

生物情報科学・細胞生物学的手法から見えてきた 植物栄養応答

3. 大規模データが解き明かす植物の土壌ストレス耐性システム

楠 和隆¹・時澤睦朋¹・小林安文^{1,2}・小林佑理子¹

¹岐阜大学応用生物科学部, ²インド工科大学グワハティ校生物工学部

キーワード 大規模データ, 土壌ストレス, オミクス解析

1. はじめに

植物は、土壌、温度、光条件、病虫害などの多種多様なストレスに曝されながら生育している。実際のストレスは単独ではなく、複数のストレス要因が複合的に作用して植物の生育を阻害することが多い。例えば、酸性土壌では低pH環境によるプロトンストレスに加えて、溶出するアルミニウム (Al) やマンガンなどの過剰害および、リン酸やホウ素などの欠乏症状が発生することが知られている。塩害地では塩による浸透圧ストレスと同時にカリウム・カルシウム欠乏が起こり、寒冷地では低温ストレスの直接的な影響に伴って土壌水分の凍結による乾燥ストレスが植物にダメージを与える。それぞれのストレス要因は植物体の複数の組織や細胞器官において様々な損傷を与える。植物はそのような複合的なストレス環境下で生存し成長を続けるために、それら複数のストレス・組織・時間軸に応じて、複雑な耐性機構を実にバランス良く制御している。本稿では、土壌環境ストレスに対して植物が備えている耐性機構の複雑性について例示した上で、それを紐解くための生物学的な大規模データの利用事例を概説し、将来の展望について述べたい。

2. 植物がもつストレス応答機構の複雑性

植物は複合的な環境ストレスに対応するために、各ストレスに対して個別の耐性機構を備えるよりも、遺伝子や代謝産物などを組み合わせる、またはそれらを多面的に機能

させる機構を発達させてきた。以下にその事例について解説する。

DREB2A (*DEHYDRATION-RESPONSIVE ELEMENT BINDING 2A*) は、乾燥、塩、高温ストレス耐性機構に重要な役割をもつ転写因子である (Sakuma *et al.*, 2006)。*DREB2A* はこれら3種のストレスに共通して発生する細胞内浸透圧の上昇に対応するために、適合溶質合成に関わる *GOLS2* (*GALACTINOL SYNTHASE 2*) の発現制御を行うほか、乾燥/塩ストレスのみに応答する *RD29A* (*RESPONSIVE TO DEHYDRATION 29A*) や、高温のみに応答する熱ショックタンパク質 (*HSP70*, *HSP18.2*) など、ストレス特異的に応答する遺伝子の制御にも重要な役割を果たしている。このようなストレス種による下流遺伝子群の違いは、転写制御因子と相互作用するコファクターの組み合わせによって行われると考えられている。また、Al耐性と低pH耐性に重要な転写制御因子である *STOP1* (*SENSITIVE TO PROTON RHIZOTOXICITY 1*) も、Alストレス下ではAl耐性を担うトランスポーター群、低pHストレス下ではpHの安定性に機能する代謝関連遺伝子群の発現を制御している (Iuchi *et al.*, 2007; Sawaki *et al.*, 2009)。

リンゴ酸トランスポーター *ALMT1* (*ALUMINUM-ACTIVATED MALATE TRANSPORTER 1*) は、Al誘導的にリンゴ酸を分泌しAlをキレート化することで植物の根に対するAl毒性を軽減することが知られている。近年の報告では、*AtALMT1* 転写がトマト斑葉細菌 (pst DC3000) や鞭毛タンパク質フラジェリン *flg22* によって誘導され、*AtALMT1* を介したリンゴ酸分泌が根圏の有用微生物を根部に誘引することで、植物に病害抵抗性を誘導することが明らかとなった (Rudrappa *et al.*, 2008; Lakshmanan *et al.*, 2012)。恒常的にリンゴ酸を放出する *AtALMT1* 過剰発現シロイヌナズナ (*Arabidopsis thaliana*) を用いた筆者らの解析においても、根圏に集積される有用微生物の増加が確認された (Kobayashi *et al.*, 2013a)。Alと*flg22*では *ALMT1* の発現誘導経路が異な

Kazutaka KUSUNOKI, Mutsutomo TOKIZAWA, Yasufumi KOBAYASHI, and Yuriko KOBAYASHI: Plant responses to nutrition revealed by bioinformatics and cell biology. 3. Big data analysis provides new insights in tolerance systems of plants to soil-stress

¹ 501-1193 岐阜市柳戸1-1

² Guwahati, 781039 Assam, India

2016年8月1日受付・2016年9月7日受理

日本土壌肥科学雑誌 第87巻 第6号 p.472~478 (2016)

るほか (Kobayashi *et al.*, 2013a), 発現部位も Al ストレス応答では Al 毒性の標的部である根の分裂, 伸長領域において強い発現を示すが (Kobayashi *et al.*, 2007), 植物病原菌の感染に対しては有用微生物を根部全体に集積するよう根部全体での発現が観察されている (Rudrappa *et al.*, 2008). 加えて *AtALMT1* の発現は低 pH や過酸化水素, 植物ホルモンであるインドール-3-酢酸やアブシジン酸といった広範なシグナルによって誘導され (Kobayashi *et al.*, 2013b), これらの解析から *ALMT1* が複数の生物学的意義を発揮するためには, 様々な環境要因に対して複雑なシグナル伝達による発現制御が適切に行われることが重要であると考えられた.

以上のように, 植物は遺伝子やタンパク質を複合的かつ多面的に利用することで, 複雑な環境に適応している. このようなストレス応答の複雑性はその解明を著しく困難にしながらも, 研究対象として非常に興味深いものになっている.

3. 生物分子情報解析の大規模化による ストレス応答機構解明への挑戦

ゲノム解読以前の環境ストレス耐性機構に関わる要因の探索は, 主動遺伝子の個別探索が主流であった. QTL (Quantitative Trait Loci) 解析やサブトラクション法などは, 形質に大きな影響を与える原因遺伝子の同定に強力な手法であり, 特に, 作物における QTL 解析は現在でもよく用いられている. 2000 年に完了したシロイヌナズナゲノム配列の解読以降は, 多くの植物でゲノム配列や遺伝子の全体像が明らかになり, 生体内分子の挙動を網羅的に捕捉できるようになった. cDNA クローンや T-DNA 挿入変異体など逆遺伝学を行うためのバイオリソースの整備と相まって, ゲノム中に存在する遺伝子の機能解析と相互関係の同定が精力的に行われるようになった. 現在では, 公共データベースに収集された多量の生物学的データや解析ツールから必要な情報を取捨選択してそれらを再解析し, これに自己の研究データを統合することで新しい知見を生み出すアプローチが取られることが多くなり, ストレス耐性機構に関わる因子のハイスループットな探索やそれらの相互関係の解明に繋がっている.

このような研究展開を可能にしたのが, 計測技術の開発とそれが生み出す大規模データの処理手法の発展である. 特に次世代シーケンス (NGS: next generation sequencing) の登場以降は, ゲノム配列と転写配列が解読された生物種が増加し, NCBI でドラフトゲノムが公開された植物種数は 150 種を超える. その結果, 複数系統の試料を用いたゲノム上の一塩基多型 (SNPs: Single Nucleotide Polymorphisms) と挿入欠失 (indels: insertions and deletions) の検出による変異解析や非モデル生物のトランスクリプトーム解析が, 一研究室でも可能になっている. 環境データについても, センシング技術の発達によって, 我々が測定できる気象・土壌・水質などに関する情報は益々広範囲で高解像度なものとなった. これ

らのデータの多くは公共のデータベースに登録され, 蓄積したデータを誰もが利用できる形になっていることが, 前述の「自他のデータを統合して知見を得る」切り口を可能にしている.

測定可能な生体分子の増加に伴って, 試料当たりに得られるデータは数万遺伝子, 数百万 SNP と大規模化が進み, 実験全体としてはギガバイト・テラバイトレベルのデータのハンドリングが珍しくなくなった. この大容量データの解析に対応するために, 多くの解析プログラムが開発され, また, 統計モデリングも正規分布を仮定した線形モデルから, ベイジアンモデリングや機械学習などのコンピュータによる計算を前提とする手法へと発展してきた. 以下では, これらの技術によって生体分子情報を取得し, 得られた「数百サンプル × 数万変数」以上の大規模マトリックスからインサイトを生み出すプロセスについて事例を交えて紹介する (図 1).

4. ゲノムワイド多型データを活用した全ゲノム解析

表現型に影響を与えるゲノム領域の探索は育種に直接適用できる利点もあり, 原因遺伝子の探索を中心にモデル植物から作物に至るまで多くの研究が行われている. 過去には主動 QTL の同定が中心であった遺伝領域の特定手法も, SNP 情報の蓄積による DNA マーカーの飛躍的な高密度化により, ゲノムワイド関連解析 (GWAS: Genome-Wide Association Study) が実用的に適用できる段階に入ってきた. GWAS は, 同一生物種の集団数百系統について形質データと数千~数百万の SNP データを取得し, 形質との関連性の高い SNP を検出することで, 集団内の形質の差をもたらす遺伝領域を特定する. シロイヌナズナやイネであれば多くの系統についての SNP 情報が公開されているため, 解析対象とする表現型さえ取得すれば GWAS を行うことができる.

筆者らの GWAS 研究例を紹介したい. シロイヌナズナ野生種 200 系統について, コントロール区に対する強弱の塩ストレス条件下での根伸長速度を塩耐性能力として評価し, 公開されていた約 20 万 SNP (当時) から原因となる遺伝領域の探索を行った (Kobayashi *et al.*, 2016). 基本的には目的形質と相関の高い SNP を選抜するといった理解で良いが, 実際の計算では遺伝的な集団構造と近縁性の補正を行って偽陽性 SNP の排除を行うため, 線形混合効果モデルを用いて各 SNP の寄与率を計算した (実際の計算は TASSEL や R などのソフトウェアを利用). 一般的な GWAS による原因遺伝領域の検出では, ボンフェローニなどの「厳しすぎる」多重検定を通過できる極端に効果の高い少数の SNP を選抜することが多い. 一方, 筆者らの場合は, 中程度の効果をもつ SNP が多く検出され, 正則化項を用いた回帰モデルによる推定結果から上位 200 個までの SNP によって塩耐性の分布の 70 % を説明できると考えられた. 検出された SNP の周辺に原因遺伝子があると考えられ, 筆者らは P 値を基準に検出した上位 200 SNP

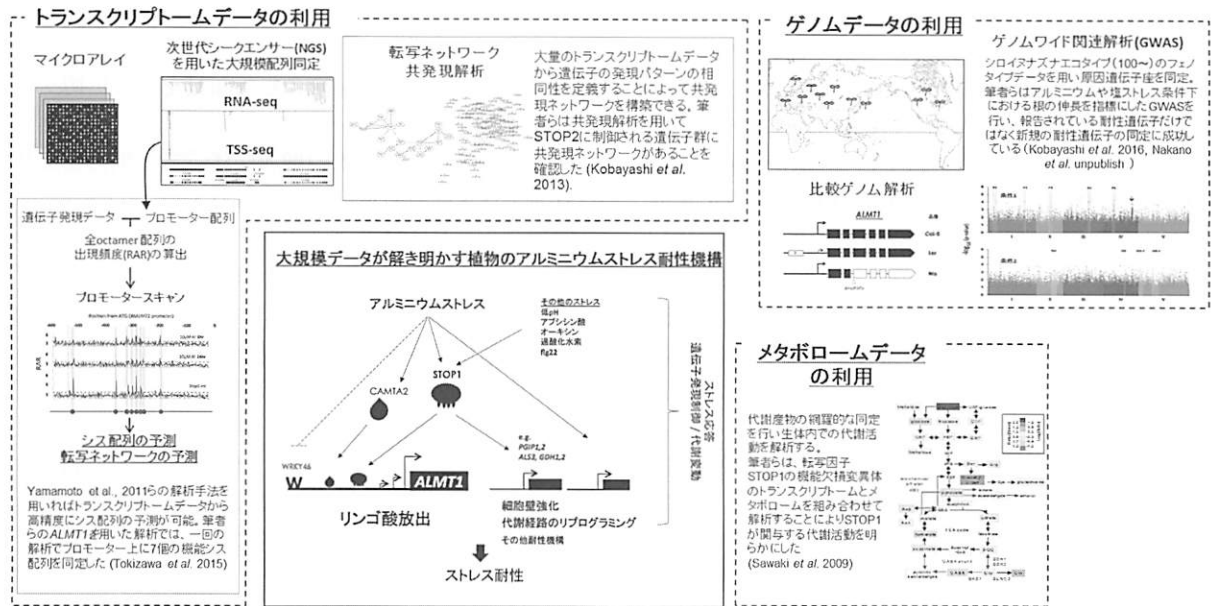


図1 各オミクス階層(ゲノム、トランスクリプトーム、メタボローム)から得られる大規模データの活用手法
アルミニウムストレス耐性機構を例として。

に関して、前後 10 kbp 内で最も近隣にある遺伝子と強い連鎖不平衡 (LD) が存在する領域にある 150 個前後の遺伝子のすべてを対象として、塩ストレス耐性に関わる候補遺伝子として選抜した。その中には、カリウムチャンネルタンパク質をコードする *KAT1* (*POTASSIUMCHANNEL IN ARABIDOPSIS THALIANA 1*) やカルシウムセンサータンパク質キナーゼをコードする *SOS3* (*SALT OVERLY SENSITIVE 3*) など既知の塩耐性遺伝子が含まれていたほか、過剰発現体を用いた耐性試験により、*RABA1c* (*RAB GTPASE HOMOLOG A1C*)、*VHA-A2* (*VACUOLAR PROTON ATPASE A2*) や *ZIP ion metal transporter* が新規な塩耐性遺伝子として同定された。このように、中程度の効果の耐性遺伝子を一括して選抜する手法としても、GWAS が有効であることが明らかになった。さらに、塩濃度の強弱で選抜された塩耐性関連遺伝子群が大きく異なっていたために、遺伝学的レベルでもストレスの強弱に応じて異なる耐性機構がはたらいっていることが示唆された。このような GWAS を用いたアプローチは、耐性機構をシステムとして理解する上で有益であると考えられた。

その他のストレス耐性に関する GWAS の報告例としては、イネの Al ストレス耐性遺伝子として *OsNRAT1* (*NRAMP1 TRANSPORTER 1*) が同定された (Famoso et al., 2011) ほか、タルホコムギのカドミウムストレス耐性について候補遺伝子を選抜した例 (Qin et al., 2015) がある。また、養分欠乏関連でも報告例が蓄積されつつあり、西アフリカのソルガムを用いたリン欠乏土壌における穀物の収量を指標とした GWAS では、Al 耐性遺伝子の一つであるクエン酸輸送遺伝子が検出され、Al 耐性との共通性が観察された (Leiser et al., 2014)。また、低窒素条件下における側根の伸長に関わる遺伝子として、ジャスモン酸応答遺伝子である *JR1* (*JASMONATE RESPONSIVE 1*) と

PHENAZINE BIOSYNTHESIS PhzC が検出され、遺伝子破壊株を用いた検証実験で関連が確認された (Gifford et al., 2013)。

得られる多型情報の高密度化に伴って、GWAS と対をなす手法としてゲノム予測による選抜 (GS: genomic selection) も注目されるようになってきた。どちらも遺伝子型 (主に SNP) と表現型の関連を解明する手法であるものの、形質の原因遺伝領域の特定を目的とする GWAS と比べて、GS は全遺伝子型情報を説明変数として各系統の形質値を予測する回帰式を構成することによって、後代系統の形質を遺伝子型から予測することを目的としており、作物育種への直接的な応用手段として研究が進んでいる (Desta and Ortiz, 2014)。GS を利用した育種の場合は、単一のストレス耐性や養分利用効率性の改善ではなく、様々な環境変動を伴う実際の農地における成長性や収量の向上を図ることが多いが、GS の過程において集積される遺伝領域の流れを追うことにより、適用する土壌環境に応じた養分利用効率関連因子やストレス耐性因子、またはそれらの組み合わせを発見できる可能性がある。

GWAS や GS を非モデル植物に適用するためには、数百から数千系統の全ゲノム SNP 情報を得るためのシーケンスコストが課題であったが、制限酵素認識サイト近隣領域のみをシーケンスする RAD-Seq を利用した GBS (Genotyping By Sequencing; Elshire et al., 2011) によって、多試料の遺伝子型を安価に特定することも可能となった。カナダの研究グループの報告では、ダイズ 304 系統の GBS で得られた 47,000 の高密度 SNP を用いた GWAS (GBS-GWAS) で、8 つの農業形質に関わる原因ゲノム領域が検出された (Sonah et al., 2015)。一方、中国の研究グループは、GBS によって抽出した 48,000 以上の SNP 情報を用いたゲノム予測 (GBS-GS) によってトウモロコ

シ交配集団における収量の子測を行い、モデルに環境条件を加味することによる予測精度の向上を報告し、水分ストレス条件下における予測精度の低下を課題として取り上げた (Zhang *et al.*, 2015). このように、GBSはLDが維持されている交配集団のみならず遺伝的背景が多様な集団においても有効であり、研究面では変異株の原因遺伝子探索に、実務面では個体識別/品種鑑定やマーカー支援育種に今後広く使われる技術であると考えられる。

5. 高精度トランスクリプトームデータの大規模解析

トランスクリプトーム研究は、マイクロアレイの登場によりオミクス階層の中で最も早く包括的な解析が普及した。トランスクリプトーム解析は今日に至るまで遺伝子の同定から遺伝子間相互作用の検出まで幅広く用いられ、植物の土壤環境ストレス研究への適用例も多い。遺伝子同定に用いられた研究例として、鉄欠乏時における経時的なトランスクリプトーム解析から鉄の吸収利用を制御する転写制御因子 *OsIRO2* の発見に繋がった例や (Ogo *et al.*, 2006)、アジサイの cDNA マイクロアレイから Al の集積や耐性に関与する Al 輸送体を同定した例が挙げられる (Negishi *et al.*, 2012). 通常ストレス処理は成長の変動を伴うことが多いために、そのトランスクリプトームを測定すると成長関連の遺伝子を含む数百個以上の変動遺伝子が検出される。そのため原因遺伝子を絞りきれない場合も多く、トランスクリプトーム解析のみで遺伝子同定に繋がった事例はそれほど多くない。

トランスクリプトーム解析が特にその真価を発揮するのは遺伝子間の因果関係の検出である。多くは転写制御因子とそれに制御される遺伝子群の関連性の解明に実績が認められる。そこでは、転写制御因子によって制御される下流遺伝子群を特定するために、転写制御因子の遺伝子破壊株や過剰発現株のトランスクリプトームを野生株と比較することはほぼ標準的に行われる (Sawaki *et al.*, 2009; Kobayashi *et al.*, 2014). 既知のストレス耐性遺伝子からその転写制御因子を探索するためにも用いられ、対象となる耐性遺伝子と共発現する遺伝子等が候補として選抜される (Ding *et al.*, 2013). さらに、トランスクリプトームデータから発現変動遺伝子群のプロモーター上で共通するモチーフを探索することで転写制御因子の結合領域であるシスエレメントを推定し、その結合配列から転写制御因子を同定することで、その関連性がより確からしいものとなる (Yamamoto *et al.*, 2011). 筆者らは、前述のシロイヌナズナの Al 耐性遺伝子である *AtALMT1* を対象として、シスエレメントと転写制御因子を同定するために、*AtALMT1* の発現が大きく変動する Al 処理と *AtALMT1* を発現制御する転写因子 *AtSTOP1* の破壊株のトランスクリプトーム情報を利用した。発現変動遺伝子群のプロモーター情報から、高頻度で検出されるオクタマー (8塩基) を検出し、*AtALMT1* のプロモーター上に8箇所のシスエレメントの存在を予測した。そのうち7つのシスエ

レメントが機能することが実験的に証明され、そのうちの一つのシスエレメントに *AtSTOP1* が結合することを *in vitro* で確認した。さらに、同定した他のシスエレメント配列の一つには CAMTA2 (CALMODULIN BINDING TRANSCRIPTION ACTIVATOR2) 転写制御因子の結合と、その機能欠損変異体において *AtALMT1* の転写活性化能が低下することを確認した (Tokizawa *et al.*, 2015).

NGS の普及以降は、RNA-Seq によってスプライシングバリエーションや非コード RNA (ncRNA: non-coding RNA) (Jin *et al.*, 2013)、ゲノムインプリンティング (Cubillos *et al.*, 2014) などのマイクロアレイでは検出できなかった種類の遺伝子発現情報が収集され、ストレス耐性との関連付けが試みられている。PacBio RSII のような超ロングリードシーケンサーを用いた Iso-Seq (Isoform-Sequencing) や、TSS-Seq (Transcription Start Site Sequencing) などの利用も新たな耐性因子の発見に有用であると考えられる。筆者らのグループでは CAGE (Cap Analysis of Gene Expression) 法によってシロイヌナズナの遺伝子の転写開始点をゲノムワイドに一括同定している (Tokizawa *et al.*, 未発表). CAGE では、転写された RNA 分子の 5' 末端に形成されるキャップ構造を手掛かりに転写開始点を含む cDNA 断片を特異的に選抜し、NGS によってゲノム上の位置を特定する。この解析でシロイヌナズナのタンパク質をコードする全遺伝子の 80% 程度の転写開始点を同定できており、前述したシスエレメントの予測などに用いている。

一方、マイクロアレイはトランスクリプトーム解析の標準方法としての座を RNA-Seq に譲り渡しつつあるが、数千枚以上に蓄積されたマイクロアレイデータはそれだけで遺伝子機能や遺伝子間の関連性を予測するための有効なデータセットとなっている。ARANET (Lee *et al.*, 2015) や ATTED-II (Aoki *et al.*, 2015) のような遺伝子共発現データベースを利用することによって、共通の生物学的機構やパスウェイに属する遺伝子群を推定することができる。筆者らのグループでは、根を対象としたマイクロアレイ 660 枚のデータを用いて遺伝子共発現ネットワーク解析を行い、14,513 遺伝子の共発現関係を構成する 16 個の共発現モジュールを形成した (Kobayashi *et al.*, 2016). 遺伝学的解析等で選抜した耐性遺伝子をこの共発現ネットワークに落とし込むことでそれらの機能推定が可能となる。前述の塩耐性に関する GWAS で検出した遺伝子群の一部は、共発現遺伝子と共にスベリン生合成に関わる遺伝子ネットワークを構成していたことから、カスバリー線の構築/維持能力が植物根における塩耐性に重要であることが推定された。

植物の土壤ストレス耐性や土壤からの水分や養分の吸収には、土壤微生物や菌根菌、根粒菌が介在していることが知られており、植物と土壤の相互作用を解明するためには土壤中の微生物群の挙動の解析も重要な要因となる。NGS の特長の一つに、混在している複数の生物種の核酸を同時

にシーケンスした後に *in silico* で各生物種の参照ゲノム配列または転写配列にマッピングすることによって、発現量を生物種ごとに測定できる点がある。土壌中または植物根表面に存在する微生物群衆に対してメタゲノムまたはメタトランスクリプトーム解析を行う、または、菌根菌や病原菌が感染した植物試料に対して dual RNA-Seq の適用により菌と宿主の遺伝子発現量を同時にプロファイリングすることで、植物の根と相互作用する微生物の両者の挙動を一括して捉えることが期待されている (Westermann *et al.*, 2016)。

6. メタボロームデータの活用事例

生体内の低分子代謝物を網羅的に検出することで得られるメタボロームデータは、速度（フロー）を扱うトランスクリプトームデータと異なり、代謝経路内に存在する化合物の蓄積量（プール）を扱うことを特徴とする。最も一般的に用いられるのは、トランスクリプトームとの併用により代謝パスウェイの変化を検出することで代謝プログラムモデルの構築を行う場合である。筆者らのグループでは、低 pH ストレス条件下で野生株と酸感受性 *stop1* 変異体のトランスクリプトームとメタボロームを比較し、細胞内 pH 調節機能をもつ代謝経路であるリンゴ酸脱水素酵素から始まる Biochemical pH Stat 経路と α -ケトグルタル酸からグルタミン酸への合成から始まる GABA-shunt 経路が、*stop1* 変異体で抑制されていることを明らかにすることで、それらの経路への STOP1 の関与を特定した (Sawaki *et al.*, 2009)。

また、土壌ストレス研究におけるメタボロームデータは、ストレスに対する代謝応答や耐性関連遺伝子の特定だけでなく (Obata and Fernie, 2012)、植物の養分状態やストレス状態のバイオマーカーとしての用途 (Sulpice *et al.*, 2013) や、mGWAS (metabolome-based-GWAS) による育種手法としての用途が期待される (Hong *et al.*, 2016)。メタボローム解析では、アミノ酸、核酸、糖、有機酸、脂質、ステロール、二次代謝物質など数百種程度の定量データが得られるが、現状では、測定する化合物の特性に合わせて複数のプラットフォーム (GC/MS, LC/MS, CE/MS など) を使い分ける必要があり、また、未同定のピークも多く検出される。今後の測定方法・機器の改良開発やアノテーション情報の充実による適用範囲の拡大が期待される。

7. フェノーム解析の重要性

ここ数年で、大規模フェノタイピングとオミクスデータ解析を組み合わせた論文を目にすることが非常に多くなっている (Slovak *et al.*, 2014; Clauw *et al.*, 2015)。近年、急速に向上した計測可能な生体内情報の範囲の拡張速度と比べて、測定できる試料と形質の数が限られていることがボトルネックになっていると認識されることとなった。このような課題を克服するために、形質測定のハイスループット化が取り組まれており、フェノミクスと呼ばれる学問領

域が形成され、数多くのフェノームプロジェクトが進行している (Houle *et al.*, 2010)。主にロボットとカメラの利用による表現型の経時的取得とそのハイスループット化が基本的なアプローチであるほか、野外においては衛星画像や LIDAR (Light Detection and Ranging) などのリモートセンシング技術を利用して、作物の葉のクロロフィル含量や土壌の物理的・化学的性質を推定することも行われている。これらのデータを天候等の時系列データと統合した上で機械学習等の統計計算を行うことにより、作物のストレス状況の計測が試みられている (Singh *et al.*, 2016)。

8. 大規模データ解析のための計算機リソース

本稿で紹介したような大規模データを解析することによって新たなインサイトを得るアプローチは、非常に魅力的な手法でありながら、以前は大規模データ解析に必要とされるリソースが、高スペック計算機の確保、大量データのハンドリング技術、統計モデリングの構築スキルなど、これまでの生物系研究室・企業が持つ資源と大幅に異なっていたために容易に取り組めない分野であった。近年では、ハードウェア/ソフトウェアの高度化、web ベースツールやクラウド環境の充実によってその垣根も随分と低くなってきたことから、多くの研究者・実務家が取り組みやすい環境が整ってきている。NGS 解析を例にとると、筆者らは古典的な手法として、目的に応じて無償のオープンソフトウェアから解析パイプラインを構築しているが、現在では、国立遺伝学研究所の DDBJ Read Annotation Pipeline (Nagasaki *et al.*, 2013)、イルミナの BaseSpace (basespace.illumina.com) などのクラウド環境の利用、CLC Genomics Workbench (CLCBio-Qiagen) や Strand NGS (Agilent Technologies) のような有償ソフトウェアの利用、国内外に多数存在している解析受託ビジネスを展開している企業の利用も有力な選択肢の一つとなる。

9. 大規模データ解析が今後の土壤肥科学にもたらすもの

植物と土壌の複雑な相互作用を解明するために、本稿で紹介したようなデータ駆動型アプローチに対する期待は益々高まっていくことが予想される。最後に、大規模データ解析が土壤肥科学の発展に貢献するにあたって、今後特に重要になると考えられる研究展開を述べたい。

一つ目は、冒頭に挙げた複数の環境ストレスに対する植物応答の理解である。これまでのオミクス解析の多くは個別のストレスについて情報が収集されてきた。しかし、同時に複数のストレスを受けた植物の遺伝子発現応答は、それぞれ単独のストレスに対する遺伝子発現の相加的な効果だけでは説明できず、比較的少数の特異的な遺伝子発現変動によって複合ストレスに応答することが分かってきた (Rasmussen *et al.*, 2013; Sewelam *et al.*, 2014)。この応答の非線形性がシグナル伝達のクロストークや多機能遺伝子を伴うストレス耐性システムを解き明かすための重要なヒントとなると考えられる。PheWAS (Phenome-wide association study)

のような発想で、個々の変異が影響する形質の組み合わせを解析することもその一助となるであろう。

二つ目は、プロファイル（パターン）による予測である。これまでの研究報告では、形質に影響を与える少数の遺伝子や化合物、環境要因を特定するために大規模データ解析が用いられている場合が多い。しかしながら、本稿で述べたようにある遺伝子や代謝物が複数の役割をもつことを考慮すると、それらの分子の絞り込みを行わずに、オミクスプロファイルをそのまま形質の診断・判定に用いる手法にも一定の合理性と利点がある。GSによる迅速な育種、メタボリック・プロファイリング法による農作物品質管理（Fujimura *et al.*, 2011）、土壌微生物のメタゲノム解析による土壌診断などが今後の農林産業などの実務を中心に拡大していくと考えられる。計測コストの低下、MinION（Oxford Nanopore Technologies）などの現場での測定が可能な機器の普及（Benítez-Páez *et al.*, 2016）、機械学習に基づく予測モデルを利用した農業生産の一般化などがこれを後押しすると考えられる。

三つ目は、分野／部門を超えた大規模データの統合である。現在、様々な環境データと生物データが世界中で収集され、それぞれのデータベースを構成している。これらのデータの統合によりそれぞれの環境データとオミクスデータを有機的に関連付けることによって、新しい知見を得ることができる。シロイヌナズナでは、野生系統の採取地気候データと高密度 SNP データを比較することで、降水量や温度と相関の高い SNP 近傍の遺伝子群が局所適応の原因として選抜された（The 1001 Genomes Consortium 2016）。また、実際に水田で栽培されたイネについて、環境データから遺伝子発現予測モデルを構築した例もある（Nagano *et al.*, 2012）。これらの統合解析から得られる知見は、農業生産や緑化における栽培品種の選定、適切な施肥管理のような実務面に直接繋げることも可能であり、農業生産の発展に大きく寄与する可能性を秘めている。

私たちに見える世界は定量可能な範囲に限定される。大規模データは未知の世界に向かって私たちの視野を拡げ、種々のデータ解析手法は羅針盤として向かうべき方向性を指し示すことで、私たちをこれまでの研究成果に基づく経験や直感に頼る世界から一歩先に連れて行ってくれる。我々はこれらのデータやツールを頼りに、目に見えない本質に迫る洞察力を手にすることができるのである。

謝 辞：本稿で紹介した研究の一部は、JSPS 科研費（22248006, 24688009）及び、特別研究員奨励費（10J04610, 13J08738）の助成を受けたものです。

文 献

Aoki, Y., Okamura, Y., Tadaka, S., Kinoshita, K., and Obayashi, T. 2015. ATTED-II in 2016: A plant coexpression database towards lineage-specific coexpression. *Plant Cell Physiol.*, 57, e5(1-9).

Benítez-páez, A., Portune, K. J., and Sanz, Y. 2016. Species-level resolution of 16S rRNA gene amplicons sequenced through

the MinION TM portable nanopore sequencer. *Gigascience*, 5, 1-9.

Clauw, P., Coppens, F., De Beuf, K., Dhondt, S., Van Daele, T., Maleux, K., Storme, V., Clement, L., Gonzalez, N., and Inzé, D. 2015. Leaf Responses to Mild Drought Stress in Natural Variants of *Arabidopsis*. *Plant Physiol.*, 167, 800-816.

Cubillos, F. A., Stegle, O., Grondin, C., Canut, M., Tisné, S., Gy, I., and Loudet, O. 2014. Extensive cis-regulatory variation robust to environmental perturbation in *Arabidopsis*. *Plant Cell*, 26, 4298-4310.

Desta, Z. A. and Ortiz, R. 2014. Genomic selection : Genome-wide prediction in plant improvement. *Trends Plant Sci.*, 19, 592-601.

Ding, Z. J., Yan, J. Y., Xu, X. Y., Li, G. X., and Zheng, S. J. 2013. WRKY46 functions as a transcriptional repressor of *ALMT1*, regulating aluminum-induced malate secretion in *Arabidopsis*. *Plant J.*, 76, 825-35.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., and Mitchell, S. E. 2011. A Robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, 6, e19379.

Famoso, A. N., Zhao, K., Clark, R. T., Tung, C. -W., Wright, M. H., Bustamante, C., Kochian, L. V., and McCouch, S. R. 2011. Genetic architecture of aluminum tolerance in rice (*Oryza sativa*) Determined through genome-wide association analysis and QTL mapping. *PLoS Genet.*, 7, e1002221.

Fujimura, Y., Kurihara, K., Ida, M., Kosaka, R., and Miura, D. 2011. Metabolomics-driven nutraceutical evaluation of diverse green tea cultivars. *PLoS One*, 6, e23426.

Gifford, M. L., Banta, J. A., Katari, M. S., Hulsman, J., Chen, L., Ristova, D., Tranchina, D., Purugganan, M. D., Coruzzi, G. M., and Birnbaum, K. D. 2013. Plasticity regulators modulate specific root traits in discrete nitrogen environments. *PLoS Genet.*, 9, e1003760.

Hong, J., Yang, L., Zhang, D., and Shi, J. 2016. Plant metabolomics : An indispensable system biology tool for plant science. *Int. J. Mol. Sci.*, 17, 767.

Houle, D., Govindaraju, D. R., and Omholt, S. 2010. Phenomics : the next challenge. *Nat. Rev. Genet.*, 11, 855-866.

Iuchi, S., Koyama, H., Iuchi, A., Kobayashi, Y., Kitabayashi, S., Kobayashi, Y., Ikka, T., Hirayama, T., Shinozaki, K., and Kobayashi, M. 2007. Zinc finger protein STOP1 is critical for proton tolerance in *Arabidopsis* and coregulates a key gene in aluminum tolerance. *Proc. Natl. Acad. Sci. U. S. A.*, 104, 9900-9905.

Jin, J., Liu, J., Wang, H., Wong, L., and Chua, N. H. 2013. PLncDB : plant long non-coding RNA database. *Bioinformatics*, 29, 1068-1071.

Kobayashi, Y., Hoekenga, O. A., Itoh, H., Nakashima, M., Saito, S., Shaff, J. E., Maron, L. G., Piñeros, M. A., Kochian, L. V., and Koyama, H. 2007. Characterization of *AtALMT1* expression in aluminum-inducible malate release and its role for rhizotoxic stress tolerance in *Arabidopsis*. *Plant Physiol.*, 145, 843-52.

Kobayashi, Y., Lakshmanan, V., Kobayashi, Y., Asai, M., Iuchi, S., Kobayashi, M., Bais, H. P., and Koyama, H. 2013a. Overexpression of *AtALMT1* in the *Arabidopsis thaliana* ecotype Columbia results in enhanced Al-activated malate excretion and beneficial bacterium recruitment. *Plant Signal. Behav.*, 8, 37-41.

- Kobayashi, Y., Sugimoto, M., Lakshmanan, V., Iuchi, S., Kobayashi, M., Bais, H. P., and Koyama, H. 2013b. Characterization of the complex regulation of *AtALMT1* expression in response to phytohormones and other inducers. *Plant Physiol.*, **162**, 732-740.
- Kobayashi, Y., Ohyama, Y., Kobayashi, Y., Ito, H., Iuchi, S., Fujita, M., Zhao, C. R., Tanveer, T., Ganesan, M., Kobayashi, M., and Koyama, H. 2014. STOP2 activates transcription of several genes for Al- and low pH-tolerance that are regulated by STOP1 in *Arabidopsis*. *Mol. Plant*, **7**, 311-322.
- Kobayashi, Y., Sadhukhan, A., Tazib, T., Nakano, Y., Kusunoki, K., Kamara, M., Chaffai, R., Iuchi, S., Sahoo, L., Kobayashi, M., Hoekenga, O. A., and Koyama, H. 2016. Joint genetic and network analyses identify loci associated with root growth under NaCl stress in *Arabidopsis thaliana*. *Plant Cell Environ.*, **39**, 918-934.
- Lakshmanan V., Kitto S. L., Caplan J. L., Hsueh Y. H., Kearns D. B., Wu Y. S., and Bais H. P. 2012. Microbe-associated molecular patterns-triggered root responses mediate beneficial rhizobacterial recruitment in *Arabidopsis*. *Plant Physiol.*, **160**, 1642-1661.
- Lee, T., Yang, S., Kim, E., Ko, Y., Hwang, S., Shin, J., Shim, J. E., Shim, H., Kim, H., Kim, C., and Lee, I. 2015. AraNet v2: An improved database of co-functional gene networks for the study of *Arabidopsis thaliana* and 27 other nonmodel plant species. *Nucleic Acids Res.*, **43**, D996-D1002.
- Leiser, W. L., Rattunde, H., Weltzien, E., Cisse, N., Abdou, M., Diallo, A., Touré, A. O., Magalhaes, J. V. and Haussmann, B. 2014. Two in one sweep : aluminum tolerance and grain yield in P-limited soils are associated to the same genomic region in West African Sorghum. *BMC Plant Biol.*, **14**, 206.
- Nagano A. J., Sato Y., Mihara M., Antonio B. A., Motoyama R., Itoh H., Nagamura Y., Izawa T. 2012. Deciphering and prediction of transcriptome dynamics under fluctuating field conditions. *Cell*, **151**, 1358-1369.
- Nagasaki, H. I., Mochizuki, T. A., Kodama, Y. U., Saruhashi, S. A., and Morizaki, S. H. 2013. DDBJ read annotation pipeline : A cloud computing-based pipeline for high-throughput analysis of next-generation sequencing data. *DNA Res.*, **20**, 383-390.
- Negishi, T., Oshima, K., Hattori, M., Kanai, M., Mano, S., Nishimura, M., and Yoshida, K. 2012 Tonoplast- and plasma membrane-localized aquaporin-family transporters in blue hydrangea sepals of aluminum hyperaccumulating plant. *PLoS One*, **7**, e43189
- Obata, T. and Fernie, A. R. 2012. The use of metabolomics to dissect plant responses to abiotic stresses. *Cell. Mol. Life Sci.*, **69**, 3225-3243.
- Ogo, Y., Itai, R. N., Nakanishi, H., Inoue, H., and Kobayashi, T. 2006. Isolation and characterization of IRO2, a novel iron-regulated bHLH transcription factor in graminaceous plants. *J. Exp. Bot.*, **57**, 2867-2878.
- Qin, P., Wang, L., Liu, K., Mao, S., Li, Z., Gao, S., Shi, H., and Liu, Y. 2015. Genomewide association study of *Aegilops tauschii* traits under seedling-stage cadmium stress. *Crop J.*, **3**, 405-415.
- Rasmussen, S., Barah, P., Suarez-Rodriguez, M. C., Bressendorff, S., Friis, P., Costantino, P., Bones, A. M., Nielsen, H. B., and Mundy, J. 2013. Transcriptome responses to combinations of stresses in *Arabidopsis*. *Plant Physiol.*, **161**, 1783-94.
- Rudrappa, T., Czymmek, K. J., Pare, P. W., Paré, P. W., and Bais, H. P. 2008. Root-secreted malic acid recruits beneficial soil bacteria. *Plant Physiol.*, **148**, 1547-1556.
- Sakuma, Y., Maruyama, K., Qin, F., Osakabe, Y., Shinozaki, K., and Yamaguchi-Shinozaki, K. 2006. Dual function of an *Arabidopsis* transcription factor DREB2A in water-stress-responsive and heat-stress-responsive gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 18822-7.
- Sawaki, Y., Iuchi, S., Kobayashi, Y., Kobayashi, Y., Ikka, T., Sakurai, N., Fujita, M., Shinozaki, K., Shibata, D., Kobayashi, M., Koyama, H. 2009. STOP1 regulates multiple genes that protect *Arabidopsis* from proton and aluminum toxicities. *Plant Physiol.*, **150**, 281-294.
- Sewelam, N., Oshima, Y., Mitsuda, N., and Ohme-Takagi, M. 2014. A step towards understanding plant responses to multiple environmental stresses : a genome-wide study. *Plant Cell Environ.*, **37**, 2024-2035.
- Singh, A., Ganapathysubramanian, B., Singh, A. K., and Sarkar, S. 2015. Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci.*, **21**, 110-124.
- Slovak, R., Goschl, C., Su, X., Shimotani, K., Shiina, T., and Busch, W. 2014. A scalable open-source pipeline for large-scale root phenotyping of *Arabidopsis*. *Plant Cell*, **26**, 2390-2403.
- Sonah, H., Donoughue, L. O., Cober, E., and Rajcan, I. 2015. Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. *Plant Biotechnol. J.*, **13**, 211-221.
- Sulpice, R., Nikoloski, Z., Tschöep, H., Antonio, C., Kleessen, S., Larhlmi, A., Selbig, J., Ishihara, H., Gibon, Y., Fernie, A. R., and Stitt, M. 2013. Impact of the carbon and nitrogen supply on relationships and connectivity between metabolism and biomass in a broad panel of *Arabidopsis* accessions. *Plant Physiol.*, **162**, 347-63.
- The 1001 Genomes Consortium. 2016. 1,135 Genomes Reveal the Global Pattern of Polymorphism in *Arabidopsis thaliana*. *Cell*, **166**, 1-11.
- Tokizawa, M., Kobayashi, Y., Saito, T., Kobayashi, M., Iuchi, S., Nomoto, M., Tada, Y., Yamamoto, Y. Y., and Koyama, H. 2015. Sensitive to proton rhizotoxicity1, calmodulin binding transcription activator2, and other transcription factors are involved in *aluminum-activated malate transporter1* expression. *Plant Physiol.*, **167**, 991-1003.
- Westermann, A. J., Förstner, K. U., Amman, F., Barquist, L., Chao, Y., Schulte, L. N., Müller, L., Reinhardt, R., Stadler, P. F., Vogel, J. 2016. Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. *Nature*, **529**, 496-501.
- Yamamoto, Y. Y., Yoshioka, Y., Hyakumachi, M., Maruyama, K., Yamaguchi-Shinozaki, K., Tokizawa, M., and Koyama, H. 2011. Prediction of transcriptional regulatory elements for plant hormone responses based on microarray data. *BMC Plant Biol.*, **11**, 39.
- Zhang, X., Pérez-Rodríguez, P., Semagn, K., Beyene, Y., Babu, R., López-Cruz, M. A., San Vicente, F., Olsen, M., Buckler, E., Jannink, J. L., Prasanna, B. M., and Crossa, J. 2015. Genomic prediction in biparental tropical maize populations in water-stressed and well-watered environments using low-density and GBSSNPs. *Heredity*, **114**, 291-299.