Analysis of the Differences in Metabolic Network Expansion between Prokaryotes and Eukaryotes

Michihiro Tanaka
mtanaka@kuicr.kyoto-u.ac.jp

Takuji Yamada
takuji@kuicr.kyoto-u.ac.jp

Masumi Itoh
itoh@kuicr.kyoto-u.ac.jp

Shujiro Okuda
okuda@kuicr.kyoto-u.ac.jp

Susumu Goto
goto@kuicr.kyoto-u.ac.jp

Minoru Kanehisa
kanehisa@kuicr.kyoto-u.ac.jp

Bioinformatics center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan

Abstract

Recent evidence points to the existence of scale-free properties in many biological networks. By topological analysis, several models including preferential attachment and hierarchical modules have been proposed to explain how these networks are organized. On the other hand, analyses using dynamics have suggested that gene expression and metabolic networks have been organized with the scale-free property by the other models such as "rich-travel-more" and "log-normal dynamics." Because most of these approaches are based on comparative genomics of extant species, and did not consider evolutionary events such as horizontal gene transfer, gene loss and gene gain, we have analyzed transition of metabolic networks from the vertical point of view of evolution. First, to identify metabolic networks of common ancestors, we applied a parsimony algorithm for the enzymatic reaction set. Then by comparing the estimated metabolic networks among common ancestors, we investigated the transition of metabolic networks along the evolutionary process. As a result, we estimated enzymatic reaction contents of 227 common ancestors from 228 extant species, and found that links of several specific metabolites have frequently changed during the course of evolution.

Keywords: metabolic network, network organization, parsimony, proportional dynamics model

1 Introduction

Recent progress in large-scale sequencing projects has resulted in the accumulation of complete genome sequence information for a number of species, and integrated pathway databases such as KEGG allow us to analyze organism-specific connectivity maps of metabolites based on the annotation of the genomes. Large-scale analyses of structural organization of such metabolic networks as well as other cellular networks including signal transduction pathways and protein-protein interaction networks revealed that they have a scale-free property [2]. In these networks, the distribution of $P(k)$, the number of metabolites participating in k reactions, follows the power-law $P(k) \sim k^{-\gamma}$, where γ is a constant. This distribution in the metabolic network implies that, whereas most metabolites are involved in only few reactions, there are a few metabolites that are involved in many reactions and serve as "hubs", e.g. ATP [6].

To explain the emergence of a scale-free network, the following two key rules have been proposed, 1) the network grows by adding new nodes, and 2) preferential attachment, where new nodes tend to link to heavily connected nodes in the network [2]. The preferential attachment model also explains evolution of other networks, such as World Wide Web and scientific collaborations [2]. Many biological networks also show hierarchical organization, where a module appears as a highly interconnected group of nodes. In such networks, each node clustered into one of the modules does not link to other nodes randomly but links to others by a unit of module [14]. These models have been proposed based

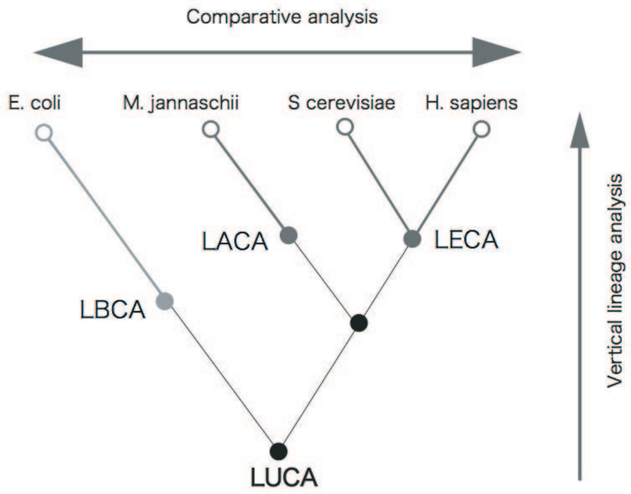


Figure 1: A schematic representation of the relative position of the four species on a simplified phylogenetic tree. In comparative genome analysis, comparison across extant species allows investigation of the evolutionary relationships among extant species (e.g. *E. coli* vs *M. jannaschii*). In vertical lineage analysis, a parsimonious algorithm, we investigated evolutionary relationships within the phylogenetic tree using the gene content information of extant species; to compare *E. coli* and *M. jannaschii*, we used the following two lineages: 1st lineage; LUCA to LBCA and LBCA to *E. coli*, 2nd lineage; LUCA to LACA and LACA to *M. jannaschii*). Abbreviations: LUCA, last universal common ancestor; LECA, last eukaryotic common ancestor; LBCA, last bacterial common ancestor; LACA, last archaeal common ancestor [11].

on purely topological analyses of the current networks in terms of the statistical analysis of degree distribution.

On the other hand, recent research has reported a growing network model based on transcriptomic and comparative genomic approaches. For example, investigations of changes in metabolite links along the evolutionary process in metabolic networks using comparative genomic approaches revealed the principle called ‘rich-travel-more’ for dynamics in metabolic networks where highly linked metabolites change their chemical links more than less linked metabolites [13, 19]. In the study by Ueda *et al.* [19], transition is calculated between species which belong to the same major domain (archaea, bacteria and eukaryote) and between 126 species without considering the phylogenetic relationship. In the work of Nacher *et al.* [13], by assuming that *M. jannaschii* is older than *E. coli*, and that *E. coli* is older than *S. cerevisiae*, two transitions from *M. jannaschii* to *E. coli*, and from *E. coli* to *S. cerevisiae* were calculated, based on absolute number changes of enzymatic reactions catalyzing a metabolite. Because their approaches were the comparative analysis of extant species, they did not consider evolutionary events including horizontal gene transfer (HGT), gene duplication, gene loss and gain. Other studies have recently demonstrated that these events occur far more frequently than previously thought [9, 16, 17]. It is, therefore, important to deal with these events to understand the evolutionary process.

Phylogenetic features of metabolic network organization have not been unexplored. Recently, two studies addressing this issue were reported. The studies focused on the evolution of preferential attachment in protein networks using four estimated ancient groups from the three major domains of life (archaea, bacteria, and eukaryote) [5]. The classification of each protein into four groups is implemented using protein conservation information between the three domains [10]. We investigated the evolution of metabolic networks by using a similar method.

We compared biological networks along the evolutionary process (named “Vertical lineage anal-

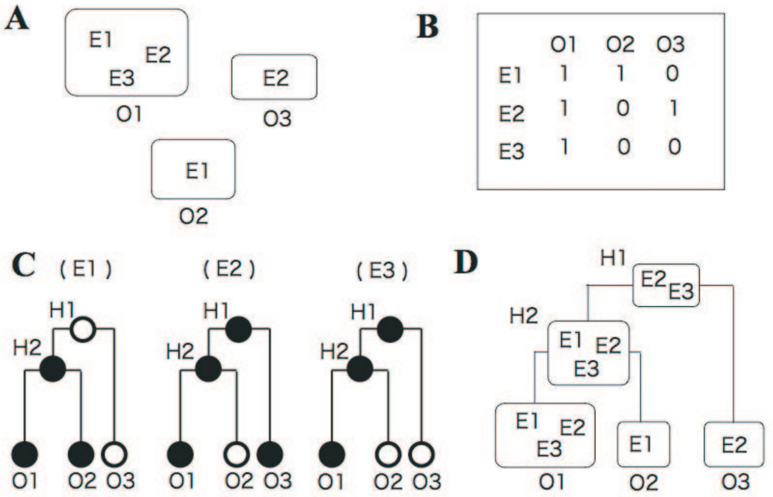


Figure 2: Schematic representation of the procedure used to reconstruct the enzymatic reaction contents of common ancestors. A: In this example, we focus on three enzymatic reactions (E1, E2 and E3). Existence of these reactions varies among species (O1, O2 and O3). Such species-specific enzymatic reaction information is stored in the KEGG PATHWAY database. B: We construct a phylogenetic profile, indicating which species have the enzymatic reactions. Presence or absence of enzymatic reactions is represented by “1” or “0,” respectively. C: We next reconstruct a parsimonious scenario by using the phylogenetic tree and profiles. Black nodes represent the presence of an enzymatic reaction while white nodes represent its absence. D: We reconstruct the enzymatic reaction contents in common ancestors (H1 and H2).

ysis” in Figure 1). To accurately trace the transition of metabolic networks during evolution, we reconstructed the enzymatic reaction contents for ancient species by the method proposed by Mirikin *et al.* [12]. They reconstructed a microbial phylogenetic network evolutionary assuming a parsimonious evolutionary scenario. We analyzed the relationship between a metabolite and the enzymatic reaction catalyzed it. Here we show that the evolutionarily expanding pattern of metabolic networks may be different between prokaryotes and eukaryotes.

2 Materials and Methods

2.1 Species Specific Metabolic Pathways

We used the KEGG database including PATHWAY and Orthology (KO). The KEGG PATHWAY database is a collection of manually drawn pathway maps for metabolism, genetic information processing, and environmental information processing for each species. The pathway information is also represented in XML format, called KEGG Markup Language (KGML). It allowed us to extract the main route of the enzymatic reaction network, i.e. excluding cofactors, and to determine each enzymatic reaction set whether an enzymatic reaction is present or absent in a given organism [7] (see Figure 2A).

2.2 Phylogenetic Profile

To represent the subset of species that contains an enzymatic reaction, we constructed a phylogenetic profile for each enzymatic reaction. This profile is a string with n entries, each represented using a bit, where n corresponds to the number of species. At each position in the profile, the presence of an enzymatic reaction in the corresponding species is indicated with 1 and its absence with 0. Using this approach, we were able to construct the phylogenetic profiles for 3032 enzymatic reactions. Each profile is comprised of 228 species consisting of 189 bacteria, 20 archaea and 19 eukaryotes. We utilized the genomic content data based on orthologous relationships called KEGG Orthology (KO). KO is developed using best-hit relations in pairwise genome comparisons and the similarity scores are stored in the SSDB database [8] (see Figure 2B).

2.3 Universal Phylogenetic Tree

In order to reconstruct the evolutionary relationships among species, we constructed a universal phylogenetic tree of the 228 species. We first constructed the NJ tree of bacteria and archaea using nucleotide sequence data for 16S ribosomal RNA in the KEGG GENES database [7] and Ribosomal RNA Project [3]. We used a multiple alignment technique based on ClustalW [18] to generate a tree. The eukaryotic part of the tree was constructed manually based on [1]. Then, we joined the phylogenetic trees of bacteria and archaea and of eukaryotes at the last common ancestor of archaea (LACA). Figure 3A is the universal phylogenetic tree in a rooted form, showing the three domains and 17 taxa based on the KEGG organisms (`http://www.genome.jp/kegg/catalog/org_list.html`).

2.4 Parsimony Algorithm for Detecting Enzymatic Reaction Contents in Common Ancestors

To identify enzymatic reaction contents of common ancestors from the phylogenetic profiles and the universal phylogenetic tree, we used the algorithm PARS using parsimony-based approach [12]. To estimate whether a specific enzymatic reaction of a common ancestor was present or absent, we first search the two direct child nodes of the ancestral node in the phylogenetic tree. If the enzymatic reaction is conserved in both two children, we conclude that it was present in the common ancestor. If it is not conserved, we search another child node with the same parent node. If the enzymatic reaction was conserved between the 1st child node and the 2nd child node, we conclude that the enzymatic reaction is conserved in a common ancestor. Starting from a leaf of the tree, we performed these processes until the status of the root was decided (Figure 2C). The algorithm was implemented in the Ruby programming language and computed on an SGI Origin3800 with 252 CPUs.

2.5 Reconstruction of Metabolic Networks for Common Ancestors

Following the method of Jeong *et al.* [6], we defined the link between two metabolites if there is a reaction from one metabolite to the other. We reconstructed a metabolic network from the links among metabolites based on the whole enzymatic reaction contents existing in a node in the universal phylogenetic tree. In this work, the reconstructed network is regarded as an undirected graph because we did not take into account of the direction of enzymatic reactions.

2.6 Calculation of Link Changes of Metabolites

To characterize network transitions, we used two functions:$|\Delta k|$ and $\beta(k)$ which were developed as the measure of transition between different states. Absolute change $|\Delta k| = |k2 - k1|$ was developed by Ueda *et al.* [19]. In the metabolic network analysis, $k1$ is the number of enzymatic reactions that catalyze a metabolite in a certain state and $k2$ is the number of such reactions in another state. These values were calculated for each metabolite on metabolic pathways [19]. $\beta(k)$ was developed by Nacher *et al.* [13]. This function defines network fluctuation as the average size of the changes between two

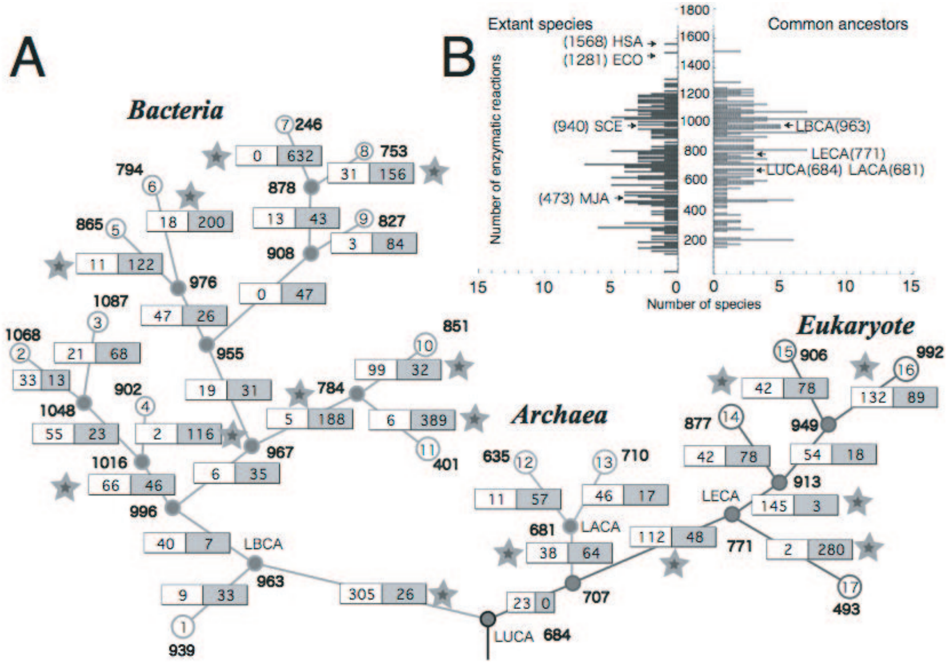


Figure 3: Estimated enzymatic reaction contents of common ancestors of major taxa. The number inside a white circle on the leaf corresponds to the following groups of extant species. Bacteria: 1, Actinobacteria; 2, Gamma and Beta proteobacteria; 3, Alpha proteobacteria; 4, Epsilon proteobacteria; 5, Clostridia; 6, Lactobacillales; 7, Mollicutes; 8, Bacteroid; 9, Spirochete; 10, Cyanobacteria; 11, Chlamydia. Archaea: 12, Crenarchaeota; 13, Euryarchaeota. Eukaryote: 14, Plants; 15, Fungi; 16, Animal; 17, Protists. The grey nodes represent common ancestors of 17 groups of extant species. The black bold numbers represent the total numbers of enzymatic reactions in common ancestors. Numbers in white box indicate the gain of enzymatic reactions, whereas numbers in gray box indicate the loss of enzymatic reactions from a common ancestor to another in the phylogenic tree. Star marks indicate that more than 100 evolutionary events occurred, where the sum of the number of gains and losses is defined as the evolutionary event. (B) The distribution of the number of enzymatic reactions is shown for extant species (left) and reconstructed common ancestors (right).

states that is represented by

$$\beta(k_1) = \frac{\sqrt{\Sigma_{i=1}^{N_{k1}} (k_{2i} - k_{1i})^2}}{N_{k_1}}, \tag{1}$$

where k_{1i} and k_{2i} are the degrees of a node i in the states 1 and 2, respectively, and N_{k1} is the number of metabolites with k_{1i} links in the state 1. In our work, to evaluate the proportional dynamics of a network, we calculated $\beta(k)$. The average degree of a node in a reconstructed metabolic network is smaller than in Ueda's network. In his network, the measured metabolites were defined as all of the registered metabolites in the KEGG database, but in our network, metabolites were eliminated such as cofactors, ions and H_2O.

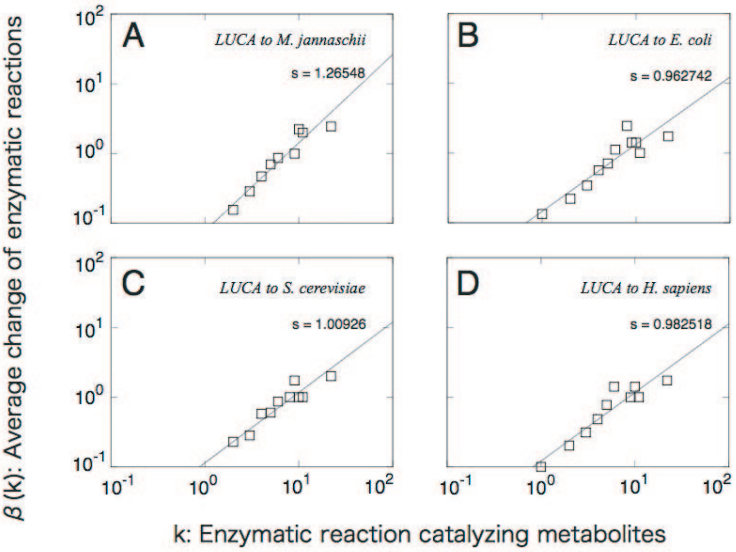


Figure 4: The change of the catalyzed enzymatic reaction of a metabolite in the evolution of metabolic networks from LUCA to *M. jannaschii* (A), *E. coli* (B), *S. cerevisiae* (C) and *H. sapiens* (D), respectively. The value S is the least squares regression coefficient calculated by Gnuplot ver. 4.0. The metabolic networks in these four species have been expanded according to the proportional dynamics proposed by Nacher *et al.* (see Materials and Methods).

3 Result

3.1 Estimation of Metabolic Networks of Common Ancestors

To estimate the enzymatic reaction contents of common ancestors, we calculated the parsimonious evolutionary scenario that take into account the presence or absence of each enzymatic reaction on a phylogenetic tree. First, 228 species were classified into 17 taxonomy groups, and then we calculated the enzymatic reaction contents of the common ancestor of each taxonomy group. Figure 3A shows the number of enzymatic reactions in 33 common ancestors 17 major taxa, and the other 16 common ancestors. The estimated number of the enzymatic reactions of the last universal common ancestor (LUCA), last eukaryotic common ancestor (LECA), last bacterial common ancestor (LBCA) and last archaeal common ancestor (LACA) were 684, 771, 963 and 681, respectively. The number of enzymatic reactions varied among common ancestors (Figure 3B).

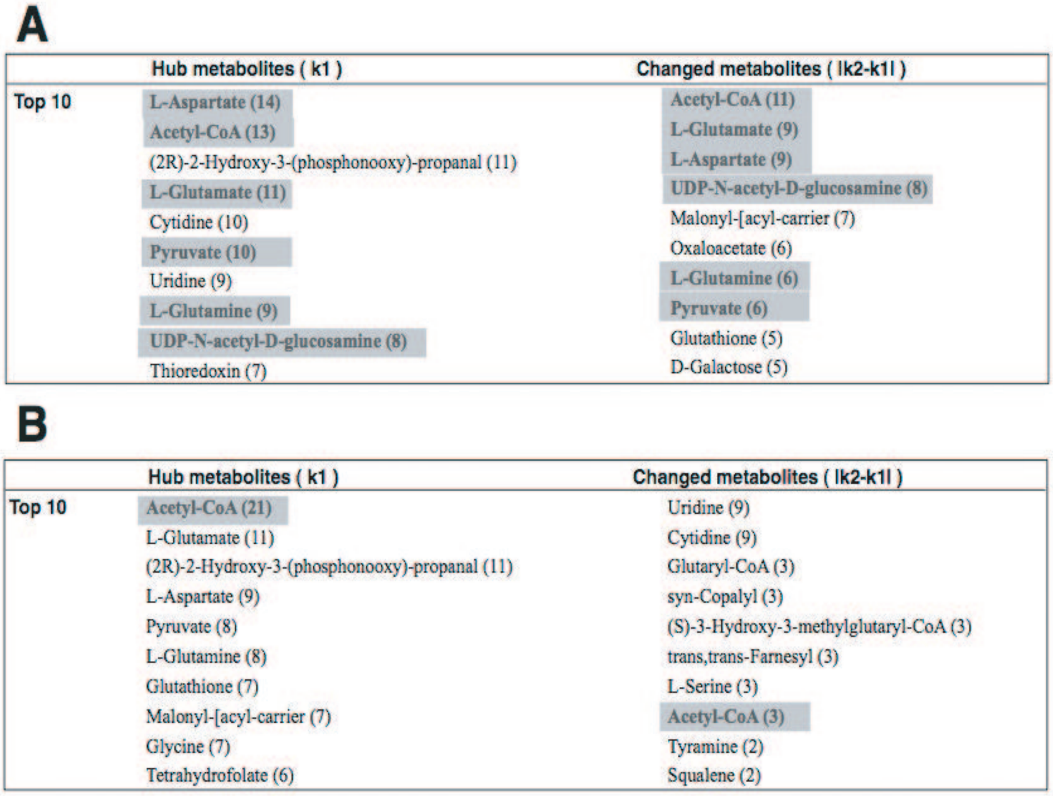
Furthermore, we examined the gain and loss of enzymatic reactions among 16 common ancestors estimated from 17 major taxa (Figure 3A). Among them, a large number of enzymatic reactions (305) were gained from LUCA to LBCA. From LUCA to the common ancestor between LACA and LECA, only 23 gain events were observed.

3.2 Transition of Metabolic Networks from Common Ancestors to Extant Species

3.2.1 Transition from LUCA to Extant Species

To investigate the features in the metabolic network transition along with the evolutionary process, we first analyzed link changes of conserved metabolites between LUCA and four selected extant species: *M. jannaschii*, *E. coli*, *S. cerevisiae* and *H. sapiens*, as representatives from archaea, bacteria, unicellular eukaryotes and multicellular eukaryotes, respectively. To quantify the link changes, we calculated $\beta(k)$, which is the average size of the changes of the metabolite with k links. Figure 4 shows the log-log plot of $\beta(k)$, and the proportional increases of $\beta(k)$ were observed in all three domains. This result

Table 1: Highly linked metabolites and highly changed metabolites. Metabolites (shaded in grey) indicate the top 10 ranked metabolites from both hub metabolites and changed metabolites. A: Top 10 ranked metabolites were observed in transition to common ancestor of Mollicutes in bacteria lineage. B: Top 10 ranked metabolites were observed in transition from LECA to the direct internal common ancestor node in eukaryote lineage. Also these relationships between the connectivity of metabolites and its change are shown on two graphs, A and H, in Figure 5.



A

	Hub metabolites (k1)	Changed metabolites (lk2-k1l)
Top 10	L-Aspartate (14)	Acetyl-CoA (11)
	Acetyl-CoA (13)	L-Glutamate (9)
	(2R)-2-Hydroxy-3-(phosphonooxy)-propanal (11)	L-Aspartate (9)
	L-Glutamate (11)	UDP-N-acetyl-D-glucosamine (8)
	Cytidine (10)	Malonyl-[acyl-carrier (7)
	Pyruvate (10)	Oxaloacetate (6)
	Uridine (9)	L-Glutamine (6)
	L-Glutamine (9)	Pyruvate (6)
	UDP-N-acetyl-D-glucosamine (8)	Glutathione (5)
	Thioredoxin (7)	D-Galactose (5)

B

	Hub metabolites (k1)	Changed metabolites (lk2-k1l)
Top 10	Acetyl-CoA (21)	Uridine (9)
	L-Glutamate (11)	Cytidine (9)
	(2R)-2-Hydroxy-3-(phosphonooxy)-propanal (11)	Glutaryl-CoA (3)
	L-Aspartate (9)	syn-Copalyl (3)
	Pyruvate (8)	(S)-3-Hydroxy-3-methylglutaryl-CoA (3)
	L-Glutamine (8)	trans,trans-Farnesyl (3)
	Glutathione (7)	L-Serine (3)
	Malonyl-[acyl-carrier (7)	Acetyl-CoA (3)
	Glycine (7)	Tyramine (2)
	Tetrahydrofolate (6)	Squalene (2)

indicates that metabolites involved in a larger number of enzymatic reactions (hubs) tend to have the larger link changes in the evolutionary process of the metabolic network.

3.2.2 Transition between Internal Nodes: Common Ancestors

We then investigated the effects of evolutionary events on metabolic networks by focusing on the metabolites that largely changed their links. Figure 5 shows the changes of such metabolites that change at least five links in some branches on phylogenetic tree. Interestingly, proportional dynamics were observed in any bacterial lineages (Figure 5A-G, $S = 0.0064 \sim 1.0147$), but not in eukaryotes (Figure 5H, $S = -0.1517$). The network transition according to the proportional dynamics indicates that metabolites with more enzymatic reactions change their links more frequently. Furthermore, there were four eukaryotic branches on the phylogenetic tree where over 100 gains and losses of enzymatic reactions happened (Figure 3A, eukaryotic branches with the star mark). However, proportional dynamics were not observed at all of the four branches. Tab 1 shows top 10 ranking of such metabolite in the A and H shown in Figure 5. First, to investigate in detail whether proportional dynamics are conserved in each branch in which a major evolutionary event occurred, we calculated the absolute link changes in all distinct parent-child pairs in the phylogenetic tree. The absolute link change for each metabolites is $|\Delta k| = |k2 - k1|$, where $k1$ is the number of a catalyzed reactions in an ancestral

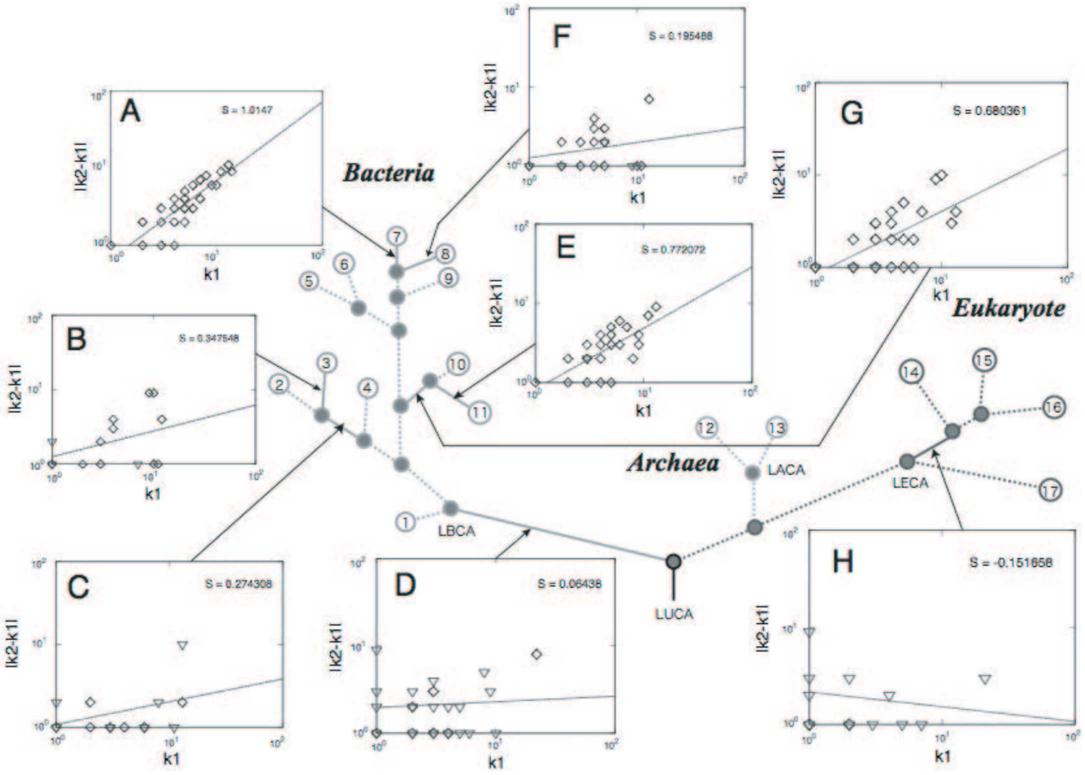


Figure 5: The change of the catalyzing enzymatic reactions of metabolites. $k1$ represents the number of enzymatic reactions involved in a metabolite at the ancestor node. The absolute change value $|\Delta k| = |k2 - k1|$ represents the number of catalyzing enzymatic reactions of the metabolite during the transition common ancestor. The node is child. Diamonds represent the decreasing change, while triangles represent the increasing change. We show the eight graphs for transition of changes of enzymatic reactions. The value S is regression coefficient of least squares methods calculated by Gnuplot (ver. 4.0), and indicates the proportional dynamics (see Materials and Methods).

species and $k2$ is those in its offspring. The distribution of $k1$ where at least five link changes have occurred.

4 Discussion

4.1 Estimation of Metabolic Networks of Ancient Species

We estimated gain and loss of enzymatic reactions and showed that the common ancestors in the bacterial lineage tend to lose a large number of enzymatic reactions. On the other hand, the common ancestor in eukaryotes and archaea does not show any differences between loss and gain of enzymatic reactions. We also showed that the number of enzymatic reactions in LUCA is not the minimum as well as in LACA, LECA and LBCA (Figure 3B). These results were obtained because we employed the algorithm for parsimonious reconstruction, and inferred the gene contents of common ancestors of the extant species, whereas many previous studies on metabolic pathway evolution are based on horizontal comparative genomics, where only differences between extant species are calculated.

One of our future goals is to verify the reconstructed gene contents of common ancestors. In previous research, a hypothetical common ancestors with a predefined number of genes or protein families was reconstructed, and then gene contents between the hypothetical and inferred data were compared [12]. This approach has an advantage in its clear statistical significance, but it is not clear whether it implies a biological significance. To address the biological significance, other methods such as network simulations are necessary. Recently, Ebenhöh *et al.* have developed an algorithm that calculates a set of compounds producible from predefined compounds in a metabolic network and called the resulting set of compounds an expansion of the network or scope [4]. Raymond *et al.* took this approach and reported the effect of oxygen in the metabolic network [15]. We propose that the algorithm can be applied to verify the reconstructed enzymatic reactions of common ancestors. Figure 3B shows that the numbers of reconstructed enzymatic reactions were categorized into the same scale in both common ancestors and extant species. The comparison of scopes in extant species with those in common ancestors may allow us to biologically verify the reconstructed metabolic networks.

4.2 Transition of Metabolite Usage in Evolutionary Process

Ueda *et al.* revealed the proportional dynamics in gene expression and metabolic networks and termed it the "rich-travel-more" model [19]. Nacher *et al.* also revealed proportional dynamics in the chemical reaction network where nodes and edges represent reactions and metabolites, respectively, and showed that it follows "log-normal dynamics" [13]. The value S in Figure 4 indicates the growth rate of the links in a network. In other words, the larger the value, the higher the change at "hub" nodes. Interestingly, the values of S in our results were much lower than those calculated by interspecies comparison. The differences are probably due to the property of the studied network. In the network used in [19], the value S ranged from 0.87 to 1.13. The network was composed of an unlimited set of metabolites including cofactors, while our network was composed of only selected metabolites. Therefore the plot $k1$ to $|\Delta k| = |k2 - k1|$ is far sparser and the variation in the degree of nodes in our network is smaller, and this is the reason why the scale of changes in the degree was lower.

These results raise another possibility, that the evolutionary pattern of metabolic networks in eukaryotes is different from that in bacteria. We observed clear difference in the calculated proportional dynamics of enzymatic reaction changes among common ancestors of major clades. In the bacteria lineage, the values ranged from 0 to 1.0, but in eukaryotes they ranged from 0.07 to 0.02 (Figure 5). These results suggest that, in bacteria, the gain and loss of enzymatic reactions have direct effects on key metabolites, such as L-aspartate and acetyl-CoA, but not in eukaryotes (Table 1). These discrepancies between bacteria and eukaryotes might reflect their differences in the characteristics of network expansion. Through evolution, the eukaryotic metabolic pathways might conserve the key metabolites and utilize other metabolites, while the bacterial metabolic pathways might utilize key metabolites to expand and shrink the size of their network.

Acknowledgments

Authors are grateful to Naoki Nariai, Oliver Ebenhöh, Thomas Handorf and Reinhart Heinrich for the ideas discussed in the graduate student internship in Berlin, and Tetsuya Sato and Keiichi Kuma for their advice of phylogenetic analysis. They also thank to Nelson Hayes for critical reading of the manuscript. This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology the Japan Science and Technology Agency. The Computational resource was provided by the Bioinformatics Center, Institute for Chemical Research, Kyoto University.

References

[1] Baldauf, S.L., Roger, A.J., Wenk-Siefert, I., and Doolittle, W.F., A kingdom-level phylogeny of eukaryotes based on combined protein data, *Science*, 290:972–977, 2000.

[2] Barabasi, A.L. and Oltvai, Z.N., Network biology: understanding the cell's functional organization, *Nat. Rev. Genet.*, 5:101–113, 2004.

[3] Cole, J.R., Chai, B., Farris, R.J., Wang, Q., Kulam, S.A., McGarrell, D.M., Garrity, G.M., and Tiedje, J.M., The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis, *Nucleic Acids Res.*, 33:D294–296, 2005.

[4] Ebenhöh, O., Handorf, T., and Heinrich, R., Structural analysis of expanding metabolic networks, *Genome Inform.*, 15:35–45, 2004.

[5] Eisenberg, E. and Levanon, E.Y., Preferential attachment in the protein network evolution, *Phys. Rev. Lett.*, 91:138701, 2003.

[6] Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., and Barabasi, A.L., The large-scale organization of metabolic networks, *Nature*, 407:651–654, 2000.

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

[8] Kanehisa, M., Goto, S., Kawashima, S., and Nakaya, A., The KEGG databases at GenomeNet, *Nucleic Acids Res.*, 30:42–46, 2002.

[9] Kunin, V., Goldovsky, L., Darzentas, N., and Ouzounis, C.A., The net of life: reconstructing the microbial phylogenetic network, *Genome Res.*, 15:954–959, 2005.

[10] Light, S., Kraulis, P., and Elofsson, A., Preferential attachment in the evolution of metabolic networks, *BMC Genomics*, 6:159, 2005.

[11] Makarova, K.S., Wolf, Y.I., Mekhedov, S.L., Mirkin, B.G., and Koonin, E.V., Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell, *Nucleic Acids Res.*, 33:4626–4638, 2005.

[12] Mirkin, B.G., Fenner, T.I., Galperin, M.Y., and Koonin, E.V., Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes., *BMC Evol. Biol.*, 3:2, 2003.

[13] Nacher, J.C., Ochiai, T., Yamada, T., Kanehisa, M., and Akutsu, T., The role of log-normal dynamics in the evolution of biochemical pathways, *Biosystems*, 83:26–37, 2006.

[14] Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., and Barabasi, A.L., Hierarchical organization of modularity in metabolic networks, *Science*, 297:1551–1555, 2002.

[15] Raymond, J. and Segre, D., The effect of oxygen on biochemical networks and the evolution of complex life, *Science*, 311:1764–1767, 2006.

[16] Snel, B., Bork, P., and Huynen, M.A., Genomes in flux: the evolution of archaeal and proteobacterial gene content, *Genome Res.*, 12:17–25, 2002.

[17] Tanaka, T., Ikeo, K., and Gojobori, T., Evolution of metabolic networks by gain and loss of enzymatic reaction in eukaryotes, *Gene*, 365:88–94, 2006.

[18] Thompson, J.D., Higgins, D.G., and Gibson, T.J., CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res.*, 22:4673–4680, 1994.

[19] Ueda, H.R. and Hogenesch, J.B., Priciples in the evolution of metabolic networks. In *E-print archive q-bio.MN/0503038.*