

発現プロファイルに基づく代謝経路の遺伝子系列ランキング

瀧川 一学 馬見塚 拓

京都大学 化学研究所 バイオインフォマティクスセンター

代謝経路における各連鎖反応を触媒するための酵素遺伝子の系列について、それらの発現情報を元にごの経路が活性化されているかをランキングする枠組みと方法を提案する。代謝経路中の反応の一次隣接性を二項関係とし発現プロファイルの類似尺度値を重みとして持つ遺伝子グラフに半順序の頂点クラスを付与しパスランキングを行うアルゴリズムを示す。この手法によって発現プロファイルを用いて興味ある代謝経路の転写活性を分析することができる。

Gene Sequence Ranking Based on Expression Profiles for Metabolic Pathway Analysis

Ichigaku TAKIGAWA and Hiroshi MAMITSUKA

Bioinformatics Center, Institute for Chemical Research, Kyoto University

A living cell contains thousands of enzymes, many of which operate at the same time. To control them, metabolism is organized and carefully regulated at many levels. For rapid reactions such as glycolysis, we can assume that genes for adjacent reactions are co-expressed. Thus, we developed a method for generating a sequence of genes that can promote a known path so that genes in each subsequence are maximally co-expressed. Based on expression similarities between genes that encode enzymes, we can analyze the transcriptional activity of possible reaction paths.

1 はじめに

生体内の細胞はすべて、自らを維持・成長・繁殖させるために、環境から物質を食物などの形で取り入れ、膨大な化学反応によって、それらを必要な物質へ分解して必要な生体分子を合成したり、それら有機分子を用いて呼吸や光合成などにより活動に必要なエネルギーを得たりする必要がある。従って、こうした生体内で起こる化学反応の総体である代謝 (metabolism) は細胞内過程を理解する上で重要な対象である [1, 2]。

細胞内で起こる化学反応はほぼ、通常ならばより高温でしか起こらない反応であり、細胞の水溶液中では酵素と呼ばれるタンパク質によって触媒されて初めて起こる。酵素によって反応性を特異的に高めることで初めて細胞内温度で反応させ、またこれら反応群を制御することができる。またエネルギー的に起りやすい反応と起りにくい反応を

共役させる反応経路を作ることで本来起りえない変換を実現できる。

環境から摂取した物質を低分子化合物へ分解する過程では、通常、いくつもの化学反応が関与する。これらの反応群は、さらに、ある反応の産物が別の反応の基質となることで複雑な反応経路のネットワークを形成しており、代謝経路 (metabolic pathway) と呼ばれ、既存の知見は代謝マップ [3] や代謝経路データベースとして KEGG[4], MetaCyc(EcoCyc)[5], WIT[6], UmBBD[7] などにまとめられている。

これらは様々な生物種から得られた知識を統合して一つのマップにしたもので反応系の静的な全体像を表したものである。本研究ではこれらの反応をどのように各生物が利用するかを、酵素遺伝子の発現情報から系の動的な振舞いとして分析するためのアプローチを示す。

遺伝子発現情報はマイクロアレイ実験等によってハイスループットに得られるようになったものの、配列情報と比べると手法の違いや条件などによって結果が変わる弱い論拠であり、従来は配列データベースのように実験をこえて横断的に使用されることが少なかった。しかし、近年のStanford Universityの大規模なデータベースStanford Microarray Database (SMD)[8]や、例えば、NCBIのGene Expression Omnibus (GEO)[9]、EBIのArrayExpress[10]など、標準化を伴う公共データベースの整備が進められ、多くのデータが利用できるようになってきた。これによって、マイクロアレイで観測された、多くの遺伝子の発現情報を、種々の生物学的、医学的、化学的、物理学的条件のもとで観察でき、遺伝子の発現情報をパスウェイの動的振舞いの理解に向けて利用することができるようになってきたと言える。

ここではある物質から目的とする最終産物に至るまでの可能な反応経路について、反応経路に対応した遺伝子系列を「系列内で隣り合う遺伝子の共発現の強さ(隣接共発現度)」によってランキングする方法を提案する。

2 発現情報による代謝経路ランキング

真核生物については転写制御の仕組みも複雑であり、反応について関与する複数の酵素遺伝子が知られている場合も多く、連鎖する一連の反応に関して酵素遺伝子がどのように転写されているかを調べることは重要である。ここではKEGG PATHWAY[4]における全生物種について知られている知見をまとめたリファレンスパスウェイと生物種個々の情報を用いてパスランキングを行う方法を示す。

2.1 代謝経路のグラフ表現

代謝経路は既存の生化学的知識の集積であり、様々なかたちのグラフ構造データとして用いられている。ここでは、生物種および出発化合物と最終化合物が与えられた場合の代謝パスのランキングのため、以下のようなグラフとして用いる。

生体化合物の集合を C 、代謝反応の集合を R 、与えられた生物種の遺伝子の集合を G とする。KEGG PATHWAYで見られるように代謝反応を化合物間の二項関係として見ると、代謝経路は自然に頂点集合を C 、辺集合を R とする有向グラフ $P := (C, R)$

とみなすことができる。

ここで反応 $r \in R$ からその反応に関与しうる遺伝子の集合 $G' \subset G$ を得るための手続きを $F: R \rightarrow 2^G$ とする(2^G は G のべき集合とする)。実際には後述するように遺伝子と反応(酵素)を結び付けるためにKEGGに登録されている様々な情報を用いる。また、遺伝子 $g \in G$ について α 個の実験条件下での発現値プロファイル(α 組の実数値)を $e(g) \in \mathbb{R}^\alpha$ とし、遺伝子の発現プロファイル間の距離(非類似度尺度)を $d(\cdot, \cdot)$ とする。

また、反応 $r \in R$ が基質 $c_1 \in C$ から産物 $c_2 \in C$ を作る場合、 $c_1 \xrightarrow{r} c_2$ などと書くことにし、反応 r を場合によって化合物間の順序ペア (c_1, c_2) と同一視することとする。

2.2 k 次隣接性による発現情報の条件付け

マイクロアレイ実験によって様々な α 個の条件で得た発現プロファイルは各々の遺伝子 $g \in G$ について、 α 次元実数値ベクトル $e(g) \in \mathbb{R}^\alpha$ となる。これをEuclid距離を計量として散布図にした例を視覚化できる $\alpha = 3$ の場合で図1に示す。

この発現プロファイル(図1中の点)の集合において、それぞれの遺伝子がコードする酵素が代謝経路上で連続する(隣り合う)反応を触媒するときその遺伝子ペア間に辺を張るとすると、図1中のようなグラフが得られる。

空間的な近さは(空間の計量に依存して)発現プロファイルの類似度を表し、辺で結合されている遺伝子間は間接的に関連している化学反応の意味で隣り合うことを意味する。発現情報を代謝経路の知識によって条件付けて分析するにはこのような代謝経路上での到達可能な隣接ステップ数 k によって制約するアプローチが興味深い。それぞれひとかたまりの遺伝子群が化学系を機能させている場合、何らかの仕組みでそれらの酵素遺伝子は巨視的には同じような時点で転写される必要があると考えられ、以下で用いる代謝パスウェイと発現情報の統合法もその一つである。本研究では代謝経路上の1次の(一つ隣り合う)隣接関係による条件付けによって発現プロファイルを調べる一方法を考察する。

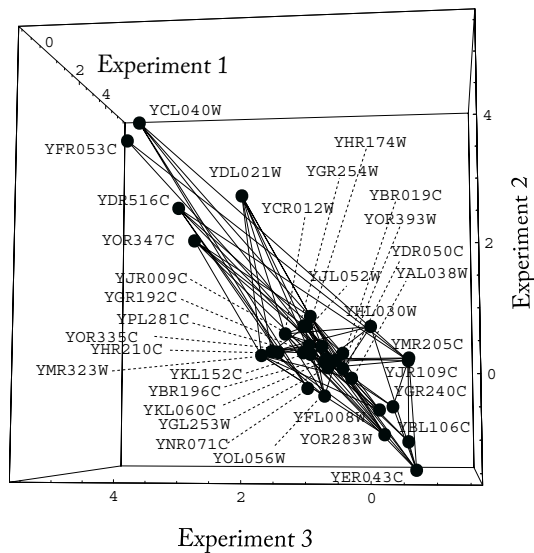


図 1: 代謝経路の 1 次隣接性による発現情報の条件づけグラフ

2.3 パスランキング

C : 化合物集合、 R : 反応集合、 F : 反応 \rightarrow 遺伝子集合間の写像、 e : 遺伝子 \rightarrow 発現プロファイル間の写像 ($e: G \rightarrow \mathbb{R}^{\alpha}$) とし、説明の具体例として図 2-a の出芽酵母の糖代謝経路の一部 (グルコース \rightarrow ピルビン酸) の場合を併せて示す。

入力 代謝経路 $G := (C, R)$ 、関連遺伝子群 G 、出発化合物 $s \in C$ 、最終産物 $t \in C$

出力 隣接共発現度により順序づけた上位 k 個の s から t への代謝反応を触媒する酵素遺伝子系列 $(g_1, g_2, \dots, g_n), g_i \in G$

アルゴリズム

1. グラフ P について全ての s - t 単純パスを列挙して、頂点をマージしたパスグラフ $\bar{P} \subset P$ を得る (図 2-b)。単純パスは同じ頂点が含まれないようなパスであり、同じ化合物を反応経路中で二度使わない反応経路のみを考える。
2. パスグラフ \bar{P} の頂点集合を V 、辺集合を E とする。このとき、このグラフに付随する新しいグラフを、反応集合 E を頂点とし、辺集合は反応 r_1 の産物と r_2 の基質が等しいかどうかの二項関係により構成する。つまり $c_1 \xrightarrow{r_1} c_2 \xrightarrow{r_2} c_3$

の場合、頂点 $r_1 \in E$ から頂点 $r_2 \in E$ に辺を張ることとする。すなわち、集合 Z

$$Z := \{(r_1, r_2) \mid c_1 \xrightarrow{r_1} c_2 \xrightarrow{r_2} c_3, c_i \in C\}$$

を辺集合としたグラフ $L(P) := (E, Z)$ を構成する。しばしば $L(P)$ はグラフ P の有向線グラフと呼ばれる。実際には出発化合物 s も始点として加え、 s から始まる反応群 $Z_s := \{(s, x) \mid x \in C\} \subset R$ へ辺を張り、同様に最終産物 t を終点として加え、化合物 t で終わる反応群 $Z_t := \{(x, t) \mid x \in C\} \subset R$ へ辺を張る。以上でパスグラフ \bar{P} からその双対グラフ

$$\tilde{P} := (E \cup \{s, t\}, Z \cup Z_s \cup Z_t)$$

を得る (図 2-c)。

3. グラフ \tilde{P} の s, t 以外の各頂点 $r \in E$ をその反応に関与する酵素遺伝子の集合 (頂点クラス) $F(r) \subset G$ で置き換える。その後、隣接する頂点クラス間がそれぞれ完全二部グラフになるように、つまり \tilde{P} の内部辺 $(r_1, r_2) \in Z$ それぞれについて、頂点集合 $F(r_1)$ と頂点集合 $F(r_2)$ が全結合になるように $F(r_1)$ の要素から辺を張る。

また、各辺 $r \in Z$ について遺伝子ペア (辺 r の始点 $\text{tail}(r)$, 終点 $\text{head}(r)$) の間の発現の非類似度 $d(e(\text{tail}(r)), e(\text{head}(r)))$ を重みとして付与する。この値は辺に相当する各遺伝子ペアの共発現の度合いが高いほど 0 に近づく正值とする。また始点 s からの辺および終点 t への辺については重みを 0 とする。このようにして、遺伝子を頂点とし半順序頂点クラスを持つ重み付き有向グラフ P' が得られる (図 3)。

4. グラフ P' において k 最短 s - t パスを列挙する。すなわち s - t パスのうち、重みの和の小ささが遺伝子系列中の隣接共発現度の強さを表し、強度の高い順に上位 k 個の遺伝子系列を得る。

注意

- ステップ 2 において得られるグラフは、反応を頂点とするグラフとしてしばしば用いられるものである。頂点 s から頂点 t へ至るすべてのパスは可能な反応経路全体を表す。

- 出発物質 s から最終物質 t までの反応経路の連鎖反応で $\dots \xrightarrow{r_1} \dots \xrightarrow{r_2} \dots$ のように反応 r_1 が反応 r_2 より前に起こる反応である場合、これを半順序関係とみて $r_1 \preceq r_2$ と書くことにする。

ステップ2において、図2-cの例で $7 \rightarrow 2$ に辺がある場合など、この半順序 \preceq の推移律が崩れる ($7 \preceq 2$ かつ $2 \preceq 6 \not\Rightarrow 7 \preceq 6$) 場合、計算量の削減のため、頂点を複製して半順序を保存するようにする。実際は、最初のパスグラフ \bar{P} を構成するためのパスマージの際に、 P の s - t 単純パスを辺ラベルでアライメントをとり対応が取れたところのみをマージすることで半順序を保存したグラフ \tilde{P} が得られる。しかし、本研究ではそのようなパスはほとんど出現しなかったため、グラフ \tilde{P} では上記の半順序 \preceq は保存されているとし上記の簡素な手続きとした。

- ステップ3において、反応集合 $E \subset R$ に関与する遺伝子群 $G_E \subset G$ について、各々の遺伝子の発現プロファイル $\{e(g) \in \mathbb{R}^a | g \in G_E\}$ が計算できるように、対象としたい G_E の遺伝子をすべて含むような対象生物種のマイクロアレイ実験のデータを α 個選んで取る。
- ステップ4において、グラフの k 最短 s - t パスは Eppstein のアルゴリズム [11] や Martins-Santos[12] のアルゴリズムなどによって効率的に計算できるが、規模が大きい場合でも頂点クラス半順序構造によりグラフ P' が有向無閉路グラフ (DAG) となり計算の効率化が期待できる。

2.4 反応→遺伝子集合間の対応付け

関数 F とおいた処理、すなわちある反応からその反応に関与する酵素遺伝子集合を得る処理には、KEGG PATHWAY 中の各反応にアノテーションされている遺伝子集合を用いる。ここではさらに、共発現の統計的有意性を考察するために、各頂点クラスタについて、その生物種が持ち、アノテーションされていない類似配列の中で最も似ているものをさらに付与することにする。

パスウェイデータを初めとする知識ベースでは、

一般に、知識の不完全さ、データのミス、入力時のミス、など各種の質にばらつきがある。このようなパラログを付与する操作により候補生成を柔軟にすることでロバスト性を向上させることが期待できる。あるいは、未知の経路や未入力経路といった対象生物種のデータに本来存在しない経路にたいしても、背景となるリファレンスパスウェイが存在すれば他の生物種の酵素遺伝子の類似配列を候補遺伝子として付与することもできる。

こうした処理は代謝経路の転写活性の動的な振舞いを調べるだけでなく、パスウェイデータの無矛盾性や検査など、知識データの維持管理のためにも有用であると考えられる。

2.5 遺伝子→発現プロファイル間の対応付け

今回データとして用いる解糖系のような反応系では反応速度が早い場合経路上の隣接遺伝子は共発現していると仮定する。特に様々な条件下で発現の振る舞いが似ていれば代謝経路上でも近い関係にあると期待できる。従って、できるだけ様々な条件下でのデータを得るため、発現プロファイルとしては GEO[9] のデータベース中の cDNA マイクロアレイ実験データから、対象遺伝子群をすべて含んだ実験を横断的に用いることとした。

3 実験と結果

3.1 データと条件

データとして KEGG PATHWAY の代謝経路データから、典型的な出芽酵母の α -D グルコースからピルビン酸に至る解糖系の経路 (パラログを含めて 33 遺伝子) を用いた。このマップが含む酵素遺伝子群をすべて含むようなマイクロアレイ実験を GEO の cDNA マイクロアレイ実験データから検索し、807 個の実験データを得た。これより発現プロファイルは遺伝子数 $33 \times$ 実験条件数 807 のデータとなった。また、発現プロファイル間の距離尺度 (非類似度尺度) としては標準的な「1.0-相関係数」を用いた。

このようにして得た GEO の様々な条件下での発現プロファイルを用いて、出芽酵母の代謝経路に上記方法を適用し、隣接共発現度の高い順に全 518400 個中、上位 21 個の遺伝子系列を得た。

図3の頂点クラスタ 4-6 にあるように、対象経路には反応 ID が異なるが遺伝子が同じものが

あり、これらは非類似度が0になるので、例えば YBR196C はどのような順番で使われてもスコアに変化を与えない。このため、3つずつ (YBR196C → YBR196×2 と YBR196 だけ) の同スコアのパターンが出現する。スコアリングとしてはこれらは同一の順位として扱い、次のものを +1 したランクとして扱った。

3.2 結果と考察

結果を表 1 に示す。これより高い隣接共発現を示した遺伝子系列では追加的に付与したパラログは用いられていないこと、複数遺伝子が関与する反応においても、頂点クラスタ 2(反応 R01786) では YGL253W が、頂点クラスタ 12(反応 R01512) では YKL152C が、頂点クラスタ 14(反応 R00200) では YAL038W がそれぞれ顕著に使われている。また変動する場合においても頂点クラスタ 7 の YMR205C/YGR240C、頂点クラスタ 10 の YJR009C/YGR192C、頂点クラスタ 13 の YGR254W/YHR174W というような候補で顕著に特徴づけられていることが分かる。

これらは個々の実験条件で見た場合、また個別のランキングを持っており、本手法によって、ある一連の実験条件と別の一連の実験条件との間の代謝経路の転写活性の差異を調べることも考えられる。

3.3 ブートストラップによる平均ランキング

酵素の反応は様々な要因によって調節されているが、速い反応系においてもっとも単純なレベルでは、連続する反応が進行するためには少なくとも関与する酵素遺伝子が転写されている必要があることより (あるいは進まないときには転写量が低い必要がある)、様々な条件下で粗視的に見て転写のされ方が代謝経路上で遠くにあるものと比べ似ていると期待できる。

ここでは 807 の実験データを重複を許して再度 807 回選びなおす (リサンプリングする) ステップを繰り返すことで、仮想的にこれらの個々の実験に依存しない全体的な傾向を調べた。このブートストラップ手続きを反復するたびに、807 個のデータの内、重複して選ばれるものや一度も選ばれないものをランダムに作り出すことで、統計的には同様の傾向を持つ、多数の発現プロファイルを擬

似的に作成することができる。

ここではブートストラップ手続きの 100 回の平均ランキングにより、平均隣接共発現度の高い順に全 518400 個中、上位 21 個の遺伝子系列を得た。結果を表 2 に示す。

上位 21 個のうち 18 個までは同じ系列であり元の k 個が統計的に有意に上位ランクの系列であることが分かる。従って、様々な条件下での実験データによる発現プロファイルによって、同様の性質を把握できることが分かる。

4 結論と今後の課題

本研究では遺伝子の発現情報を用いて、代謝経路の挙動を動的に分析する一手法を提示した。代謝パスランキングは、用いるマイクロアレイ実験の条件を吟味することで、より広い範囲で応用できる。

今後の課題として様々な距離尺度における結果比較、転写レベルの考慮、タイムコースデータにおける時間遅れの考慮、BRENDA[13] 等の反応データベース情報を用いた反応速度などの考慮 (例えば、反応列中の律速となっている酵素 (キーエンザイム) をコードする遺伝子の考慮) などが挙げられる。このように関連酵素遺伝子の発現情報と既存の代謝経路の生化学的知識を併せて、代謝の遺伝子の観点からの仕組みの理解や、薬物代謝・先天的代謝異常など特別な反応における個体差解析への応用等を目指したい。

謝辞

本研究を行うにあたって、度々にわたり様々な貴重なご教示をいただきました京都大学大学院薬学研究科 奥野恭史助手、および、キリンビール株式会社 米谷隆研究員に深く感謝します。

参考文献

- [1] V. Hatzimanikatis., C. Li, J. A. Imita, C. S. Henry, M. D. Jankowski, and L. J. Broadbelt, Metabolic networks: enzyme function and metabolite structure, *Current Opinion in Structural Biology*, **14**, 300–306, 2004.
- [2] J. A. Papin, N. D. Price, S. J. Wiback, D. A. Fell, and B. O. Palsson, Metabolic

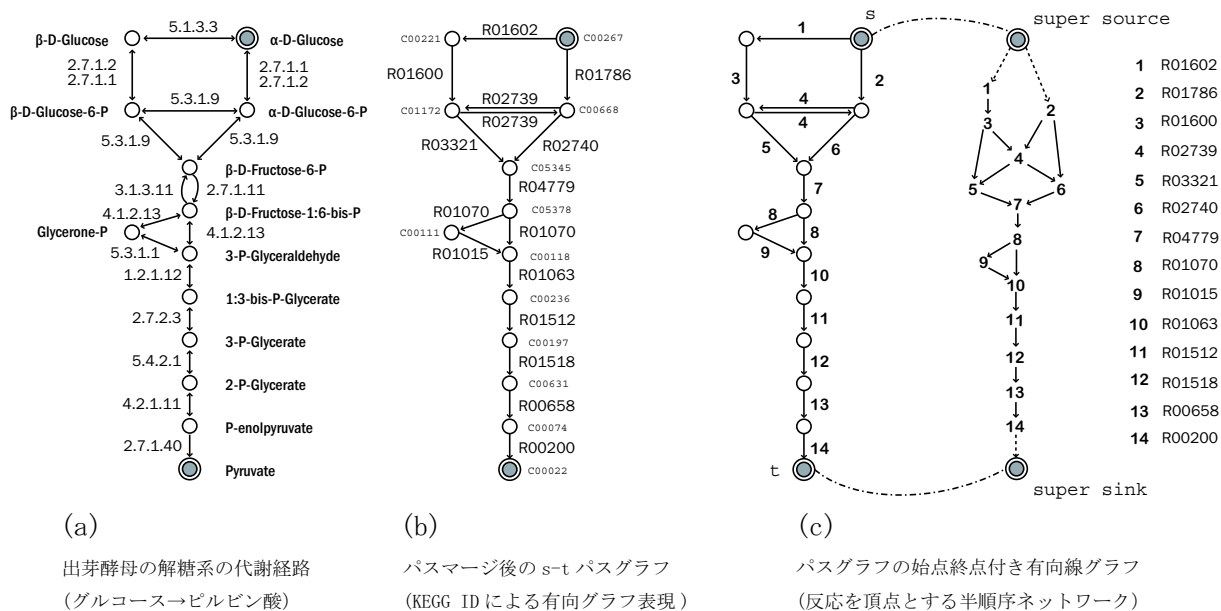
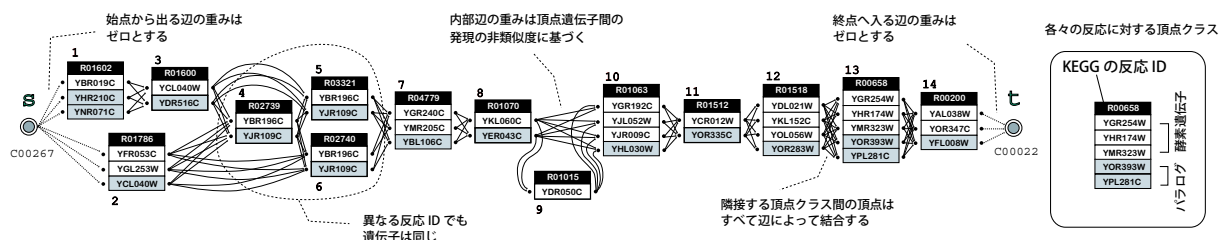


図 2: 対象とする代謝経路、化合物パスグラフ、および付随する反応グラフ



pathways in the post-genome era, *Trends in Biochemical Sciences*, **28**(5), 250–258, 2003

[3] G. Michal, *Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology*. John Wiley & Sons, 1998.

[4] M. Kanehisa, S. Goto, M. Hattori, K. F. Aoki-Kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki, and M. Hirakawa, From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Research*, **34**, D354–357, 2006.

[5] I. M. Keseler, J. Collado-Vides, S. Gama-Castro, J. Ingraham, S. Paley, I. T. Paulsen, M. Peralta-Gil, P. D. Karp, EcoCyc: a comprehensive database resource for Escherichia coli. *Nucleic Acids Research*, **33**, D334–337, 2005.

[6] R. Overbeek, N. Larsen, G. D. Pusch, M. D’Souza, E. Selkov Jr., N. Kyrpides, M. Fonstein, N. Maltsev, E. Selkov, WIT: integrated system for high-throughput genome sequence analysis and metabolic reconstruction. *Nucleic Acids Research*, **28**(1), 123–125, 2000.

表 1: 隣接共発現度の上位 21 個までの遺伝子系列

ランク		非類似度尺度: 1.0-Pearson Correlation																						
スコア		パス																						
1	(2.809)	s	→	YGL253W	→	YBR196C	(→	YBR196C)	→	YMR205C	→	YKL060C	→	YJR009C	→	YCR012W	→	YKL152C	→	YGR254W	→	YAL038W	→	t
2	(2.825)	s	→	YGL253W	→	YBR196C	(→	YBR196C)	→	YMR205C	→	YKL060C	→	YJR009C	→	YCR012W	→	YKL152C	→	YHR174W	→	YAL038W	→	t
3	(2.889)	s	→	YGL253W	→	YBR196C	(→	YBR196C)	→	YMR205C	→	YKL060C	→	YGR192C	→	YCR012W	→	YKL152C	→	YGR254W	→	YAL038W	→	t
4	(2.905)	s	→	YGL253W	→	YBR196C	(→	YBR196C)	→	YMR205C	→	YKL060C	→	YGR192C	→	YCR012W	→	YKL152C	→	YHR174W	→	YAL038W	→	t
5	(2.929)	s	→	YGL253W	→	YBR196C	(→	YBR196C)	→	YGR240C	→	YKL060C	→	YJR009C	→	YCR012W	→	YKL152C	→	YGR254W	→	YAL038W	→	t
6	(2.945)	s	→	YGL253W	→	YBR196C	(→	YBR196C)	→	YGR240C	→	YKL060C	→	YJR009C	→	YCR012W	→	YKL152C	→	YHR174W	→	YAL038W	→	t
7	(3.009)	s	→	YGL253W	→	YBR196C	(→	YBR196C)	→	YGR240C	→	YKL060C	→	YGR192C	→	YCR012W	→	YKL152C	→	YGR254W	→	YAL038W	→	t

表 2: 隣接共発現度の上位 21 個までの遺伝子系列 (100 回ブートストラップ平均ランク)

ランク		元ランク		非類似度尺度: 1.0-Pearson Correlation																				
				パス																				
2.39	(2)	s	→	YGL253W	→	YBR196C	(→	YBR196C)	→	YMR205C	→	YKL060C	→	YJR009C	→	YCR012W	→	YKL152C	→	YHR174W	→	YAL038W	→	t
2.79	(1)	s	→	YGL253W	→	YBR196C	(→	YBR196C)	→	YMR205C	→	YKL060C	→	YJR009C	→	YCR012W	→	YKL152C	→	YGR254W	→	YAL038W	→	t
4.82	(4)	s	→	YGL253W	→	YBR196C	(→	YBR196C)	→	YMR205C	→	YKL060C	→	YGR192C	→	YCR012W	→	YKL152C	→	YHR174W	→	YAL038W	→	t
4.97	(6)	s	→	YGL253W	→	YBR196C	(→	YBR196C)	→	YGR240C	→	YKL060C	→	YJR009C	→	YCR012W	→	YKL152C	→	YHR174W	→	YAL038W	→	t
5.28	(3)	s	→	YGL253W	→	YBR196C	(→	YBR196C)	→	YMR205C	→	YKL060C	→	YGR192C	→	YCR012W	→	YKL152C	→	YGR254W	→	YAL038W	→	t
5.48	(5)	s	→	YGL253W	→	YBR196C	(→	YBR196C)	→	YGR240C	→	YKL060C	→	YJR009C	→	YCR012W	→	YKL152C	→	YGR254W	→	YAL038W	→	t
8.24	(-)	s	→	YGL253W	→	YBR196C	(→	YBR196C)	→	YGR240C	→	YKL060C	→	YGR192C	→	YCR012W	→	YKL152C	→	YHR174W	→	YAL038W	→	t

- [7] L. B. Ellis, D. Roe, L. P. Wackett, The University of Minnesota Biocatalysis/Biodegradation Database: the first decade. *Nucleic Acids Ressearch*, **34**, D517–521, 2006.
- [8] C. A. Ball, I. A. Awad, J. Demeter, J. Gollub, J. M. Hebert, T. Hernandez-Boussard, H. Jin, J. C. Matese, M. Nitzberg, F. Wymore, Z. K. Zachariah, P. O. Brown, G. Sherlock, The Stanford Microarray Database accommodates additional microarray platforms and data formats. *Nucleic Acids Ressearch*, **33**(1), D580–582, 2005.
- [9] T. Barrett, T. O. Suzek, D. B. Troup, S. E. Wilhite, W. C. Ngau, P. Ledoux, D. Rudnev, A. E. Lash, W. Fujibuchi, R. Edgar, NCBI GEO: mining millions of expression profiles—database and tools. *Nucleic Acids Ressearch*, **33**(1), D562–566, 2005.
- [10] H. Parkinson, U. Sarkans, M. Shojatalab, N. Abeygunawardena, S. Contrino, R. Coulson, A. Farne, G. G. Lara, E. Holloway, M. Kapushesky, P. Lilja, G. Mukherjee, A. Oezcimen, T. Rayner, P. Rocca-Serra, A. Sharma, S. Sansone, A. Brazma, ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic Acids Ressearch*, **33**(1), D553–555, 2005.
- [11] D. Eppstein, Finding the k shortest paths. *SIAM Journal on Computing*, **28**(2), 652–673, 1998.
- [12] E. Q. V. Martins and E. Santos, A new shortest paths ranking algorithm, *Investigação Operacional*, **20**(1), 47–62, 2000.
- [13] I. Schomburg, A. Chang, C. Ebeling, M. Gremse, C. Heldt, G. Huhn, D. Schomburg, BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Ressearch*, **32**, D431–433, 2004.