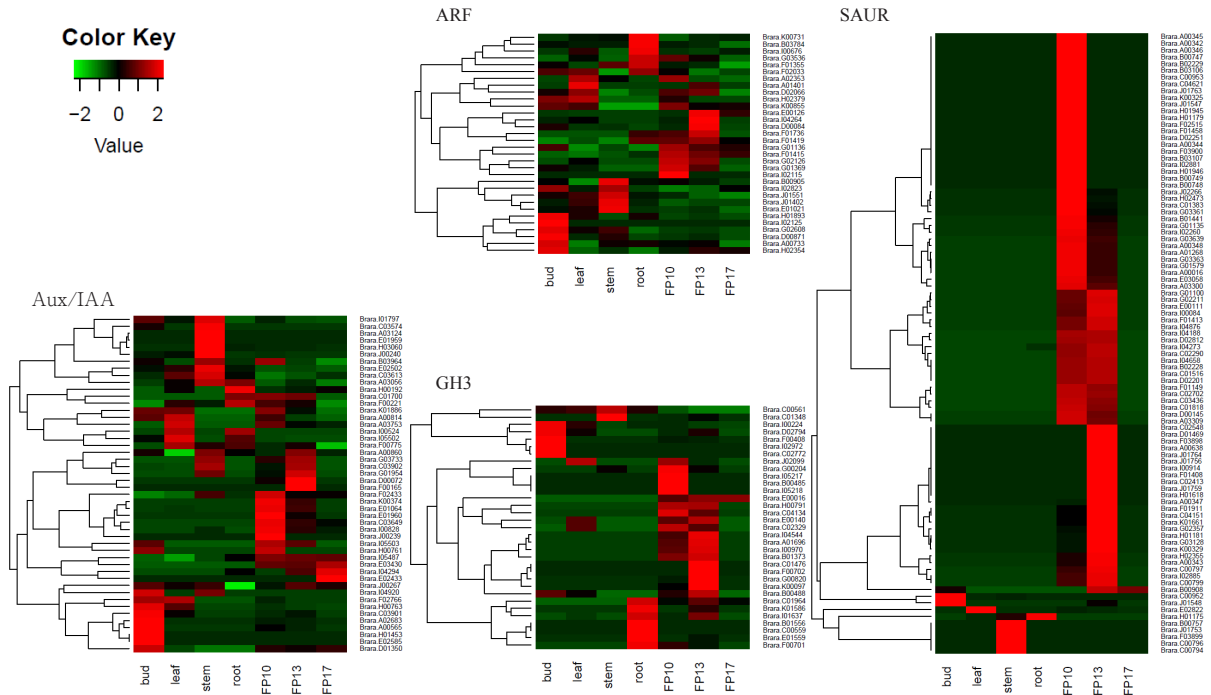


Fig. 4. Expression profiles of the auxin-responsive genes. Heatmap showing expression profiles of auxin-responsive genes based on RNA-seq data in various tissues. Values represent the Z-score of FPKM.

子特異的遺伝子発現に関しては, *cysteine proteinases protein*, *dihydroneopterin aldolase* および *K-box region and MADS-box transcription factor protein* などが発現していた。交配10, 13日後の種子は交配17日後に比べ, 遺伝子発現数も多く, 他の組織に比べて特異的発現する遺伝子数も多かった。これは, 交配10日後から13日後は Fast Plants では胚乳が発達している時期であり, 交配17日後は種子の成熟期になり胚乳が消失している時期である。このことから胚乳特異的発現遺伝子の存在が, この遺伝子数の違いに影響していることが考えられた。

オーキシンは胚発生期において重要な役割を持つ植物ホルモンである。Fast Plants におけるオーキシン応答遺伝子の発現様式を明らかにするため, 公開されている Fast Plants のアノテーション情報からオーキシン応答遺伝子を検索した。その結果, 35個の *ARF* 遺伝子, 38個の *Aux/IAA* 遺伝子, 137個の *SAUR* 遺伝子, 37個の *GH3* 遺伝子が見出された。'Chiifu' 系統では *ARF* 遺伝子が31個, *Aux/IAA* 遺伝子が55個同定されているが (9, 10), Fast Plants の遺伝子数とは異なっていることから, 系統によるオーキシン応答遺伝子の応答性の違いが系統の表現型にも影響している可能性が考えられた。同定した遺伝子発現を調べたところ, 7個の *ARF*, 6個の *Aux/IAA*, 26個の *SAUR*, 4個の *GH3* 遺伝子では全ての組織において発現がみられなかった。さらに, 各組織における発現量を比較した結果 (Fig. 4), オーキシン応答遺伝子は,

組織特異的な発現を示すものが多くみられた。 *SAUR* 遺伝子の突然変異体では, 組織特異性を示す遺伝子が多くみられたが, 交配10, 13日後の種子において発現している遺伝子が特に多いことが明らかとなった。 *SAUR* の機能については未解明なことが多いが, 近年の分子生物学的解析, 遺伝学的解析, 生化学的解析により植物の成長発達に関わっていることが示されてきている (14)。シロイヌナズナの *SAUR36* 遺伝子突然変異体は, 種子発芽に関わることが示されているが (15), 種子発達時期に働く他の *SAUR* 遺伝子の機能については未解明のままであることから, さらなる解析が必要であると考えられる。

本研究室では, *B. rapa* を用いて発達中の種子の組織である胚乳で発現するゲノムインプリント遺伝子の同定を行っている。ゲノムインプリンティングとは, 2つの対立遺伝子が両親のどちらから由来するかによって遺伝子の発現のオン・オフを決定する機構であり, ゲノムインプリント遺伝子は, 父親由来または, 母親由来の対立遺伝子のみを発現する遺伝子のことである。植物では, ゲノムインプリンティングは胚乳組織でおこなうことが知られている (16)。本研究で得られたデータセットは, 様々な植物組織での発現動向をみており, ゲノムインプリント遺伝子の発現様式解析を行う際, 有用な情報となるであろう。

謝辞

本研究を実施するにあたり、貴重なご意見や様々な支援を頂きました京都産業大学総合生命科学部生命資源環境学科の河邊昭准教授、木村成介教授に感謝いたします。また、次世代シーケンサーによる塩基配列解析を行った機関である京都産業大学の降旗初佳氏に感謝します。

要約

Brassica rapa は、葉菜類のハクサイやコマツナ、根菜類であるカブ、油糧作物であるアブラナなどを含み、世界中で栽培・消費されており農業上重要な野菜のひとつである。*B. rapa* は本来、自家不和合性を持ち遺伝学的解析を行うには困難が伴うが、Wisconsin Fast Plants として知られる Rapid cycling *Brassica rapa* は、自家和合性であり、その上、生育期間が約40日短く、研究室内で生育させるのに適した性質をもつ。本研究では、Fast Plants の5つの組織（蕾、葉、莖、根および交配10、13、17日後の種子）においてRNA-seq解析を行い、組織特異的な遺伝子およびオーキシン関連遺伝子の探索を行った。アノテーションされている遺伝子のうち87.7%の遺伝子が発現しており、569個の遺伝子が全ての組織で恒常的に発現していることが明らかとなった。

次に、組織特異的な遺伝子の同定を行ったところ、蕾、葉、莖、根でそれぞれ1,641個、875個、428個、965個の組織特異的な発現する遺伝子を同定することができた。オーキシン応答遺伝子の発現様式について解析したところ、組織特異的な発現を示す遺伝子が多く見られたが、*SAUR* 遺伝子は交配10、13日後の種子において組織特異的な発現を示すものが多くみられた。本研究で得られたデータセットは、Fast Plants において遺伝子発現様式解析を行う際、有用な情報となるであろう。

引用文献

- 1) Nagaharu U. 1935 Jpn J Bot. 7, 389-452.
- 2) Wang X., Wang H., Wang J., Sun R., Wu J., *et al.* 2011 Nat. Genet. 43, 1035-1039.
- 3) Liu S, Liu Y, Yang X., Tong C., Edwards D., Parkin I.A., *et al.* 2014, Nat. Commun. 5, 3930.
- 4) Chalhoub B., Denoeud F., Liu S., Parkin I.A., Tang H., *et al.* 2014, Science 345, 950-953.
- 5) Bayer P.E., Hurgobin B., Golicz A.A., Chan C.K., Yuan Y., *et al.* 2017, Plant Biotechnol J. 15, 1602-1610.
- 6) Yang J., Liu D., Wang X., Ji C., Cheng F., *et al.* 2016, Nat. Genet. 48, 1225-1232
- 7) Lauffer H.B., Lauffer D. Williams P. 2012, Wisconsin Fast Plants Program
- 8) Chapman E.J. and Estelle M. 2009, Annu. Rev. Genet. 43:265-285.
- 9) Mun J.H., Yu H.J., Shin J.Y., Oh M., Hwang H.J. 2012, Mol. Genet. Genomics. 287, 765-784.
- 10) Paul P., Dhandapani V., Rameneni J.J., Li X., Sivanandhan G., *et al.* 2016, PLoS one 11, e0151522.
- 11) Pertea M., Kim D., Pertea G.M., Leek J.T., Salzberg S.L., *et al.* 2016, Nat. Protoc. 11, 1650-1667.
- 12) Trapnell C., Roberts A., Goff L., Pertea G., Kim D., *et al.* 2012, Nat. Protoc. 7, 562-578.
- 13) Yanai I., Benjamin H., Shmoish M., Chalifa-Caspi V., Shklar M., *et al.* 2005, Bioinformatics. 21, 650-659.
- 14) Ren H. and Gray W.M. 2015, Mol Plant. 8, 1153-1164.
- 15) Narsai R., Law S.R., Carrie C., Xu L., Whelan J. 2011, Plant Physiol. 157, 1342-1362.
- 16) Gehring M. 2013, Annu. Rev. Genet. 47, 187-208.