



筑波大学遺伝子実験センター
形質転換植物デザイン研究拠点
研究セミナー(18)



日時: 3月13日 14:00 - 16:00

場所: 遺伝子実験センター内セミナー室 (2階)

新規有用遺伝子探索に向けた大規模オミックス情報の活用法

矢野健太郎 准教授 (明治大学・農学部)

大規模なオミックス情報が、新規有用遺伝子の探索や代謝パスウェイの推定、遺伝子発現制御機構の解明を促進することは明らかである。近年、多くのモデル植物において、ゲノム、トランスクリプトーム、メタボローム、フェノームなどのオミックス情報が Web データベースに蓄積し、利用可能となっている。同時に、Ontology に代表されるように、遺伝子やタンパク質などのアノテーションの枠組みも整備が進んでいる。また、情報基盤の整備に伴い、オミックス情報やアノテーション情報を解析するためのツールも多く開発・公開されている。これらの有用なデータベースとツールの特性を記述した Web サイトとして、SHared Information of GENetic Resources (SHIGEN) と *Nucleic Acids Research* 誌の Database Summary Paper, Bioinformatics Links Directory などがある。ここで、類似のオミックス情報を提供するデータベースが複数存在する場合には、提供情報の差異に留意すべきである。たとえば、シロイヌナズナの代謝パスウェイ情報は KEGG と AraCyc から利用できる。しかし、両者の提供情報は同一ではない。公開されているオミックス情報を十分に活用するためには、複数のデータベースの利用が不可欠である。また、データベースが提供する二次的な解析結果(例:塩基配列のアセンブルやマッピング、相同性検索)を利用する場合、解析した時期・更新情報にも留意が必要である。

大規模なオミックス情報から有用な情報を効率的に引き出すためには、大規模情報を汎用的な計算機リソースでも、短時間で処理できるバイオインフォマティクス解析基盤の整備が求められる。高速シーケンサーやマイクロアレイのハイスループット化は、網羅的なトランスクリプトーム解析を実現する。しかし、従来のバイオインフォマティクス解析手法では、これらの実験手法から得られた大規模情報を効率的に扱うことができない。さらに、解析結果も規模が大きいため、迅速な解釈が困難である。たとえば、遺伝子発現データ解析において広く用いられている階層的クラスタリングは、遺伝子(プローブ)数の増加に伴い、非現実的な計算リソース(計算時間や計算機メモリーなど)を要するため、データの大規模化に対応できない。階層的クラスタリングから得られるデンドログラムとヒートマップについても、その全体を計算機モニタなどで俯瞰できない。計算量やデンドログラムの大きさは、データ規模を小さくすることによって低減できる。そこで、データ規模を縮小するために、解析対象とする遺伝子を事前に選抜するなどの前処理が頻繁に適用される。しかし、このアプローチでは、目的とする遺伝子も解析対象から除去する危険を伴う。そこで、当プロジェクトでは、大規模トランスクリプトーム情報を汎用的なワークステーションでも短時間で解析し得る統計手法と視覚化法を開発している。これらの手法により、形質関連遺伝子群や発現パターンが類似する遺伝子群を迅速かつ簡便に抽出できる。

大規模なオミックス情報とバイオインフォマティクス解析手法は、生命システムの全体像の解明に向けたシステムズ・バイオロジー研究の基盤となる。植物科学や育種などのさらなる発展のためにも、バイオインフォマティクス研究者の育成と解析手法の開発が強く望まれる。

世話人 溝口 剛 mizoguchi@gene.tsukuba.ac.jp