

## 薬用植物培養物の有用物質生産に関わる遺伝子群の網羅的解析

原光二郎, 川上寛子

秋田県立大学大学院生物資源科学研究科

リンドウ科のセンブリ (*Swertia japonica*) の培養系を材料として、NGS 技術を活用して二次代謝物質の生産に関連する遺伝子の発現を網羅的に解析することを目的とした。同時に、データの品質管理など、バイオインフォマティクスの手法を改善し、学内で情報共有を図った。センブリカルスより RNA を抽出し、ライブラリーを作製後、片側 100 bp のペアエンドの塩基配列を決定した。3 種類の異なるプログラムを用いて低品質のリードを除去し、Trinity プログラムにより *de novo* アセンブルを行った。その結果、約 20 万本のコンティグが得られ、およそ 63% に相同性がみられなかった。二次代謝が活性化される発根処理によって 200 倍以上の高発現を示した遺伝子が 24 個見つかったが、二次代謝には無関係と思われた。アノテーションの結果、393 本のシトクロム p450、142 本の WRKY 転写調節因子をコードすると考えられるコンティグが見つかり、そのうちの 3 本の p450 と 1 本の WRKY を示すコンティグが発根処理に伴って高発現しており、その相同性より、二次代謝への関与が示唆された。

**キーワード:** RNA-seq, センブリ, 二次代謝, NGS, シトクロム p450

薬用植物は医薬品・化粧品の開発において重要な生物資源であるが、長期間の栽培が必要な場合が多く、天候不順・異常気象による生産・品質面でのリスクを受けやすい。さらに近年では、国際的な需要の高まりによる野生種の乱獲や供給体制の未整備などのグローバルな影響も受けるようになってきた。一方、植物細胞培養系を用いた有用物質生産の場合、均一な細胞を無菌的かつ最適化された環境におくため、生産性・品質・安全性を高めることが可能となる。植物細胞培養を用いる際の問題点としては、最適な培養環境や生産条件を見いだす為に多くの試行錯誤が必要であること、培養中に遺伝的な変異が生じて生産性が不安定になる可能性があることが挙げられる。多くの場合、有用物質の最適生産条件は植物種に依存し、また、非生産性・低生産性の原因を一般化することもできないのが現状である。しかし、近年の次世代シーケンシング (Next-Generation

Sequencing, 以下 NGS) 技術の発展により、遺伝子レベルの変化を網羅的に解析することが可能になってきたことから、有用物質の生合成に直接関わる酵素遺伝子群の情報が得られやすくなった。さらに、生合成に関わる直接的な因子だけでなく、前駆体生産・分解などの生産、輸送・蓄積などの流通、転写調節・フィードバックなどの制御といった、間接的因子を含めた物質生産性の細胞内での全体像を捉えられることが可能になりつつある。本研究では、古来より民間薬として利用されてきたリンドウ科のセンブリ (*Swertia japonica*) の培養系を材料として、その含有成分であるスウェルチアマリン (図 1) などのセコイリドイド類およびキサントン類の生産を指標として、NGS 技術を活用して物質生産に関連する遺伝子の発現を網羅的に解析する (*de novo* RNA-seq 解析) ことを目的とした。これまでに、物質生産性の異なるカルス・不定根で発現している遺

伝子のカタログ化がすでに終了しており（川上, 2015）、本研究では、発根促進物質の影響とともに、NGS データの品質管理やアセンブル条件の最適化など、学内での NGS 利用の普及に向けたバイオインフォマティクス的手法に検討した。



図 1 センブリのカルス（写真左）と不定根（写真右）。スウェルチアマリン（右図）は不定根で産生される。

## 材料と方法

### 培養

秋田県産のセンブリの種子を無菌播種して発芽させ、得られた芽生え由来のカルスを 3% sucrose, 1% agar, 10  $\mu$ M naphthylacetic acid (NAA) と 1  $\mu$ M kinetin (KIN) を含む Woody Plant 基本培地 (WP, Lloyd & McCown, 1981) で継代培養した。不定根の発根作用が知られる制御物質を添加したカルスを NGS 解析用のサンプルとした。発根処理として制御物質の入った溶液にカルスを浸し、固体培地上で 10 日間培養後、3% sucrose, 0.01  $\mu$ M NAA, 0.01  $\mu$ M KIN を含む 1/8 WP 液体培地に植え換え、110 rpm, 25  $^{\circ}$ C, 明条件下で 10 日間振盪培養した。

### RNA 抽出と NGS サンプルの調製, NGS 解析

凍結乾燥後のサンプルをマルチビーズショッカー（安井器械）で破砕し、RNeasy Plant Mini Kit (QIAGEN) を用いて DNase I 処理と全 RNA 抽出を行った。1  $\mu$ g の全 RNA から TruSeq RNA Sample Prep Kit ver.2 (Illumina) を用いて NGS 解析用のライブラリーを調製した（岡野と小黒-岡野, 2013）。核酸の精製には AMPure XP (Beckman Coulter) を、逆転写反応には PrimeScript Reverse Transcriptase (TaKaRa) をそれぞれ用いた。ライブラリーは Bioanalyzer (Agilent

Technologies) および KAPA Library Quantification Kit（日本ジェネティクス）で定量し、4 nM に希釈した後、最終的に 5 pM または 10 pM のサンプルとした。本学バイオテクノロジーセンターの HiSeq 1000 (Illumina) を用いて、次世代シーケンス解析（片側 100 bp のペアエンド）を行った。

### 計算機環境

得られたリードの解析は、本学バイオテクノロジーセンターに設置の Linux ワークステーション（CentOS 6.5, Intel Xeon E5-2670v2 [2.50 GHz / 20 cores], メモリ 512 GB）、または Mac Pro (Apple Inc., Intel Xeon E5 [3.50 GHz / 6 cores], メモリ 64 GB) を組み合わせて行った。各種プログラムは個別にインストールしたほか、パッケージ管理システム（Homebrew, <http://brew.sh>) も利用した。

### de novo アセンブリ

まず、リードの高品質化には、FASTX-Toolkit (ver. 0.0.14, fastq\_quality\_filter), Prinseq (ver. 0.20.4), Trimmomatic (ver. 0.32) の 3 つのプログラムを用い、FastQC (ver. 0.11.2) で品質を確認した。次に、*de novo* アセンブリを Trinity (ver. 2.0.6, 最大メモリを 496 GB, 使用 CPU 数を 20 個, 最小コンティグ長を 200 bp) を用いて行った。重複したコンティグを除くために CD-HIT プログラム (ver. 4.6, cd-hit-est, 相同性が 95% 未満を除去) で処理し、300 bp 未満の短いコンティグを除いたものを最終的なコンティグとした。各種プログラムは、あらかじめ環境設定やパラメータを記載したシェルスクリプトを作成し、半自動的に解析した。これらのスクリプト類は、学内サーバで公開している ([www.lichen.akita-pu.ac.jp/~hara](http://www.lichen.akita-pu.ac.jp/~hara))。

### アノテーション

各コンティグのアノテーション（特徴付け）および遺伝子オントロジー (gene ontology, GO) 解析は、Blast2GO (ver. 3.0, BioBam 社) および PC にインストールした NCBI BLAST (ver. 2.2.30+) を用いて行った。Blast2GO では、シロイヌナズナのタンパク質データベースに対する blastx ( $1^{-10} < e\text{-value}$ ) の結果を用いた。

## 発現変動遺伝子群 (DEGs) の解析

各コンティグへのリードのマッピングには bowtie プログラム (ver. 1.1.0) を、リードカウントには RSEM プログラム (ver. 1.2.19) を、発現変動遺伝子 (differentially expressed gene, DEG) の検出には edgeR プログラムをそれぞれ用いた。

## 結果と考察

### リードの品質管理

全サンプル合わせて 86,274,280 ペアのリードが得られた。リードの品質が重要であることより、今回は FASTX-Toolkit, Prinseq, Trimmomatic の 3 種類の品質管理プログラムで処理した。その結果、75,705,145 ペア (除去率 12.25%) の高品質リードが得られ、以降の解析に用いた。

### de novo アセンブル

材料のセンブリは参照ゲノム配列が存在しないため、de novo アセンブルプログラムの Trinity を使用した。このプログラムは、多くのメモリを必要とするものの、本学の Linux ワークステーションにより比較的短時間 (約 31 時間) で解析できることが分かった。

de novo アセンブリの結果、およそ 20 万本のコンティグ (転写産物) が得られた (表 1)。近縁の *Gentiana macrophylla* ではコンティグが約 43000 本と報告されていること (Hua, 2014) と、コンティグの平均長・N50 が短いことより、比較的短い配列が多いことが予想され、今後、アセンブルプログラムの選択やパラメータの最適化が必要と考えられた。

表 1 アセンブルのまとめ

項目	値
コンティグ (遺伝子) 数	136,058
コンティグ (転写産物) 数	197,122
平均長 (bp)	1,048.28
最大長 (bp)	16,220
N50 (bp)	1,637
GC含量 (%)	39.27

## アノテーション

得られたコンティグ (転写産物) を、モデル植物のシロイヌナズナのタンパク質データに対して blastx 検索を行った結果、およそ 63%に相同性がみられなかった。二次代謝のようなその植物固有の機能を解析するには、多様な生物の遺伝子との比較が必要である。しかし、全生物のタンパク質データ (nr) を用いる場合、時間がかかりすぎるため、今後は、アノテーション付きのデータベース (UniProt など) を一次解析として用いるなど、工夫が必要である。

## 発現変動遺伝子群 (DEGs) の解析

発根処理・未処理の 5 サンプル間で、各コンティグの発現比を調べた結果、それぞれ 200 倍以上の発現比を示した発現変動遺伝子 (DEGs) が 44 個見つかった。このうち、発根処理により誘導された DEGs が 24 個、抑制された DEG が 20 個であった (図 2)。誘導された DEGs は、アミノ酸代謝や複製・分裂に関わる遺伝子が多く含まれており、二次代謝に関わる遺伝子は見つからなかった。

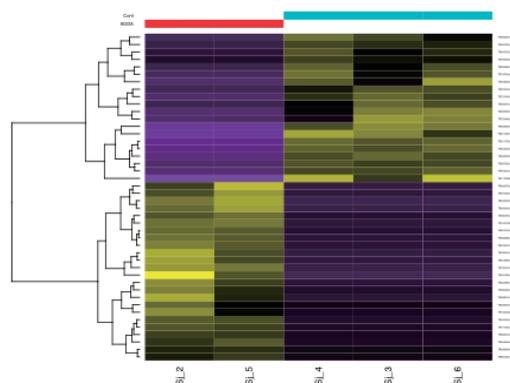


図 2 DEGs のクラスター分析。発根処理 (赤)・未処理 (青) 間で、発現が誘導 (黄) または抑制 (紫) された遺伝子を発現様式で分類した。

## 二次代謝関連遺伝子の解析

これまでの解析 (川上, 2015) により、シトクロム p450 と WRKY 転写調節因子が二次代謝と関連遺伝子の転写制御にそれぞれ関わる事が予想されている。これらの遺伝子を探索した結果、393 本の p450, 142 本の WRKY にそれぞれ相同性のあるコンティグが見つかった。シロイヌナズナのゲノム中には 244

個の p450 遺伝子が存在することが報告されている (Bak *et al.*, 2011). センブリにはより多くの p450 遺伝子が存在することが予想されるものの、今回得られたコンティグの多くは選択的スプライシングによる類似配列と考えられるため、最終的な遺伝子数を比較するためにはゲノム自体の解析が必要である。

シトクロム p450 と WRKY 転写調節因子をコードすると予想されたコンティグのうち、発根処理により高発現したものが合わせて 4 本見つかった(表 2)。p450 とアノテーション付けられた 3 本のコンティグは、いずれも CYP72A サブファミリーに属する p450 タンパク質をコードすると予想された。Catharanthus roseus の CYP72A1 は、セコイリドイド類の代謝経路のうち、ロガニンからセコロガニンへの開環反応を触媒することが報告されており (Irmeler *et al.*, 2000), この遺伝子がセンブリのセコイリドイド生合成に関わることが示唆された。

表 2 発根処理で高発現した DEGs

Contig	Fold change	Length (bp)	Type
cytochrome p450			
TR64154 c0_g1.i1	139.6	890	CYP72A
TR97823 c1_g1.i5	46.3	1,097	CYP72A
TR97823 c1_g1.i16	43.7	1,828	CYP72A
WRKY transcription factor			
TR79104 c0_g1.i2	27.1	2,012	WRKY33

一方、高発現した WRKY は一つだけで、シロイヌナズナの WRKY33 と相同性 (65%) があつた。この WRKY33 は病原菌感染時の防御応答に際し、ファイトアレキシンの生合成に関わる p450 などの遺伝子を制御していることが報告されている (Birkenbihl *et al.*, 2011) ことから、センブリにおいても二次代謝産物の生合成酵素遺伝子の制御に関わっていることが示唆された。今後、これらの候補遺伝子の発現様式を別の方法でも明らかにすることで、二次代謝への関与をより明確にすることができる。

センブリ培養物が産するキサントン類はシキミ酸由来と考えられ、骨格形成や配糖化に関わる遺伝子は不明である。Blast2GO プログラムの KEGG パスウェイ解析ツールの結果より、フラボノイドの生合成

に関わると予想されるコンティグは 66 本、モノテルペノイドは 45 本がそれぞれ関わると予想された。今後、ある代謝経路の特有の遺伝子の探索だけでなく、代謝経路全体の遺伝子発現の動きを捉えることが有用物質生産のために重要であり、このようなパスウェイ解析が非常に有効な手段となる。

## 文献

- Bak, S., Beisson, F., Bishop, G., Hamberger, B., Höfer, R., Paquette, S., & Werck-Reichhart, D. (2011). Cytochromes p450. *The Arabidopsis Book / American Society of Plant Biologists*, 9, e0144.
- Birkenbihl, R. P., Diezel, C., & Somssich, I. E. (2012). Arabidopsis WRKY33 Is a Key Transcriptional Regulator of Hormonal and Metabolic Responses toward *Botrytis cinerea* Infection. *Plant Physiology*, 159(1), 266-285
- Hua, W., Zheng, P., He, Y., Cui, L., Kong, W., & Wang, Z. (2014). An insight into the genes involved in secoiridoid biosynthesis in *Gentiana macrophylla* by RNA-seq. *Molecular Biology Reports*, 41(7), 4817-4825
- Irmeler, S., Schroder, G., St-Pierre, B., Crouch, N. P., Hotze, M., Schmidt, J., Strack, D., Matern, U., Schroder, J. (2000). Indole alkaloid biosynthesis in *Catharanthus roseus*: new enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. *The Plant Journal*, 24(6), 797-804
- 川上寛子 (2015). 「センブリの組織培養と二次代謝物質産生に関する研究」『秋田県立大学大学院博士論文』
- Lloyd G, & McCown B (1981). Commercially-feasible micropropagation of mountain laurel, *Kalmia latifolia*, by use of shoot-tip culture. *International Plant Propagators' Society*, 30, 421-427
- 岡野桂樹, 小黒-岡野美枝子 (2013). 「イルミナ TruSeq RNA Preparation Kit を用いた RNA-Seq 用サンプル調製法の検討」『ヤマザキ学園大学雑誌』 3, 49-67

〔 平成 27 年 6 月 30 日 受付  
平成 27 年 7 月 31 日 受理 〕

## RNA-seq Analyses for Production of Secondary Metabolites in Medicinal Plant Cultures

Kojiro Hara, Hiroko Kawakami

*Graduate School of Bioresource Sciences, Akita Prefectural University*

RNA-seq analyses were performed to elucidate the regulation of secondary metabolism-related genes in culture of *Swertia japonica* (common name: Senburi), which is used as traditional folk medicines in Japan. RNA was extracted from calli treated with or without a rooting agent and subjected to cDNA library synthesis and next-generation sequencing. High-quality paired-end reads processed with three programs were used for *de novo* assembly with the Trinity program, which generated approximately 200,000 contigs. Of these, about 63% were unannotated. Under the treated condition, differentially expressed transcripts with high fold-change values of >200 were identified. Twenty-four transcripts were found to be upregulated; however, these transcripts were not associated with secondary metabolism. Among 393 and 142 contigs annotated as the cytochrome p450 and WRKY transcription factor, three p450 and one WRKY transcripts were upregulated, respectively. These contigs were assigned to CYP72A or WRKY33, suggesting that these play important roles in secondary metabolism.

**Keywords:** RNA-seq, *Swertia japonica*, secondary metabolism, NGS, cytochrome p450