

機関番号：11301

研究種目：若手研究(B)

研究期間：2009～2010

課題番号：21770035

研究課題名(和文)

シロイヌナズナの遺伝子共発現データベース ATTED-III の構築

研究課題名(英文)

Construction of ATTED-III, Gene Coexpression Database for Arabidopsis

研究代表者

大林 武 (OBAYASHI TAKESHI)

東北大学・大学院情報科学研究科・助教

研究者番号：50397048

研究成果の概要(和文)：

シロイヌナズナの共発現データベース ATTED-II は、未知の機能グループを推察する手段として広く利用されてきた。しかしこの遺伝子共発現情報は遺伝子間の機能的関係の存在を支持するものの、それが具体的にどのような関係なのかについての情報をもたらさない。本研究では、共発現データの具体的な関係を明らかにするために、条件特異的共発現データを作成し、これらの共発現データの値と構造を調べるために2つのタイプのビューアと共に ATTED-II 上で公開した。

研究成果の概要(英文)：

ATTED-II, a coexpressed gene database for Arabidopsis, has been used to find functionally related gene groups. However, this traditional coexpression information cannot provide the detail of gene relationships. In this study, I constructed condition-specific coexpression data and two types of viewers to reveal the situation when and where the genes function together.

交付決定額

(金額単位：円)

| | 直接経費 | 間接経費 | 合計 |
|--------|-----------|-----------|-----------|
| 2009年度 | 1,500,000 | 450,000 | 1,950,000 |
| 2010年度 | 1,900,000 | 570,000 | 2,470,000 |
| 年度 | | | |
| 年度 | | | |
| 年度 | | | |
| 総計 | 3,400,000 | 1,020,000 | 4,420,000 |

研究分野：生物学

科研費の分科・細目：基礎生物学・植物分子生物・生理学

キーワード：環境応答、バイオインフォマティクス

1. 研究開始当初の背景

ゲノムワイドに遺伝子機能を推定する方法として遺伝子配列の類似性による機能推定が広く行われているが、近年になって大量の遺伝子発現情報を利用した機能推定法が注目されてきた。マイクロアレイ技術の普及に伴い、公共のデータベースには多くの遺伝子発現データが蓄積しており、特にシロイヌナズナ、ヒト、マウス、ラットでは大量の発

現データを利用することができる。これらの遺伝子発現データはそれ自体が非常に有用な情報であるが、多数の実験サンプルを用いることで遺伝子発現パターンの類似性を考えることができる。遺伝子発現パターンが同じ遺伝子群(すなわち共発現遺伝子群)は、同じ機能を有していると推測することができることから、このような共発現遺伝子群を用いて実験ターゲットを絞り、遺伝子破壊等

で機能同定を行うアプローチが最近増えつつある。

遺伝子共発現の計算にはバイオインフォマティクスの技術が必要なため、広く実験系研究者が利用できるよう、これまでに申請者はシロイヌナズナ及び動物の共発現データベースを構築し運用してきた。特に本申請のシロイヌナズナの共発現データベース ATTED-II は、申請当初においても一日千件を超えるアクセスがあり、共発現利用分野において世界をリードしていた。

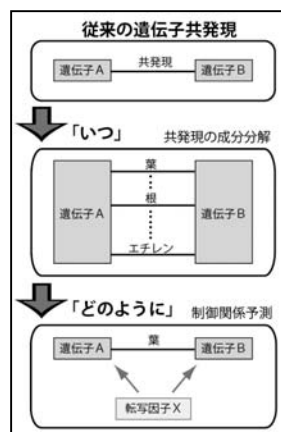
2. 研究の目的

一方で遺伝子共発現法の弱点として、共発現により支持された関係の生物学的意味が不明確であることが挙げられ、そのような生物学的意味は共発現データ利用者の知識・洞察力に拠っていた。この問題を克服すべく、これまでに申請者は共発現ネットワークへ幾つかの情報を加える試みを行ってきた。それらの情報とは、既知のタンパク質間相互作用データ、保存共発現データ、代謝パスウェイデータならびに転写因子のアノテーションである。これらにより共発現ネットワークの理解は容易になった。しかしながら未だに根本的な問題として残っていることは、注目している遺伝子共発現関係が「いつ」「どのようにして」形成されるのかが不明であることである。例えば現状の遺伝子共発現ネットワークにおいては、遺伝子機能の静的関係性を示すものの各状況に応じて機能パートナーが変化するという動的関係性を知ることができないが、この問題こそが本申請で提案する「いつ」「どのようにして」を明らかにすることで解決できる問題である。本研究では遺伝子共発現の生物学的意味を、遺伝子制御の観点、すなわち(1)どのような組織や植物ホルモンによって導入される関係か、(2)どの転写因子によって制御されている関係か、を推定し、シロイヌナズナの遺伝子共発現データベース ATTED-II を大幅に発展させた ATTED-III 構築を目的とする。

3. 研究の方法

申請者がこれまでに作成してきた共発現情報は、遺伝子間の機能的関係の存在を支持するものの、それが具体的にどのような関係なのかは分からない。

そこでまず、いつ・どのように、これらの遺伝子が共に機能するのか



を明らかにするため、共発現を計算するための遺伝子発現行列に成分分解の統計学的手法である主成分分析を適応し、この共発現に寄与している因子（生物学的背景）を明らかにする。

次に共発現関係がどのような共制御に基づいて誘起されるのかを明らかにすることで共発現遺伝子群の位置づけを明確にする。そのために、転写因子-非制御遺伝子の予測を行う。

最後にこれらの詳細な共発現関係を一度に提示・比較するビューアを開発し、データベースとして公開する。

4. 研究成果

まず当初計画通り、公共のデータベース (TAIR, <http://arabidopsis.org/>) に登録されているマイクロアレイデータから、代表的な5組織ならびに7つの植物ホルモンに関する実験を手動で選別し、その主成分分析により12の仮想サンプルを作成した。しかし、実際に行ってみると手動選別が難しいこと、ならびに主成分分析で綺麗に分かれないことが分かった。一方で全発現データを一度に主成分解析する方法は共発現解析を深化させる一アプローチとして有効であることが分かり、これを論文として発表した

(Kinoshita and Obayashi, 2009)。しかし、主成分分析の結果は各成分が独立となるため、数学的には解析しやすくなるが、各実験条件との対応が複雑であり、各種成分の意味づけは容易ではなかった。

この為、各実験条件を強く反映するよう、シロイヌナズナのマイクロアレイプロジェクトの AtGenExpress から主要な5つ分類(成長段階、生物的处理、非生物的处理、植物ホルモン処理、光照射)を用いて条件特異的共発現データを作成した。更に各条件特異的共発現データを一度に比較できるビューアを作成したところ、従来の共発現データからは得ることのできない、共発現関係の意味づけを容易に行うことが可能になった。

従来の共発現指標(発現パターンの相関係数)では、異なる共発現データの値を直接比較することはできないため、図1のような横並びの表示をしても意味をくみ取ることができなかった。我々は共発現指標を比較できない根本的な理由が、比較したいデータ間の相関係数の分布がそろっていないことであることを見だし、相互共発現順位MRを新たな共発現指標として用いることにした

(Obayashi and Kinoshita, 2009)。従来の共発現データ(図1Cの4列目(MR(all))だけでは、様々な遺伝子が並んでいるだけだったが、条件特異的共発現データを併記することによって、共発現を誘起している条件は遺伝子によって異なることが分かる。すなわち

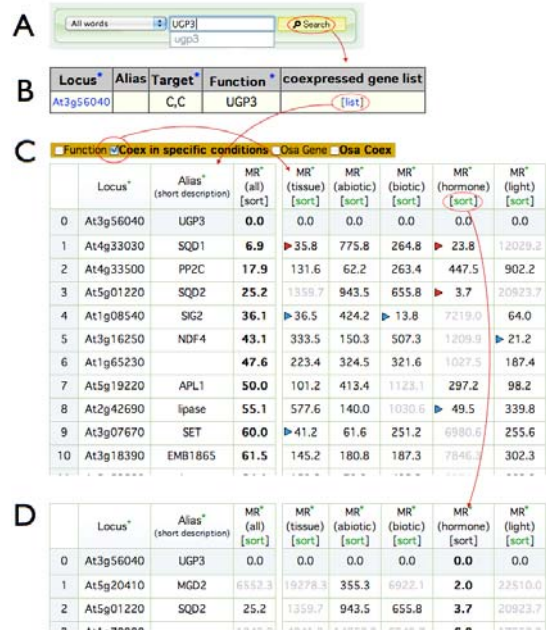
図1の例では、全ての共発現遺伝子が一度に働くのではなく、条件によって共に働く遺伝子パートナーが代わっていくことを示している。また、図1のツールとは別にEdgeAnnotationと名付けたツールを作製し、注目している遺伝子群の全組み合わせの条件特異的共発現データを取得することが可能にした。

図1のビューアを用いることで、共発現遺伝子がまとまっているか、様々であるかを見て取れるが、更に理解を深めるために図2のビューア NetworkDrawerを開発した。このビューアを用いることで、共発現遺伝子の構造を直接比較することが可能になる。図2Aは従来の共発現データに基づいて構築した共発現ネットワークである。この共発現ネットワークは、B, C, D, Eの各条件特異的共発現下において、ネットワークの密度が変化し、また場合によってはネットワーク間を結ぶ共発現関係も観察することが可能になった。また、更にネットワーク解析を容易にするために、インタラクティブな解析ができるCytoscape Webシステムを実装した(図2F)。

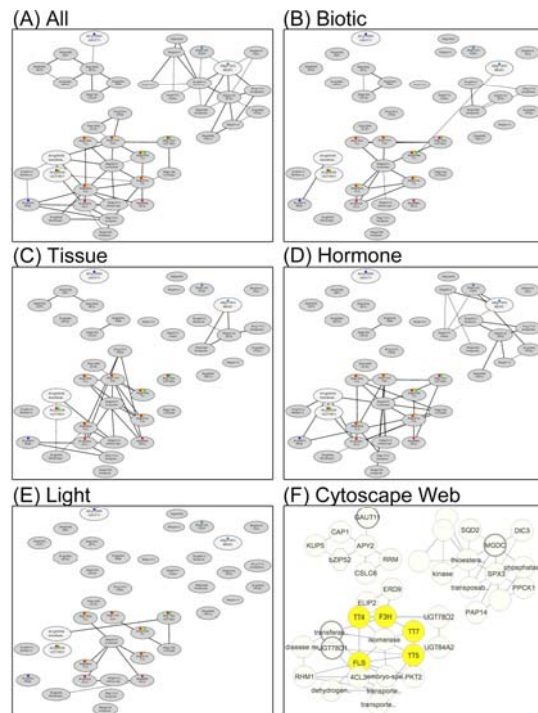
次に、これら成果に基づき共発現を導入する共通の転写因子の予測方法の研究も進めたが、予測の擬陽性が高く直ぐに実用化することはできなかった。高い擬陽性の一要因として、共発現データの不確実性が考えられたため、データの信頼性を向上される目的で、複数種の共発現データを比較するアプローチを試みており、中間的なデータではあるが擬陽性率を改善できる見込みである。種の比較は様々な植物への共発現データの利用の観点からも重要であることから、値を比較するビューアを開発し、ATTED-IIに実装した。(図3)

従来の共発現指標では、異種間の共発現を比較することはできなかったが、新たに開発した相互共発現順位MRを用いることで、直接比較を可能にした。複数の種で支持されている共発現関係はより重要であると推測ができ、機能推定に役に立つ。そのみならず、研究段階であるが共通の制御因子の解析にも有効であることを見いだした。

これらの共発現関係を深化させる研究、ならびに ATTED-II への実装により、データベース利用件数は 2008 年の申請当初から着実な伸びを続け、今年度では 2 倍の伸びとなる見込みである。(図4)



(図1: 比較共発現ビューア。A: 遺伝子の検索、B: 検索結果、C: 従来の条件非特異的共発現遺伝子リストと新規に作成した5種類の条件特異的共発現遺伝子リスト、D: 条件特異的共発現の一つで遺伝子を並び替えた状態)

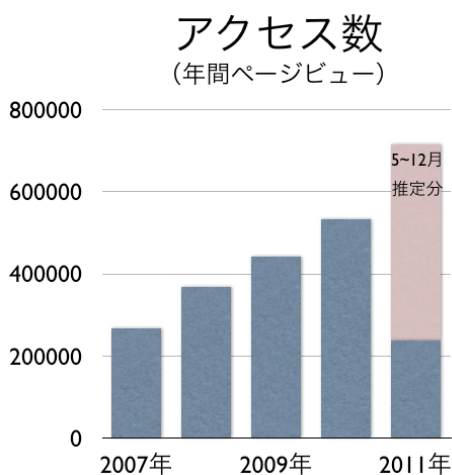


(図2: 条件特異的共発現の構造を比較するビューア: NetworkDrawer)

Top 300 coexpressed genes to At1g44575 (Ath.c4-1, Osa.c2-0)

| Function Coex in specific conditions Osa Gene Osa Coex | | | Osa MR [*] for Os01g0869800 | Osa MR [*] for Os04g0690800 |
|--|--|------------------------------|--------------------------------------|--------------------------------------|
| Locus [*] | Alias [*] (short description) | MR [*] (all) [sort] | | |
| 0 | At1g44575 | NPQ4 | 0.0 | 262.5 |
| 1 | At1g12900 | GAPA-2 | 1.0 | 0.0 |
| 2 | At3g55800 | SBPASE | 3.5 | 18.5 |
| 3 | At1g42970 | GAPB | 3.7 | 103.9 |
| 4 | At1g54780 | thylakoid lumen 18.3 kDa | 5.5 | 111.5 |
| 5 | At4g38970 | FBA2 | 8.5 | 535.4 |
| 6 | At1g20340 | DRT112 | 9.2 | 26.5 |
| 7 | At4g01150 | | 9.5 | 167.8 |
| 8 | At1g32060 | PRK | 10.9 | 71.8 |
| 9 | At1g67740 | PSBY | 12.2 | 8.5 |
| 10 | At3g26650 | GAPA | 13.4 | 42.2 |
| 11 | At1g20340 | | 15.5 | 320.9 |
| 12 | At1g20340 | | 18.5 | 103.9 |

(図3: 共発現の異種比較ビューア)



(図4: ATTED-II のアクセス数の推移)

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計6件)

- Obayashi T, Nishida K, Kasahara K, Kinoshita K ATTED-II updates: condition-specific gene coexpression to extend coexpression analyses and applications to a broad range of flowering plants. *Plant Cell Physiol*, 52, 213-9 (2011) 査読あり
- Obayashi T, Kinoshita K COXPRESdb: a database to compare gene coexpression in seven model animals. *Nucleic Acids Res*, 39, D1016-22 (2011) 査読あり
- 大林武、木下賢吾 「共発現情報から見る遺伝子機能ネットワーク」 *化学と生物* [KASEAA] (2011), 49, p48-56、査読なし
- Obayashi T, Kinoshita K. Coexpression landscape in ATTED-II: usage of gene list and gene network for various types of pathways. *J. Plant Research*, 123,

311-9 (2010)、査読あり

- Kinoshita K, Obayashi T Multi-dimensional correlations for gene coexpression and application to the large-scale data of Arabidopsis. *Bioinformatics*, 25, 2677-84 (2009)、査読あり
- Obayashi T, Kinoshita K Rank of correlation coefficient as a comparable measure for biological significance of gene coexpression. *DNA Research*, 16, 249-60 (2009)、査読あり

[学会発表] (計6件)

- Obayashi T "How to measure the similarity of gene expression patterns to identify gene function" QBIC2011 Bio-Informatics Workshop 2011年3月11日、東京理科大学
- Obayashi T and Kinoshita K, "Rank of Correlation Coefficient Is a Comparable Measure of Gene Coexpression", The 2010 Annual Conference of Japanese Society for Bioinformatics、2010年12月13日、九州大学
- Obayashi T, Kinoshita K "COXPRESdb: a gene coexpression database for 7 animal species" 4th Asian Young Researchers Conference on Computational and Omics Biology 2010年12月1日、シンガポール、シンガポール共和国
- Obayashi T, Kinoshita K "Rank of Correlation Coefficient as a Comparable Measure for Biological Significance of Gene Coexpression" Asian Young Researchers Conference on Computational and Omics Biology 2010年3月10日、台南、台湾
- Obayashi T, Kinoshita K "Number of microarray samples to construct effective gene coexpression data"、第51回日本植物生理学会年会、2010年3月20日、熊本大学
- Obayashi T, Kinoshita K "COXPRESdb: A database of coexpressed gene networks for human, mouse and rat" International congress of physiological sciences、2009年7月27日、京都

6. 研究組織

(1) 研究代表者

大林 武 (OBAYASHI TAKESHI)

東北大学・大学院情報科学研究科・助教

研究者番号: 50397048