World Academy of Science, Engineering and Technology
International Journal of Biomedical and Biological Engineering
Vol:1, No:11, 2007

Molecular Evolutionary Analysis of Yeast Protein Interaction Network

Soichi Ogishima, Takeshi Hase, So Nakagawa, Yasuhiro Suzuki, and Hiroshi Tanaka

***Abstract*—**To understand life as biological system, evolutionary understanding is indispensable. Protein interactions data are rapidly accumulating and are suitable for system-level evolutionary analysis. We have analyzed yeast protein interaction network by both mathematical and biological approaches. In this poster presentation, we inferred the evolutionary birth periods of yeast proteins by reconstructing phylogenetic profile. It has been thought that hub proteins that have high connection degree are evolutionary old. But our analysis showed that hub proteins are entirely evolutionary new. We also examined evolutionary processes of protein complexes. It showed that member proteins of complexes were tend to have appeared in the same evolutionary period. Our results suggested that protein interaction network evolved by modules that form the functional unit. We also reconstructed standardized phylogenetic trees and calculated evolutionary rates of yeast proteins. It showed that there is no obvious correlation between evolutionary rates and connection degrees of yeast proteins.

***Keywords*—**Protein interaction network, evolution, modularity, evolutionary rate, connection degrees.

I. Introduction

BIOLOGY has entered a new era. The vast amounts of accumulating biological data and knowledge have completely overwhelmed our ability to understand it. In a new era, it is necessary to understand how the components which involve in biological systems interact and work together as biological systems from the various biological data and knowledge of components at molecular level. To understand life as biological system, evolutionary understanding is indispensable. It reveals the structure of biological systems and leads us to the "ontological" comprehension of biological systems. Comprehensive data of protein interactions have been also accumulating [1][2][3] and are suitable for system-level evolutionary analysis.

We have analyzed yeast protein interaction network by both mathematical and biological approaches. In this poster presentation, we inferred the evolutionary birth periods of yeast proteins by reconstructing phylogenetic profile [4]. It has been thought that hub proteins that have high connection degree are evolutionary old [5]. But our analysis showed that hub proteins are entirely evolutionary new. We also examined evolutionary processes of protein complexes. It showed that member proteins of complexes were tend to have appeared in the same evolutionary period. Our results suggested that protein interaction network evolved by modules that form the functional unit. It is consistent with the previous suggestion from the facts that two proteins tend to interact with each other if they are in the same or similar evolutionary categories [6].

Manuscript received May 20, 2005. This work was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

S. O. Author is with the Computational Biology Department, Graducate School of Tokyo Medical and Dental University, Yushima 1-5-45, Bunkyo-ku, Tokyo 113 8510 Japan (phone: +81 3 5803 5840; fax: +81 3 5803 0247; e-mail: ogishima@sysevo.org).

T. H. Author is with the Bioinformatics Department, Graduate School of Tokyo Medical and Dental University, Yushima 1-5-45, Bunkyo-ku, Tokyo 113 8510 Japan (e-mail: hase@bioinfo.tmd.ac.jp).

S. N. Author is with the Computational Biology Department, Graducate School of Tokyo Medical and Dental University, Yushima 1-5-45, Bunkyo-ku, Tokyo 113 8510 Japan (e-mail: so@bioinfo.tmd.ac.jp).

Y. S. Author is with the Complex Systems Science Department, Graducate School of Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601 Japan (e-mail: ysuzuki@is.nagoya-u.ac.jp).

H. T. Author is with the Computational Biology Department, Graducate School of Tokyo Medical and Dental University, Yushima 1-5-45, Bunkyo-ku, Tokyo 113 8510 Japan (e-mail: tanaka@cim.tmd.ac.jp).

II. Materials and Methods

A. Materials

We collected yeast (*Saccharomyces cerevisiae*) protein interaction data from MIPS (Munch Information Center for Protein Sequences) CYGD (Comprehensive Yeast Genome Database) at http://mips.gsf.de/genre/proj/yeast/index.jsp [1], into were integrated data from Y2H (Yeast 2 hybrid), TAP (Tandem affinity purification) and immunocoprecipitation experiences. It contains 4610 proteins and 8972 interactions (Aug. 12, 2003) and 8503 complexes (Nov. 17, 2003). We also collected proteome data (amino acid sequences) of *Escherichia coli*, yeast (*Saccharomyces cerevisiae*), *Schizosaccharomyces pombe* and *Arabidopsis thaliana* from NCBI Entrez genome database at http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Genome [7].

B. Methods

We inferred the evolutionary birth periods of yeast proteins and examined relationships between their birth periods and connection degrees. Evolutionary birth periods were inferred by reconstructing phylogenetic profile of yeast with regard to *E.coli*, *S.pombe* and *A.thaliana*.

1) Phylogenetic Profile

Phylogenetic profile is a molecular evolutionary profile which indicates presence/absence of orthologous genes. We employed BLASTP [8] screening on NCBI Entrez genome data

(E-value threshold was set to 1.0×10^{-6}) and finally identfied computational orthologues which have over 60% global similalities by ClustalW [9] multiple alignments.

2) Evolutionary rates

We reconstructed standardized phylogenetic trees [10] to calculate evolutionary rates of yeast genes and then examined relationships between evolutionary rates and connection degrees.

III. Results

A. Evolutionary Old Proteins do not have High Connection Degrees

We first inferred the evolutionary birth periods of yeast proteins by reconstructing phylogenetic profile. Phylogenetic profile is a profile of the presence/absence of orthologous proteins to the correspondent protein. Number of proteins emerged in each evolutionary period is shown in Fig. 1. It showed that number of proteins was increased suddenly both in the common ancestor of eukaryotes and in the lineage of *S.cerevisiae*. It suggested that lots of novel genes were emerged in the common ancestor of eukaryotes to archive the complex and diverse systems of eukaryotes and half of yeast genes were gained in the yeast specific lineage to adopt its specific environment.

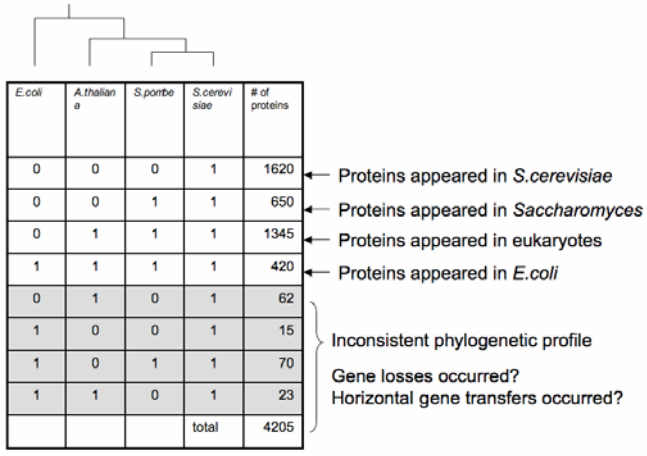

Fig. 1 Number of proteins emerged in each evolutionary period. One(1)/zero(0) indicates the presence/absence of protein in the correspondent species, respectively. In this poster presentation, we excluded inconsistent phylogenetic profile: e.g., proteins whose phylogenetic pattern is "1101" were thought to be lost in the *S.pombe* specific lineage. We considered that inconsistency in phylogenetic profile was caused by both gene losses and horizontal gene transfers

We then examined relationships between the evolutionary birth period of yeast proteins and their connection degrees [Fig. 2]. It has been known that overall distribution of yeast proteins with regard to their connection degrees obeys scale-free distribution. We found that each distribution of proteins whose evolutionary birth periods are the same obeys scale-free distribution.

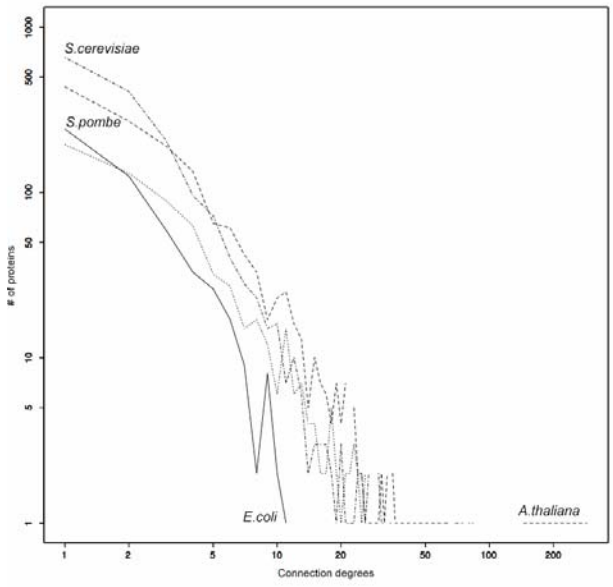

Fig. 2 Relationships between the evolutionary birth period of yeast proteins and their connection degrees. Solid line indicates *E.coli*, dashed line indicates *A.thaliana*, dotted lines indicates *S.pombe* and dotdash line indicates yeast (*S.cerevisiae*)

We also showed that the connection degrees of proteins appeared in the old evolutionary period (*E.coli*) are not high, whereas those of proteins appeared in the second old evolutionary period (*A.thaliana*) are dramatically high. It has been thought that hub proteins that have high connection degree are evolutionary old [5]. Because it has been thought that protein interaction networks evolved by preferential attachments. But our results showed that hub proteins are entirely evolutionary new and evolutionary old proteins do not have high connection degrees.

B. Member Proteins of Complexes Tend to Appear in the same Evolutionary Period

We inferred evolutionary processes of protein complexes by inferring each evolutionary birth period of member proteins of complexes. Relationships between rates of the most/secondary populated evolutionary period were shown in Fig. 3. The most/second populated means the most/second major proteins emerged in the same evolutionary period, respectively. The sum of rates of the most/secondary populated evolutionary period tends to close to 1. It showed that member proteins of complexes tend to appear in the same evolutionary period. For example, majority of member proteins of DNA metabolic complex 410.30 (MIPS complex ID) were appeared in the same evolutionary period [Fig. 4]: 80% of all the proteins were emerged in eukaryotes, 10% were emerged in *Saccharomyces* and the last 10% were emerged in yeast.

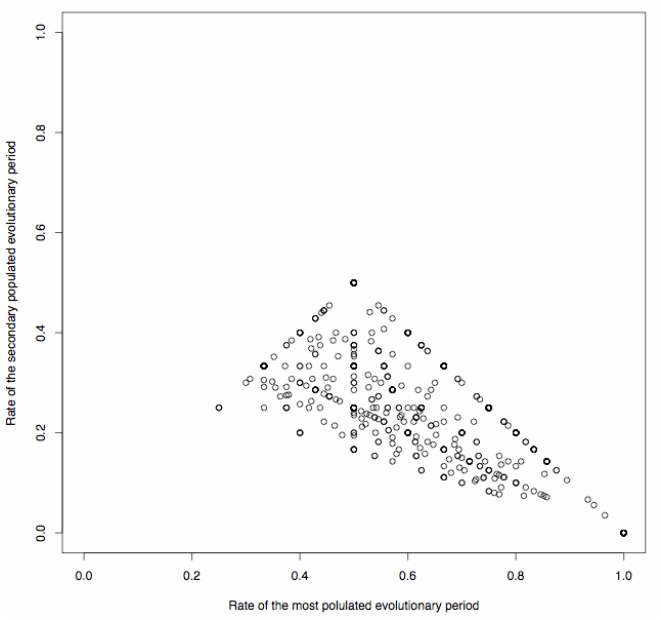

Fig. 3 Relationships between rates of the most/secondary populated evolutionary period. The sum of rates of the most/secondary populated evolutionary period tends to close to 1

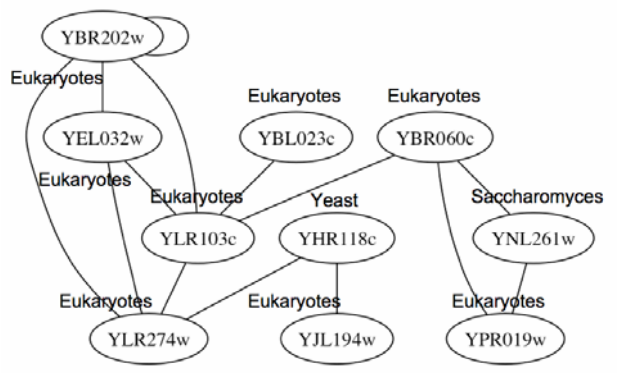

Fig. 4 e.g., DNA metabolic complex 410.30 (in MIPS complex ID) and its member proteins (in ORF name) with the evolutionary birth periods

C. No Obvious Correlation between Evolutionary Rate and Connection Degrees

We reconstructed standardized phylogenetic trees and calculated evolutionary rates of yeast proteins. The average/ standard deviations of branch length were shown in Fig. 5. Distribution of evolutionary rate was also shown in Fig. 5. +3σ, +2σ and +σ indicate that the corresponding branch length was exceeded the average branch length by +3σ, +2σ and +σ (accelerated evolution), where -3σ, -2σ and -σ indicate that the corresponding branch length was below the average branch length by -3σ, -2σ and –σ (decelerated evolution).

Averages

	Yeast	S.pombe	A.thaliana	E.coli
Yeast	0.000	0.481	0.582	0.666
S.pombe	0.481	0.000	0.574	0.659
A.thaliana	0.582	0.574	0.000	0.642
E.coli	0.666	0.659	0.642	0.000

Standard deviations

	Yeast	S.pombe	A.thaliana	E.coli
Yeast	0.000	0.149	0.133	0.112
S.pombe	0.149	0.000	0.139	0.116
A.thaliana	0.133	0.139	0.000	0.121
E.coli	0.112	0.116	0.121	0.000

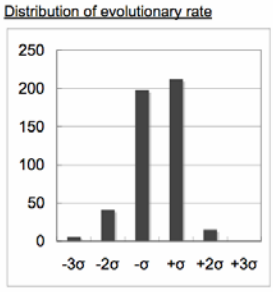

Fig. 5 Average/standard deviations of branch length and distribution of evolutionary rate

We then examined relationships between evolutionary rates and connection degrees of yeast proteins [Fig. 6]. It showed that there is no obvious correlation between evolutionary rates and connection degrees, because distribution of proteins that indicate accelerated/decelerated evolution obeys scale-free distribution.

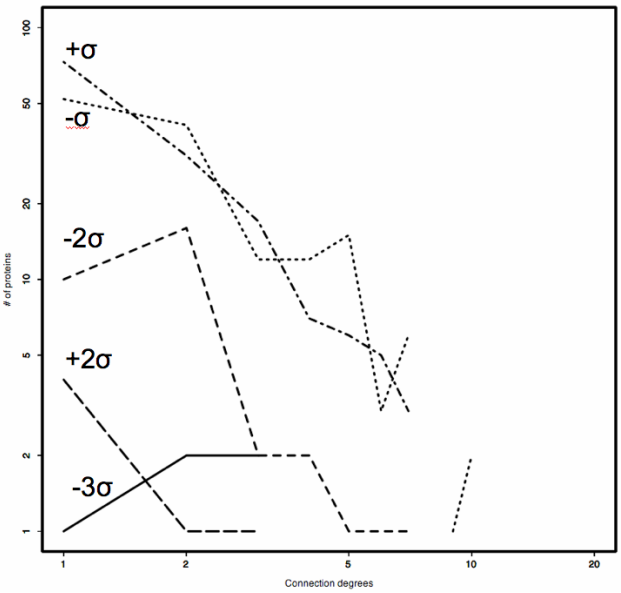

Fig. 6 Relationships between evolutionary rates and connection degrees. Solid line indicates -3σ, dashed line indicates -2σ, dotted lines indicates –σ, dotdash line indicates +σ and longdash line indicates +2σ

We also examined the relationships between evolutionary rates and function of yeast proteins [Fig. 7]. It showed that proteins that related to transportation tend to indicate high evolutionary rates, whereas proteins that related to metabolism tend to indicate low evolutionary rates. It suggested that proteins that related to metabolism were essential in surviving and thus were highly conserved through the evolutionary processes, whereas proteins that related to transportation have roles in intracellular regulation to adapt the extracellular environment and thus were highly diverged to acquire novel functions.

List of proteins whose evolutionary rates were high (+1σ over average)

ORF	Species 1	Species 2	Molecular distance	Average	S.D.	GO(Biological process)
YIL149c	*Yeast*	*S.pombe*	0.847	0.481	0.149	transport
YIL149c	*Yeast*	*E.coli*	0.924	0.666	0.112	transport
YIL149c	*Yeast*	*A.thaliana*	0.859	0.582	0.133	transport
YDL199c	*Yeast*	*S.pombe*	0.841	0.481	0.149	unknown
YMR128w	*Yeast*	*E.coli*	0.904	0.666	0.112	ribosome biogenesis and assembly
YBR041w	*Yeast*	*S.pombe*	0.789	0.481	0.149	transport

List of proteins whose evolutionary rates were low (-2σ under average)

ORF	Species 1	Species 2	Molecular distance	Average	S.D.	GO(Biological process)
YPL135w	*Yeast*	*E.coli*	0.289	0.666	0.112	coenzyme and prosthetic group metabolism
YOR226c	*Yeast*	*E.coli*	0.281	0.666	0.112	coenzyme and prosthetic group metabolism
YJR121w	*Yeast*	*E.coli*	0.314	0.666	0.112	coenzyme and prosthetic group metabolism
YJR009c	*Yeast*	*E.coli*	0.317	0.666	0.112	carbohydrate metabolism
YJL052w	*Yeast*	*E.coli*	0.299	0.666	0.112	carbohydrate metabolism
YGR192c	*Yeast*	*E.coli*	0.311	0.666	0.112	carbohydrate metabolism

Fig. 7 List of yeast proteins whose evolutionary rates were high/low. Evolutionary rates above were calculated between Species 1 and Species 2

IV. Discussions

It was suggested that yeast protein interaction network evolved by modules that form the same functional unit as shown in Fig. 7. The representative functional unit is protein complex. Proteins do not function by themselves, thus proteins should evolve with modularity. It is consistent with the suggestion from the facts that two proteins tend to interact with each other if they are in the same or similar evolutionary categories [6].

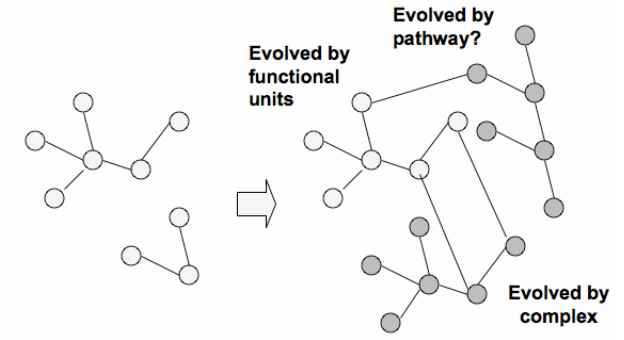

Fig. 8 Schematic illustration modular evolution of yeast protein interaction network

Interestingly, it has been thought that there are two types of hubs from the point of view of temporal behaviors: “party” hubs and “date” hubs [11]. Party hubs interact with most of their partners simultaneously and function inside modules, whereas 'date' hubs bind their different partners at different times or locations and organize the proteome, connecting biological processes or modules to each other. It suggests that party hubs and their party forms modules such as complex or pathways and evolved with modularity, whereas date hubs and their different partners evolved by preferential attachments.

Though there have been controversies on whether there is correlation or not, we revealed that there is no obvious correlation between evolutionary rates and connection degrees. That is, proteins that have high connection degrees do not evolved conservatively, whereas proteins that have low connection degrees do not evolved under positive selection. It suggests that hub proteins are essentially indispensable but are not highly conserved and it does not contradict our finding that hub proteins are not evolutionary old.

References

[1] U. Guldener, M. Munsterkotter, G. Kastenmuller, N. Strack, J. van Helden, C. Lemer, et al., “CYGD: the Comprehensive Yeast Genome Database”, *Nucleic Acid Research,* vol. 33 (Database issue), pp. D364-D368, Jan, 2005.

[2] L. Giot, J.S. Bader, C. Brouwer, A. Chaudhuri, B. Kuang, Y. Li, et al., “A protein interaction map of Drosophila melanogaster”, *Science*, vol. 5, no. 302, pp. 1727-1736, Dec, 2003.

[3] S. Li, C.M. Armstrong, N. Bertin, H. Ge, S. Milstein, M. Boxem, et al: “A map of the interactome network of the metazoan C. elegans”, *Science*, vol. 23, no. 303, pp. 540-543, Jan, 2004.

[4] M. Pellegrini, EM. Marcotte, MJ. Thompson, D. Eisenberg, TO. Yeates. “Assigning protein functions by comparative genome analysis: protein phylogenetic profiles”, *Proceeding of National Academy of Sciences USA*, vol. 96, no. 8, pp. 4285-4288, Apr, 1999.

[5] E. Eisenberg, E.Y. Levanon, “Preferential attachment in the protein network evolution”, *Physycal Review Letter*, vol. 91, no. 13, pp. 138701, Sep, 2003.

[6] H. Qin, H.H. Lu, W.B. Wu, W.H. Li, “Evolution of the yeast protein interaction network”, *Proceeding of National Academy of Sciences USA*, vol. 100, no. 22, pp. 12820-12824, Oct, 2003.

[7] NCBI Entrez genome database (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Genome).

[8] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, "Basic local alignment search tool", *Journal of Molecular Biology*, vol. 215, no. 3, pp. 403-410, Oct, 1990.

[9] W.R. Pearson, D.J. Lipman, “Improved tools for biological sequence comparison”, *Proceeding of National Academy of Sciences USA*, vol. 85, no. 8, pp. 2444-2448, Apr, 1988.

[10] T. Endo, S. Ogishima, H. Tanaka, “Standardized phylogenetic tree: a reference to discover functional evolution”, *Journal of Molecular Evolution*, vol. 57 (supplement 1), pp. S174-181, 2003.

[11] J.D. Han, N. Bertin, T. Hao, D.S. Goldberg, G.F. Berriz, L.V. Zhang, “Evidence for dynamically organized modularity in the yeast protein-protein interaction network”, *Nature*, vol. 430, no. 6995, pp. 88-93, Jul, 2004.