

ダイズ遺伝資源の多様性に基づく代謝制御機構の解明

澤田有司*

独立行政法人理化学研究所環境資源科学研究センター

Metabolic Systems Research Based on the Diversity of Gene Resource in *Glycine Max*

Yuji SAWADA*

RIKEN Center for Sustainable Resource Science, Yokohama 230-0045

ABSTRACT

Diverse plant metabolites are thought to be important for plant growth and physiology, and metabolome analysis has been useful in many research fields. Thus, we have established a practical metabolomics strategy: widely targeted metabolomics. We have already reported the metabolome quantitative locus (mQTL) analysis by linkage mapping strategy. Using the leguminous plants *Lotus japonicus* and *Glycine max*, we successfully identified the mQTLs. This linkage mapping is based on the significant differences in parent lines, which are the origin of recombinant inbred lines. Thus, linkage mapping is not suitable for large scale bioresources and diverse bioresources. Using high speed sequence methods, a huge number of single nucleotide polymorphisms (SNPs) have been reported in plant genomes. The genome-wide SNPs were also used for QTL analysis based on association mapping, and this was termed genome-wide association study (GWAS). In this study, we conducted metabolome analysis for 80 lines of the worldwide *G. max* core collection. Finally, we provide a matrix as follows: 175 metabolites accumulation data versus 240 samples (80 lines by 3 biological replicates). Using analysis of variance (ANOVA) and principal component analysis (PCA), the metabolome data suggested no trend. Using the Pearson correlation coefficient (PCC > 0.9), we successfully elucidated the specific network groups: amino acid, flavonoid, nucleotide and sugar. These metabolite accumulation patterns in *G. max* lines will be used for mQTL analysis based on the GWAS. *Soy Protein Research, Japan* **17**, 8-12, 2014.

Key words : metabolomics, linkage mapping, association mapping, Pearson correlation coefficient, multivariate analysis.

* 〒230-0045 横浜市鶴見区末広町1-7-22

ダイズは重要な食糧資源であり、収量や品質に関わる表現形質の育種研究が進んでいる。近年では数百種から数千種のダイズ遺伝資源が整備され、各系統に特徴的な表現形質の利活用が検討されている¹⁾。これらの表現形質は遺伝子の多型によって生じていると考えられる。従って、より多くの遺伝子多型を含む遺伝資源を利用すれば、従来育種では作出困難であった高機能ダイズの開発や効率的な交配育種が可能になると期待されている。ダイズの重要な農業形質の多くは収量、形、色などの可視の表現型である。申請者は不可視な表現型である代謝産物の蓄積様式に注目して研究を行っている。しかし、従来の手法では多様な物性を示す代謝産物の蓄積様式の解明は困難であった。

代謝産物の蓄積様式は転写産物、たん白質を経て生合成される生命現象の最終段階として極めて重要な情報を含んでいると考えられる²⁾。この代謝産物を広範囲に検出する手法としてメタボローム解析が提唱されている。初期のメタボローム解析は予め標的を設定しない非ターゲット解析が主な手法であった。この手法では標品化合物を入手できない未同定化合物を検出できる。その反面、大規模な情報解析を必要とするため数百種の遺伝資源を定量的に比較することは困難であった。そこで、我々は予め標準化合物の超高感度検出条件を最適化し、これを広範囲に検出するワイドターゲット解析を開発した³⁾。

申請者はこれまでに水溶性の700種以上の代謝産物（アミノ酸、有機酸、糖、フラボノイドなど）が検出可能なメタボローム解析技術を開発し、モデルマメ科植物ミヤコグサおよびダイズの特定制代謝産物にかかわる量的形質座同定に成功している^{4,5)}。これらの連鎖解析では異なる親株二種の交配システムを利用するため、2種のゲノム間の差に基づく代謝研究が可能である。一方、ダイズの遺伝資源は数千種存在するため、交配システムの作出とマーカー情報の整備が必要な連鎖解析では、一部の系統間の差のみが検出可能である。すなわち、連鎖解析ではダイズ遺伝資源の多様性を生かした代謝研究は困難であった。

そこで、多様な遺伝資源の遺伝子マーカーと表現型を比較する関連解析が注目されている。近年のシーケンス技術の向上により、シロイヌナズナ、イネでは数十万から数百万の一塩基多型情報が取得され可視の表現型のゲノムワイドな関連解析に用いられている^{6,7)}。これと同様に、ダイズの遺伝資源の一塩基多型情報が公開されつつある。さらに、これらの一塩基多型情報に基づく関連解析によってダイズ種子のたん白質や脂質の量に関する関連解析が報告されている。

本研究ではダイズ遺伝資源の多様性に基づく効率的な代謝育種をめざし、NIASから配布されている世界のダイズコアコレクションの特徴的な代謝プロファイルを取得した。さらに、地域別に代謝プロファイルと比較し、14の国と地域で採取されたダイズに顕著に蓄積する代謝産物を発見した。これらの系統間の差は遺伝子の有無や発現量の差によって制御されていることが考えられる。これらに加えマメ科の薬粧植物カンゾウの主要成分であるグリチルリチンを同時に検出し、ダイズの含有の有無を検討した。

方 法

植物種子材料

ダイズ種子は、農業生物資源ジーンバンクの世界のダイズコアコレクション2012年度産を購入し、実験に利用した。コアコレクションリストは96種であるが、種子増幅が完了した全80種を利用した。

分析試料の作成方法

全80系統のダイズ種子は、1粒づつ葉包紙で包んで木槌で叩いて予備破碎した。予備破碎したダイズ種子は2 mL チューブ（ハードチューブ、BMS社）に直径5 mmジルコニアビーズ（アズワン社）を入れたのちビーズショッカー（シェイクマスターネオ、BMS社）で1500 rpm, 5 min の条件で粉末化した。各系統について独立した種子を3粒づつ利用し、計240サンプル調製した。このダイズ種子粉を4 mg秤量し、1 mLの抽出バッファー（80%MeOH, 0.1%ギ酸）を加えて抽出した（粉末化条件と同様）。抽出液（4 mg/mL）を1,500 rpm, 10 minの条件で遠心分離し、上清を以降のサンプル処理に用いた。

抽出物の前処理

抽出液を含んだ2 mLチューブを自動分注システム（Microlabstar, ハミルトン社）にセットし、8チャンネルの分注ヘッドで25 μ Lを96穴プレート（Thermo社）に分注し、各ウェルに225 μ Lの抽出バッファーを加えて10倍希釈した。次に、5倍に希釈された抽出液を25 μ L再度分注し、乾燥窒素吹き付け型の溶媒乾燥機（Ultra vap, Povair社）でドライアップした（室温, 30 min）。乾固されたサンプルに250 μ LのLC-MSグレードの純粋を加え再溶解した（0.08 mg/mL）。次に384 well filter plate（Watman社）でフィルター濾過し（60 μ Lづつ二回）、384穴プレート（Waters社）に回収し（約100 μ L）、メタボローム解析に利用した。

サンプルの分析

抽出サンプルは、液体クロマトグラフィーにタンデム四重極型質量分析装置をハイフネーションしたUPLC-TQS (Waters社) で分析した。分析結果は同社の解析ソフト (マスリンクス, Waters社) で解析し、各代謝産物のクロマトグラフピークの面積値を算出した。次にPerlスクリプトを利用し、代謝産物とサンプルの情報を含んだエクセルファイルを作成した。代謝産物の分離条件と検出条件は既報の方法で行った⁸⁻¹⁰⁾。予備分析を行い、検出対象は540種の高感度検出条件 (Selected reaction monitoring, SRM) から208種の標準化合物と2種の内部標準物質を対象としたSRMを選択した。

データ解析

各サンプルの分析結果から得られたメタボローム情報 (240サンプル×208種の代謝産物検出強度) を利用した。まず、3回測定した抽出バッファーコントロールの検出強度の平均をノイズ (N)、各サンプルの検出強度をシグナル (S) と定義し、シグナルとノイズ比が10を超える175代謝産物を選択した。次に多群のサンプルに適用可能な統計解析 (ANOVA) と多変量解析 (主成分分析) を行った。さらに、各代謝産物のピアソンの相関係数を総当たりで算出し、これに基づくネットワーク解析を行った。

結果と考察

実験デザイン

本研究は今後公開されるSNP情報に基づくGWASを行うために、より多様なダイズ遺伝資源のメタボローム情報を取得した。まず、世界のダイズコアコレクション80種を入手し、生物学的な反復を得るために独立した3粒づつ計240種の種子粉を調製した。また、各ラインに付与された栽培地情報 (Fig. 1) を利用し、栽培地と代謝様式の影響を考察した。特に栽培地別の影響が強い集団を利用したGWASでは偽陽性が生じやすいため、事前にこれらの集団構造を予測する目的がある。

メタボロームデータの取得と標的化合物の選択

240サンプルの抽出液を利用し、まず556種の代謝産物の検出条件を利用したLC-MS/MS分析を行った。この予備分析で検出可能な代謝産物として210種を予め設定した。次に、240サンプルとコントロールとして抽出バッファーのみを入れた3サンプルをLC-MS/MS分析した。検出強度がノイズレベルの10倍以上 (S/N

≥ 10) の171種のSRMを選択し、以降の解析に利用した。今回、ダイズの代謝多様性を検討するために、マメ科の薬用植物カンゾウの主要成分であるグリチルリチンを測定したが、顕著な蓄積を示す系統は見られなかった。

統計解析

ダイズのコアコレクションは多群のサンプルとして統計解析可能なAnalysis of variance (ANOVA) を利用し、有意な変化を示す代謝産物を列挙した。しかし、各ライン間の差 (Fig. 2A)、栽培地間の差 (Fig. 2B) はいずれも大きく、ANOVAで代謝の傾向を示すことはできなかった。

主成分分析による全体像の確認

統計解析では顕著な差が見られなかったため、全代謝産物法を利用し最も分散の大きな変数を調べる主成分解析 (PCA) を行った。この結果、韓国 (KOR) およびネパール (NEP) の一部ラインで特徴は示されたが、全体としては明確なグループは示されていない (Fig. 3)。

相関解析

同調的に制御される代謝産物は各サンプル間の代謝産物の蓄積量が正に相関すると推定される。そこで、171種の代謝産物間のPearson's Correlation Coefficient (PCC) を総当たりで算出した。さらに相関関係の全容を解明するためにPCC>0.9以上の顕著な相関値を示す代謝産物群についてネットワーク解析を行った。この結果、アミノ酸、糖、核酸、フラボノイド、イソフラボノイドなどが顕著なグループを示した。これらの結果から一次代謝産物である、アミノ酸、糖、核酸などは生合成上離れていても恒常的に量が制御され、この結果高い相関値を示すと考察される。一方、フラボ

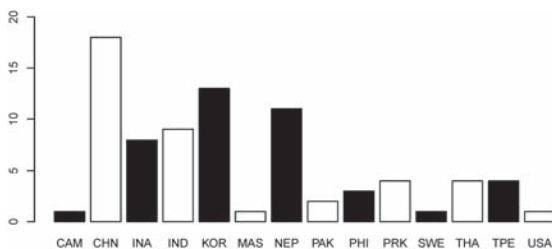


Fig. 1. Origin of 80 lines in *G. max* core collection origins. The three-letters abbreviation codes for countries and regions were based on the International Olympic Committee (IOC).

謝 辞

メタボローム情報の取得では環境資源科学研究センターの技術員である坂田あかね（サンプル抽出）、佐藤心郎（LC-MS/MS分析）、山田豊（情報処理）の支援を得た。また、薬用植物カンゾウとの比較では同センターの澤井学、澤田朋美の支援を得た。本研究にご協力いただいた皆様に感謝いたします。

今後の展開

多様な遺伝資源に基づく代謝育種は今後植物科学の中心的な課題になると考えられている。特に高密度な SNP 情報による標的遺伝子の同定はメタボローム解析技術を利用した統合オミックス解析において強力な手法になると期待されている。今後のダイズ遺伝資源のゲノムシーケンスの発展に期待している。

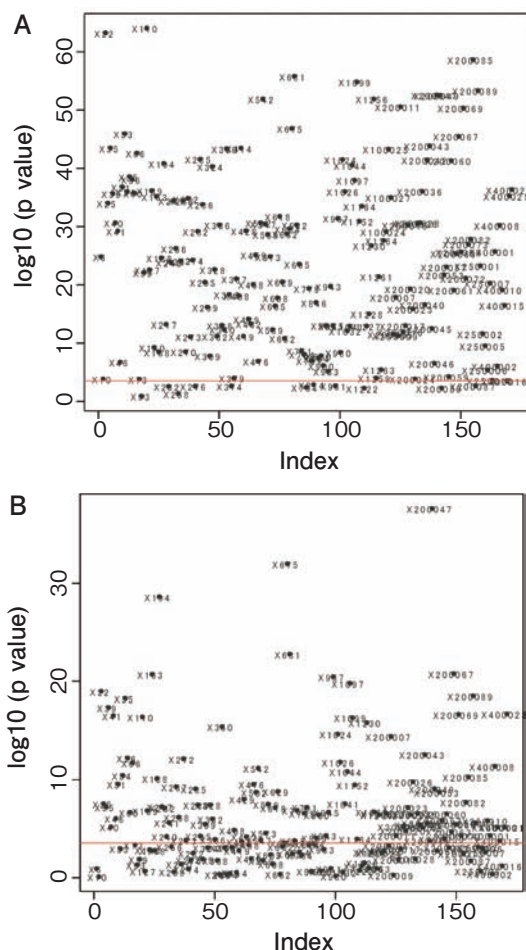


Fig. 2. ANOVA p value plots of metabolome data. (A) The 240 samples were grouped by 80 lines. (B) The 240 samples were grouped by 14 groups based on the countries and regions. Fails discovery rate value (adjusted p-value by Bonferroni method) was shown by red line.

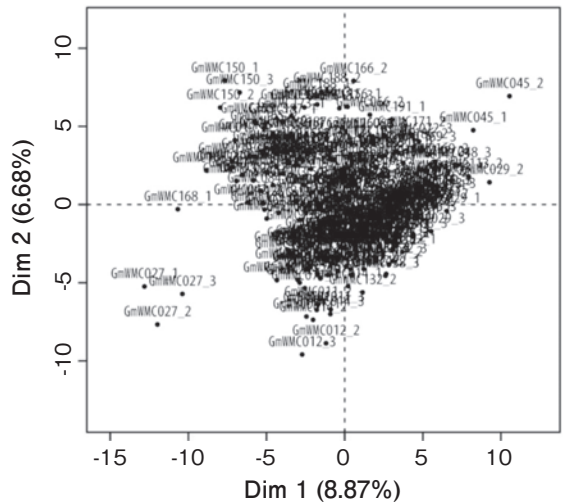


Fig. 3. PCA analysis for metabolome data.

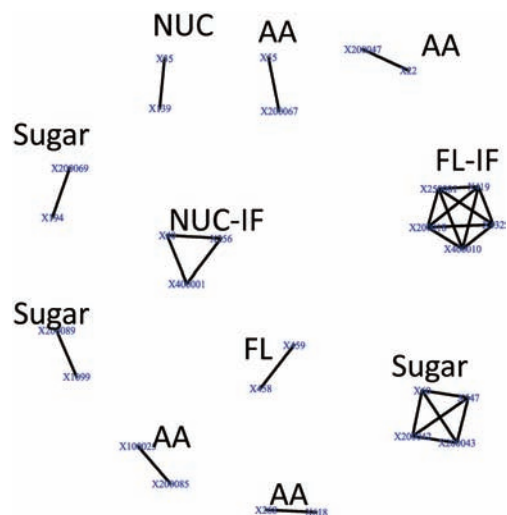


Fig. 4. PCC network graph for metabolome data. The node was metabolites and edge were shown by PCC > 0.9. Abbreviation: AA, amino acid; NUC, nucleotide; FL, flavonoid; IF, isoflavonoid.

要 約

近年、多様なダイズ遺伝資源の一塩基多型情報 (SNP) が大規模に取得可能になり、各種表現形質と関連解析を行うゲノムワイドアソシエーション解析 (GWAS) が注目されている。本研究ではNIASから配布されている世界のダイズコアコレクションを対象としてメタボローム解析を行い、メタボロームGWASを可能にする基盤情報を取得した。さらに、ダイズ遺伝資源の代謝システムを評価するために以下の検討を行った。まず、他群間の比較が可能なANOVAを行い、80種の系統間と14の栽培地を指標とした解析では顕著な傾向がみられないことが分かった。次に、171種の代謝産物情報を利用した主成分解析を行い、韓国とネパールの一部品種に特徴がみられるが、全体としては顕著な傾向がみられないことを示した。最後に171種の代謝産物のPCC値を総当たりで算出し、これらに基づくネットワーク解析をした結果、アミノ酸、糖、核酸、フラボノイド、イソフラボノイドが顕著なグループを示した。今後、ダイズのSNP情報を利用したメタボロームGWASを行い、これらの代謝制御にかかわる分子機構を解明する。

文 献

- 1) Kaga A, Shimizu T, Watanabe S, Tsubokura Y, Katayose Y, Harada K, Vaughan DA and Tomooka N (2012): Evaluation of soybean germplasm conserved in NIAS genebank and development of mini core collections. *Breed Sci* **61**, 566-592.
- 2) Ryan D and Robards K (2006): Metabolomics: The greatest omics of them all? *Anal Chem* **78**, 7954-7958.
- 3) Sawada Y, Akiyama K, Sakata A, Kuwahara A, Otsuki H, Sakurai T, Saito K and Hirai MY (2009): Widely targeted metabolomics based on large-scale MS/MS data for elucidating metabolite accumulation patterns in plants. *Plant Cell Physiol* **50**, 37-47.
- 4) Sawada Y, Nakabayashi R, Yamada Y, Suzuki M, Sato M, Sakata A, Akiyama K, Sakurai T, Matsuda F, Aoki T, Hirai MY and Saito K (2012): RIKEN tandem mass spectral database (ReSpect) for phytochemicals: a plant-specific MS/MS-based data resource and database. *Phytochemistry* **82**, 38-45.
- 5) Sawada Y and Hirai MY (2013): Integrated LC-MS/MS system for plant metabolomics. *Comput and Struct Biotechnol J* **4**, (20130101)
- 6) Atwell S, Huang YS, Vilhjalmsdottir BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, Jiang R, Mulyati NW, Zhang X, Amer MA, Baxter I, Brachi B, Chory J, Dean C, Debieu M, de Meaux J, Ecker JR, Faure N, Kniskern JM, Jones JD, Michael T, Nemri A, Roux F, Salt DE, Tang C, Todesco M, Traw MB, Weigel D, Marjoram P, Borevitz JO, Bergelson J and Nordborg M (2010): Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465**, 627-631.
- 7) Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Lin Z, Buckler ES, Qian Q, Zhang QF, Li J and Han B (2010): Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet* **42**, 961-967.
- 8) Sawada Y, Akiyama K, Sakata A, Kuwahara A, Otsuki H, Sakurai T, Saito K and Hirai MY (2009): Widely targeted metabolomics based on large-scale MS/MS data for elucidating metabolite accumulation patterns in plants. *Plant Cell Physiol* **50**, 37-47.
- 9) Sawada Y, Kuwahara A, Nagano M, Narisawa T, Sakata A, Saito K and Hirai MY (2009): Omics-based approaches to methionine side chain elongation in *Arabidopsis*: characterization of the genes encoding methylthioalkylmalate isomerase and methylthioalkylmalate dehydrogenase. *Plant Cell Physiol* **50**, 1181-1190.
- 10) Sawada Y, Toyooka K, Kuwahara A, Sakata A, Nagano M, Saito K and Hirai MY (2009): *Arabidopsis* bile acid:sodium symporter family protein 5 is involved in methionine-derived glucosinolate biosynthesis. *Plant Cell Physiol* **50**, 1579-1586.