Original Paper

Reaction Structure Profile: A Comparative Analysis of Metabolic Pathways Based on Important Substructures

Yuta Ashida,[†1] Tomonobu Ozaki[†2] and Takenao Ohkawa[†1]

Comparative analysis of organisms with metabolic pathways gives important information about functions within organisms. In this paper, we propose a new method for comparing the metabolic pathways with reaction structures that include important enzymes. In this method, subgraphs from pathways that include 'choke point' or 'load point' are extracted as important "reaction structures," 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 in the KEGG database, we succeed in reconstructing a phylogenetic tree, and confirm the effectiveness of the method.

1. Introduction

Organisms take material such as food into their bodies. These materials become compounds and energy necessary for sustaining the activity of the organisms by various chemical reactions. The whole of such chemical reactions is called metabolism. This chemical reaction is the reaction of enzymes that convert a certain compound into another compound. The compound before being converted by the enzyme reaction is called a substrate, and after conversion is called a product. A chain of reactions, in which the product of one enzyme reaction becomes the substrate of another enzyme reaction, happens within an organism. The large-scale network composed by the chain reaction is called a metabolic pathway. Information about metabolic pathways is stored on pathway databases such as KEGG (Kyoto Encyclopedia of Genes and Genomes) [10].

Metabolic pathways contain important information on the function of organisms. Analysis of metabolic pathways gives hints about the evolutionary process. Furthermore, comparative analysis of metabolic pathways among species is an effective means of obtaining information about the functional relation of organisms. Therefore, a variety of comparative analysis techniques for metabolic pathways has been explored in recent years. For instance, comparison techniques that use the network topology of the metabolic pathway [8),11),18)] and methods based on the enumeration of each enzyme reaction included in the metabolic pathway [9),16)] have been proposed. However, the former pays attention only to the network structure of the metabolic pathway and does not use biological assumption. Because the latter treats a metabolic pathway as a set of the enzyme reactions, it does not use constructional information of the metabolic pathway such as order relation of reactions.

On the other hand, a comparison technique of organisms based on single genes such as rRNA [6)] is used in phylogenetic systematics to aim at clarification of phylogenetic relation. However, the technique by genome array cannot present phylogenetic relation accurately because of the fault that phenomena such as horizontal diffusion of the gene are not reflected [5)]. Consequently, a technique from a viewpoint different from genome array is needed to complement traditional phylogenetic systematics.

In this paper, we propose an analysis method of comparing metabolic pathways that uses a "reaction structure profile." Our method is developed for obtaining new insight into phylogeny by focusing attention on a partial reaction structure including an important enzyme on the metabolic pathway. We call these partial structures local structure.

An enzyme used frequently in metabolism or an enzyme indispensable for organisms is regarded as an important enzyme. It is possible to focus attention on an important "reac-

†1 Graduate School of Engineering, Kobe University
†2 Organization of Advanced Science and Technology, Kobe University

tion structure" by extracting the reaction structure including important enzymes for organisms. Note that the reaction structure means a subgraph extracted from a metabolic pathway. The reaction structure profile represents how much important reaction structures overlap between organisms. Reaction structure allows treating important reaction structures as features of organisms, and phylogenetic relation of functions between species can be clarified.

This paper is organized as follows. Section 2 discusses related research. The proposed method is described in Section 3, and Section 4 presents results obtained using the proposed method in an actual comparison of metabolic pathways. Finally, Section 5 summarizes problems and future works.

2. Related Work

An early study of comparison of pathways has been done by June, et al. [11]. In their method, a metabolic pathway is expressed in a directed graph in which each node and edge represent enzyme and compound respectively. The similarity in the network structure of the metabolic pathway is calculated by using the exponential graph kernel. However, this method pays attention only to a global network structure of the metabolic pathway. Thus, local structural features of metabolic pathway are not considered.

One such method is that proposed by Tohsato [16]. Metabolic pathways are handled as sets of enzyme reactions. Based on the presence or absence of metabolic reactions, the metabolic pathway of an organism is represented by a bit string comprised of the digits "1" and "0", called the "reaction profile." The similarity between organisms is evaluated by comparing reaction profile strings. However, her method has the problem in the point that structural information of the metabolic pathway is not considered.

José also proposed a method for comparison of pathways [2]. In José's method, metabolic pathways are compared via pseudo-alignment. Pseudo-alignment is a mapping of each reaction in one pathway to the most similar reaction in another pathway. Similarity between reactions is computed with enzyme and compound similarity. Enzyme similarity is chosen from hierarchical, information content and gene ontology similarity, and compound similarity is denoted as '1' or '0' representing identical or distinct respectively. José's method focuses on each enzyme reaction, enzyme, and compounds related to that reaction. On this point, José's method differs from our method.

In addition, there are pathway alignment methods to find approximate pathways of the query pathway [12] and pathway analysis with three alternative comparative methods [3]. They relate to our method in that the former can find approximate substructures of metabolic pathways and the latter can compare metabolic pathways from various perspectives.

3. Comparative Analysis with Reaction Structure Profile

Metabolism bears a function within an organism. Functions within an organism are different depending on features of the organism such as diet and environment. Thus, metabolic pathways that express the process of metabolism with a network have a structure peculiar to each organism and represent a feature of the organism. Hence, metabolic pathways allow comparison of functions among species. The distance between organisms is calculated based on the feature extracted from the metabolic pathway. Then, phylogenic relation on the function of the organism can be clarified by hierarchical clustering [7].

Some enzymes are indispensable to certain organisms or are used frequently during metabolism in metabolic pathways. For instance, the enzyme adenine phosphoribosyltransferase (EC:2.4.2.7) works in the purine metabolism of Homo sapiens. If this enzyme is missing, adenine phosphoribosyltransferase deficiency is caused. Therefore, EC:2.4.2.7 is indispensable and an important enzyme for Homo sapiens. The structure of metabolism related to this important enzyme is represented as a subgraph including the enzyme.

We call the subgraph on metabolic pathways a "reaction structure." Furthermore, we regard reaction structures that include important enzymes as important reaction structures. An example of an extracted reaction structure is shown in **Fig. 1**. We assume a metabolic pathway to be an undirected graph whose node is the enzyme. The metabolic pathway in Fig. 1

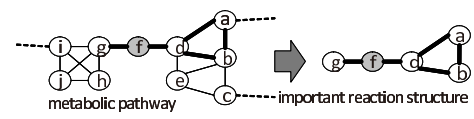

Fig. 1 An example of reaction structure extraction.

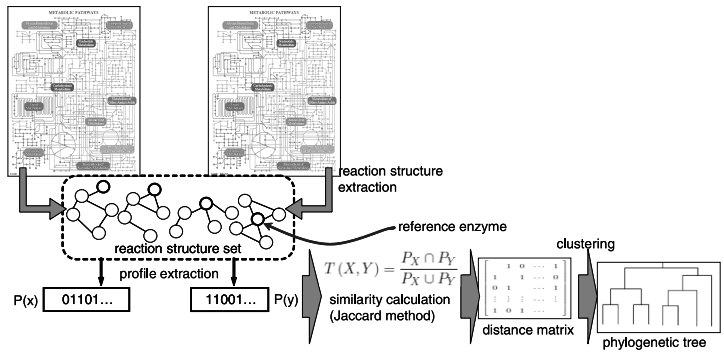

Fig. 2 The procedure for comparative analysis with reaction structure profile.

shows the surroundings of the "reference enzyme" (node f) of the metabolic pathway and the broken line is an edge connected with the node that isn't shown in this figure. The reaction structure connected in the heavy line is one example of important reaction structures being extracted. More than one reaction structure is actually extracted from one reference enzyme.

To calculate the distance between organisms, we extract important reaction structures, and calculate the "reaction structure profile," which represents how many important reaction structures overlap between metabolic pathways of different organisms. Distance between organisms is evaluated based on the reaction structure profiles. This procedure is applied with all combinations of organisms to get a distance matrix that is used for hierarchical clustering. The overall flow of the proposed method is shown in **Fig. 2**.

3.1 Reaction Structure Profile

The surrounding part that includes important enzymes shows the flow of metabolism related to the important enzyme. Such a part on the metabolic pathway is preserved during the process of evolution. Many of such parts exist between different species without substantial change. Therefore, they become good indices for comparisons between species. Organisms that are closely related phylogeneticly have important parts of their metabolic pathways in common. On the other hand, distant organisms have little. Reaction structure profiles focus on important substructures of metabolic pathways.

Then, to facilitate reaction structure extraction, we extract reaction structures including an important enzyme for the organism. A reaction structure including an important enzyme is an important part for an organism and is sure to influence evolution greatly. To treat this important reaction structure as a feature of a function, we construct a reaction structure profile.

Note that, only reaction structures that contain an important enzyme will be used. Therefore, the reaction that was not important but peculiar could be ignored. However, our method will be available because important reaction structures tend to be conserved during the process of evolution as I have already stated.

3.2 Reference Enzyme Search

It is necessary to decide a reference point where extraction of important reaction structure starts from. We call such a point the "reference enzyme." We use the search method of 'load point' and 'choke point' that Syed, et al.[14] proposed for determining the reference enzyme.

The 'load point' is an enzyme that concentrates the shortest path obtained from the combination of all enzymes on the metabolic pathway and is used a lot in metabolism. Then, the 'load value' of a certain enzyme is decided depending on the number of enzymes that are adjacent to that enzyme and on the shortest path that passes that enzyme. The enzyme that has high a 'load value' is defined as a 'load point.' The 'load value' is defined as follows[14].

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

where M means the number of enzymes on the metabolic pathway, p_m represents the number of shortest paths that contain enzyme m, k_m means the number of enzymes adjacent to enzyme m on the metabolic pathway.

On the other hand, the 'choke point' is an en-

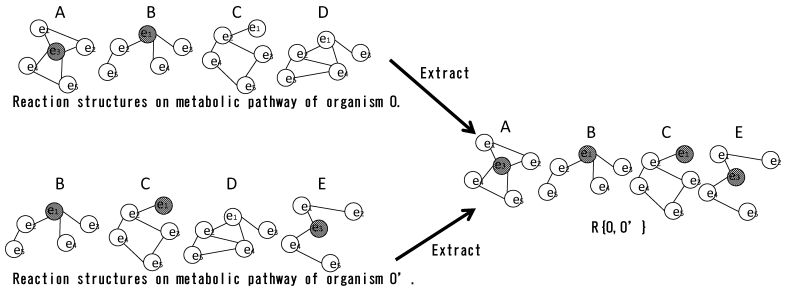

Fig. 4 Example of reaction structure.

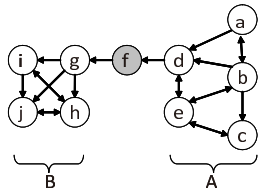

Fig. 3 Example of 'choke point' and 'load point'.

zyme whose loss affects the maintenance of life. The shortest path that has many 'load points' plays a big role in metabolism. The enzyme on which a lot of such shortest paths pass is especially important for the organism. Thus, if a lot of 'load points' exist on the shortest path that passes a certain enzyme, the enzyme is assumed to be a 'choke point.' The number of 'load points' existing on the shortest path that passes a certain enzyme is the 'choke value.' In this paper, we defined 'choke point' as the top five enzymes that have a high 'choke value.'

In **Fig. 3** for example, gray node 'f' becomes a 'choke point.' The shortest path that links nodes included in part B to the node included in part A must pass node 'f.' Therefore, the number of 'load points' on the shortest path that passes node 'f' is large. If enzyme 'f' is lost, the flow of metabolism that connects part A to part B is severed.

The tool for searching this 'load point' and 'choke point' is provided as a part of the function of the Pathway Hunter Tool [13], which is available to the public on the Web [⋆1]. In our method, we extract a 'choke point' with the Pathway Hunter Tool and use it as the reference enzyme.

3.3 Construction of Reaction Structure Profile

First of all, important enzymes, namely the reference enzymes, are retrieved on the metabolic pathway. Then, all reaction structures that have an exactly constant number of nodes are extracted within a range that is reachable in a specific number of steps from the reference point enzyme. If we employ a small number as the number of nodes in a reaction structure to be obtained, the number of extracted reaction structures is too great to specify a feature of the organism. The importance of a reaction structure lessens if the number of nodes is large. Moreover, when the number of steps from the reference enzyme is large, even an enzyme far from the reference enzyme is included in a reaction structure, and the importance of the reaction structure lessens. However, it is difficult to determine these parameters on ahead. Therefore, we decide the number of nodes and the number of steps from the reference enzyme by a preliminary experiment.

Let G_O and $G_{O'}$ be sets of important reaction structures extracted from organisms O and O'. Then, set union $R\{O, O'\}$ of the important reaction structure extracted from these two organisms is denoted by $R\{O, O'\} = G_O \cup G_{O'} = \{r_1, r_2, \cdots, r_n\}$.

The reaction structure profile to O' of O shows whether reaction structures included in $R\{O, O'\}$ exist on the metabolic pathways of O by a multi-dimensional vector, and is defined as follows.

$$P_o\{O'\} = [p_{o1}\ p_{o2}\ \cdots\ p_{on}]$$
$$\begin{cases} p_{oi} = 1 & (r_i \preceq N_O) \\ p_{oi} = 0 & \texttt{otherwise} \end{cases} \quad (1 \le i \le n) \quad (2)$$

$P_o\{O'\}$ denotes the reaction structure profile to O' of O because they contain reference enzymes. $r_i \preceq N_O$ means reaction structure $r_i\,(1 \le i \le n)$ is included on metabolic pathway N_O of organism O.

The examples of reaction structures are shown in **Fig. 4**. Each node is an enzyme, the gray one is a reference enzyme. A reaction

⋆1 http://pht.tu-bs.de

structure that has no reference enzyme is not an important reaction structure. For instance, it is assumed that organism O has four reaction structures A, B, C, and D in Fig. 4 on the metabolic pathway, and O' has B, C, D, and E on the metabolic pathway. Then, A and B are extracted as important reaction structures for O, and B, C, and E are extracted as important for O'. Although D appears in both O and O', D isn't included in $R\{O, O'\}$ because D is not important reaction structure for O and O'. Therefore, $R\{O, O'\}$ is $\{A, B, C, E\}$ and reaction structure profiles between these two organisms are specified as follows, $P_O\{O'\} = [1\ 1\ 1\ 0]$, $P_{O'}\{O\} = [0\ 1\ 1\ 1]$.

3.4 Distance between Organisms

The similarity between X and Y, denoted as $T(X, Y)$, is defined according to the Jaccard coefficient measure as follows.

$$T(X,Y) = \frac{P_X\{Y\} \otimes P_Y\{X\}}{P_X\{Y\} \oplus P_Y\{X\}} = \frac{N_z}{N_x + N_y - Nz} \quad (3)$$

where N_x and N_y are the number of occurrences of '1' in reaction structure profiles $P_X\{Y\}$ and $P_Y\{X\}$ respectively. N_z is the occurrence of '1' in both $P_X\{Y\}$ and $P_Y\{X\}$. The value of $T(X, Y)$ is always between 0 and 1. The nearer to 1 the value is, the higher the degree of similarity between two reaction structure profiles is, and the nearer to 0 the value is, the lower the degree of similarity between two reaction structure profiles is.

In the example of the previous section, reaction structure profiles between two organisms are $P_O\{O'\} = [1\ 1\ 1\ 0]$, $P_{O'}\{O\} = [0\ 1\ 1\ 1]$, and then the similarity is calculated as $T(O, O') = \frac{2}{4} = 0.5$.

$D(X, Y)$, distance between X and Y, is defined as follows.

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

4. Experiment

To confirm the effectiveness of the proposed method, we conducted an experiment for comparative analysis of metabolic pathways among species.

Conventional phylogeny with rRNA sequence analysis has the possibility to include mistakes. However, conventional is established in large part. Therefore, in this paper, we use conventional phylogenetic tree in NCBI taxonomy, and define that construction of phylogenetic tree which is similar to conventional phylogenetic tree is worthwhile. Our goal isn't reconstruction of conventional phylogenetic tree. However, the method which can reconstruct phylogenetic tree from metabolic pathway will be available for phylogenetic analysis of unknown organisms.

Sixty-four organisms used in the experiment are shown in **Table 1**. We calculated the distance by using reaction structure profiles regarding all combinations of organisms used for the comparison and constructed a phylogenetic tree by hierarchical clustering. We used the 'choke points' as reference enzymes in the experiments. Clustering and construction of the phylogenetic tree were done with statistical processing software R ver.2.4.1. We used the furthest neighbor method for clustering in consideration of preliminary experiments and the feature of clustering methods.

First, we did a preliminary experiment to show that numeric settings of the number of nodes and the number of steps from the reference enzyme for extracting reaction structures were suitable. We show the comparative analysis only in Archaea changing each numerical value in **Fig. 5**. In (a) and (b), we configure the number of nodes to 5. Then we set the number of steps from the reference enzyme of (a) to 5 and the number of steps from the reference enzyme of (b) to 2. In (c) and (d), the number of steps from the reference enzyme is set to 3, and the number of nodes of (c) is set 3 and the number of nodes of (d) is set to 7. In (e), the number of steps from the reference enzyme is set to 3 and the number of nodes is set to 5.

As a result, organisms of Crenarchaeota and Euryarchaeota are classified the most similarly to NCBI taxonomy in (e). To construct a phylogenetic tree that complements conventional phylogeny, we set the number of steps from the reference enzyme to 3 and the number of nodes to 5.

Next, we constructed a phylogenetic tree based on random chosen reaction structure to confirm effectiveness of focusing important reaction structure (**Fig. 6**). We use random chosen enzymes as reference enzyme. In Fig. 6, Crenarchaeota and Euryarchaeota aren't classified at all. Thus, availability to use important reaction structure is confirmed.

Then we constructed a phylogenetic tree in three domains. The result is shown in **Fig. 7**. Most organisms are divided into three domains

Table 1 The 64 organisms included in the phylogenetic analysis. Full scientific names were abbreviated into three character notation (Abbr.) and their domain information[17] in phylogeny were also represented.

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

(Archaea, Eubacteria, and Eukaryote). This result shows that the three domain theory that C. R. Woese advocated[17] is supported from the viewpoint of metabolic function. Moreover, the cluster of Eubacteria is divided early, and Archaea and Eukaryote are divided into another cluster afterwards. This supports the hypothesis that organisms first divided into Eubacteria and Archaea[5], and Eukaryote evolved from Archaea.

4.1 Discussion

By experiment in three domains, it turns out that a phylogenetic tree that is similar to the conventional phylogenetic tree[⋆1] is constructed with the proposed method. The conventional phylogenetic tree is based on biological assumptions such as the similarity of phenotype. Therefore, the phylogenetic tree constructed with our method has a small part different from biological assumption, and was able to show phylogenetic relation on function not deriving from biological assumption.

However, there is a part that is greatly derived from biological assumption in particular. In Fig. 7, Aquifex aerolicus of Eubacteria and Arabidopsis thaliana (ath) of Eukaryote are included in the cluster of Archaea. Such a part contradicts the biological assumption. We think this problem may be due to lack of data in the pathway database. Not only the metabolic pathway, but information on organisms is not complete and a lot of information is insufficient[1),4)]. If data concerning the enzyme chosen as a reference enzyme is imperfect, none of the reaction structure is extracted from the reference enzyme. Such a case was frequently caused while in the actual experiment. As a result, the reaction structure profile becomes imperfect.

⋆1 NCBI Taxonomy [http://www.ncbi.nlm.nih.gov/Taxonomy/]

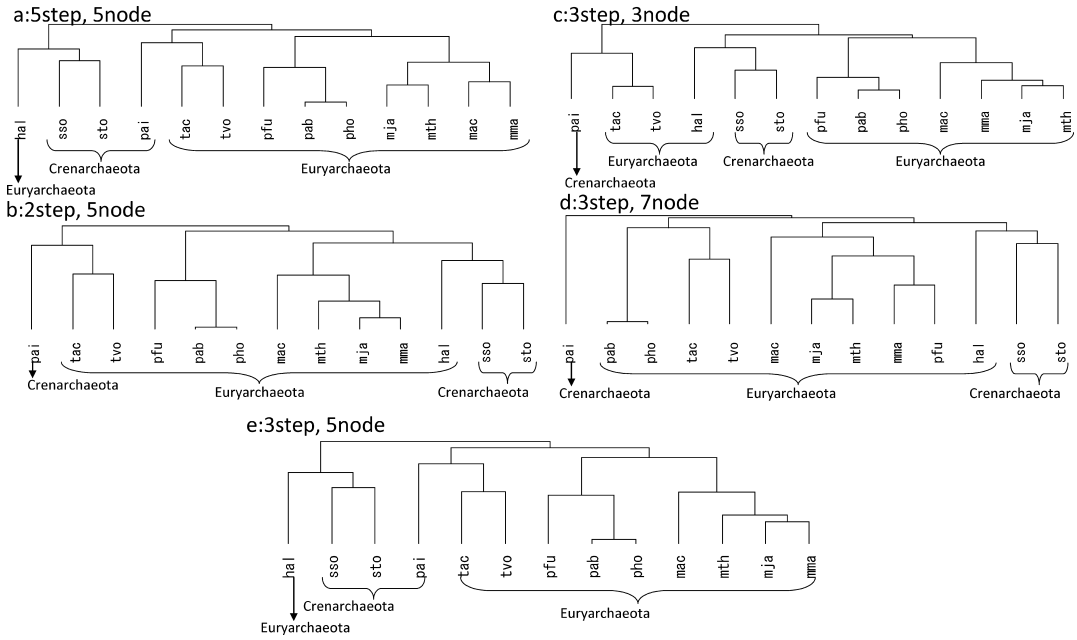

Fig. 5 Result of preliminary experiments.

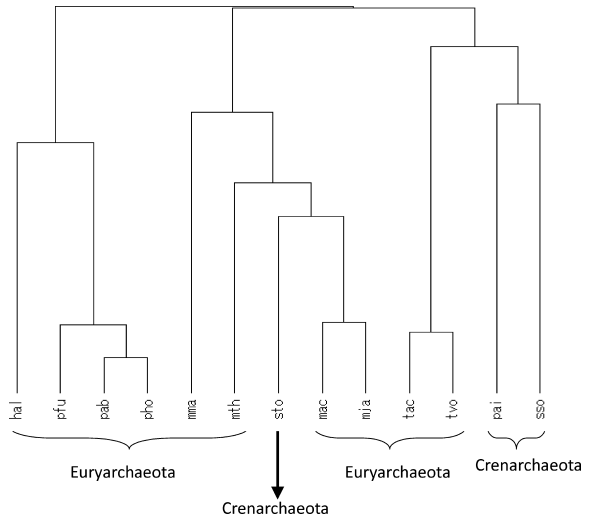

Fig. 6 Phylogenetic tree in Archaea based on random chosen reaction structure.

4.2 Comparison with Related Works

We use the NCBI taxonomy tree as the gold standard in comparison with related works. To compare our method with other method, we use the second cousins similarity to NCBI taxonomy, but it is necessary to note that datasets of organisms used in each method are not identical.

The second cousins similarity [15] is similarity between trees based on second cousin pairs. A sibling is a cousin of degree 0, a nephew is a cousin of degree 0.5, a first cousin is a cousin of degree 1 and so on.

The second cousins similarity to NCBI taxonomy of the phylogenetic tree based on our method in Fig. 7 is 0.2102426.

The method of June, et al. [11] completely classifies three domains in more organisms in comparison with our experiments. Moreover, the method reproduces three domain theory from the viewpoint of function as well as the proposed method does. However, the method doesn't reproduce the hypothesis [5] that Eukaryote evolved from Archaea.

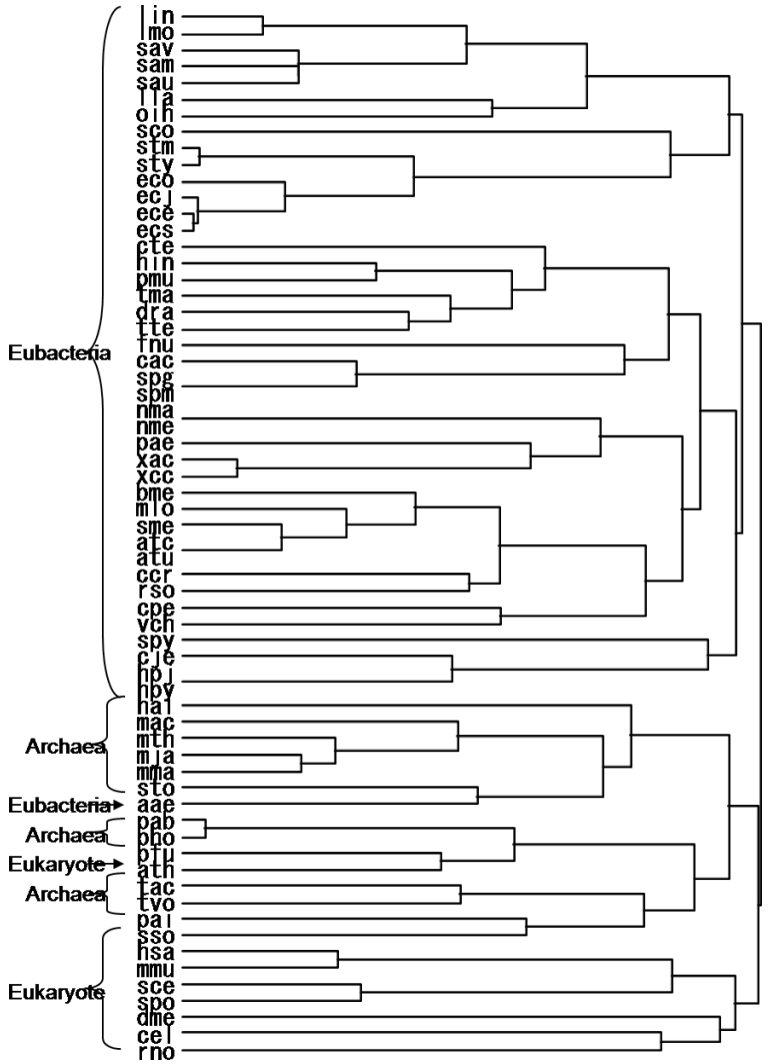

Fig. 7 Phylogenetic tree from 64 organisms.

The proposed method is a little inferior to the method in the domain classification in all domains. However, the second cousins similarity to NCBI taxonomy of the phylogenetic tree based on the method of June, et al. [11] is 0.1792829. Thus, the proposed method is near a conventional phylogeny on the classification in the domain and better than the method in that the proposed technique supplements conventional phylogeny. Moreover, on the point of reproducing the hypothesis that Eukaryote evolved from Archaea [5], the proposed method captures a feature of function more effectively and is able to derive phylogenetic relation on function among species.

The proposed method is better than the method by Tohsato [16] in classifying three domains with more organisms. Moreover, the method by Tohsato doesn't reproduce hypothesis [5]. However, the second cousins similarity to NCBI taxonomy of the phylogenetic tree based on method of Tohsato [16] is 0.2458101.

In common to all results of these three methods, the classification of Eubacteria isn't made well. This is because organisms included in Eubacteria are numerous and have highly varied metabolic systems.

Moreover, 'sso' and 'sto,' which are Crenarchaeota, and 'hal,' which is Euryarcheaota, are classified into the vicinity by a lot of experiment results including an existing method. For these reasons, it can be predicted that 'hal,' 'sso' and 'sto' have a very similar structure in an important metabolic part.

5. Conclusion

In this paper, we proposed a method for effective extraction of features on function by making reaction structure profiles that paid attention to an important reaction structure on the metabolic pathway. Phylogenetic relation on function was able to be derived without contradicting biological assumption by using reaction structure profiles. Moreover, the proposed method was able to reproduce the three domain theory in which organisms are generally classified into three domains (Archaea, Eubacteria, and Eukaryote) and the hypothesis that organisms first divided into Archaea and Eubacteria, and Eukaryote evolved from Archaea from the viewpoint of the function, which is not reproduced in preceding research. We used the NCBI taxonomy tree as the gold standard, but our goal is to calculate similarity between organisms on important reaction structures and to compliment the conventional phylogenetic tree, not to reconstruct it. Our method constructed an approximate phylogenetic tree, not an identical one. Thus, our method has the possibility to complement conventional phylogeny.

As future work, it is possible to limit metabolic pathways used for the comparison analysis to a specific metabolic system, and to change the metabolic system gradually. In the current method, all metabolic pathways are used. However, there are metabolic systems (such as the glycolytic system) that are developed in all organisms and those (such as metabolic systems that are related to photosynthesis) that are developed only in specific groups of organisms. It can be thought that it is possible to obtain good results on domain classification if we limit the metabolic system to the former [3]. It is possible to obtain good results on classification within domain if we limit it to the latter. The feature on function is captured by changing the metabolic part for the purpose of classification, and phylogenetic relation on function can be shown more effectively.

References

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) Clemente, J.C., Satou, K. and Valiente, G.: Phylogenetic reconstruction from non-genomic data, *Bioinformatics*, Vol.23, pp.e110–e115 (2006).
3) Dandekar, T., Schuster, S., Snel, B., Huynen, M. and Bork, P.: Pathway alignment: Application to the comparative analysis of glycolytic enzymes, *Biochemical Journal*, Vol.343, pp.115–124 (1999).
4) Félix, L. and Valiente, G.: Validation of metabolic pathway databases based on chemical substructure search, *Biomolecular Engineering*, Vol.24, pp.327–335 (2007).
5) Feng, D.F., Cho, G. and Doolittle, R.F.: Determining divergence times with a protein clock: update and reevalution, *Proc. National Academy of Science of USA*, Vol.94, pp.13028–13033 (1997).
6) Fitch, W.M. and Margoliash, E.: Construction of phylogenetic trees, *Science*, Vol.155, No.3760, pp.279–284 (1967).
7) Forst, C.V. and Schulten, K.: Phylogenetic analysis of metabolic pathways, *Journal of Molecular evolution*, Vol.52, No.6, pp.471–489 (2001).
8) Heymans, M. and Singh, A.K.: Deriving phylogenetic trees from the similarity analysis of metabolic pathways, *Binformatics*, Vol.19, pp.i138–i146 (2003).
9) 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).
10) 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).
11) 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).
12) Pinter, R.Y., Rokhlenko., O., Yeger-Lottrm, E. and Ziv-Ukelson, M.: Alignment of metabolic pathways, *Bioinformatics*, Vol.21, No.16, pp.3401–3408 (2005).
13) Rahman, S.A., Advani, P., Schunk, R., Schrader, R. and Schomburg, D.: Metabolic pathway analysis web service (Pathway Hunter Tool at CUBIC), *Bioinformatics*, Vol.21, No.7, pp.1189–1193 (2005).
14) 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).
15) Shasha, D., Wang, J.T.L. and Zhang, S.: Unordered tree mining with applications to phylogeny, *Proc. 20th International Conference on*

Data Engineering, pp.708–719 (2004).
16) Tohsato, Y.: A method for species comparison of metabolic networks using reaction profile, *Inf. Proc. Soc. Japan Trans. Bioinformatics*, Vol.47, No.SIG 17(TBIO 1), pp.42–47 (2006).
17) Woese, C.R., Kandler, O. and Wheelis, M.L.: Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya, *Proc. National Academy of Sciences*, Vol.87, pp.4576–4579 (1990).
18) 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).

(Received September 21, 2007)
(Accepted November 19, 2007)

(Communicated by *Susumu Goto*)

Yuta Ashida received the B.E. degree from Kobe University. He is currently a student of M.E. course at the Graduate School of Engineering, Kobe University. His major research interests include Bioinformatics and Graph mining.

Tomonobu Ozaki received his Ph.D. in Media and Governance from Keio University in 2002. Now he is an assistant professor of Organization of Advance Science and Technology, Kobe University. He is a member of JSAI.

Takenao Ohkawa received his B.E., M.E., and Ph.D. degrees from Osaka University in 1986, 1988, and 1992, respectively. He is currently a Professor at the Department of Computer Science and Systems Engineering, Graduate School of Engineering, Kobe University. His research interests include intelligent software and bioinformatics. He is a member of the IEEE, the Institute of Electronics, Information, and Communication Engineers, the Institute of Electrical Engineers in Japan, and the Japanese Society for Artificial Intelligence.