

# 代謝経路を用いた遺伝子発現情報の視覚化システム

田丸 良子<sup>†</sup> 天笠 俊之<sup>†</sup> 金谷 重彦<sup>†</sup> 植村 俊亮<sup>†</sup>

<sup>†</sup> 奈良先端科学技術大学院大学 情報科学研究科 〒 630-0192 奈良県生駒市高山町 8916-5

E-mail: †{r-tamaru,amagasa,skanaya,uemura}@is.aist-nara.ac.jp

あらまし 近年, DNA マイクロアレイ技術は飛躍的に発展し, それにより遺伝子発現情報を網羅的にモニターすることが出来るようになってきている. よく似た発現パターンを持つ遺伝子のセットはよく似た機能を持つと言われていることから, DNA マイクロアレイの情報から未知遺伝子の機能を推測することが可能である. しかしながら, 現状では生物学的に意味のあるデータを抽出することは簡単ではない. 遺伝子発現情報を解析するための一つのアプローチとして遺伝子発現パターンの代謝経路へのマッピングがあげられる. 遺伝子は互いに相互作用し, 複雑なネットワークを気づいていることが知られている. このようなネットワークは, ノードが遺伝子を, エッジが遺伝子の相互作用を現すようなグラフとして表現できる. 生体内相互作用ネットワークは非常に複雑で, 高度に再現することは難しく, 現在はこれらの関係を視覚化する標準的なパッケージは存在しない. 本論文で, 我々は遺伝子発現情報を代謝経路図にマッピングするシステムを提案する. 代謝経路の視覚化には, XML 準拠の SVG を用いた. SVG は JavaScript などを用いることにより, ダイナミックな表示が可能となる. 本システムを用いることにより, 代謝経路における遺伝子発現の動きを包括的に観察することが可能になる.

キーワード 遺伝子発現, DNA マイクロアレイ, 代謝経路, SVG

## A System for Visualizing Gene Expressions using Metabolic Networks

Ryoko TAMARU<sup>†</sup>, Toshiyuki AMAGASA<sup>†</sup>, Shigehiko KANAYA<sup>†</sup>, and Shunsuke UEMURA<sup>†</sup>

<sup>†</sup> Nara Institute of Science and Technology, Graduate School of Information Science

Takayama 8916-5, Ikoma, Nara, 630-0192 Japan

E-mail: †{r-tamaru,amagasa,skanaya,uemura}@is.aist-nara.ac.jp

**Abstract** It has become possible to monitor gene expressions comprehensively with the rapid progress of the DNA microarray technique. Microarray data is useful for estimating the functions of a gene because a set of genes is considered to have similar functions if the genes have similar gene expression profiles. However, it is difficult to extract biologically meaningful data from such data. One promising approach for browsing gene expression data is to map them onto genetic networks. It is a well-known fact that genes interact with each other and form complex networks. Genetic networks are modeled as graphs where a node represents a gene and an edge represents an interaction of the genes at the ends. Estimating genetic networks using gene expression data is one of the most important tasks in the field of bioinformatics. In this paper, we propose a system for visualizing gene expressions data by mapping them onto metabolic pathways. In our implementation, we employ SVG for drawing graphs to support dynamic redrawing of metabolic networks. As a result, when you use our system, you can monitor movement in gene expressions comprehensively.

**Key words** Gene Expressions, DNA Microarray, Metabolic Pathway, SVG

### 1. はじめに

#### 1.1 研究の背景

ワトソンとクリックが 1953 年に DNA 二重螺旋構造を解明 [1] してからわずか 50 年で, ヒトゲノムの完全解読が宣言された. またこれまでに 100 種類以上もの生物のゲノムが読まれるなど, 遺伝子を取り巻く分子生物学はめまぐるしい進化を見せている.

ゲノム解読研究は, 現在, 遺伝子発現の分子制御などに関与するいわゆる調節領域の塩基配列情報の解読に焦点が移っている. 今後は, 複数の遺伝子と蛋白質の相互作用が織りなす生体反応をより深く理解するために, RNA と蛋白質の発現, 蛋白・蛋白の相互作用やシグナルパスウェイなどの研究が重要となる. 遺伝子発現による全 mRNA をゲノムワイドに検索するための主力な解析手法の一つとして, DNA マイクロアレイが開発さ

れた。遺伝子発現のデータは、発現プロファイルの数が数千から数万にもなる高次元のデータである。最近の主流は、遺伝子発現パターン類似性により遺伝子を分類するクラスタリングなどの統計解析手法であるが、多くの研究者はアルゴリズムを十分理解できておらず、統計学が必須の高次元のデータ解析に頭を悩ませている [2]。2003 年 8 月の Nature にもあるように、これまでの論文の解析には、間違った結論を出しているものが少なくないと言われている [3]。

機能未知の遺伝子に対しては、遺伝子発現パターンを類似したものに分類する処理 (クラスタリング) がよく行われている。また、代謝・シグナル伝達など既知のパスウェイに関しては、そのネットワークを視覚化し、その上に遺伝子発現パターンをマッピングする方法がある。近年、遺伝子や蛋白質を初めとする各種データベースを統合したデータベースが増加している。しかし、高度に統合されたデータベースであっても、データの多くが文字情報のみを提供しているに過ぎず、研究者にとって利用し易いとは言いがたい。特に化合物の構造やネットワークなどは視覚的に表現しないと良いアイデアは浮かびにくい。今後はどのようにこれらのデータを可視化していくかが問題となっている。

## 1.2 システムの概要

今回我々は遺伝子発現パターンを代謝経路 (パスウェイ) にマッピングするシステムを提案する。遺伝子発現情報には、生物学的な現象や特徴を対応付けしていく必要があり、これらの注釈情報を付け加えられた実験データをどのように視覚化して表現するかが重要である。本システムでは、Web 上で公開されている電子化された代謝データベースの中から、XML 形式で記述された枯草菌の代謝に関する情報を利用している。まず、これらの代謝データから必要な情報を抽出し、グラフ描画ツールを用いて座標計算を行う。次に、出力された SVG 画像には、マウスイベントを起こすために、新たな要素やスクリプトなどを追加し、代謝マップが完成する。本システムではマウスをノードに重ねることにより、ノードの情報 (ラベルされている遺伝子名、ノードの属する代謝名) を画面に表示したり、任意の代謝経路を強調するなどのイベントを起こすことができる。また、ノードをクリックすることにより、ラベルされている遺伝子の配列などの詳細情報を記した別画面が起動する。本システムを利用して遺伝子発現情報を代謝経路にマッピングすることにより、ある経路上でつながりのある遺伝子間の相互作用を網羅的に観察することができる。

## 2. 基本事項と関連研究

本論文で登場する 遺伝子発現, DNA マイクロアレイ, 代謝, XML と SVG について説明し, 関連研究について述べる。

### 2.1 遺伝子発現

ある生物の持つ DNA に含まれる全遺伝情報、またはこの遺伝情報を担う DNA をゲノムという。ゲノムの中で生命活動に必要な機能を持つタンパク質や RNA を規定している部分を遺伝子という。ゲノムとは遺伝情報の伝達の単位であり、遺伝子とは情報発現の単位である。DNA にコードされた遺伝情報は、

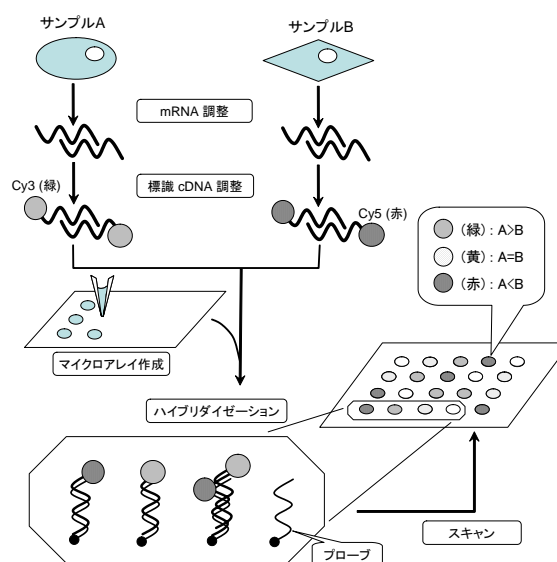


図 1 DNA マイクロアレイ

Fig. 1 DNA microarray

複製の過程で DNA から DNA に遺伝情報が伝えられ、転写と翻訳の過程で遺伝子から mRNA を介してタンパク質へと情報が伝えられる。このようにして遺伝子の情報が発現すること、すなわち、転写や翻訳が起こることを「遺伝子発現」と言う。すべての生物は自分自身の遺伝子の発現を巧妙に制御している。遺伝子にはタンパク質に関する情報はコードされているがその発現の制御に関する情報はコードされていない。したがって、遺伝子部分を解読し、タンパク質の立体構造や基本的な機能を決定しただけでは、生命のシステムは理解することはできない。生命システムを理解するには、どのような時期、状態の場合に遺伝子のスイッチがオン / オフになるのかという遺伝子の発現を制御する調節領域の配列を解釈しなければならない

### 2.2 DNA マイクロアレイ

マイクロアレイとは、特殊加工したスライドガラスなどの固相基板上にプローブと呼ばれる塩基配列の異なる何千何万種類の DNA を格子状に固定させたものである [4]。蛍光標識した試料 DNA または RNA を基板上の DNA と結合させ、各々のプローブからのシグナルを検出し、そのデータをコンピューターで解析する。オリゴ DNA や cDNA をプローブとしたマイクロアレイに、mRNA から逆転写酵素反応により蛍光標識を取り込ませたターゲット cDNA をハイブリダイズさせて、検出器で蛍光強度を検出する。一枚のマイクロアレイからは数千から数万の配列データが得られるため、数値解析、生物学的に意味のある情報を抽出などの過程において、大量のデータを対象とした統計解析や遺伝子情報処理が必要不可欠である。

コントロール細胞とサンプル細胞から mRNA サンプルを抽出し、蛍光標識した cDNA を同一マイクロアレイ上で競合的ハイブリダイゼーションを行い、両者の蛍光色素を測定・比較する方法を二蛍光標識法という (図 1)。蛍光色素には現在 Cy3 (緑) と Cy5 (赤) が最もよく用いられている。この方法は相対的な変化を見るのにきわめて有効であり、蛍光色素 Cy3, Cy5 はそれ自体には色は付いていないが、スキャン後ソフトウェア

的に強度の大きさに応じてそれぞれ緑色、赤色がコンピュータモニター上で着色される。したがって、ある遺伝子がコントロール細胞ではほとんど発現せず、サンプル細胞では過剰に発現しているような場合、その遺伝子に対応するスポットは赤色に見えることになる。緑に見えるスポットは正常細胞でのみ発現している遺伝子であり、黄色に見えるスポットは2つの細胞で同程度、共に発現しているものである。両方の細胞で共に発現がない遺伝子に関してはスポットは黒色に見えることになる [5]。

### 2.3 代謝

生体内において、酵素を触媒としてある化合物（基質）が別の化合物（生成物）に変化する化学変化を代謝という。代謝反応においてある反応の生成物が別の反応の基質となることから、代謝反応は複雑なネットワークを形成している。酵素は遺伝子産物であることから、ある代謝に関わる酵素をコードする遺伝子の発現を調べることで、その代謝の制御の様子を知ることができる。

### 2.4 XML と SVG

#### 2.4.1 XML

Extensible Markup Language: 拡張可能なマークアップ言語 (XML) は、ISO が標準化した SGML (Standard Generalized Markup Language) から派生した汎用的なデータ記述言語で、1998年2月にW3Cにより勧告された [6]。XMLは、利用者が独自のタグを使ってデータを囲むことによって、データの属性情報や論理構造を定義し、データを構造化することができる。データの意味と値が明確であることから、アプリケーション間のデータの交換など、厳密に定義されたデータが必要な場面で利用できる。W3Cにより標準化されていること、扱いやすいことなどの理由から、XMLは、現在ではシステム/アプリケーション開発における必須の要素技術となっている。

データの種類や形式の異なる多数のバイオ情報データベースを相互に連携させるため、XMLをベースとするデータ標準形式の設計が進められている。代表的なデータベースの例としては、DDBJ-ML [7]、PDBj-ML [8]、KEGG [9] などが挙げられる。

#### 2.4.2 SVG

SVGは、Scalable Vector Graphicsの略で、XML準拠の二次元ベクトルグラフィックスのフォーマットである [10]。2002年4月、W3Cにより勧告候補として公開された。SVGにはベクトルグラフィック（直線や曲線など）、イメージ、テキストという三種類のグラフィックオブジェクトがある。これらのグラフィックオブジェクトはグループ化、スタイル化、変形、合成することでレンダリングする前にオブジェクトを作成できる。また、変形のネスト、クリッピングパス、アルファマスク、フィルタ効果、テンプレートオブジェクトなどの機能がある。SVGを利用する利点として、アプリケーションからスクリプト言語のDocument Object Model (DOM) APIなどを用いてアクセスすることが可能であり、エレメント、アトリビュート、プロパティを操作することが可能である。マウスが重なった、クリックされたなどの色々なイベントハンドリングはどのようなSVGのグラ

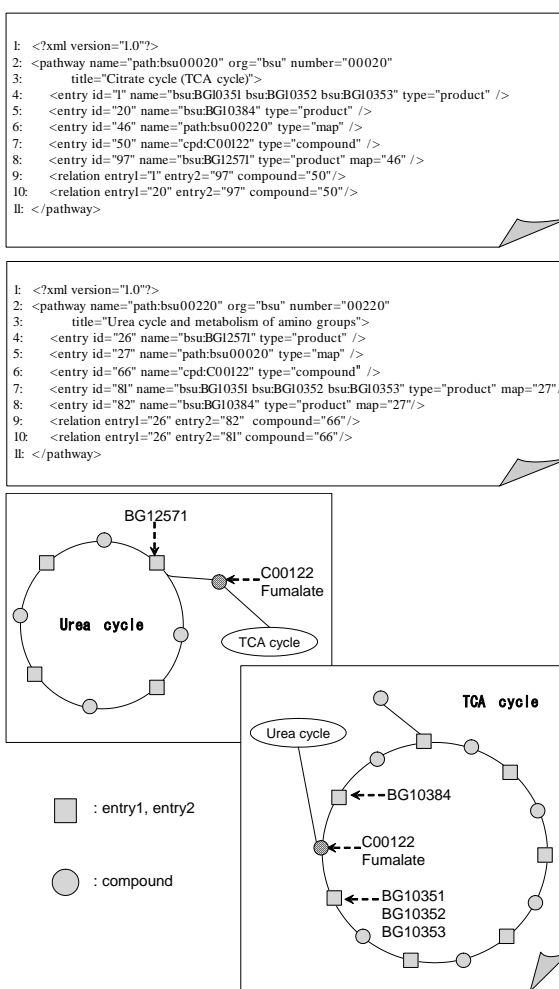


図2 KEGG PATHWAY の例

Fig. 2 A sample of KEGG PATHWAY data

フィカルなオブジェクトにでも割り当てることができる。

### 2.5 関連研究

代謝経路に遺伝子発現パターンをマッピングする研究には、生命システム情報統合データベース (Kyoto Encyclopedia of Genes and Genomes [KEGG]) [11], Encyclopedia of Escherichia coli K12 Genes and Metabolism (EcoCyc) [12] などがあげられる。KEGGのなかのPATHWAYデータベースでは、生体反応を構成する過程が単位ごとに描かれた代謝経路の図が生物種別に提供されている。代謝が各単位ごとに分割されて提供されているために、すべての代謝系を観察することは非常に困難である。また、表示される図を手書きで作成しているため、複雑な生体反応を網羅的に蓄積するには膨大な手間を必要とする。EcoCycは大腸菌に特化した代謝経路に関するデータベースである。生体反応を自動的に描画しているため、生体反応情報の更新に対して容易に対応できることが特徴である。現在のところ大腸菌のみに対応しているが、時系列のマイクロアレイデータをアニメーションで表示する機能がある。

本システムでは、代謝経路を一枚のマップで表現することで、一度に全代謝を網羅できる。代謝経路の描画はデータベースの情報を元に自動描画することで、データの更新に柔軟に対応できる。また、各代謝ごとに強調する機能を付け加えることで、

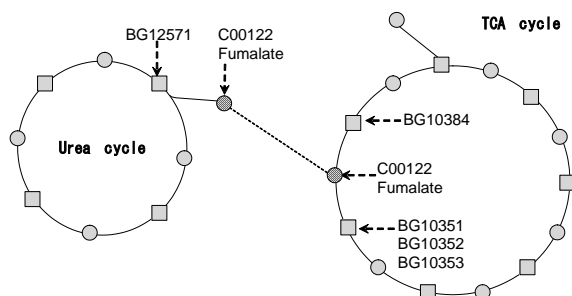


図3 代謝経路の結合

Fig.3 connect metabolic pathways

どの代謝で起きている変化であるかという情報を同時に提供することができる。

### 3. システムの構築

#### 3.1 代謝経路の自動描画

システムの構築の流れを以下に示す。本章ではこの流れに沿って詳しく説明していく。

- (1) KEGG PATHWAY から XML 形式で公開されている枯草菌の代謝データを取得する。
- (2) 代謝経路の描画に必要な情報を取り出し、GraphViz の入力ファイルを生成する。
- (3) オプションの調節をして GraphViz を用いてグラフを作成する。
- (4) SVG を加工してシステムに組み込む。

##### 3.1.1 KEGG PATHWAY の特徴

取得した KEGG PATHWAY のデータの一部を図2に示す。今回用いた枯草菌の代謝データは現在 86 のパスウェイに分けられ、XML 形式で提供されている。XML 形式のデータを用いた理由の一つに、XML には、データモデルとそのアクセスのための標準 API を定めた DOM (Document Object Model) という規格が利用できる点が挙げられる。DOM は特定のプログラミング言語に依存しない汎用的な仕様になっている。このように XML データ操作のための API が標準化されているため XML を利用するアプリケーションの構築が容易になり、プログラムの移植性も高くなるという利点がある。また DOM 以外にもデビッド・メギンソン氏らによって開発され、実質的に‘標準’と呼ばれるほどになった SAX (Simple API for XML) [13], DOM や SAX をラップして抽象化する Java の標準 API, JAXP (Java API for XML Processing) [14] などの API が存在する。

代謝経路は化合物をノード、触媒する酵素はエッジのラベルとして表現されるのが一般的であるが、KEGG PATHWAY では、化合物も酵素も同じレベルのノードとして表現している。このようにして酵素も化合物と同レベルのノードとして扱うことにより、代謝経路上の複雑な反応を表すハイパーエッジを表現しやすくしている。KEGG PATHWAY では、ノードとエッジをそれぞれ ‘entry’, ‘relation’ という要素で定義している。entry 要素に定義されている type 属性により、ノードが化合物であるか、酵素であるかを判別する。relation では、酵素 - 化合物 - 酵素 という酵素-酵素反応で表現される三項関係を定義している。

#### 3.1.2 代謝経路の結合

KEGG PATHWAY では解糖系、クエン酸回路、尿素回路のように各单位ごとに代謝を分割して提供しているが、本システムではこれらの代謝経路を一枚の代謝マップ上で表現する。KEGG PATHWAY では、各代謝に共通して存在する化合物によってそのつながりを表現している。図2で例を挙げているクエン酸回路と尿素回路はフマル酸という共通した化合物でつながっている。一枚の代謝として描画する場合、各代謝のフマル酸は同じノードとして扱われるが、本システムでは図3に示すように、各代謝を区別するために各代謝ごとにフマル酸に対応するノードを作り、これら二つをエッジで結合することにより、各代謝のつながりを表現した。

##### 3.1.3 グラフオプションの調節

本システムでは、KEGG PATHWAY の中でグラフのエッジに相当する情報を記述している relation 要素から、ノードのつながりを表す記述を抽出し、DOT 形式と呼ばれる GraphViz 入力データを作成した。relation 要素の中で定義されていない、どのノードともエッジでつながれていないノードに関しては代謝経路から除いた。GraphViz は AT&T 研究所が開発しているオープンソースのグラフ描画ツールのセットである [15]。要素間の関係式を並べると自動的に座標計算をしてグラフを描画する機能を持つ。要素間の関係式には、ノードの長さや、フォントの大きさ、ノードの形や線のスタイルといった様々な属性が指定できる。このようなグラフオプションを調節せずにグラフを描画すると、ノードが中心部に集まり、中心から離れるほどノード間のスペースが空いてしまう。そこで、パス同士をつなげる化合物を結ぶエッジは長く、化合物を表すノードのラベルフォントを小さくし、全体的にノードが散らばるようにオプションを調節する必要があった。GraphViz は入力データを元に自動的に座標計算を行い、SVG で出力する。

#### 3.2 SVG の加工

GraphViz から出力された SVG に遺伝子発現パターンをマッピングしたり、マウスイベントを起こすためにスクリプトを埋め込み、要素や属性を追加する(図4)。

ノードのラベル部分は遺伝子名に書き換えた。KEGG PATHWAY では各遺伝子に独自のエンタリーをつけており、代謝データから自動的に生成した代謝経路ではラベル部分がエンタリーになっている。エンタリーはデータを処理する際には識別子として扱えるため便利ではあるが、生物学実験には用いられることは少ない。生物学実験者を対象としているシステムを構築しているため、ノードの表示部分は遺伝子名に置き換えた。

各パスウェイごとに強調させるために、同じパスウェイに属するノードとエッジの上にパスウェイを表す要素を付け加えてグループ化する。図4下で示すように、加工を施す前の SVG の構造はノードとラベルがすべて同レベルで存在している。グループ化してパスウェイ要素を追加することで、各パスウェイごとにノードとエッジを区別することができる。パスウェイ同士をつないでいるエッジはどのパスウェイにも属さず、パスウェイ要素と同じレベルに構造化される。

酵素を表すノードには、ノードをクリックした場合に、配列

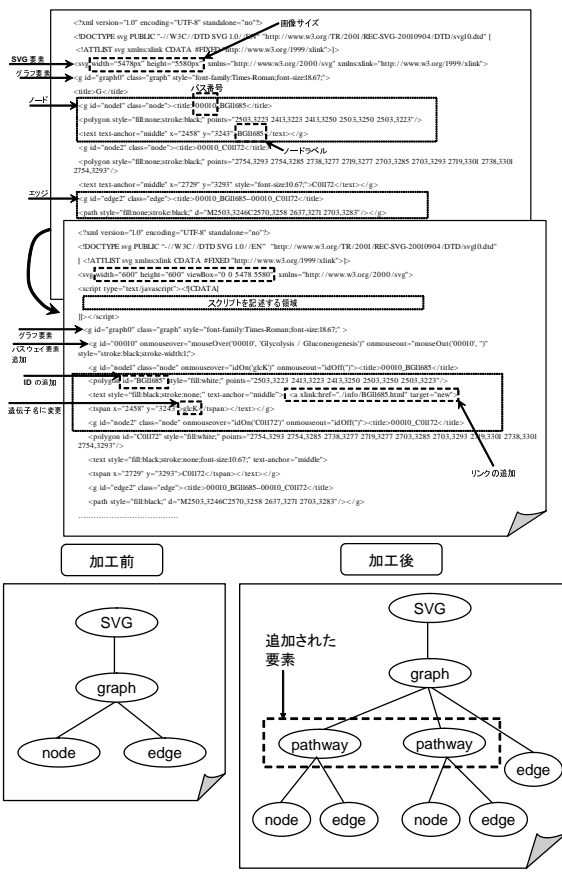


図 4 SVGの加工

Fig. 4 Process for mouse event

などの詳細情報を表す画面を開くためにリンク先を追記した。また、マッピングのためにノードに識別子を追加した。

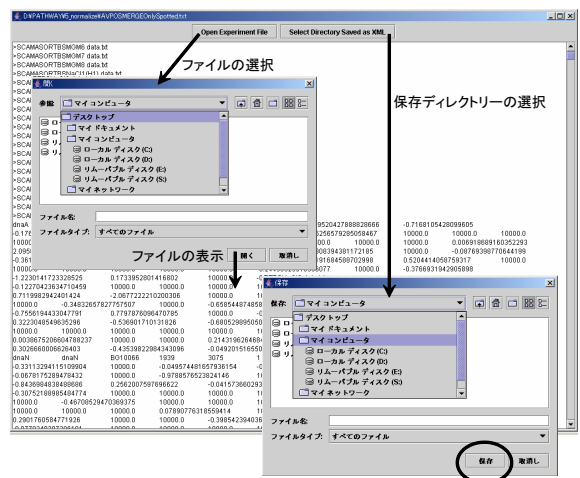
## 4. システムの概要

### 4.1 入力データの準備

本システムでは、二蛍光標識法を用いた DNA マイクロアレイの実験を対象としている。あるスポットに対して得られた二種類の蛍光色素の強度の比を正規化 (normalization) したデータを用いる。正規化は、ある実験におけるコントロール細胞とサンプル細胞での発現量の比の平均値が各実験間で等しくなるような処置を施すことを目的としている。正規化の作業は Microsoft社の Excel などのグラフ作成ソフトを用いて行うこともできるが、本研究で扱っている実験データの正規化には、奈良先端科学技術大学院大学 情報科学研究科 金谷重彦助教教授により開発された MicroArrayInfomatics システムを用いた。遺伝子発現データを XML 形式で記述しようとする活動があり、MAGE-ML という標準形式が存在するが、実際生物学者に認識され、利用されてるとは言い難い [16]。また我々のシステムでは、遺伝子発現量のみを利用し、実験条件などの情報は考慮していないため、今回は独自に文書構造を規定した。

#### 4.1.1 操作手順

図 5 上に DNA マイクロアレイデータを XML に変換する手順を示し、下に入力する実験データ形式と変換後の XML データ形式を示す。データ変換画面の上部には二つのボタンがある。



```

>SCAMASORTBSnaerobic1 data.txt
>SCAMASORTBSnaerobic2 data.txt
>SCAMASORTBSnaerobic3 data.txt

```

スポットID	遺伝子名	遺伝子ID	ゲノム上の位置と向き	補正対数比 (A)	(A)の順番と対応する
dnaA	dnaA	BG10065	410 1750	-0.49520427888828666	-0.716805428099605
dnaN	dnaN	BG10066	1939 3075	10000.0	10000.0
yanA	yanA	BG10067	3206 3421	0.4709723843824544	0.07867103540940873
				0.18291408729360997	

```

<?xml version="1.0" ?>
<!--offset-->
<!--title--> SCAMASORTBSnaerobic </title>
<!--value spoID="dnaN" geneName="dnaN" keggID="BG10066" pos_from="1939" pos_to="3075"
direction="T" color="white" nodeID="BG10066 BG10083 BG10095 BG10642 BG10737 BG10752 BG12656"
10000.0 -->

```

遺伝子に対応する SVG 上のノード ID を追加

図 5 入力ファイル作成画面とデータ変換後の XML ファイル

左のボタンで実験データを選択するとその内容がボタンの下のテキスト領域に表示される。選択されたデータの内容を確認したら右のボタンで保存ディレクトリーを選択し、保存ボタンを押す。

正規化後の DNA マイクロアレイデータは、上から順に正規化に用いた実験データがあり、その後、スポット ID、遺伝子名、遺伝子 ID (KEGG で使われているエントリー)、ゲノム上の位置と方向、補正対数比の順で並んだデータが記述されている。このように複数の実験データを一度に扱うことが可能となっている。

#### 4.1.2 属性の追加

XML 形式に変換されたデータには、元のデータに加えて、正規化された数値から対応するノードに対する色を決定するカラー属性と、マッピングするノードを識別するためのノード ID が追加される。正規化された測定値は、二種類の蛍光色素の強度の比 (Cy5 (赤)/Cy3 (緑)) であり、本システムでは比が三倍以上のものを有意な値とし、三倍以下、三倍以上、五倍以下と、五倍以上でノードの色に変化を持たせた。すなわち、Cy5 の値が Cy3 に比べて五倍以上の発現強度を示す場合には赤、三倍以上五倍以下の場合はピンクを、Cy3 の値が Cy5 の値に比べて五倍以上発現している場合は緑、三倍以上五倍以下の場合は黄緑を、両者の比が三倍以下なら黄色を、測定不能の場合はノードの色に変化はない (白色のまま)。本システムでは、スクリプトでこの追加属性の二つを読み込み、SVG 上の対応するノード

を認識し、スタイル属性に記述されているノードの色を書き換えることで発現変化を表現する。

#### 4.2 システムのレイアウトと基本操作

本システムのレイアウトを図 6 に示す。画面全体は XHTML で作成し、SVG を埋め込んでいる。代謝経路は画面左側に、操作部分は右側に配置した。本システムでは、利用者が Browse ボタン、またはセレクトメニュー項目から任意のマイクロアレイデータ (XML 形式) を選択すると、代謝経路上の対応するノードの色が変化する仕組みになっている。この時、発現状況を調べたい代謝がある場合には、画面右下に分類されている代謝名にチェックをつけておくと、その代謝に属するノードとエッジが強調される仕組みがある。代謝経路上にマッピングされた様子を見て、あるノードについて調べたい場合、ノードにマウスを重ねると、操作部分のテキスト領域にノードにラベルしている遺伝子名と、そのノードが属するパスウェイの名前が表示される。更に、ノードをクリックするとラベルされている遺伝子の配列や他のデータベースへのリンクが貼られた別ウィンドウが開く。また、ズームボタンを用いると、任意の部分を拡大・縮小することが可能である。

### 5. システムの実装

枯草菌を用いた DNA マイクロアレイのデータを、構築したシステムの代謝経路にマッピングすることにより、嫌気条件下においた枯草菌の代謝経路における遺伝子発現パターンを検証する。DNA マイクロアレイのデータは、奈良先端科学技術大学院大学 情報科学研究科の小笠原研究室から提供を受けた。

#### 5.1 実験対象とその特徴

枯草菌 (*Bacillus subtilis*) は、通性嫌気性<sup>(注1)</sup>、孢子形成菌で、分子生物学の研究対象として重要なモデル生物の一つである。枯草菌のエネルギー代謝の中心は他の生物と同様 ATP (アデノシン三リン酸) である。呼吸や醗酵によりエネルギーを獲得して ATP は生成される。ここで呼吸とは有機化合物を細胞外の電子受容体 (例えば酸素) を介して酸化してエネルギーを獲得する一連の反応系 (代謝) のことであり、酸素の要求性の違いにより、好氣的呼吸と嫌氣的呼吸に分かれる。通性嫌気性菌である枯草菌は、酸素存在下では好氣的呼吸を、酸素非存在下では嫌氣的呼吸を行う。

ATP が生成される主な経路には、酸素を必要としない解糖系・エムデンマイヤーホフ経路と酸素を必要とするクエン酸回路 (TCA 回路) の二つが挙げられる。前者は好気性菌・嫌気性菌のいずれでも起こるが、後者は嫌気性菌では起こらない。枯草菌を含む通性嫌気細菌はどちらの条件下でも生育できるが、酸素の無い条件下ではクエン酸回路は機能しない。この場合クエン酸回路を構成する酵素は不必要になるため、酵素の生成は抑制される方がエネルギーを効率的に利用できると思われる。

#### 5.2 システムとその評価

枯草菌を嫌気状態においた場合の遺伝子発現のデータを本システムの実装に用いた。枯草菌は通性嫌気性菌であるため、

嫌気条件下においた場合にはそれまで活性化していた解糖系、クエン酸回路のうち、クエン酸回路が機能しなくなる。実装には嫌気条件下においた細胞について一定時間ごとにサンプルを取った 6 点 (0, 15, 30, 60, 90, 120 min) の時系列データを用いた。

各サンプルにおいてクエン酸回路に分類されているノードの色の変化を図 6 に示す。実験結果を選択し、Carbohydrate Metabolism に分類されている Citrate cycle (TCA cycle) にチェックすると、クエン酸回路が強調され、関係する遺伝子発現の変化が把握しやすくなる。正常細胞の mRNA は緑で、嫌気条件下におかれた細胞の mRNA は赤で標識され、二つの細胞の遺伝子発現量の比がノードの色としてマッピングされる。このようにしてクエン酸回路を観察すると、すべての実験においてほとんどの遺伝子産物においてノードの色が緑色または薄緑を示しているため、その発現が抑制されていることが解る。このように、代謝経路の上にマッピングすることにより、ノードの色の変化で遺伝子発現変化に関する情報を得ることができると同時に、遺伝子とパスウェイの関係についての情報も視覚的に得ることができる。

### 6. まとめ

本論文では、XML で記述されたデータを元に代謝経路を自動描画し、DNA マイクロアレイ実験から得られた遺伝子発現パターンを代謝経路にマッピングするシステムを提案した。本研究では、電子化された代謝データベースの内容をもとに、分割して提供されていた代謝経路を一枚のパスウェイマップに自動描画することができた。一枚のマップですべての既知の代謝を網羅しているため、各代謝ごとにパスウェイマップを見る必要がなくなる。パスウェイマップの画像形式に SVG を採用することにより、拡大・縮小のための画像を別に用意する必要がない、ラスタ画像に比べて画像サイズが小さい、スクリプトを組み込んで動的な表示が可能、などの利点が得られた。実際に同じ代謝マップを JPEG 形式で出力した結果、約 1 KB のサイズであった SVG に対して約四倍の大きさになった。JPEG 形式の場合は、更に拡大した場合の図を用意する必要があるが、SVG の場合はこのような作業は必要としない。

DNA マイクロアレイは、大量の遺伝子発現を同時に測定できることから、ゲノムワイドな機能解析への新しい技術として注目されている。ヒトは五感の中で視覚から得られる情報量は全体の約八割を占めるとも言われている。扱う遺伝子数も、個々の遺伝子に関する情報量も膨大な DNA マイクロアレイの実験結果は適切な視覚化をすることが非常に重要である。我々の構築したシステムでは、代謝経路にマッピングすることで個々の遺伝子の情報をたどりながら属するパスウェイを調べるのではなく、視覚的にパスウェイを判断でき、同じパスウェイの中の遺伝子発現パターンを観察することができる。

#### 6.1 今後の課題

本システムの開発にあたり、明らかになった問題点、およびシステムにおける改良の余地がある機能などについて今後の展望を以下に述べる。

- 代謝パスウェイのレイアウト：本システムではパスウェイ

(注1)：酸素の有無にかかわらず生育が可能

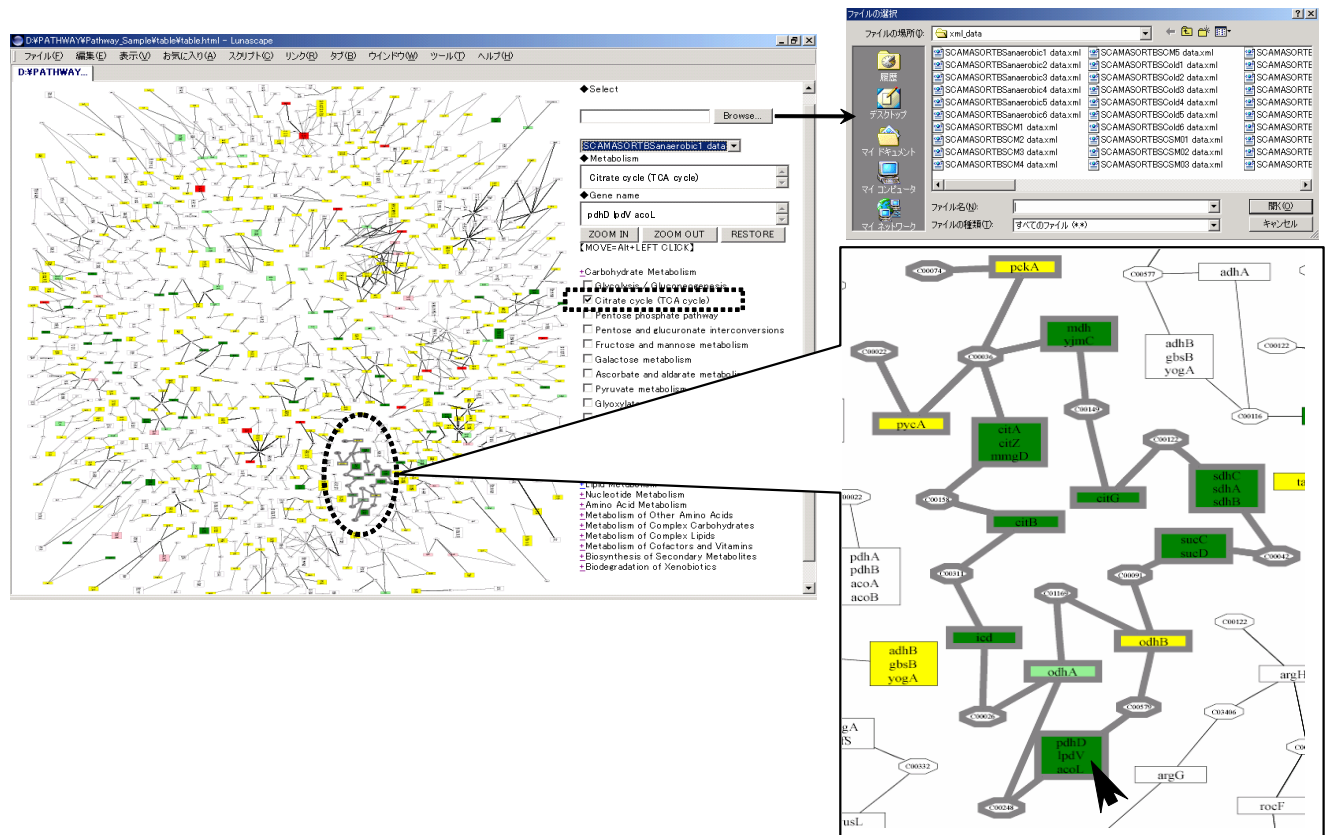


図 6 遺伝子発現閲覧システムの画面レイアウト

Fig. 6 An example of Gene Expression Browsing System

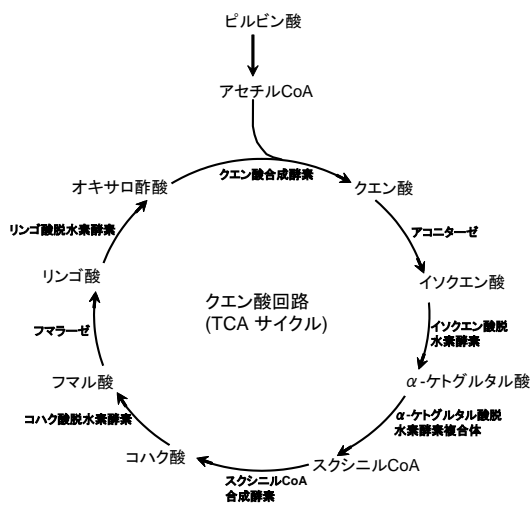


図 7 クエン酸回路 (TCA サイクル)

Fig. 7 Citrate cycle (TCA cycle)

イの自動描画に GraphViz の neato パッケージを用いた。生化学の教科書などで用いられている生物学者になじみの深い代謝パスウェイのレイアウトは独特で、neato パッケージで使用されているレイアウトアルゴリズムではその独特な代謝経路の形を表現しきれない。独特なレイアウトとは、例えば、図 6 で強調されている クエン酸回路 (TCA 回路) は、図 7 に示すように、全体的に丸く、時計回りのサークルで描かれることが殆どで、この回路の出発物質であるピルビン酸は円の上部から回路

に合流する。このように代謝経路は円アルゴリズムや階層アルゴリズムなど複数のアルゴリズムが組み合わせて使われている様に見えるため、現在の自動描画の手法では対応できていない。現在グラフ描画のための Java ライブラリーである yFiles [17] を用いた描画方法を検討中である。

- ファイルのアップロード：現在一枚の実験ファイルをアップロードし、スクリプトで読み込むなどの作業には数秒で処理できるが、ファイルは一枚ずつしかアップロードすることができない。ある状態の細胞に対して一定時間ごとにサンプルを取り遺伝子発現のパターンを調べるような場合には、それらをアニメーションで表示し、時間を追って変化の様子を見たいという要求があるのでこのような表示方法にも対応していきたい
- 発現の有意差：二蛍光標識法を用いたマイクロアレイでは、二つのサンプルの発現比をもとめ、発現レベルが有意に異なる遺伝子を見つけてくる方法である。しかし、どの発現レベルが「有意」に異なるかの判断基準はまだ確立されていない。本システムでは、実験データを XML に変換する際に発現レベルの差が五倍以上のものを「有意」な値として XML データ中にカラー属性を追加しているが、この判断が正しいという保証はない。「有意」とするレベルは各研究者が独自に判断しているのが現状であり、本システムでも、ユーザーが有意と思われるレベルを選択できるようにするか、色の変化を発現レベルごとにグラデーションで表すなどする必要がある。
- 複数の遺伝子が関与する酵素の発現レベルの取り扱い：

遺伝子産物を表すノードには複数の遺伝子が関与している場合が少なくない。これらの遺伝子は必ずしも同じ遺伝子発現パターンを示すとは限らない。しかし本システムではノードの色は一色でしか表現できないのが現状である。このように複数の遺伝子の関与したノードの色をどのように表現するかが今後の大きな課題の一つである。枯草菌はオペロン<sup>(注2)</sup>をもっているため、それを一つの単位としてまとめるという方法も考えられる。また、SVG にはグラデーション機能を持つため、この機能をうまく取り入れることで複数の遺伝子の発現を一つのノードで観察できるかどうか検証する必要がある。

#### 謝 辞

この研究の一部は、21 世紀 COE プログラム「フロンティアバイオサイエンスへの展開 Exploiting New Frontiers in Bioscience」によるものである。ここに記して謝意を表す。

#### 文 献

- [1] Watson, J.D. and Crick, F.H.C., "A Structure of Deoxyribose Nucleic Acid" *Nature* 171, 1953.
- [2] Steen Knudsen 著, 塩島 聡, 松本 治, 辻本 豪三 監訳, "わかる! 使える! DNA マイクロアレイ解析入門", 羊土社, 2002.
- [3] Tilstone, C., "DNA microarrays: Vital statistics" *Nature*, 424, 610-612, 2003.
- [4] Brown, P.O. and Shalon, T.D., "Methods for fabricating microarrays of biological samples" *USP* 5, 807, 522.
- [5] 松村 正明, 那波 宏之, "DNA マイクロアレイと最新 PCR 法", 秀潤社, 2000.
- [6] World Wide Web Consortium. Extensible Markup Language (XML) 1.0 (Second Edition). <http://www.w3.org/TR/REC-xml>, October 2000. W3C Recommendation 6 October 2000.
- [7] DDBJ : <http://www.ddbj.nig.ac.jp>, (2004 年 2 月 URL 確認).
- [8] PDBML : <http://www.pdbj.org/>, (2004 年 2 月 URL 確認).
- [9] Kanehisa, M, Goto, S, "KEGG: Kyoto Encyclopedia of Genes and Genomes" *Nucleic Acids Res.*, 28, 27-30, 2000.
- [10] World Wide Web Consortium. Scalable Vector Graphics (SVG) 1.1 Specification. <http://www.w3.org/TR/SVG11/>, January 2003. W3C Recommendation 14 January 2003.
- [11] KEGG : <http://www.genome.ad.jp/kegg/>, (2004 年 2 月 URL 確認).
- [12] Karp, P., Riley, M., Paley, S. and Pellegrini-Toole, A., "EcoCyc: Electronic Encyclopedia of E. coli Genes and Metabolism" *Nucleic Acids Research*, 24, 32-40, 1996.
- [13] SAX : <http://www.saxproject.org/>, (2004 年 2 月 URL 確認).
- [14] JAXP : <http://java.sun.com/xml/jaxp/>, (2004 年 2 月 URL 確認).
- [15] GraphViz : <http://www.graphviz.org/>, (2004 年 2 月 URL 確認).
- [16] MAGE-ML : <http://www.geml.org/omg.htm>, (2004 年 2 月 URL 確認).
- [17] Wiese, R. et al, "yFiles: visualization and automatic layout of graphs" *In Proceedings of the 9th International Symposium on Graph Drawing*, LNCS, Springer, 2001, <http://www.yworks.de/>.

---

(注2): 同時に転写調節され、ひとつながりの mRNA として転写される遺伝子のまとめ