

オミックスアプローチと高速DNAシーケンシング技術が解き明かす次世代の植物科学

The Next-generation Plant Science Revealed by the Omics Approaches and High-throughput DNA Sequencing Technologies

大柳 一*

Hajime Ohyanagi

近年の実験分子生物学における手法・技術のブレイクスルーによって、広くライフサイエンス分野全般において新しい研究スタイルが確立しつつある。本稿では「オミックスアプローチ」と「高速DNAシーケンシング技術」を解説し、植物科学分野でこれらの技術を活用して得られた成果の中から筆者の関わったものをいくつか紹介したい。

According to the recent breakthrough in experimental molecular biology, new fashions for research project, in terms of broad life sciences, have been established. Here I summarize the Omics Approaches and High-throughput DNA Sequencing Technologies, and introduce a couple of my research outcomes achieved by taking advantage of these technologies.

1. まえがき

我が国を象徴する穀物であり最重要農業植物とされるイネのゲノム学研究においては、(独)農業生物資源研究所を中心とした国際共同研究プロジェクトによる*Oryza sativa* ssp. *japonica* cv. Nipponbareの高精度全ゲノム塩基配列解読達成⁽¹⁾ならびにその網羅的遺伝子アノテーションの成果^{(2), (3), (4)}が広く植物科学全般研究におけるゲノム情報基盤として普及し、その基盤に基づいたポストゲノム研究の時代に入って久しい。加えて近年のオミックス解析技術とDNA配列解読技術の飛躍的発展により、多種多様な種・亜種・栽培品種のオミックス情報と全ゲノム情報が比較的容易に得られるようになり、植物科学研究はまた新たな局面を迎えていると考える。本稿ではまずオミックスアプローチならびに高速DNAシーケンサによるゲノム解読の基礎知識を解説し、その技術を活用した研究プロジェクトの例として、筆者が取り組んでいるプロジェクトの中からダイズとイネに関する植物科学研究を紹介したい。

2. オミックスとは何か

今日のライフサイエンスの爆発的発展は、「ゲノムDNA」の配列解読達成を抜きにしては語れない。しかし我々ヒトを含むすべての生物種の体内・細胞内で日々起こっている生化学反応の大半は、ゲノムDNAそのものではなくそこから転写・翻訳されて生じた「メッセンジャーRNA」や「タンパク質」と呼ばれる分子が存在

してこそ成り立つものである(図1)。ゲノムDNAはいわばただの「設計図」であり、そこから実際の機能を持つ「部品」であるタンパク質を、適期に適量だけ生産する仕組みこそが生命体という複雑なシステムを支える重要な原理である。オミックスは「総体」を意味する「-ome」という英語接尾辞から造られた造語であり(図1)、ゲノムDNAのみならずメッセンジャーRNAやタンパク質、さらには生体内代謝の結果生じた小分子(メタボライト)など、各種生体内分子の種類・量を最新の解析機器を用いて網羅的に直接計測し、その動態から生命現象を直接理解しようとする学問領域のことである。オミックス各層の情報を串刺しかつ網羅的に統合・閲覧することができれば(図2)、生命現象をより深く理解できるようになると期待される。

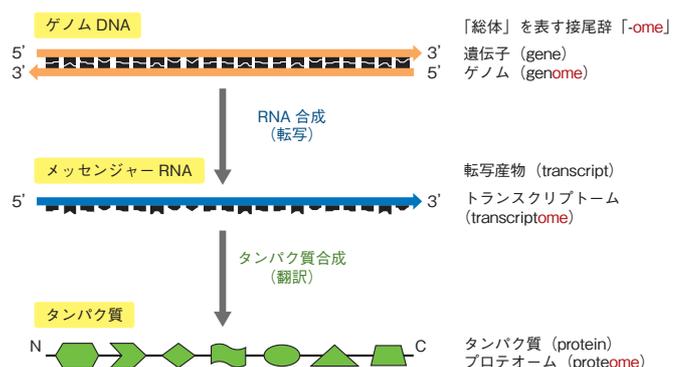


図1 遺伝情報のセントラルドグマ(中心原理)とオミックス階層

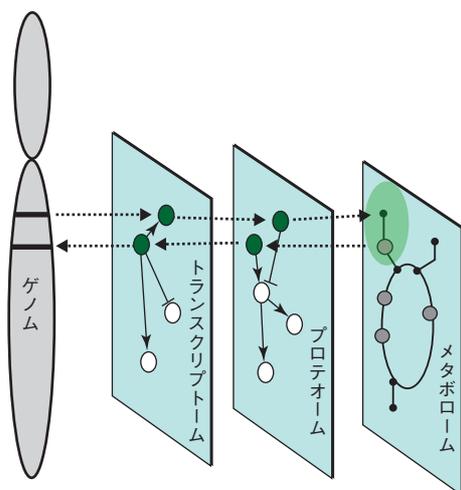


図2 オミックス階層を越えた情報の統合

3. オミックスアプローチによるダイズ湿害耐性研究

3.1 ダイズの湿害問題

ダイズは日本の食文化に欠かせない重要農業植物でありながらも、その国内自給率の低さ（約5%）がかねてより指摘されており、これは我が国農業における重要な問題の一つとされている。農業技術的にダイズは出芽期の冠水（栽培畑に水が過剰な状態）に弱く、播種期が梅雨期と重なるうえに水持ちの良すぎる傾向にある水田転換畑で転作作物としてダイズが栽培されることの多い我が国では、この湿害は自給率向上のために克服されるべき課題である。

3.2 ダイズプロテオームデータベース構築とオミックステーブルへの発展

筆者が参加する共同研究グループは、ダイズが冠水ストレスを受けた際のタンパク質群（プロテオーム）の量的変動を測定し、湿害耐性をダイズに与えるために必要な因子についての知見を得ることを目的に、その解析結果をWeb上にデータベースとしてまとめ、広く一般に公開した^{(5), (6)}(図3)。さらに我々は、メッセンジャーRNA（トランスクリプトーム）と生体内小分子（メタボローム）の量的変動も測定し、オミックステーブルとして串刺しでの情報閲覧が可能なインターフェースを開発した(図3)。今後はオミックス各層の情報網羅性を高め、また冠水ストレス以外の条件で得られた情報や新たな実験技術も活用しながら、さらなるデータの蓄積と有用因子の発見を狙っていく。

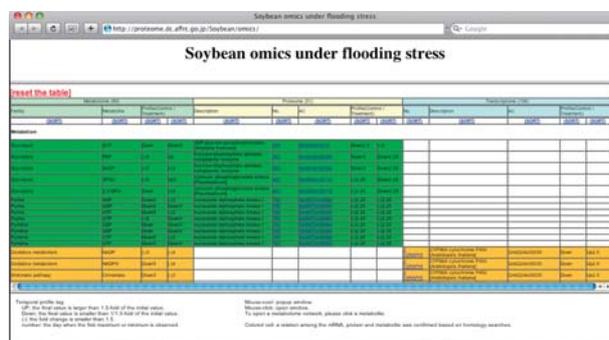
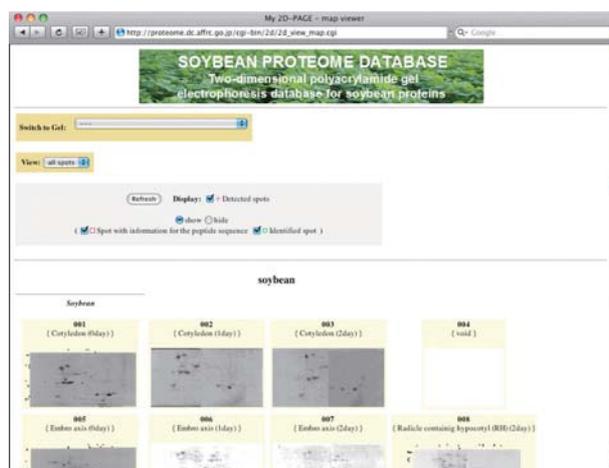


図3 ダイズプロテオームデータベースの主画面（上）とオミックステーブル（下）
Soybean proteome databaseホームページ
(<http://proteome.dc.affrc.go.jp/Soybean/>) より引用

4. 高速DNAシーケンサによるゲノム解読の基礎知識

4.1 リファレンスマッピング

DNAシーケンサはオミックス階層(図1)におけるDNA配列(ゲノム)を解読するための解析装置であるが、人工的なRNA→DNA逆転写技術を利用してメッセンジャーRNA配列を解読することも可能であり、またゲノム中のDNAメチル化部位や転写因子結合部位探索等の高次生命現象に関連した網羅的解析にも活用できることから、現在のライフサイエンスにおける最重要技術の一つであることは間違いない。今日の最新鋭機器の原形となるような自動DNAシーケンサは1980年代にすでに実用化されていたが、2005年頃から時間当たり・コスト当たりの解読可能塩基数に飛躍的な向上が見られている。この高速DNAシーケンサによる解読リード長は長鎖化の方向に進んでいるものの、2012年時点の技術ではまだ短鎖配列(数百bp以下)を大量に解読する能力にこそ長けており、よってゲノム解読研究においては「リファレンスマッピング」の手法を用いた解析との相性が非常に良い(図4)。イネ分野においては前述のNipponbareゲノムをリファレンスとして利用し、栽培品種間や亜種・近縁種

間のゲノム多型情報解析による進化学・育種学研究が数多く進められるようになってきた。

4.2 デノボアセンブル

ただし非モデル植物などの近縁ゲノムが得られない種については、仮にある程度近縁なゲノム解読済みの種をリファレンスとして準備しても、ゲノムDNA同士の進化距離がそれほど近くないためリードマップ率が低下し、上述のリファレンスマッピングの手法がそれほど有効でない場合が多い。この場合、高速DNAシーケンサのみでのゲノム復元を目指して、もう一つの代表的手法である「デノボアセンブル」を採ることになる(図4)。こちらの手法はDNA断片配列同士の相互類似性を検索し、それらを「のりしろ」として元のゲノム配列を復元していくことを基本原理としているため、本来は短鎖配列を利用した解析には不向きなものであるが、現在ではインサートサイズの異なる複数のライブラリ(ペアエンド配列・メイトペア配列)の組み合わせ手法、そして優れたアセンブルアルゴリズムと強力な計算機資源を利用して、数多くのチャレンジがなされつつある。

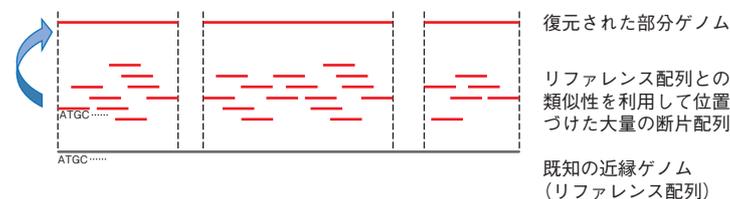
5. 野生イネゲノム解読への挑戦

5.1 リファレンスマッピングによる近縁野生イネゲノム解読

現在人類によって栽培されているイネ*Oryza sativa*は、長い歴史の中で的人為的な栽培化と選抜の過程によって、今の優れた農業形質(多収性・病害や冷害への抵抗性など)を獲得してきたと考えられている。イネ栽培化の地理的起源とその後の拡散については科学的論争があり定説が得られていなかったが、今回筆者の所属する国際共同研究グループは、主にリファレンスマッピングの技術を活用して、栽培イネ(*Oryza sativa*) 1,083系統とその直接の祖先種とされる野生イネ(*Oryza rufipogon*) 446系統のゲノムを大規模解読・進化系統解析し、イネ栽培化は「中国の珠江中流域」にて一度だけ起こったことを科学的に示すことに成功した⁹⁾。これは長年の論争に終止符を打つ、大変インパクトの大きい研究成果であった。

「リファレンスマッピング」と「デノボアセンブル」

リファレンスマッピング



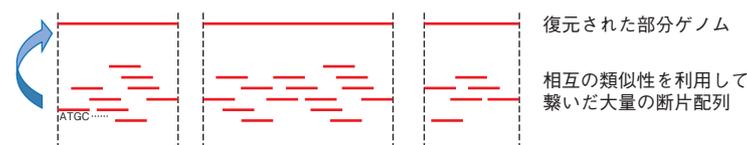
長所

- ・比較的少なめの断片配列からゲノム全域を復元できる可能性がある
- ・要求される計算機パワーが少なめである

短所

- ・近縁のゲノムが既知でなくてはならない
- ・ギャップが残ってしまう

デノボアセンブル



長所

- ・近縁の既知ゲノムが不要で、手がかりの全くないゲノムを復元できる可能性がある

短所

- ・精度よく復元するためには断片配列を数多く解読する必要がある
- ・強力な計算機、特に大容量メモリ(256~512GB以上)が要求される
- ・ギャップが残ってしまう上、それぞれの部分ゲノムの位置が不明

図4 高速DNAシーケンサによる2つのゲノム解読手法

5.2 デノボアセンブルによる遠縁野生イネゲノム解読

さらに我々は、育種的遺伝子リソースとしての役割を期待し、病害抵抗性をはじめとする有用農業形質に関連する遺伝背景の解明・応用を目指して、栽培イネ*Oryza sativa*からは遠縁で、CC型と呼ばれるゲノムを持つ野生イネ*Oryza officinalis*の全ゲノム塩基配列解読に取り組んでいる。*Oryza officinalis*とその近縁の野生イネ種は、その育種学的意義に加えて広い地理的分布や一部の種の四倍体性などから進化的にも興味深く、そのリファレンスゲノムを確立することは次世代の育種学研究と進化学研究にとって意義深い成果であると考えられる。現在（2012年末）までに、我々は6種類のインサート長の異なるライブラリ（ペアエンド配列・メイトペア配列）をilluminaシーケンサによって解読し、Ray[®]やPlatanus[®]といった最新のアセンブルプログラム等を用いてデノボアセンブル手法でゲノム配列を復元している。現在までの検証作業により、このアセンブル結果はかなりの正確性を持っていることが示唆されている⁽¹⁰⁾。今後は新たなライブラリ追加によるアセンブル改善と、米国アリゾナ大学を中心に推進されている*Oryza*属の全ゲノム配列解読を目指した国際プロジェクト「I-OMAP (International *Oryza* Map Alignment Project)」との共同作業により、*Oryza officinalis*全ゲノム配列を完成に近づけていく予定である。

6. むすび

高速DNAシーケンサの技術革新はまだまだ続き、時間当たり・コスト当たりの解読塩基数は今後も急増していくことが予想される。またオミックスアプローチについてもその技術発展は続くであろう。これを受けて将来的におそらく今より多くのライフサイエンス分野の研究者が、これらの手法が産出する極めて大規模なデータと向き合っていくことになるだろう。この新しい研究スタイルは伝統的なライフサイエンス分野の科学者が持つ能力とは異なる対応力（情報処理技術・解析技術）を要求するものであり、そのデータハンドリング・解析におけるソリューションビジネスは拡大の方向にむかうことが予想される。今後も筆者は良き共同研究者とのパートナーシップを発展させつつ、お客様の研究活動進展ならびに広く社会貢献をミッションとして、国際的にも一流と評価される高いレベルの研究活動をソリューションとして提供していきたいと考えている。

共同研究者である 国立遺伝学研究所・(独) 農業生物資源研究所・(独) 農研機構 作物研究所・(独) 放射線医学総合研究所・(公) 前橋工科大学・明治大学・米国アリゾナ大学・米国メイン州立大学・中国科学院の関係者様、ならびに社内関係各位に感謝する。

用語解説

[*Oryza sativa* ssp. *japonica* cv. Nipponbare]

(独) 農業生物資源研究所を中心とした国際共同研究プロジェクトが高精度全ゲノム塩基配列決定時に採用した栽培品種（日本晴・にっぼんばれ）。ジャポニカ種に分類される。かつては我が国の作付面積第1位を誇る品種であったが、近年作出された他の優良品種に押され、現在の作付面積割合は大きくない。

[*Oryza officinalis*]

栽培イネの代表種である*Oryza sativa*とは同じ属（*Oryza*属）に含まれるものの、かなり遺伝的に遠縁と考えられている野生イネ種。CC型と呼ばれるゲノムをもち、またこの*officinalis*と遺伝的関連が示唆されている野生イネ種がいくつかあるため、遠縁野生イネ研究の第一歩として筆者らのグループが注目している。

[illuminaシーケンサ]

2012年末現在で最も普及している高速DNAシーケンサ。米国illumina社の製品群。1塩基合成反応（SBS：Sequencing by Synthesis）と名付けられた新技術により前世代のDNAシーケンサとは桁違いの解読能力を実現している。現在も新製品のリリースを重ねており、解読能力は年々向上している。

参考文献

- (1) The International Rice Genome Sequencing Project. (2005) The map-based sequence of the rice genome. *Nature*, 436, 793–800.
- (2) Ohyanagi, H. *et al.* (2006) The Rice Annotation Project Database (RAP-DB) : hub for *Oryza sativa* ssp. *japonica* genome information. *Nucleic Acids Res.*, 34, D741–D744.
- (3) Rice Annotation Project. (2007) Curated genome annotation of *Oryza sativa* ssp. *japonica* and comparative genome analysis with *Arabidopsis thaliana*. *Genome Res.*, 17, 175–183.
- (4) Rice Annotation Project. (2008) The Rice Annotation Project Database (RAP-DB) : 2008 update. *Nucleic Acids Res.*, 36, D1028–D1033.
- (5) Ohyanagi, H. *et al.* (2009) Soybean proteome database: a data resource for plant differential omics. *J Proteome Res.*, 8, 3539–3548
- (6) Ohyanagi, H. *et al.* (2012) Soybean Proteome Database 2012: update on the comprehensive data

-
-
-
-
-
-
-
-
-
-
- repository for soybean proteomics. *Front Plant Sci.*, 3, 110
- (7) Huang X. *et al.* (2012) A map of rice genome variation reveals the origin of cultivated rice. *Nature*, 490, 497–501.
- (8) Boisvert, S. *et al.* (2010) Ray: Simultaneous Assembly of Reads from a Mix of High-Throughput Sequencing Technologies. *J Comput Biol.* 17: 1519–1533.
- (9) Platanus: <http://platanus.bio.titech.ac.jp/> (An assembler for highly heterozygous genomes, by Tokyo Institute of Technology)
- (10) 大柳 一 (2012)「次世代育種に向けた取り組み～高速DNAシーケンシングとデータ解析の遺伝学～」日本育種学会第122回講演会・ワークショップ・次世代の遺伝学と育種 (招待講演)