

令和 5 年 6 月 8 日現在

機関番号：24302

研究種目：基盤研究(C)（一般）

研究期間：2020～2022

課題番号：20K06043

研究課題名（和文）植物遺伝子機能予測のための統合的学習アルゴリズムの開発と応用

研究課題名（英文）Development and application of integrative learning algorithms for gene function prediction in plants

研究代表者

福島 敦史（Atsushi, Fukushima）

京都府立大学・生命環境科学研究科・教授

研究者番号：80415281

交付決定額（研究期間全体）：（直接経費） 3,000,000円

研究成果の概要（和文）：本研究課題では、以下の研究開発を行った。(1) アンサンブル学習によるDRAの統合解析手法の開発、(2) 客観的かつ持続可能な植物ネットワークオントロジーの構築、(3) 得られたgold standardによる客観的な遺伝子機能予測精度の評価、(4) メタボロームデータの遺伝子機能予測への利活用法の探索、(5) 植物メタボロームデータをLinked Open Dataとして共有。

研究成果の学術的意義や社会的意義

公的データベースへのデータ公開と共有は、公開データからの知識の再発見および検証可能な仮説構築を促すと考えられる。これまでの研究財産である公的利用可能なオミクスデータ群が宝の山であるとするならば、新規ネットワーク解析アルゴリズムの開発が新規遺伝子機能予測の向上と機能解析への有力な指針を与える。将来的に、ストレス耐性付与等の有用形質獲得に寄与しうる未解明の遺伝子制御ネットワーク同定が、バイオ技術イノベーションから社会還元への道筋に多様性を与える。

研究成果の概要（英文）：Achievements; (1) Development of integrative DRA with Ensemble learning, (2) Construction of an objective and sustainable 'plant network ontology' as a gold standard, (3) Evaluation of the overall accuracy of the functional prediction using the constructed gold standard, (4) Exploring to use metabolome data for gene function prediction, and (5) Sharing of plant metabolome data as a Linked Open Data.

研究分野：植物ゲノム情報学

キーワード：バイオインフォマティクス

1. 研究開始当初の背景

これまで研究代表者は、トランスクリプトームデータについて、2群間で有意に異なる遺伝子共発現パターンを示す遺伝子群を同定するディファレンシャル共発現 (共発現差異、以降 differential coexpression, DC) に着目してきた。その過程で開発した DiffCorr (Fukushima, *GENE*, 2013) の単純な計算では、依然として DC ネットワークを構成する遺伝子ペアが多すぎて、機能予測するまでの落としこみがきわめて困難であった。一方で、Differential Regulatory Analysis (DRA) アプローチが複数開発されてきており、それぞれで特徴があることがわかってきた。またこれまで使用されてきた gold standard は、マニュアルキュレーションによる既存データベース依存だということに気づいた。総合すると、自動ネットワークオントロジー (NeXO) 構築および DRA アンサンブル学習の組み合わせで、遺伝子機能予測を飛躍的に改善できるのではないかとこの着想に至った。

2. 研究の目的

本研究では、統合ディファレンシャルネットワーク情報を利活用した植物遺伝子機能予測のための統合的学習アルゴリズムの開発を行う。そのコア技術として、発現差異 (DEGs)、ディファレンシャル共発現 (DC) および転写因子情報を組み合わせた DRA アプローチを使う。客観的かつ持続可能な遺伝子機能予測の性能評価のために、gold standard としてネットワークオントロジーを利活用する。さらに予測機能向上のため、アンサンブル学習により複数 DRA を統合する。

3. 研究の方法

本研究では、研究期間内で以下 4 項目を段階的に進め、統合ディファレンシャルネットワーク情報を利活用した客観的かつ持続的な遺伝子機能予測手法を開発する：

- (A) アンサンブル学習によるディファレンシャルネットワークの統合解析手法の開発
- (B) 客観的かつ持続可能な植物ネットワークオントロジー (gold standard) の構築
- (C) 項目 B で得られた gold standard による客観的な遺伝子機能予測精度の評価
- (D) 開発手法によるストレス条件下における作物種トマトの重要転写因子探索
- (E) 上記で得られたデータは、ウェブの国際標準である Resource Description Framework (RDF) でオープンデータ化

4. 研究成果

(1) アンサンブル学習による DRA の統合解析手法の開発

DRA アプローチは複数亜種があり、これらを情報統合することが有用である。アンサンブル学習による性能向上のために、複数の DRA 手法を R パッケージとして独立に実装し、GitHub (<https://github.com/afukushima/>) で公開した。

(2) 客観的かつ持続可能な植物ネットワークオントロジーの構築

gold standard データの客観性担保と構築の自動化のため、Cliques Extracted Ontology (CliXO, Kramer et al. *Bioinformatics*, 2014) に基づく NeXO を生成した。

(3) 得られた gold standard による客観的な遺伝子機能予測精度の評価

上記 NeXO をトレーニングデータとし、アンサンブル学習型の DRA 手法をブラッシュアップする。RIF、TFactS、DCGL v2.0 といった複数の DRA 手法による遺伝子機能予測を統合した統計値を算出し、ストレス条件およびコントロール条件の二群間で異なる遺伝子制御ネットワーク上で重要な因子の同定を試みた。また、本アプローチによる遺伝子機能予測は、複数作物種データについて展開中である。

(4) これまで陽に用いられていなかったメタボロームデータについて、遺伝子機能予測へのデータ利活用が可能かどうか予備調査した。すでに得られていた公共利用可能なシロイヌナズナ変異体 50 種の代謝表現型（メタボタイプ）データセット（Fukushima et al. *Plant Physiol*, 2014）を再解析し、評価した。階層クラスタ解析の結果、構築したデンドログラムから植物ホルモン ABA の生合成およびシグナル伝達に関わる遺伝子群がクラスタ化していた。つまり、これらの変異体の代謝表現型が互いによく似ていることを示しており、遺伝子機能予測にも利活用できることを示唆している。

(5) このようなメタボロームデータの再解析・再利用にはメタデータが重要である。理研環境資源科学研究センター（CSRS）で取得した植物メタボロームデータを体系的に整理し、RDF に基づいて公開した（<http://metabobank.riken.jp/>, Fukushima et al. *Plant Cell Physiol*, 2022）。

5. 主な発表論文等

〔雑誌論文〕 計7件（うち査読付論文 6件/うち国際共著 0件/うちオープンアクセス 5件）

1. 著者名 Kusano Miyako, Worarad Kanjana, Fukushima Atsushi et al.	4. 巻 11
2. 論文標題 Transcriptomic, Hormonomic and Metabolomic Analyses Highlighted the Common Modules Related to Photosynthesis, Sugar Metabolism and Cell Division in Parthenocarpic Tomato Fruits during Early Fruit Set	5. 発行年 2022年
3. 雑誌名 Cells	6. 最初と最後の頁 1420 ~ 1420
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/cells11091420	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Fukushima Atsushi, Takahashi Mikiko, Nagasaki Hideki, Aono Yusuke, Kobayashi Makoto, Kusano Miyako, Saito Kazuki, Kobayashi Norio, Arita Masanori	4. 巻 63
2. 論文標題 Development of RIKEN Plant Metabolome MetaDatabase	5. 発行年 2021年
3. 雑誌名 Plant and Cell Physiology	6. 最初と最後の頁 433 ~ 440
掲載論文のDOI (デジタルオブジェクト識別子) 10.1093/pcp/pcab173	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Abdelrahman Mostafa, Nishiyama Rie, Tran Cuong Duy, Kusano Miyako, Nakabayashi Ryo, Okazaki Yozo, Matsuda Fumio, Ch?vez Montes Ricardo A., Mostofa Mohammad Golam, Li Weiqiang, Watanabe Yasuko, Fukushima Atsushi, Tanaka Maho, Seki Motoaki, Saito Kazuki, Herrera-Estrella Luis, Tran Lam-Son Phan	4. 巻 118
2. 論文標題 Defective cytokinin signaling reprograms lipid and flavonoid gene-to-metabolite networks to mitigate high salinity in <i>Arabidopsis</i>	5. 発行年 2021年
3. 雑誌名 Proceedings of the National Academy of Sciences	6. 最初と最後の頁 e2105021118
掲載論文のDOI (デジタルオブジェクト識別子) 10.1073/pnas.2105021118	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Nagai Keisuke et al.	4. 巻 584
2. 論文標題 Antagonistic regulation of the gibberellic acid response during stem growth in rice	5. 発行年 2020年
3. 雑誌名 Nature	6. 最初と最後の頁 109 ~ 114
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41586-020-2501-8	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Utsumi Yoshinori, Tanaka Maho, Utsumi Chikako, Takahashi Satoshi, Matsui Akihiro, Fukushima Atsushi, Kobayashi Makoto, Sasaki Ryosuke, Oikawa Akira, Kusano Miyako, Saito Kazuki, Kojima Mikiko, Sakakibara Hitoshi, Sojikul PUNCHPAT, Narangajavana Jarunya, Seki Motoaki	4. 巻 -
2. 論文標題 Integrative omics approaches revealed a crosstalk among phytohormones during tuberous root development in cassava	5. 発行年 2020年
3. 雑誌名 Plant Molecular Biology	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s11103-020-01033-8	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Shinozaki Yoshihito et al.	4. 巻 117
2. 論文標題 Fruit setting rewires central metabolism via gibberellin cascades	5. 発行年 2020年
3. 雑誌名 Proceedings of the National Academy of Sciences	6. 最初と最後の頁 23970 ~ 23981
掲載論文のDOI (デジタルオブジェクト識別子) 10.1073/pnas.2011859117	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Desmet Sandrien, Saeys Yvan, Verstaen Kevin, Dauwe Rebecca, Kim Hoon, Niculaes Claudiu, Fukushima Atsushi, Goeminne Geert, Vanholme Ruben, Ralph John, Boerjan Wout, Morreel Kris	4. 巻 19
2. 論文標題 Maize specialized metabolome networks reveal organ-preferential mixed glycosides	5. 発行年 2021年
3. 雑誌名 Computational and Structural Biotechnology Journal	6. 最初と最後の頁 1127 ~ 1144
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.csbj.2021.01.004	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計3件 (うち招待講演 1件 / うち国際学会 0件)

1. 発表者名 福島敦史, 高橋みき子, 長崎英樹, 小林誠, 草野都, 斉藤和季, 小林紀郎, 有田正規
2. 発表標題 植物 メタボロームデータの標準化と共有に向けた統合メタデータベースの開発
3. 学会等名 第38回日本植物バイオテクノロジー学会
4. 発表年 2021年

1. 発表者名 福島敦史
2. 発表標題 メタボロームデータの標準化と共有のための統合メタデータベース開発
3. 学会等名 第15回メタボロームシンポジウムプログラム（招待講演）
4. 発表年 2021年

1. 発表者名 福島敦史, 高橋みき子, 長崎英樹, 小林誠, 草野都, 斉藤和季, 小林紀郎, 有田正規
2. 発表標題 植物 メタボロームデータの標準化と共有に向けた統合メタデータベースの開発
3. 学会等名 第63回日本植物生理学会年会
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関