

## 反応構造プロファイルによる異種生物種間の 代謝パスウェイ比較解析

芦田 優太<sup>†</sup> 尾崎 知伸<sup>††</sup> 大川 剛直<sup>†</sup>

<sup>†</sup> 神戸大学大学院工学研究科 <sup>††</sup> 神戸大学自然科学系先端融合研究環

本研究では代謝パスウェイ上の重要な酵素を含む部分反応構造に着目した代謝パスウェイの比較解析手法を提案する。生物種比較の際、生物にとって重要な酵素を含む反応構造が生物種間でどれくらい重複して含まれるかを調べ、反応構造プロファイルとし、生物種間の距離を算出する。古細菌、真正細菌、真核生物の3つのドメインに含まれる64種の生物に対して提案手法を適用し、系統樹を構築することで本手法の有効性を確認した。

## A Comparative Analysis of Metabolic Pathways Based on Reaction Structure Profile

Yuta Ashida<sup>†</sup> Tomonobu Ozaki<sup>††</sup> Takenao Ohkawa<sup>†</sup>

<sup>†</sup> Graduate School of Engineering, Kobe University <sup>††</sup> Organization of Advanced Science and Technology, Kobe University

We propose a new method for comparing the metabolic pathways with reaction structures that include important enzymes. In this method, important “reaction structures” are extracted and a “reaction structure profile,” which represents whether extracted reaction structures are observed in the metabolic pathway of other organisms, is created. Distance regarding function within organisms between species is defined using the “reaction structure profile.” By applying the proposed method to the metabolic networks of 64 representative organisms selected from archaea, eubacteria and eukaryote, we succeed in reconstructing a phylogenetic tree, and confirm the effectiveness of the method.

### 1 はじめに

生物は物質を食物等の形で体内に取り入れる。これらの物質は体内の様々な化学変化を通してエネルギーや生命維持に必要な化合物へと変換される。このような化学変化の総体が代謝と呼ばれる。

この化学反応はある化合物を他の化合物へと変換する酵素反応であり、変換前の化合物を基質、変換後の化合物を生成物と呼ぶ。ある酵素反応の生成物が他の酵素反応の基質となる連鎖反応が生体内で生じている。この連鎖反応で構成される大規模なネットワークが代謝パスウェイである。この代謝パスウェイに関する情報はKEGG (Kyoto Encyclopedia of Genes and Genomes)<sup>8)</sup>等のデータベースで収集、公開されている。

代謝パスウェイは生物の機能に関する重要な情報を含んでいる。さらに、異種生物種間での代謝パスウェイ比較解析は生物機能における類縁関係に

ついで重要な情報を得る有効な手段となる。それゆえ、近年様々な代謝パスウェイ比較手法が研究されている。代謝パスウェイのネットワークポロジを用いた手法<sup>15, 9, 6)</sup>や代謝パスウェイに含まれる各酵素反応の有無に基づく比較手法<sup>7, 13)</sup>等がその一例である。

一方、系統分類学ではrRNAなどを用いたゲノム配列による生物種の比較手法<sup>4)</sup>が用いられるが、進化の過程で起こる遺伝子の水平伝播等の現象が反映されないという欠点があり、生物の進化類縁関係を正確に表すことはできない<sup>3)</sup>。それゆえ、従来の系統分類学を補完するようなゲノム配列とは異なった視点からの手法が必要とされている。

そこで、本研究では反応構造プロファイルを用いた代謝パスウェイ比較手法を提案する。提案手法は代謝パスウェイ上の重要な酵素を含む部分的な反応構造に着目することで、系統分類における

新しい知見を得ることを目的とする。

代謝中で頻繁に用いられる酵素や生物にとって不可欠な酵素は重要な酵素と見做すことができる。この重要な酵素を含む反応構造を抽出することで重要な反応構造に着目することができる。ここで、反応構造とは代謝パスウェイ上の部分グラフを意味する。反応構造プロファイルは異種生物種間で重要な反応構造がどれだけ重複して存在しているかを表す。反応構造プロファイルを用いることで重要な反応構造を特徴として扱うことができ、生物の機能に関する類縁関係を明らかにすることができる。

本論文は以下の構成となっている。2章で関連研究について、3章で提案手法について述べる。そして4章で提案手法による代謝パスウェイ比較解析結果を示し、5章で本研究の結論と今後の課題について述べる。

## 2 関連研究

パスウェイ比較手法に関する既存研究に June らによる研究がある<sup>9)</sup>。June らの手法では代謝パスウェイは酵素をノード、化合物をエッジとした有向グラフで表わされ、指数グラフカーネルによって代謝パスウェイのネットワーク構造上の類似性が求められる。しかし、この手法は代謝パスウェイの全体的なネットワーク構造にのみ着目しており、代謝パスウェイの部分的な構造上の特徴が考慮されない。

既存手法としては他に Tohsato による手法<sup>13)</sup>がある。Tohsato による手法では代謝パスウェイを酵素反応の集合としてとらえている。代謝パスウェイを各代謝反応の有無によって“1”と“0”の2値ビット列で表現し、反応プロファイルとする。生物種間の類似度はこの反応プロファイルを比較することで算出される。しかし、この手法では代謝パスウェイの構造的特徴が考慮されない。

## 3 反応構造プロファイルを用いた比較解析

代謝は生物の機能を担っており、生物の機能は食餌や生存環境などの特徴によって異なる。ネットワークによって代謝プロセスを表現している代謝パスウェイは、生物種固有の構造を持ち、生物の特徴を表していると考えられる。そのため、代謝パスウェイを用いることで異種生物種

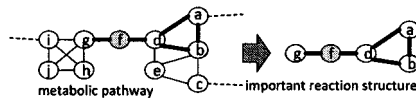


Fig. 1 An example of reaction structure extraction.

間における機能上の比較が可能となる。生物種間の距離は代謝パスウェイから抽出した特徴によって求められ、階層的クラスタリングを用いて生物の機能における類縁関係が導かれる<sup>5)</sup>。

代謝パスウェイ中には生物にとって不可欠な酵素や代謝中で頻繁に用いられる酵素が存在する。例えば、adenine phosphoribosyltransferase (EC:2.4.2.7) はヒトのプリン代謝で用いられる酵素であるが、この酵素が失われれば adenine phosphoribosyltransferase 欠乏症が引き起こされる。よって EC:2.4.2.7 はヒトにとって不可欠であり、重要な酵素であると言える。このような重要な酵素に関わる代謝部分はその酵素を含む部分グラフとして表わされる。

本論文では代謝パスウェイ上の部分グラフを反応構造 (reaction structure) と呼ぶ。さらに、重要な酵素を含む反応構造を重要な反応構造として扱う。反応構造抽出の例を Fig. 1 に示す。本論文では代謝パスウェイを酵素をノードとする無向グラフとして扱う。Fig. 1 中の代謝パスウェイは基準点酵素 (ノード f) の周辺を表しており、破線は図中になくノードに接続しているエッジである。太線で接続されている反応構造が抽出される重要な反応構造の一例である。実際には一つの基準点酵素から複数の反応構造が抽出される。

生物種間の距離を求めるために重要な反応構造を抽出し、反応構造プロファイルを作成する。反応構造プロファイルは異なる生物の代謝パスウェイ間でどれだけ重要な反応構造が重複して存在するかを表し、生物種間の距離は反応構造プロファイルを元に算出される。この操作を生物の全組み合わせについて行い、階層的クラスタリングに用いる距離行列を作成する。提案手法の全体図を Fig. 2 に示す。

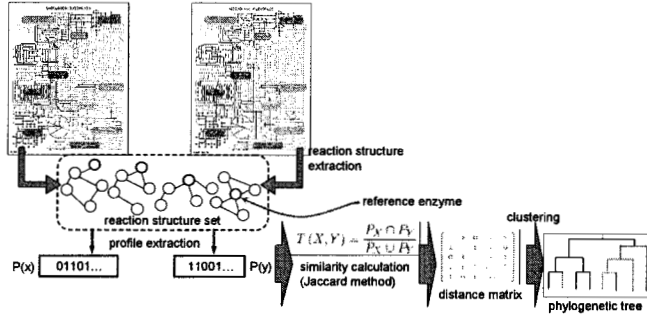


Fig. 2 The procedure for comparative analysis with reaction structure profile.

### 3.1 反応構造プロフィール

重要な酵素の周辺部は重要な酵素が関わる代謝の流れを表している。代謝パスウェイ上のそのような部分は進化の過程で保存され、異種生物間で大きな変化なく存在している。そのため、それらは異種生物種間比較の際により指標となる。系統分類上で近隣な生物ほど代謝パスウェイ上に重要な部分を共通して持っており、遠い生物ほど少ないと考えることができる。そこで、反応構造プロフィールは代謝パスウェイ上の重要な酵素に着目する。

重要な酵素を含む反応構造は生物にとって重要な部分であり、生物の進化に大きくかかわっている。この重要な反応構造を機能に関する特徴として扱うため、反応構造プロフィールを作成する。

なお、反応構造プロフィールの作成に用いられるのは重要な酵素を含む反応構造のみである。そのため、重要な反応ではないが、生物固有であるような反応は無視される恐れがある。しかし、すでに述べたように、重要な反応構造は進化間で保存される傾向があるため本手法は有効であると言える。

### 3.2 基準点酵素探索

反応構造プロフィール作成のため、まず反応構造抽出の起点となる酵素、すなわち基準点酵素の決定が必要である。基準点酵素決定にはSyedらによる‘load point’、‘choke point’探索手法<sup>11)</sup>を用いた。

‘load point’は代謝パスウェイ上の全酵素の組み合わせから得られた最短経路が集中している酵

素であり、代謝中で頻繁に用いられる酵素を表す。‘load point’を決定する‘load value’は隣接している酵素数と自身を通る最短経路の数から算出され、高い‘load value’を持つ酵素が‘load point’となる。‘load value’は以下の式で求められる。

$$L_m = \ln \left[ \frac{p_m/k_m}{\sum_{i=1}^M p_i / \sum_{i=1}^M k_i} \right], \quad (-\infty < L_m < \infty) \quad (1)$$

ここで  $M$  は代謝パスウェイ上の酵素数を表し、 $p_m$  は酵素  $m$  を通る最短経路の数を、 $k_m$  は  $m$  と隣接している酵素数を意味する。

一方、‘choke point’とはその酵素の欠損が生命の維持に大きな影響を与えるような酵素のことである。‘choke point’を多く含む最短経路は代謝中で大きな役割を担っており、そのような最短経路が多く通る酵素もまた大きな役割を担っていると考えることができる。よって、ある酵素を多くの‘load point’を含む最短経路が通っている場合、その酵素を‘choke point’とする。ある酵素を通る最短経路上の‘load point’の数によってその酵素の‘choke value’が求められ、高い‘choke value’を持つ酵素が‘choke point’となる。本論文では高い‘choke value’を持つ酵素上位5個を‘choke point’とした。

例えば Fig. 3 では、灰色のノード‘f’が‘choke point’となる。Bの部分に含まれるノードとAの部分に含まれるノードの最短経路は‘f’を必ず通るので、‘f’を通る最短経路上の‘load point’の数は多くなる。酵素‘f’が失われればAの部分とBの部分繋ぐ代謝の流れは遮断されることとなる。

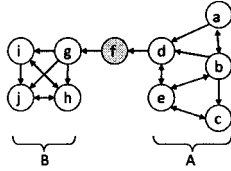


Fig. 3 Example of 'choke point' and 'load point'.

この 'load point' と 'choke point' の探索は Pathway Hunter Tool<sup>10)</sup> の機能の一部として実装され、Web 上に公開されている<sup>1)</sup>。本論文では Pathway Hunter Tool を用いて 'choke point' を探索し、基準点酵素として用いている。

### 3.3 反応構造プロファイルの作成

まず基準点酵素となる重要な酵素を探索し、その後基準点酵素から特定範囲内で、ある一定のノード数の反応構造を抽出する。これはノード数が小さすぎれば抽出される反応構造が多くなりすぎ、大きすぎれば反応構造の重要性が薄まるからである。また、基準点酵素からの範囲が広すぎれば基準点酵素から離れすぎてしまい、重要性が薄まってしまう。しかし、事前にこれらのパラメータを適切に決定することは困難であるため、予備実験を行い抽出する反応構造のノード数と基準点酵素からの探索範囲を決定した。

生物種  $O$  と  $O'$  から抽出した重要な反応構造の集合をそれぞれ  $G_O$  と  $G_{O'}$  とする。そして、 $G_O$  と  $G_{O'}$  の和集合を  $R\{O, O'\} = G_O \cup G_{O'} = \{r_1, r_2, \dots, r_n\}$  とする。

生物種  $O$  の  $O'$  に対する反応構造プロファイルは  $R\{O, O'\}$  に含まれる反応構造が  $O$  の代謝パスウェイ上にいくつ存在するかを表し、以下の式で求められる。

$$P_o\{O'\} = [p_{o1} \ p_{o2} \ \dots \ p_{on}]$$

$$\begin{cases} p_{oi} = 1 & (r_i \leq N_O) \\ p_{oi} = 0 & \text{otherwise} \end{cases} \quad (1 \leq i \leq n) \quad (2)$$

$P_o\{O'\}$  は  $O$  の  $O'$  に対する反応構造プロファイルを表し、 $r_i \leq N_O$  は反応構造  $r_i$  ( $1 \leq i \leq n$ ) が  $O$  の代謝パスウェイ  $N_O$  上に含まれることを意味する。

<sup>1</sup> <http://pht.tu-bs.de>

反応構造の例を Fig. 4 に示す。それぞれのノードが酵素を示し、灰色のノードが基準点酵素を表す。基準点酵素を含まない反応構造は重要な反応構造ではないことを意味する。例えば、生物種  $O$  の代謝パスウェイ上に Fig. 4 中の A, B, C, D の 4 つの反応構造を持っており、 $O'$  が B, C, D, E の反応構造を持っていたとする。この場合 A と B が  $O$  の重要な反応構造として抽出され、B, C, E が  $O'$  の重要な反応構造として抽出される。反応構造 D は  $O$  と  $O'$  両方の代謝パスウェイ上に存在するが、 $R\{O, O'\}$  には含まれないので D は  $O$  と  $O'$  の重要な反応構造ではない。よって  $R\{O, O'\}$  は  $\{A, B, C, E\}$  となり、これら 2 生物種の反応構造プロファイルは以下になる。 $P_o\{O'\} = \{1 \ 1 \ 1 \ 0\}$ ,  $P_{o'}\{O\} = \{0 \ 1 \ 1 \ 1\}$

### 3.4 距離算出

生物種  $X, Y$  間の類似度  $T(X, Y)$  は Jaccard 係数法を用いて以下のように決定する。

$$T(X, Y) = \frac{P_X\{Y\} \otimes P_Y\{X\}}{P_X\{Y\} \oplus P_Y\{X\}}$$

$$= \frac{N_z}{N_x + N_y - N_z} \quad (3)$$

ここで  $N_x$  と  $N_y$  はそれぞれの反応構造プロファイル  $P_X\{Y\}$  と  $P_Y\{X\}$  中の '1' の出現数であり、 $N_z$  は  $P_X\{Y\}$  と  $P_Y\{X\}$  双方で '1' となるビットの数である。 $T(X, Y)$  は常に 0 と 1 の間の数値となる。1 に近いほど二つの反応構造プロファイル間の類似度が高く、0 に近いほど類似度が低いことを表す。

前章での例では 2 生物種間の反応構造プロファイルは  $P_o\{O'\} = \{1 \ 1 \ 1 \ 0\}$ ,  $P_{o'}\{O\} = \{0 \ 1 \ 1 \ 1\}$  であり、よって類似度は  $T(O, O') = \frac{2}{4} = 0.5$  となる。

$X$  間の  $Y$  距離  $D(X, Y)$  を以下のように定義する。

$$D(X, Y) = 1 - T(X, Y) \quad (4)$$

## 4 実験

提案手法の有効性を確認するため、異種生物種間での代謝パスウェイ比較解析実験を行った。

rRNA 配列解析による従来の系統分類学は誤りを含む可能性があるが、大部分は確立されている。そのため、本論文では NCBI taxonomy による既

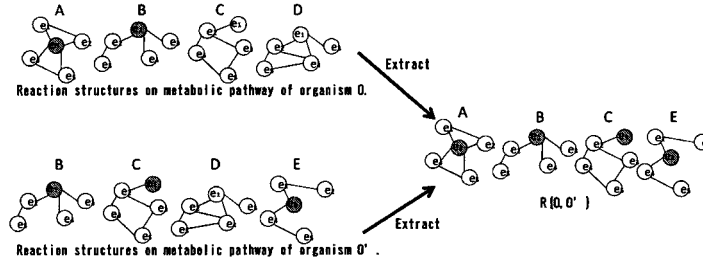


Fig. 4 Example of reaction structure.

存の系統樹を評価に用い、それに近い系統樹の構築に価値があるとした。

Table 1 に実験に用いた 64 種の生物種名を示す。比較に用いる生物種の全組み合わせにおいて反応構造プロファイルを用いて距離を算出し、階層的クラスタリングによって系統樹を構築した。基準点酵素には 'choke point' を用いている。クラスタリングと系統樹の構築は統計処理ソフトウェア R ver.2.4.1. を用い、クラスタリング手法は予備実験とクラスタリング手法の特徴から最長距離法を用いた。

まず、反応構造抽出の際に必要な反応構造のノード数と基準点酵素からのステップ数の数値設定のための予備実験を行った。古細菌に属する生物種のみを解析に用い、各変数値を変更した実験結果を Fig. 5 に示す。(a) と (b) はともにノード数を 5 とし、(a) はステップ数を 5、(b) はステップ数を 2 と設定した結果である。(c) と (d) はともにステップ数を 3 とし、(c) はノード数を 3、(d) はノード数を 7 と設定した。そして、(e) はステップ数を 3、ノード数を 5 と設定した結果である。

結果として (e) の設定でクレン古細菌とユーリ古細菌が NCBI taxonomy に最も類似した分類がなされた。従来の系統樹を補完する系統樹構築のため、以降の実験では抽出する反応構造のノード数を 5、基準点酵素からのステップ数を 3 と設定した。

次に、3 種のドメインに属する生物種を用いて系統樹を構築した (Fig. 6)。Fig. 6 では大部分の生物種が古細菌、真正細菌、真核生物のドメインに分かれている。この結果は C. R. Woese による 3 ドメイン説<sup>14)</sup>を代謝機能の観点から再現してい

る。さらに、真正細菌のクラスタがまず分岐し、その後、古細菌と真核生物が分かれている。これは古細菌と真正細菌が最初に分岐して進化し、古細菌から真核生物が進化したという生物学の仮説<sup>3)</sup>を支持する結果である。

#### 4.1 考察

3 つのドメインによる実験から提案手法を用いて従来の系統樹<sup>2)</sup>に類似した系統樹が構築された。従来の系統樹は表現型による類似性のような生物学的前提に基づいている。そのため、提案手法による系統樹は生物学的前提から大きく外れることなく、生物の機能における系統関係を示しているということができる。

しかし、部分的には生物学的前提から外れている部分が存在する。Fig. 6 では真正細菌の *Aquifex aerolicus*(*aae*) と真核生物の *Arabidopsis thaliana*(*ath*) が古細菌のクラスタに含まれており、そのような部分は生物学的前提と矛盾している。これはパスウェイデータベースの情報の不足によって起こる可能性がある。代謝パスウェイだけでなく、生物に関する情報は完全ではなく、多くの情報が不十分である<sup>2, 1)</sup>。基準点酵素として選んだ酵素に関するデータが完全でなければ、その酵素から抽出されるべき反応構造が得られない恐れがある。実際の実験においてもそのような事態は生じた。その結果、反応構造プロファイルが不完全になった可能性も考えられる。

#### 4.2 関連研究との比較

関連研究との比較において NCBI taxonomy による系統分類を指標として用い、各手法による系統樹

<sup>2</sup> NCBI Taxonomy [http://www.ncbi.nlm.nih.gov/Taxonomy/]

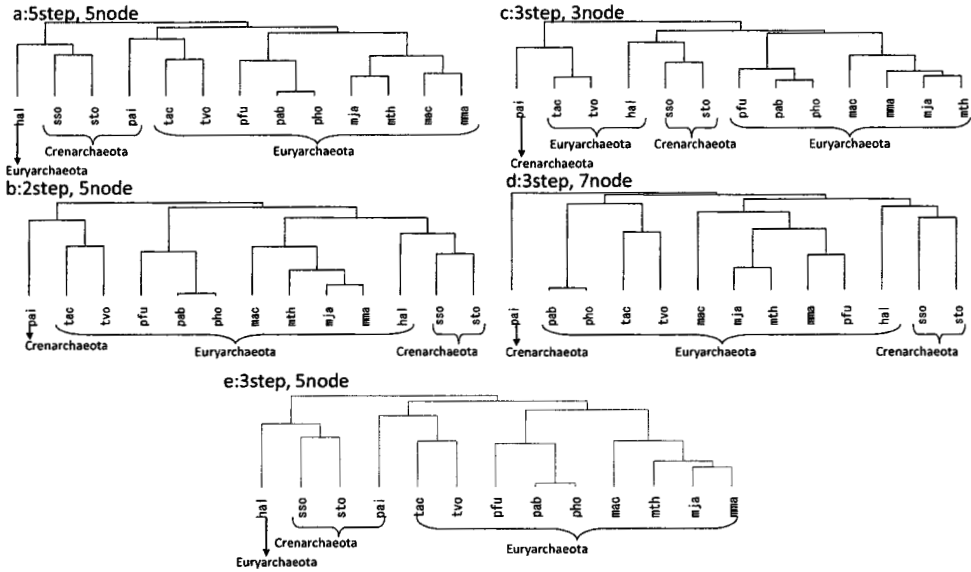


Fig. 5 Result of preliminary experiments.

との second cousins similarity を算出した。second cousins similarity<sup>12)</sup> は second cousin ペアに基づいた木構造データ間の類似度である。なお、それぞれの手法で用いている生物種のデータベースは同一でない。

提案手法による系統樹 (Fig. 6) の NCBI taxonomy との second cousins similarity は約 0.210 である。提案手法はドメインの分類において June らの手法<sup>9)</sup> による系統樹にやや劣っている。しかし、June らの手法による系統樹の NCBI taxonomy との second cousins similarity は約 0.179 であり、提案手法よりもやや低い数値となっている。よって提案手法は各ドメイン内における分類は従来の系統分類に近く、既存の系統分類を補完する手法として優れているということが出来る。また、提案手法のみ古細菌から真核生物が進化したという仮説<sup>3)</sup> を支持しているという点が大きく異なっている。

提案手法による系統樹はより多くの生物種で 3 つのドメイン分類が行えているという点で Tohsato による系統樹<sup>13)</sup> よりも優れている。また Tohsato による系統樹も古細菌から真核生物が進化したという仮説<sup>3)</sup> を再現していない。しかし、NCBI taxonomy に対する second cousins similarity は約 0.246

であり、提案手法よりも高い数値となっている。

これら 3 種の手法に共通して、真正細菌の分類に誤りが多いことが挙げられる。これは真正細菌に含まれる生物種は多数存在し、多様な代謝システムを持っていることが原因であると考えられる。

さらに、ユリー古細菌の 'hal' やクレン古細菌である 'sso' と 'sto' が関連研究を含め数多くの実験で近隣に分類されていることから、これらは代謝システム、特に重要な代謝部分において非常に類似した構造を持っているのではないかと考えられる。

## 5 結論

本論文では代謝パスウェイ上の重要な反応構造に着目した反応構造プロファイルを作成することで、効果的に生物の機能的特徴を抽出する手法を提案した。反応構造プロファイルを用いることで生物の機能的類縁関係を生物学的前提と矛盾することなく導くことが出来た。更に、生物は古細菌、真正細菌、真核生物のドメインに分類されるという 3 ドメイン説と、生物はまず古細菌と真正細菌に分かれて進化し、古細菌から真核生物が進化したという仮説を生物の機能上の観点から再現することが出来た。本論文では NCBI taxonomy を指

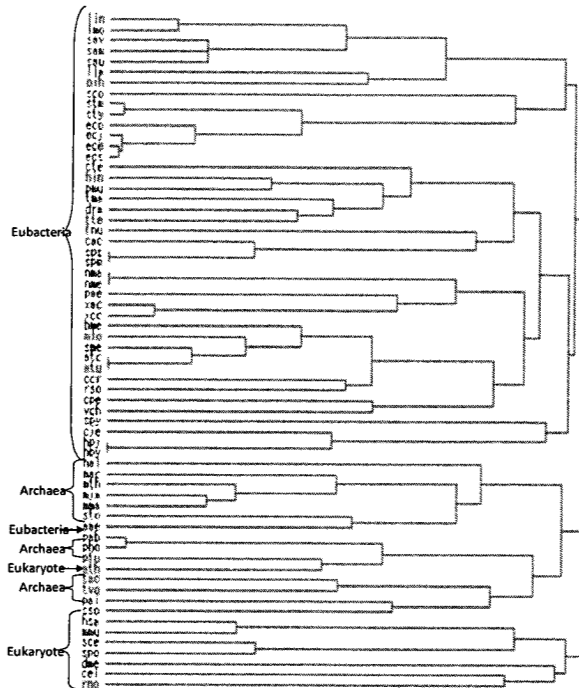


Fig. 6 Phylogenetic tree from 64 organisms.

標としたが、本手法の目的は従来の系統樹の再現ではなく、異種生物種間の重要な代謝部分構造による類似度を算出し、従来の系統分類を補完することである。提案手法により既存の系統樹と類似しているが同一ではない系統樹を構築したことで、提案手法は従来の系統樹を補完することが可能であると考えられる。

今後の課題として比較解析に用いる代謝パスウェイを特定の代謝系に限定し、使用する代謝系を段階的に変化させることが考えられる。本手法では比較解析を行う際、比較する生物が持つ全ての代謝パスウェイを用いている。しかし、代謝パスウェイ上には多くの生物において発達している代謝系(解糖系等)と、特定の生物の集団でのみ発達している代謝系(光合成に関わる代謝系等)が存在している。前者のような代謝系に限定して比較解析を行えばドメイン分類において良い結果を得ることができ、後者のような代謝系に限定して比較解析を行えばドメイン内での分類等において良い結果

を得ることができると考えられる。このように分類の目的毎に使用する代謝部分を変更することで生物の機能的特徴を段階的に捉え、より効果的に生物の機能的類縁関係を現すことを考えている。

#### 参考文献

- 1) Clemente, J. C., Satou, K. and Valiente, G.: Finding Conserved and Non-Conserved Reactions Using a Metabolic Pathway Alignment Algorithm, *Genome Informatics*, Vol. 17, No. 2, pp. 46-56 (2006).
- 2) Félix, L. and Valiente, G.: Validation of metabolic pathway databases based on chemical substructure search, *Biomolecular Engineering*, Vol. 24, pp. 327-335 (2007).
- 3) Feng, D. F., Cho, G. and Doolittle, R. F.: Determining divergence times with a protein clock:update and reevaluation, *Proceedings of the National Academy of Science of USA*, Vol. 94, pp. 13028-13033 (1997).
- 4) Fitch, W. M. and Margoliash, E.: Construction of phylogenetic trees, *Science*, Vol. 155, No. 3760, pp. 279-284 (1967).

Table 1 The 64 Organisms included in the phylogenetic analysis. Full scientific names were abbreviated into three character notation (Abbr. ) and their domain information<sup>14)</sup> in phylogeny were also represented. )

| Abbr.                    | Organism                             |
|--------------------------|--------------------------------------|
| <b>Domain:Archaea</b>    |                                      |
| hal                      | Halobacterium sp.                    |
| mac                      | Methanosarcina acetivorans           |
| mja                      | Methanococcus jannaschii             |
| mma                      | Methanosarcina mazei                 |
| nth                      | Methanobacterium thermoautotrophicum |
| pab                      | Pyrococcus abyssi                    |
| pai                      | Pyrobaculum aerophilum               |
| pfu                      | Pyrococcus furiosus                  |
| pho                      | Pyrococcus horikoshii                |
| sso                      | Sulfolobus solfataricus              |
| sto                      | Sulfolobus tokodaii                  |
| tac                      | Thermoplasma acidophilum             |
| tvo                      | Thermoplasma volcanium               |
| <b>Domain:Eubacteria</b> |                                      |
| aae                      | Aquifex aerolicus                    |
| atc                      | Agrobacterium tumefaciens C58 Cereon |
| atu                      | Agrobacterium tumefaciens C58 UWash  |
| bme                      | Brucella melitensis                  |
| cac                      | Clostridium acetobutylicum           |
| ccr                      | Caulobacter crescentus               |
| cje                      | Campylobacter jejuni                 |
| cpe                      | Clostridium perfringens              |
| cte                      | Chlorobium tepidum                   |
| dra                      | Deinococcus radiodurans              |
| ece                      | Escherichia coli O157 EDL933         |
| ecj                      | Escherichia coli K-12 W3110          |
| eco                      | Escherichia coli K-12 MG1655         |
| ecs                      | Escherichia coli O157 sakai          |
| fnu                      | Fusobacterium nucleatum              |
| hin                      | Haemophilus influenzae               |
| hpj                      | Helicobacter pylori J99              |
| hpy                      | Helicobacter pylori 26695            |
| lin                      | Listeria innocua                     |
| lla                      | Lactococcus lactis                   |
| lmo                      | Listeria monocytogenes               |
| mlo                      | Mesorhizobium loti                   |
| nma                      | Neisseria meningitidis serogroup A   |
| nme                      | Neisseria meningitidis serogroup B   |
| oih                      | Oceanobacillus iheyensis             |
| pae                      | Pseudomonas aeruginosa               |
| pinu                     | Pasteurella multocida                |
| rso                      | Ralstonia solanacearum               |
| sam                      | Staphylococcus aureus MW2            |
| sau                      | Staphylococcus aureus N315           |
| sav                      | Staphylococcus aureus Mu50           |
| sco                      | Streptomyces coelicolor              |
| sme                      | Sinorhizobium meliloti               |
| spg                      | Streptococcus pyogenes M3            |
| spm                      | Streptococcus pyogenes M18           |
| spy                      | Streptococcus pyogenes               |
| stm                      | Salmonella typhimurium               |
| sty                      | Salmonella typhi                     |
| tma                      | Thermotoga maritima                  |
| tte                      | Thermoanaerobacter tengcongensis     |
| vch                      | Vibrio cholerae                      |
| xac                      | Xanthomonas axonopodis               |
| xcc                      | Xanthomonas campestris               |
| <b>Domain:Eukaryote</b>  |                                      |
| ath                      | Arabidopsis thaliana                 |
| cel                      | Caenorhabditis elegans               |
| dme                      | Drosophila melanogaster              |
| hsa                      | Homo sapiens                         |
| mmu                      | Mus musculus                         |
| rno                      | Rattus norvegicus                    |
| sce                      | Saccharomyces cerevisiae             |
| spo                      | Schizosaccharomyces pombe            |

- 5) Forst, C. V. and Schulten, K.: Phylogenetic analysis of metabolic pathways, *Journal of Molecular evolution*, Vol. 52, No. 6, pp. 471-489 (2001).
- 6) Heymans, M. and Singh, A. K.: Deriving phylogenetic trees from the similarity analysis of metabolic pathways, *Binformatics*, Vol. 19, pp. i138-i146 (2003).
- 7) Hong, S. H., Kim, T. Y. and Lee, S. Y.: Optimal Auction Design, *Phylogenetic analysis based on genome-scale metabolic pathway reaction content*, Vol. 65, pp. 203-210 (2004).
- 8) Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. and Hattori, M.: The KEGG resource for deciphering the genome, *Nucleic Acids Reserch*, Vol. 32, pp. D277-D280 (2004).
- 9) Oh, S. J., Joung, J. G., Chang, J. H. and Zhang, B. T.: Construction of phylogenetic trees by kernel-based comparative analysis of metabolic networks, *BMC Bioinformatics*, Vol. 7, No. 284, pp. 1471-2105 (2006).
- 10) Rahman, S. A., Advani, P., Schunk, R., Schrader, R. and Schomburg, D.: Metabolic pathway analysis web service (Pathway Hunter Tool at CU-BIC), *Bioinformatics*, Vol. 21, No. 7, pp. 1189-1193 (2005).
- 11) Rahman, S. A. and Schomburg, D.: Observing local and global properties of metabolic pathways: 'load points' and 'choke points' in the metabolic networks, *Bioinformatics*, Vol. 22, No. 14, pp. 1767-1774 (2006).
- 12) Shasha, D., Wang, J. T. L. and Zhang, S.: Unordered tree mining with applications to phylogeny, *Proceedings of the 20th International Conference on Data Engineering*, pp. 708-719 (2004).
- 13) Tohsato, Y.: A method for species comparison of metabolic networks using reaction profile, *Information Processing Society of Japan Transactions on Bioinformatics*, Vol. 47, No. SIG 17(TBIO 1), pp. 42-47 (2006).
- 14) Woose, C. R., Kandler, O. and Wheelis, M. L.: Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya, *Proceedings of the National Academy of Sciences*, Vol. 87, pp. 4576-4579 (1990).
- 15) Zhu, D. and Qin, Z. S.: Structural comparison of metabolic networks in selected single cell organisms, *BMC Bioinformatics*, Vol. 6, No. 8, pp. 1471-2105 (2005).