HOKKAIDO UNIVERSITY

Title	発現プロファイルの利用
Author(s)	遠藤，剛；高林，厚史
Description	6章 バイオインフォマティクス 2
Relation	光合成研究法．北海道大学低温科学研究所，日本光合成研究会共編
Citation	低温科学，67，663-667
Issue Date	2009-03-31
Doc URL	https://hdl.handle.net/2115/39208
Type	departmental bulletin paper
File Information	67-092.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

2．発現プロファイルの利用

遠藤　　剛[1)]，高林　厚史[2)]

高等植物，とりわけシロイヌナズナにおいて，公開マイクロアレイデータの蓄積が進み，最近ではそれらデータを光合成研究に取り入れる動きも出てきている．本稿ではそれらマイクロアレイデータの光合成研究への利用について概説する．

Application of public microarray data to a photosynthesis study

Tsuyoshi Endo, Atsushi Takabayashi

Recently, a lot of microarray data is available in the public repositories recently.　Here, we briefly introduce you to how to use it and give you examples of how it is used in photosynthesis studies.

光合成においてモデル植物として用いられてきたのは，エンドウやホウレンソウなど，Ⅰ）入手や栽培が容易，II）無傷葉緑体の単離が容易，Ⅲ）サイズが大きい，等の条件を充たした植物であった．これは光合成研究の手法として主に物理化学的な測定や生化学的な解析が用いられてきた事が大きな理由であると考えられる．その後，形質転換が容易なタバコも光合成研究の材料として用いられるようになってきたが，状況が一変したのは2000年に高等植物として初めてシロイヌナズナの全ゲノムが解読されてからである．それ以来，光合成研究のモデル植物としてシロイヌナズナを選択し，その充実した変異株リソースを用いた逆遺伝学的に解析するアプローチが盛んに行われるようになってきた．さらに，マイクロアレイの発現データなど膨大なゲノムリソースが大量に公開されていることもシロイヌナズナを用いる利点と言えよう．

とりわけ光合成研究者にとって興味深い点は，核ゲノムコードの光合成遺伝子群は転写レベルで共発現（co-expressed）する傾向が見られる事である．これは他の代謝経路と比べて例外的なレベルといえ，光合成遺伝子群の発現調節を考える上でも大きな特徴であると共に，使い方しだいでは新規光合成遺伝子の同定につながる大変有益な情報である．実際，近年になって，蓄積された公開マイクロアレイデータを利用して，新規光合成関連タンパク質を同定したとの論文も報告されている（2.3参照）．そこで本稿では，光合成研究者を対象とし，公開マイクロアレイデータの利用法について概説する．

2.1章では，実際に公開マイクロアレイデータを用いて解析するために必要な3つのステップ，すなわち，Ⅰ）公開マイクロアレイデータの取得，II）データの前処理，Ⅲ）遺伝子発現解析，について概説する（図1）．

2.2章では，すでに解析済みの遺伝子発現データを公開しているデータベースを利用する事で，自ら解析する事なしで，公開マイクロアレイデータを利用する方法について紹介する．

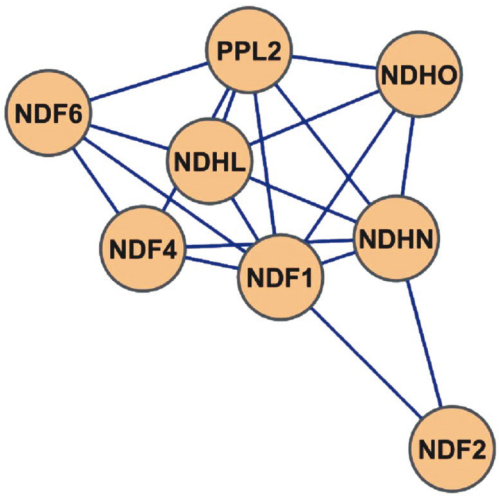

図1：新規 *NDH* 遺伝子候補の発現ネットワーク（r＞0.85）

1）京都大学大学院生命科学研究科
2）名古屋大学遺伝子実験施設（日本学術振興会特別研究員 PD）

2.3章では，実際に公開マイクロアレイデータを光合成研究に利用する際の参考に，筆者らの研究を含め，実際の研究例を紹介する．

2.1 公開マイクロアレイデータを用いた遺伝子発現解析について

様々な生物種においてゲノム情報が充実してきた今日では，ポストゲノム時代に向けてゲノムワイドな解析が広く行われるようになってきた．中でも良く用いられているのがDNAマイクロアレイを用いた包括的な遺伝子発現解析（トランスクリプトーム解析）である．DNAマイクロアレイは発現解析のみならず，ChIP-on-chip，SNP解析等，非常に幅広く用いられているが，ここでは発現解析に絞って簡単に紹介したい．

DNAマイクロアレイは2種類のタイプに大別される．Stanford大学のPat Brownらの研究グループによって開発されたcDNAスポットアレイ[1]と，Affymetrix社によって開発されたオリゴヌクレオチドアレイ[2]である．後者は一枚のアレイで発現が検出できる遺伝子数が多く，また，ターゲット配列に対して完全に相補的なPerfectMatch（PM）プローブと中央の一塩基のみをミスマッチさせたMissMatch（MM）プローブを比較する事で非特異的なハイブリダイゼーションの影響を排除できる事から，現在では主流となっている．ただし，一昔前に比べれば安価にマイクロアレイを利用できるようになってきたとはいえ，気軽に実験できる価格とは言えないため，筆者らの研究グループでは公開マイクロアレイデータを自分の研究目的のために利用させて頂いている．

2.1.1 公開マイクロアレイデータの利用法

多くのジャーナルでは，マイクロアレイを用いた大規模発現解析のデータを投稿するにあたって，MGED (Microarray Gene Expression Data) Societyが作成したMIAME（minimal information. about a microarray experiment: http://www.mged.org/Workgroups/MIAME/miame.html）ガイドラインに従い情報公開をする事が必要とされている．公開されたマイクロアレイデータの公的レポジトリとしてはアメリカNCBIのGEO（Gene Expression Omnibus），欧州EMBLのArrayExpress等が利用さられており，それらレポジトリにアクセスする事で登録されているマイクロアレイデータを入手する事が可能である．シロイヌナズナの公開アレイデータに関して言えば，NASCArrays（Nottingham Arabidopsis Stock Centre Transcriptomics Service）[3]がよく用いられている．

シロイヌナズナの公開マイクロアレイデータとして広く用いられているのが，AtGenExpressプロジェクトによって作成・公開されたデータセットである．AtGenExpressはシロイヌナズナの遺伝子発現データを網羅的に収集するための国際的なプロジェクトで日本の理化学研究所も中核的な役割を担った．このプロジェクトではAffymetrixのGeneChipであるATH1アレイが用いられており，2万以上のシロイヌナズナ遺伝子の発現を解析する事が可能になっている．AtGenExpressのデータは，TAIRもしくはNASCから入手可能である．また，NASCは現在でもAffymetrixのGeneChipを用いたシロイヌナズナの解析データを精力的に集積・提供し続けており，これら（AtGenExpress以外の）NASCArrayデータもよく用いられている．なお，現段階では，Affymetrix社以外の製品を用いたデータはほとんど登録されていない．

登録されたマイクロアレイデータを利用するときには，多くの場合バックグラウンド補正，および正規化等の前処理を行う必要があり，そのあとで発現解析を行う事になる．これらデータ解析のための商用ソフトとしてはGeneSpringやSpotfire等が知られているが，フリーのBioconductor（http://www.bioconductor.org/）もよく用いられるようになった．最近ではBioconductorについての日本語ドキュメントも多くなってきたため，解析は容易になってきたと言える．

2.1.2 前処理

マイクロアレイデータを用いて発現解析を行う際には，バックグラウンド補正や正規化などの前処理が必要となる．バックグランド補正はスポットのバックグランドの蛍光強度の影響を考慮し真のシグナル強度を検出するために必要であり，また正規化はデータの偏りを補正する事で，複数のマイクロアレイデータを相互に比較し定量的な検討を行うために必要な作業である．

前処理のためのアルゴリズムとしては，RMA（Robust Multi-array Analysis），MAS5（Microarray Analysis Suite），dChip（DNA-Chip Analyzer）などの数々の手法が知られており，それらはBioconductor等の解析ソフトを用いる事で容易に試す事ができる．なお，どのアルゴリズムを用いるか判断に迷った場合には，解析結果を比較した論文[4,5]が有益と思われる．いずれにせよ処理す

るアレイの枚数が多くなれば，実行にはその分だけ多くのメモリと時間を必要となるため，本格的な解析には相応のマシンパワーが必要になる．

2.1.3　遺伝子発現解析

マイクロアレイの一般的な使用法としては，まず興味のある遺伝子についての時系列や組織別の発現変動，ストレス応答など，ある実験条件で発現が有意に増減する遺伝子群を選抜する事が考えられる．これらの解析に関しては，例えば Bioconductor を用いる事で比較的容易に対応できる．

一方で，様々な実験条件で同じような発現プロファイルを示す遺伝子群は機能的に関連している可能性が高いと考えられるため，それを利用した未知遺伝子のアノテーションも試みられている．その試みの１つがクラスタリングであり，遺伝子群を発現プロファイルの類似性を指標にしてクラスタに分け，それらのクラスタ内の遺伝子は機能的に関連していると考える解析手法である[6]．具体的なクラスタリングの手法としては，階層的なクラスタリングや自己組織化マップ（self-organizing map；SOM）等があり，実際に光合成研究にも利用されている[7]．

いずれの手法においてもその類似性の比較に発現強度間の違いを基にした距離を利用しており，その遺伝子発現プロファイルの相似度の指標としては，ユークリッド距離だけでなく，ピアソンの相関係数やスピアマンの順位相関係数が用いられている．ユークリッド距離を用いる場合には発現データ値の大小が反映されるが，相関係数を用いる場合には相対的な増減・変動パターンの類似性が反映される．生物学的には発現レベルの絶対値（ユークリッド距離）よりも変化の方向（相関係数）が重要である場合も多く，そのような場合には発現プロファイルの相似度の指標として相関係数が用いられる．一方で，転写因子のように刺激に応答して発現強度が大きく変動するような遺伝子に着目する際には，ユークリッド距離が有用と思われる．

また，ある遺伝子と機能的に類似する遺伝子群を探索する目的であれば，単純にその遺伝子と相関係数が高い遺伝子群を探す事で充分である事も多い．実際に，筆者らはこの手法を用いて新規 *ndh* 変異株を同定している（後述）．

2.2　データベースの利用

実際に実験研究者が公開マイクロアレイデータを利用して大量解析を行おうとすると，数学やコンピュータの知識が要求される場面が少なくない．この障壁を乗り越えるためには，それ相応の時間が必要であり，その時間が取れない人は少なくないと思われる．しかし，最近のデータベースの充実により，公開データベースの利用により自ら解析せずとも典型的なデータ解析の多くは達成できるようになってきた．

典型的なデータベースでは，AtGenExpress 等のデータセットを用いて図１のような手順を踏んで大規模なデータ解析を行い，解析済みの遺伝子発現データを公開している．とりわけ，シロイヌナズナに関しては個性的なデータベースが次々と公開されているため，自分の目的に応じて使い分ける事が重要と思われる．

例えば，ある特定のシロイヌナズナの遺伝子の発現プロファイルを調べたいのであれば，eFP Browser（http://www.bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi）や Genevestigator（https://www.genevestigator.ethz.ch/）等を利用するのが容易であり，とりわけ Genevestigator を利用した解析は論文の Figure としてもよく利用されている印象がある．また，ある実験条件で発現強度が変化する遺伝子群を選抜したい場合には，DART（http://tabacum.agr.nagoya-u.ac.jp/dart/）が利用できる．

また，興味のあるシロイヌナズナの遺伝子と共発現している遺伝子群を選抜する目的であれば，ATTED-II（http://www.atted.bio.titech.ac.jp/）や CSB. DB（http://csbdb.mpimp-golm.mpg.de/），Cress Express（http://www.cressexpress.org/）などのデータベースが利用可能である．それらデータベースでは，興味のある遺伝子名を入力するだけでその遺伝子と高い相関係数を示す遺伝子群のリストを得る事ができる．とりわけ，光合成関連遺伝子においては機能的に関連する遺伝子群が高い相関係数を示すため，これらデータベースの利用は非常に有効である．例えば，実際に光化学系 II の遺伝子名を入れると，相関係数が高い遺伝子群のなかに光化学系 II の関連遺伝子群が含まれることが確認できるはずである．

いずれにせよ，データベースの進歩は日進月歩であり，すぐに情報は古びてしまうため，筆者は Nucleic Acid Research の Database issue などをチェックし，新しいデータベースや更新されたデータベースを自分で実際に試している．

2.3 光合成研究における発現プロファイルの利用法

ここでは実際に光合成研究において発現プロファイルを利用した研究について紹介する．

光合成研究に発現プロファイルを利用した先駆的な研究として Max-Planck 研究所の Dario Leister らの研究グループによる *pgrl1* 変異株の単離・同定が挙げられる．まず彼らはシロイヌナズナの公開マイクロアレイデータを利用してクラスター解析を行い，発現プロファイルの類似性を指標にしてシロイヌナズナの光合成遺伝子群のグループ分けを行った[7]．その結果，光合成関連遺伝子群と葉緑体リボソーム遺伝子群は共発現している傾向が見られ，しかも他の一般的な葉緑体タンパク質をコードする遺伝子群と異なった発現制御を受けている事を見出している．さらに彼らは，新規光合成関連と予想された遺伝子について変異株を利用して逆遺伝学的な解析を行い，光化学系 I 循環的電子伝達経路で機能している *PGR5* 遺伝子と相互作用する *PGRL1* 遺伝子の同定に成功している[8]．

PGRL1 遺伝子はシロイヌナズナで2コピー（*PGRL1A*，*PGRL1B*）存在し，両方の遺伝子機能を破壊する事で初めて表現型が現れるため，通常の変異株スクリーニングでこの遺伝子にたどり着くのは困難であると思われる．一方，Leister らの研究グループでは共発現解析であらかじめ遺伝子数を絞り込み *pgl1a*/*pgrl1b* の二重変異体を作成する事で，その変異株が *pgr5* と同様の表現型を示すことを見出した．この研究は光合成研究に発現プロファイル解析を組み合わせる事で研究効率を上げた好例と言えるだろう．

また，我々の研究グループでは光化学系 I 循環的電子伝達経路で機能する NAD (P) H Dehydrogenase（NDH）の未同定サブユニットの探索にこのアイディアを適用し，成果を挙げている．具体的には，まず，ATTED-II データベースを利用して既知の *ndh* 遺伝子（ndhL，ndhN，ndhO）と相関係数が高い遺伝子群のリストを作成した．また同時に，系統プロファイル法[9] を組み合わせ，シロイヌナズナ全タンパク質のうち，NDH 複合体を持っている生物種のゲノムにホモログが存在し，NDH 複合体を持たない生物種のゲノムにホモログが存在しないタンパク質のリストを作成した．

次に，作成したリストに含まれている遺伝子について，T-DNA が挿入されたシロイヌナズナ変異株を取り寄せ，実際に NDH 活性を測定した．その結果，NDH 活性が失われた新規変異株を5ライン得ることができた[10,11,12]．これら変異株の原因遺伝子群は相互に共発現しており（図 2），生化学的な解析からも NDH の新規サブユニットの可能性が高いと考えている．

これらの例は発現プロファイル，特に共発現に着目する事で，興味ある機能を有する光合成関連遺伝子の同定が従来法よりも迅速・簡便に行う事ができる可能性を示唆している．今後は，シロイヌナズナ以外の植物種においても公開マイクロアレイデータの数が増える事が予想され，ますますこの手法が重要になると期待される．

参考文献

1）M. Schena, D. Shalon, R. Heller, A. Chai, P. O. Brown, & R. W. Davis, Proc. Natl. Acad. Sci. USA. **93** (1996) P. 10614.

2）D. J. Lockhart, H. Dong, M. C. Byrne, M. T. Follettie, M. V. Gallo, M. S. Chee, M. Mittmann, C. Wang, M. Kobayashi, H. Horton, & E. L. Brown, Nat. Biotechnol. **14** (1996) P.1675.

3）D. J. Craigon, N. James, J. Okyere, J. Higgins, J. Jotham, & S. May, Nucleic Acids Res. **32** (2004) P.D575.

4）F. F. Millenaar, J. Okyere, S. T. May, M. van Zanten, L. A. Voesenek, & A. J. Peeters, BMC Bioinformatics. **7** (2006) P.137.

5）W. K. Lim, K. Wang, C. Lefebvre, & A. Califano, Bioinformatics. **23** (2007) P.i282.

6）J. Gollub, & G. Sherlock, Methods Enzymol. **411** (2006) P.194.

7）A. Biehl, E. Richly, C. Noutsos, F. Salamini, & D. Leis-

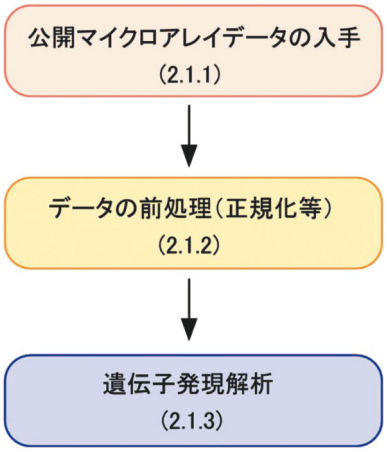

図 2：発現解析のフローチャート

ter, Gene. **344** (2005) P.33.
8）G. DalCorso, P. Pesaresi, S. Masiero, E. Aseeva, D. Schünemann, G. Finazzi, P. Joliot, R. Barbato, & D. Leister, Cell. **132** (2008) P.273.
9）M. Pellegrini, E. M. Marcotte, M. J. Thompson, D. Eisenberg, & T. O. Yeates, Proc. Natl. Acad. Sci. USA. **96** (1999) P.4285.
10）A. Takabayashi, N. Ishikawa, T. Obayashi, S. Ishida, J. Obokata, T. Endo, & F. Sato, Plant J. accepted (2008) doi: 10.1111/j.1365-313X.2008.03680.x
11）S. Ishihara, A. Takabayashi, K. Ido, T. Endo, K. Ifuku, & F. Sato, Plant Physiol. **145** (2007) P.668.
12）N. Ishikawa, A. Takabayashi, S. Ishida, Y. Hano, T. Endo, & F. Sato, Plant Cell Physiol. **49** (2008) P.1066.