

# Development and Application of Statistical Methods for Biological Network Inference from Gene Expression Profiles

ウォン, プイ, シャン

<https://doi.org/10.15017/1932015>

---

出版情報 : 九州大学, 2017, 博士 (農学), 論文博士  
バージョン :  
権利関係 :



氏 名 : ウォン プイ シャン

論文題名 : Development and Application of Statistical Methods for Biological Network Inference from Gene Expression Profiles  
(遺伝子発現データからのネットワーク推定に資する統計手法の開発と適用)

区 分 : 乙

## 論 文 内 容 の 要 旨

次世代シーケンサー (NGS) の発展により、細胞内に存在する DNA や RNA 塩基配列データを高速かつ効率的に取得することが可能になっており、様々な生物種を対象として、ゲノムシーケンス情報や様々な条件下での細胞内遺伝子発現情報や経時的に起こる細胞変化情報を網羅的に測定した時系列データが数多く取得されている。一方、測定された NGS データから生物学的知見を得るための解析手法については十分に開発されていない。そこで、本研究では、NGS データから生物学的知見を得るための有効な手法を開発し、さらに、開発した手法をバイオ燃料生産の有用株とされている海洋微細藻類 *Fistulifera solaris* (*F. solaris*) の NGS データ解析に適用して、トリアシルグリセリド蓄積時における遺伝発現動態の解析を行い、開発した手法の有効性を検証した。

まず、低窒素条件下での *F. solaris* のトリアシルグリセリド高蓄積能の原因を明らかにするため、*F. solaris* とその近縁種である *Phaeodactylum tricornutum* (*P. tricornutum*) の遺伝子発現比較を行った。この解析において、ゲノム配列データから同定した遺伝子オントロジーと RNA-Seq データで測定された遺伝子発現量を組み合わせた解析手法を適用した。*F. solaris* と *P. tricornutum* の両方に存在しながらも、低窒素条件下において遺伝子発現量が有意に異なる遺伝子群を同定し、遺伝子発現変動パターンによるグループ分けを行い各グループの機能を解析した。その結果、*F. solaris* と *P. tricornutum* では低窒素状態に適応するために異なる遺伝子が発現していることが明らかになった。特に、*F. solaris* では酸化ストレス反応に関連した遺伝子が発現する事が明らかになり、窒素欠乏状態において、低窒素条件に対応するストレス応答遺伝子群の発現は抑制され、光合成関連遺伝子と酸化ストレス関連遺伝子の発現増加が起こることが明らかになった。

次に、*F. solaris* におけるトリアシルグリセリド合成のメカニズムを明らかにするため、グラフアルゴリズムを組み込んだ新規のパスウェイ解析手法を開発し、時系列データへ適用した。トリアシルグリセリド蓄積時の発現変動遺伝子群に対して、Gene Set Enrichment Analysis (GSEA) 解析を行った結果、光合成・脂肪酸合成、アミノ酸合成の代謝経路に関連する遺伝子群が優位に発現変動していることが明らかになった。そこで、これらの発現変動している代謝経路群が協同して活性化しているかどうかを調べるため、KEGG データからこれらの代謝経路の化合物間の反応情報を抽出し、その 2 項関係を連結した包括的な代謝パスウェイモデルを再構成し、時間ごとに活性化している代謝経路を明確にした。さらに、作成したパスウェイグラフ上で、最も少ない反応数でブドウ糖からトリアシルグリセリドに至る経路を最短経路探索のグラフアルゴリズムを適用して解析した結果、トリアシルグリセリド生合成に関連する遺伝子群は、長時間協調して反応経路を活性化しているのではなく、それぞれの時間で、必要な経路の

みを活性化していることが明らかになった。

次に、*F. solaris*におけるトリアシルグリセリド合成の制御機構を解明するために、少数の時系列遺伝子発現データから、遺伝子間の制御関係を推定する手法を開発し、ネットワーク推定を行った。この手法では、遺伝子発現量を数値データから-1、0、+1の3パターンのカテゴリカルデータに変換し、各遺伝子の遺伝子発現をカテゴリカルデータのベクトルとして表現した。すべての遺伝子はこの3つの数値の組み合わせで表わされたベクターパターンのいずれかに分類される。各パターンベクトルをグラフ上で一つの頂点として捉え、頂点間をつなぐエッジについて、ベクトルの要素が一つだけ異なっているベクトル間には相互作用があると定義して、そのベクトル間を結合し、推定ネットワークとした。開発したネットワーク推定手法の有効性をシロイヌナズナの開花までの時系列データを用いて検証した結果、既知の転写因子が開花に向けて順番に制御されている事が示され、開発した手法を適用することで、限定的な時系列データからの制御ネットワークが推定可能であることが判明した。続いて、開発した手法を*F. solaris*の時系列データへ適用した結果、大きく2つのサブグラフを構成することができ、推定したネットワークは、バイオ燃料生産に重要な制御因子の推定に利用可能であることが判明した。

以上の結果をまとめると、本研究では、NGS データから生物学的知見を得るための3種類の新規手法を開発し、それぞれの有効性を検証した。さらに、*F. solaris*におけるトリアシルグリセリド高蓄積能の要因や、その遺伝子制御システムを明らかにした。この結果は、海洋微細藻類によるバイオ燃料生産の制御システム解明に貢献するものである。