Functional Classification of Coiled-Coil Proteins in Multiple Genomes

Akiyasu C. Yoshizawa[1] acyshzw@kuicr.kyoto-u.ac.jp
Shujiro Okuda[1] okuda@kuicr.kyoto-u.ac.jp
Yuki Moriya[1] moriya@kuicr.kyoto-u.ac.jp
Masumi Itoh[1] itoh@kuicr.kyoto-u.ac.jp
Toshiaki Katayama[2] ktym@hgc.jp
Shin Kawano[1] kawano@kuicr.kyoto-u.ac.jp
Shinobu Okamoto[1] okamoto@kuicr.kyoto-u.ac.jp
Shuichi Kawashima[2] shuichi@kuicr.kyoto-u.ac.jp
Minoru Kanehisa[1,2] kanehisa@kuicr.kyoto-u.ac.jp

[1] Bioinformatics Center, Institute for Chemical Research, Kyoto University
Gokasho, Uji, Kyoto 611-0011, Japan
[2] Human Genome Center, Institute of Medical Science, University of Tokyo
4-6-1 Shirokane-dai, Minato-ku, Tokyo 108-8639, Japan

Keywords: coiled-coil, prediction, gene ontology, ortholog cluster, comparative genomics

1 Introduction

The coiled-coil structure is a typical hyper-secondary structure formed by the mutual intertwining of multiple alpha-helices, and well known for its variety of biological applications, such as in the structure of keratin of the intermediate filament, desmosomal proteins, motor proteins, and some proteins related to the dynamic structure of membranes such as SNAREs. In order to predict the existence of this structure from the amino acid sequence, several algorithms have been proposed: e.g. Lupas algorithm (1982,1991), the pairwise correlation algorithm (1995), and another prediction method using hidden Markov model (2002).

We previously collected complete or draft genomic sequences of more than 160 organisms, compared the numbers of the predicted coiled-coil containing molecules in these genomes, and showed the following [6]: (1) Organisms of some taxa (e.g. clostridia and mollicutes) have more coiled-coil molecules than organisms of other taxa (e.g. actinobacteria, alpha-proteobacteria), (2) in the eukaryotes, both the number and the ratio of coiled-coil molecules show a sudden increase compared to prokaryotes. Here we examined the molecular function of these coiled-coil containing molecules in 197 complete and draft genomic sequences by assigning Gene Ontology terms.

2 Methods

2.1 Prediction Method and Genomic Data

For the prediction of coiled-coil domains, we used three programs, COILS2.2 [2], Multicoil [5], and Marcoil [1], as described in the reference [6]. All the amino acid sequences were obtained from KEGG GENES and KEGG DGENES (Release 32.0).

2.2 Gene Ontology Assignment

The original GO annotation data were obtained from the Gene Ontology Association and NCBI. We used RefSeq to retrieve the associations between GO and LocusLink. Based on these associations, we assigned GO terms to the entries of KEGG GENES. If more than 90% of the high-level terms (“GO_slim”) of the molecular function ontology of the sequences in one of the Ortholog Clusters [3] are identical, we assumed all the sequences

in this ortholog cluster should be assigned the same GO_slim terms [4].

3 Results and Discussion

The ratio of coiled-coil molecules to the whole genome in the eukaryotes is about twice that of those in the prokaryotes. The coiled-coil containing "catalytic" molecules are present in all organisms and the tendencies of their ratio both to the whole genome and to all the "catalytic" molecules are similar (Fig. 1(a)). On the contrary, coiled-coil containing "enzyme regulator" molecules are present mostly in the eukaryotes as shown in Fig.1(b). Interestingly, few prokaryotes that contain them, such as bacillales (Staphylococcus) and chlamydia, are intracellular parasitic.

Generally, the ratio of GO_slim terms assigned to eubacteria is similar to the ratio assigned to archaebacteria. On the other hand, the ratio of eukaryotic "unknown" (could not be assigned by GO_slim terms) molecules is about twice that of the ratio of prokaryotic molecules.

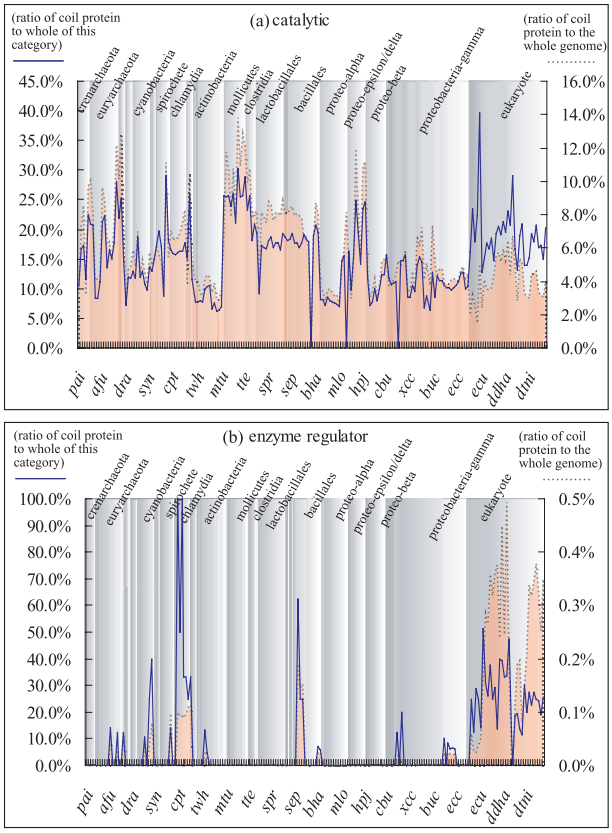

Figure 1: The dotted line indicates the ratio of (a) catalytic and (b) enzyme regulator molecules to the whole genome. The continuous line indicates the ratio of coiled coil molecules in each category.

Acknowledgements

This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology, the Japan Society for the Promotion of Science, and the Japan Science and Technology Corporation. The computational resource was provided by the Bioinformatics Center, Institute for Chemical Research, Kyoto University.

References

[1] Delorenzi, M. and Speed, T., An HMM model for coiled-coil domains and a comparison with PSSM-based predictions, *Bioinformatics*, **18**:617-625, 2002.

[2] Lupas, A., Van Dyke, M. and Stock, J., Predicting coiled coils from protein sequences, *Science*, **252**: 1162-1164, 1991.

[3] Moriya, Y., Katayama, T., Nakaya, A., Itoh, M., Yoshizawa, A.C., Okuda, S. and Kanehisa, M., Automatic generation of the KEGG Ortholog cluster and its assignment to draft genomes, submitted.

[4] Okuda, S., Yoshizawa, A.C., Moriya, Y., Itoh, M., Katayama, T., Goto, S. and Kanehisa, M., Functional category assignment of multiple genomes using ortholog gene clusters, submitted.

[5] Wolf, E., Kim, P.S. and Berger, B., MultiCoil: A program for predicting two- and three-stranded coiled coils, *Protein Sci.*, **6**:1179-1189, 1997.

[6] Yoshizawa, A.C., Okuda, S., Itoh, M., Katayama, T., Kawashima, S. and Kanehisa, M., Prediction and Comparison of Coiled-Coil Proteins in Multiple Genomes, *Genome Informatics* IBSB 2004 Poster Abstracts, 77-78, 2004.