

## 大豆のサポニン含量を制御する転写制御因子の探索

關 光\*・北村 実紗子・世戸口 貴宏・鈴木 隼人

大阪大学大学院工学研究科生物工学専攻

### Gene Mining for Candidate Transcription Factors that Regulate Soyasaponin Content in Soybean

Hikaru SEKI\*, Misako KITAMURA, Takahiro SETOGUCHI and Hayato SUZUKI

Division of Advanced Science and Biotechnology, Graduate School of Engineering, Osaka University, Osaka 565-0871

#### ABSTRACT

Many oleanane-type triterpenoid saponins, called soyasaponins, have been isolated and characterized from soybean seeds. Recent studies have identified various enzymes, including cytochrome P450 monooxygenases and UDP-dependent glycosyltransferases, required for the biosynthesis of soyasaponins in soybean. However, the regulatory mechanisms that control their expression remain unknown. Previously, we identified a transcription factor (GubHLH3), belonging to subclade IVa of basic helix-loop-helix (bHLH) proteins, that directs soyasaponin biosynthesis in the medicinal legume *Glycyrrhiza uralensis* (licorice). This transcription factor could help us understand the conserved mechanisms regulating saponin biosynthesis in legumes. To test the hypothesis that bHLH-mediated signaling is conserved in legumes, we investigated the function of GubHLH3 as a regulator of saponin biosynthesis in soybean. We generated transgenic soybean hairy roots overexpressing *GubHLH3*. RNA-seq analysis of *GubHLH3*-overexpressing (GubHLH3-OX) soybean roots confirmed that the expression levels of soyasaponin biosynthetic genes were increased in the GubHLH3-OX lines compared with the control lines. These results suggest that the mechanism regulating soyasaponin biosynthesis is similar in licorice and soybean. To identify a functional ortholog of *GubHLH3* from soybean, we performed homology searches against a soybean gene database, and identified two putative orthologs belonging to subclade IVa of bHLH proteins. To obtain additional candidate *GubHLH3* orthologs, we performed gene co-expression analysis using known genes encoding soyasaponin biosynthetic

\* 〒565-0871 大阪府吹田市山田丘2-1

enzymes as guide genes. This analysis identified two additional candidate *GubHLH3* orthologs belonging to subclade IVa of bHLH proteins. Functional analysis of these candidate transcription factors is needed. *Soy Protein Research, Japan* **23**, 46-51, 2020.

Key words : saponin, biosynthesis, cytochrome P450 monooxygenase (CYP), UDP-dependent glycosyltransferase (UGT), bHLH-type transcription factor

ソヤサポニンとはマメ科植物が共通して産生するオレアナン型トリテルペン配糖体の総称である。大豆はアグリコン部ならびに糖鎖の構造が異なる50種類以上のソヤサポニンを産生する。ソヤサポゲノールBをアグリコン部とするDDMPサポニンとその分解産物には抗高脂血症作用、大腸がん細胞増殖抑制作用などの機能性が報告されている一方で、ソヤサポゲノールAをアグリコン部とするグループAサポニンは不快味の原因成分であると報告されている<sup>1)</sup>。そのため、ソヤサポニンの組成や量は大豆の機能性や味に大きく影響する重要な形質である。

ソヤサポニンの生合成には、単純なイソプレレン単位の繰り返し構造をもつ鎖状の2,3-オキドスクアレンから環状骨格のβ-アミリンを生成するβ-アミリン合成酵素 (BAS), β-アミリンの酸化修飾によりアグリコン部を生成するシトクロムP450モノオキシゲナーゼ (CYP), アグリコン部への糖転移反応を触媒するUDP糖依存型配糖化酵素 (UGT) が関わる。近年、ダイズにおいてこれらの反応を触媒する酵素の多くが同定さ

れている<sup>26)</sup> (Fig. 1)。その一方で、大豆サポニンの生合成制御機構に関する知見は少ない。そこで本研究では、大豆サポニン生合成酵素遺伝子の発現制御に関わる転写制御因子の同定を目的として、候補となる転写因子の探索を下記の2つのアプローチにより進めた。

私たちは、マメ科に属する薬用植物カンゾウ (甘草) においてソヤサポニン生合成酵素遺伝子の転写を活性化するbHLH (basic helix-loop-helix) 型転写因子であるGubHLH3を先に同定した<sup>7)</sup>。そこで、GubHLH3と構造的に類似する転写因子がダイズにおいてもソヤサポニン生合成制御に関わっている可能性を検証するため、GubHLH3を高発現する組換えダイズ毛状根組織を作出しRNA-seq解析を行うことでソヤサポニン生合成酵素遺伝子を含む遺伝子発現全体におよぼす効果を調べた。これと並行して、公開されているダイズの大規模遺伝子発現データを利用した「遺伝子共発現解析」により、既知のソヤサポニン生合成酵素と遺伝子発現パターンが類似する転写因子を探索した。

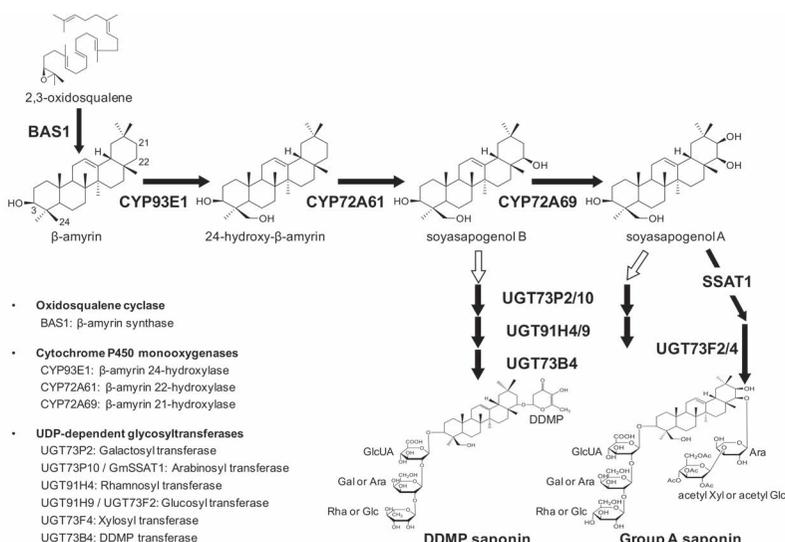


Fig. 1. Proposed soyasaponin biosynthetic pathways in soybean. Filled and open arrows indicate reaction steps catalyzed by known and unknown enzymes, respectively. Ara, arabinose; Gal, galactose; Glc, glucose; GlcUA, glucuronic acid; Rha, rhamnose; Xyl, xylose.

## 方 法

### 組換えダイズ毛状根の誘導

滅菌したダイズ品種エンレイの種子を0.8%アガー上に播種し、23°C暗黒下で5日間静置した後、さらに明所16時間/暗所8時間で1週間栽培した。GFP遺伝子を選抜マーカーとして有する*GubHLH3*高発現用プラスミドコンストラクト<sup>7)</sup>、あるいは空ベクターを保持する*Agrobacterium rhizogenes* (K599株)の菌体をニードルでかきとり幼植物体の子葉の下2 cmほどの箇所を針接種した。接種した幼植物体の地上部を切りとりGamborg B5寒天培地上に移し約3週間共存培養することにより毛状根の誘導を行った。

### 組換えダイズ毛状根の培養

約3週間の共存培養後、蛍光顕微鏡観察でGFP蛍光が確認できた毛状根を単離し、終濃度0.5 g/LのMES (2-morpholinoethanesulfonic acid, monohydrate) および300 mg/Lのセフトキシムを含むGamborg B5液体培地に移し26°C、90 rpm、暗黒下で増殖させた。

### RNA-seqデータ解析

各組換えダイズ毛状根ラインから調製したトータルRNAを用いて、HiSeq2500 (イルミナ) ペアエンド解析によるRNA-seq解析 (1サンプルあたり3,000万リード)を行った。Trimmomatic (v0.39.1)<sup>8)</sup>によるアダプター配列と低クオリティ配列の除去、STAR (v2.7.3.a)<sup>9)</sup>によるダイズゲノムへのマッピング、featureCounts (v2.0.0)<sup>10)</sup>によるリードカウントを行った。得られたカウントデータを用いてTCC-GUI<sup>11)</sup>で発現変動遺伝子群 (DEGs) を検出した。

## 結果と考察

### *GubHLH3*を高発現するダイズ毛状根のRNA-seq解析

*GubHLH3*高発現ダイズ毛状根 (*GubHLH3*-OX) とコントロールでの2群間比較 (N=2) を行った結果、ダイズゲノム (Wm82.a2.v1) でアノテートされている56,044個のダイズ遺伝子のうち、1,839個のDifferentially expressed gene (DEG) が検出された。そのうち、*GubHLH3*-OXラインで有意に発現が上昇した遺伝子は878遺伝子であり、逆に発現量が減少したのは961遺伝子であった (Fig. 2A)。さらに、変動幅の大きい上位100遺伝子の発現量をもとに主成分分

析 (PCA) を行った結果、*GubHLH3*高発現毛状根とコントロール毛状根は明確に区別された (Fig. 2B)。また、既知の12個のソヤサポニン生成酵素遺伝子のうち8遺伝子が*GubHLH3*高発現により有意に発現が上昇していることが判明した (Table 1)。

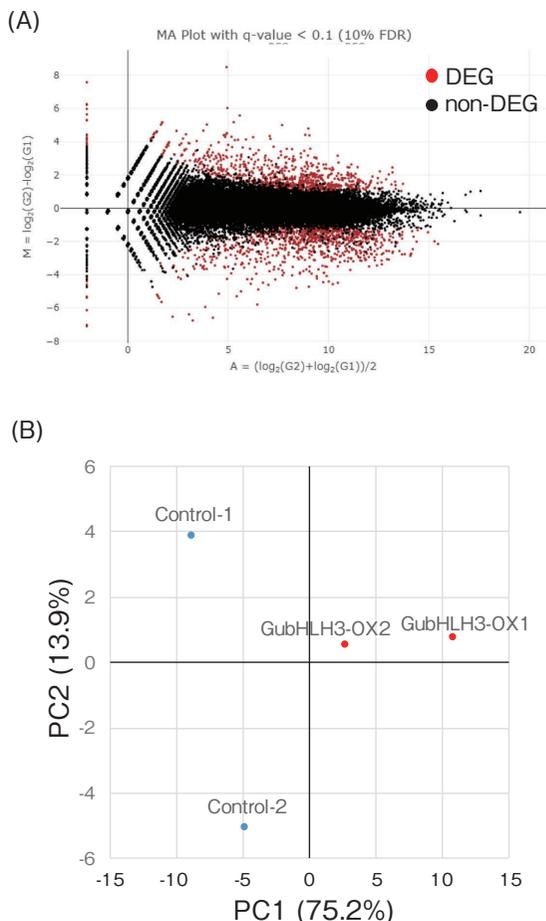


Fig. 2. Changes in gene expression in *GubHLH3*-overexpressing (*GubHLH3*-OX) hairy roots. (A) MA plot of gene expression in *GubHLH3*-OX hairy roots and control hairy roots. Genes upregulated in *GubHLH3*-OX hairy roots have positive M values. Each dot represents a transcript and its log2 fold change in expression (*GubHLH3*-OX/Control). Red dots indicate significantly differentially expressed genes (DEGs, q-value < 0.1). (B) Principal component analysis (PCA) of the transcriptomes of two *GubHLH3*-OX hairy root lines and two control lines. Each point indicates one RNA-seq library.

Table 1. Identification of differentially expressed genes (DEGs) and normalized read counts for saponin biosynthetic genes in *GubHLH3*-overexpressing (*GubHLH3*-OX) hairy roots.

Gene ID	Annotation	Enzyme activity	DEG	TMM-Normalized expression level			
				Control-1	Control-2	GubHLH3-OX1	GubHLH3-OX2
GLYMA_07G001300	GmBAS1	$\beta$ -amyrin synthase	1	2,137	2,973	4,907	9,757
GLYMA_08G350800	CYP93E1	$\beta$ -amyrin 24-hydroxylase	1	7,524	8,598	13,057	31,178
GLYMA_08G238100	CYP72A61	$\beta$ -amyrin 22-hydroxylase	0	3,920	5,563	6,625	11,884
GLYMA_15G243300	CYP72A69	$\beta$ -amyrin 21-hydroxylase	1	601	938	2,256	2,958
GLYMA_11G053400	UGT73P2	C3, Galactosyl transferase	1	2,988	4,424	7,498	14,177
GLYMA_01G046300	UGT73P10	C3, Arabinosyl transferase	1	322	615	2,188	6,157
GLYMA_08G181000	UGT91H4	C3, Rhamnosyl transferase	1	3,060	5,040	10,418	22,516
GLYMA_10G104700	UGT91H9	C3, Glucosyl transferase	1	497	783	1,258	1,958
GLYMA_16G033700	UGT73B4	C22, DDMP transferase	1	2,327	2,701	6,198	9,544
GLYMA_08G181000	GmSSAT1	C22, Arabinosyl transferase	0	749	852	1,119	1,761
GLYMA_07G254600	UGT73F2/4	C22, Xylosyl/Glucosyl transferase	0	836	1,556	1,949	3,217

Normalized read counts for saponin biosynthetic genes in *GubHLH3*-OX1 and -OX2 hairy root lines (highlighted with black background) compared with control lines (Control-1 and -2). DEGs and non-DEGs are indicated by a 1 and 0, respectively, in the DEG column.

### 分子系統樹解析

上記結果から、*GubHLH3*に対して構造的に類似性が高い転写因子がダイズにおいてもソヤサポニン生合成制御に関わる可能性が示唆された。植物のbHLH型転写因子はその構造から26種のサブファミリーに分類される<sup>12)</sup>。その中で、カンゾウの*GubHLH3*およびマメ科タルウマゴヤシにおいてサポニン生合成に関与することが報告されているTSAR1およびTSAR2<sup>13)</sup>はいずれもIVaサブファミリーに属する。ダイズに存在する319個のbHLH型転写因子のうち、24個がIVaサブファミリーbHLHに属する<sup>14)</sup>。そこで、最尤法により、24個のダイズIVaサブファミリーbHLHと上記の既知のサポニン生合成関連bHLHの分子系統樹解析を行った。これにより、*GubHLH3*およびTSAR2と同ジクレイドに属する2種の候補bHLH転写因子(Candidate TF1およびCandidate TF2)を見出した(Fig. 3)。これら2種の転写因子は*GubHLH3*に対して55%のアミノ酸配列同一性を示した。

### 遺伝子共発現解析

*GubHLH3*に対する構造類似性に基づく候補探索に加え、遺伝子共発現解析による候補転写因子の探索も進めた。ダイズの遺伝子共発現データベースであるSoyNet (<https://www.inetbio.org/soynet/>)を用いて12種の既知のソヤサポニン生合成酵素遺伝子(Fig. 1)と発現相関を示すダイズ転写因子を探索した。その結果、2,758個のダイズ転写因子の中から少なくとも1種の生合成酵素遺伝子と発現相関を示す13種の転写

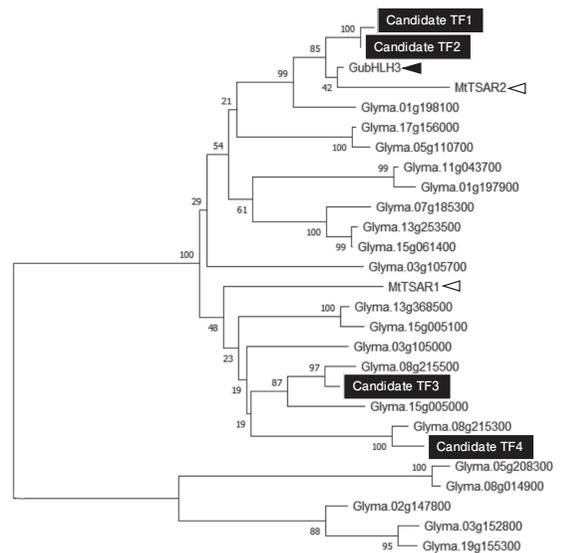


Fig. 3. Phylogenetic tree of soybean transcription factors (TFs) belonging to subclade IVa of the basic helix-loop-helix (bHLH) proteins. A neighbor-joining tree was constructed based on the full-length protein sequences of 24 soybean TFs belonging to subclade IVa of the bHLH proteins. *GubHLH3* of *Glycyrrhiza uralensis* (filled arrowhead) and TSAR1 and TSAR2 (open arrowheads), which direct triterpene saponin biosynthesis in *Medicago truncatula*, are included as references. Candidate TFs (TF1-4) are highlighted with black background.

因子を見出した (Fig. 4). このうちCandidate TF3はCYP72A61およびUGT73P2と、Candidate TF4はBAS1およびUGT73P2と発現相関を示した (Fig. 4). さらに、これら2種の転写因子はいずれもIVaサブファ

ミリー bHLHに属することから (Fig. 3), ソヤサポニン合成を制御する転写因子の有力な候補と考えられる.

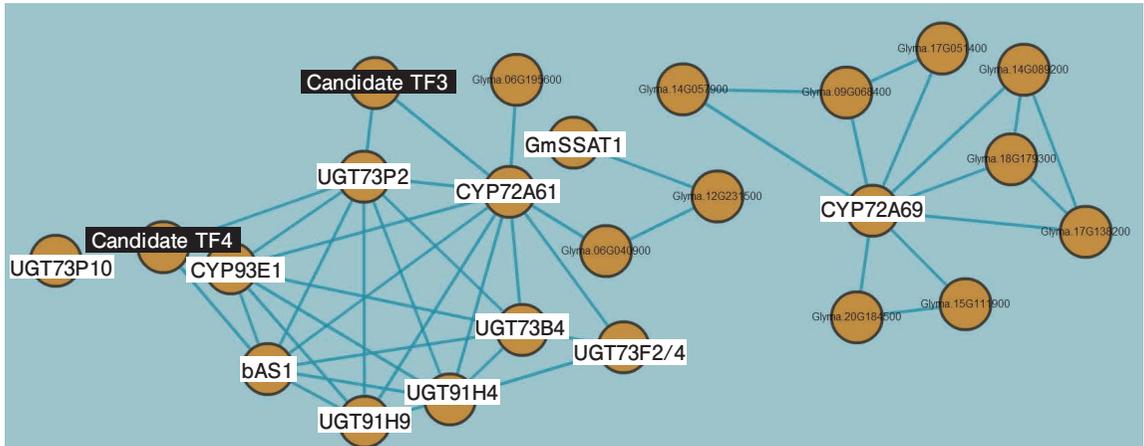


Fig. 4. Co-expression networks of transcription factor (TF) genes and saponin biosynthetic enzyme genes. Networks representing the correlations between 13 TF genes and previously characterized genes encoding soya saponin biosynthetic enzymes. Colored circles represent genes and lines represent correlations among genes. Two candidate TFs (TF3 and TF4) belonging to subclade IVa of basic helix-loop-helix proteins are highlighted with black background.

## 要 約

本研究では、大豆サポニン生合成酵素遺伝子の発現制御に関わる転写因子の同定を目的として、候補となる転写因子遺伝子の探索を下記の二つの手法で進めた。私たちは、マメ科の薬用植物カンゾウ（甘草）においてソヤサポニン生合成酵素遺伝子の転写を活性化するbHLH型転写因子（GubHLH3）を先に同定した。そこで、GubHLH3と構造的に類似する転写因子が大豆サポニン生合成制御に関わっている可能性を検証するため、GubHLH3を高発現する組換えダイズ毛状根を作成しRNA-seq解析を行った。その結果、GubHLH3高発現ダイズ毛状根において、大豆サポニン生合成に関わる既知の生合成酵素遺伝子の発現量がコントロール区と比べて増高していることが判明した。そこで、相同性検索によりGubHLH3と構造的に類似するダイズ転写因子を探索した結果、2種の候補bHLH型転写因子を見出した。さらに、上記と並行して、公開されているダイズの大規模遺伝子発現データを活用した「遺伝子共発現解析」を行い、更なる候補転写因子の探索を進めた。これにより、2,758個のダイズ転写因子遺伝子の中から、既知の大豆サポニン生合成酵素と遺伝子発現パターンが類似する2種のbHLH型転写因子を新たに見出した。今後、これら計4種のbHLH型転写因子の機能解析を進めていく必要がある。

## 文 献

- 1) 塚本知玄, 吉城由美子 (2010): サポニンの機能性, 「大豆のすべて」, (喜多村啓介編) (サイエンスフォーラム, 東京), pp. 281-287.
- 2) Takagi K, Nishizawa K, Hirose A, Kita A and Ishimoto M (2011): Manipulation of saponin biosynthesis by RNA interference-mediated silencing of  $\beta$ -amyrin synthase gene expression in soybean. *Plant Cell Rep*, **30**, 1835-1846.
- 3) Seki H, Tamura K and Muranaka T (2015): P450s and UGTs: key players in the structural diversity of triterpenoid saponins. *Plant Cell Physiol*, **56**, 1463-1471.
- 4) Yano R, Takagi K, Takada Y, Mukaiyama K, Tsukamoto C, Sayama T, Kaga A, Anai T, Sawai S, Ohyama K, Saito K and Ishimoto M (2017): Metabolic switching of astringent and beneficial triterpenoid saponins in soybean is achieved by a loss-of-function mutation in cytochrome P450 72A69. *Plant J*, **89**, 527-539.
- 5) Krishnamurthy P, Tsukamoto C and Ishimoto M (2020): Reconstruction of the evolutionary histories of UGT gene superfamily in legumes clarifies the functional divergence of duplicates in specialized metabolism. *Int J Mol Sciv*, **21**, 1855.
- 6) Louveau T, Orme A, Pfalzgraf H, Stephenson MJ, Melton R, Saalbach G, Hemmings AM, Leveau A, Rejzek M, Vickerstaff RJ, Langdon T, Field RA and Anne Osbourn A (2018): Analysis of two new arabinosyltransferases belonging to the carbohydrate-active enzyme (CAZY) glycosyl transferase family1 provides insights into disease resistance and sugar donor specificity. *Plant Cell*, **30**, 3038-3057.
- 7) Tamura K, Yoshida K, Hiraoka Y, Sakaguchi D, Chikugo A, Mochida K, Kojoma M, Mitsuda N, Saito K, Muranaka T and Seki H (2018): The basic helix-loop-helix transcription factor GubHLH3 positively regulates soyasaponin biosynthetic genes in *Glycyrrhiza uralensis*. *Plant Cell Physiol*, **59**, 783-796.
- 8) Bolger AM, Lohse M and Usadel B (2014): Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114-2120.
- 9) Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M and Gingeras TR (2013): STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15-21.
- 10) Liao Y, Smyth GK and Shi W (2014): featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923-930.
- 11) Su W, Sun J, Shimizu K and Kadota K (2019): TCC-GUI: a Shiny-based application for differential expression analysis of RNA-Seq count data. *BMC Res Notes*, **12**, 133.
- 12) Pires N and Dolan L (2010): Origin and diversification of basic-helix-loop-helix proteins in plants. *Mol Biol Evol*, **27**, 862-874.
- 13) Mertens J, Pollier J, Bossche RV, Lopez-Vidriero I, Franco-Zorrilla JM and Goossens A (2016): The bHLH transcription factors TSAR1 and TSAR2 regulate triterpene saponin biosynthesis in *Medicago truncatula*. *Plant Physiol*, **70**, 194-210.
- 14) Hudson KA and Hudson ME (2015): A classification of basic Helix-Loop-Helix transcription factors of soybean. *Int J Genomics*, **2015**, 603182.