

研究開発課題 事後評価結果

➤ 課題情報

研究開発課題名 「非モデル植物のための遺伝子ネットワーク情報活用基盤」

研究代表名 大林 武（東北大学 大学院情報科学研究科 教授）

➤ 事後評価結果：

本研究開発課題は、植物遺伝子の共発現データベース ATTED-II を基盤として、遺伝子の共発現情報から非モデル植物の未知の遺伝子機能の解明へつなげることを目的とし、共発現解析の高精度化、サンプル条件の可視化、種間比較などの機能強化のほか、公開基盤や他のデータベースとの連携機能の整備を重点的に実施するものである。

RNA-seq 共発現の解析パイプラインを再構築し、定量法をアライメントベースに変更することで、共発現情報の高精度化を図った。2024 年 9 月に公開した ATTED-II version 12.0 では、新たな植物種として主要穀類であるコムギとオオムギを追加し、対象植物種を 9 種から 11 種に拡張した。また、主成分解析に基づくサンプル再構成・可視化を実現する「CoexViewer」と「PC View」を開発・実装し、主成分ごとの生物学的説明を KEGG エンリッチメント解析と生成 AI により自動生成する機能を追加することで、共発現モジュールが機能するサンプル条件を提示できるようにした。マクロスケールの CoexMap (UMAP) では、外れ値補正・スケーリング調整、種選択インターフェイス、並列表示等を実装し、視認性を向上させ、異なる植物種間での共発現構造の比較しやすい形にした。メゾスケールの細胞内局在別 CoexMap は、現在、開発版が実装段階になっており、葉緑体・核等の局所モジュールにおいて、種間対応性が高いことが確認されている。植物ゲノムデータベース Plant GARDEN への直リンク、種固有の遺伝子 ID と Entrez ID の対応表の提供、Zenodo リポジトリによる恒久保存、API/RDF 整備等を実施し、外部データベースとの連携機能を向上させた。公開環境の整備では、新しくサーバを導入して開発系サーバと公開系サーバを物理的に分離し、SSD 運用、OS/WEB の更新、Ajax による非同期通信の導入等により、応答性・耐障害性を大幅に改善した。生成 AI を用いた主成分軸の自動説明文生成機能の追加は、当初計画になかった実施内容であり、将来的なデータベースの継続性や安定的な稼働につながる重要な成果となった。

研究開発期間終了までに、当初計画通り、共発現解析対象となる生物種を 19 種 20 系統にまで拡大し、主要作物およびモデル植物群を包括的にカバーできるようにするとともに、メゾスケールの種間比較機能の実装により、細胞内局在別の CoexMap を複数種で整列して、全遺伝子 CoexMap と並列表示する機能から、種間で保存された細胞機能モジュールを俯瞰的に評価できる解析環境を提供する予定である。

上述の通り、進捗状況と今後の成果の見込みは期待通りで優れている。RNA-seq 共発現データに基づいて共発現モジュールが機能するサンプル条件を提示出来るようにした点、ならびにモデル植物と非モデル植物の共発現モジュールの対応関係の比較を実現した点を高く評価する。データベースへのアクセス数の増加とともに、多数の論文で引用されるなど、産業応用を含めた利活用拡大が進んだことを評価する。

以上