Classification and Analysis of Eukaryotic ABC Transporters in Complete Eukarya Genomes

Yoshinobu Igarashi and Minoru Kanehisa

The ABC(ATP-Binding Cassette) transporters form a major class of active transporters which are widespread in archaea, bacteria, and eukarya. Most of eukaryotic ABC transporters are involved in exporting antibiotic drugs. In this study, first, we identified the eukaryotic ABC transporters in three complete eukarya genomes, *S.cerevisie*, *C.elegans* and *D.melanogaster*. Next, we classified the eukaryotic ABC transporters into the clusters of orthologs and paralogs using the hierarchical cluster analysis. In addition, we built Hidden Markov Models (HMMs) for individual clusters, and identified the orthologs of individual clusters in othrer genomes by using the HMMs.

Keywords: ABC transporter / Hidden Markov Model / Bioinformatics

Accumulation of genome information is increasing at an accelerative tempo, recent years, and already, the genomes of 6 archaea, 28 eubacteria, and 3 eukarya were determined. These sequences facilitate the analysis of the genome comparison, the analysis of evolutionary relationships and the reconstruction of pathways. However, frequently, the annotation of the proteins and genes sequenced by the genome projects are still based on only the simple sequence similarity. The systematic analysis of genome comparison or evolutionary relationships can carry out producing the more detailed annotations which are never obtained only from the simple sequence similarity. Furthermore, such systematic analysis can be used to avoid the wrong annotations which tend to be produced from the evaluation of only sequence similarity.

The structure of a prokaryotic ABC transporter usually consists of three components. A typical transporter consists of two integral membrane proteins each having six transmembrane segments, two peripheral proteins that bind and hydrolyze ATP, and a periplasmic (or lipoprotein) substrate binding protein. Many of the

MOLECULAR BIOLOGY AND INFORMATION — Biological Information Science —

Scope of research

This laboratory aims at developing theoretical frameworks for understanding the information flow in biological systems in terms of genes, gene products, other biomolecules, and their interactions. Toward that end a new database is being organized for known molecular and genetic pathways in living organisms, and computational technologies are being developed for retrieval, inference and analysis. Other studies include: functional and structural prediction of proteins from sequence information and development of sequence analysis tools.

Prof
KANEHISA, Minoru
(D Sc)

Assoc Prof
GOTO, Susumu
(D Eng)

Instr
NAKAYA, Akihiro
(D Sc)

Students:
PARK, Keun-joon (DC)
IGARASHI, Yoshinobu (DC)
KATAYAMA, Toshiaki (DC)
NAKAO, Mitsuteru (DC)
YOSHIZAWA, Akiyasu (DC)
OKUJI, Yoshinori (DC)
KATO, Masaki (MC)
ITOH, Masumi (MC)
LEVCHENKO, Maria (RS)
Research Fellow:
SATO, Kazushige (RF)
HATTORI, Masahiro (RF)

genes for the three components form operons as in fact observed in known archaea and bacteria genomes[1].

On the other hand, in a typical eukaryotic ABC transporter, the membrane spanning protein and the ATP-binding protein are fused, forming a polypeptide with the membrane spanning domain and the ATP-binding domain.

In prokaryotic ABC transporters, the ATP-binding protein component is the most conserved, the membrane protein component is somewhat less conserved, and the substrate-binding component is most divergent in terms of the sequence similarity[2][3]. Therefore, in eukaryotic ABC transporters, it could also be expected that ATP-binding domain is the most conserved domain.

In this analysis, we first searched and compared the eukaryotic ABC transporters in three eukarya complete sequenced genomes, *S.cerevisiae*, *C.elegans* and *D.melangaster*. We identified ATP-binding domains and other domains in the candidate sequences using the hidden Markov model in Pfam5.0. Then, in order to confirm whether these candidate sequences also have the membrane spanning domain simultaneously, we predicted the membrane spanning domains using SOSui transmembrane prediction program[4]. The prediction results of SOSui program were used as the reference for removing false positive eukaryotic ABC transporters from candidate sequences by manual operation. Then, eukaryotic ABC transporters of 24 in *S.cerevisiae*, 46 in *C.elegans*, 50 in *D.melanogaster* were identified (Table 1). We counted alternatively spliced proteins as one entry.

Next, they were classified into orthologs and paralogs from sequence similarity and domain structure according to the hierarchical cluster analysis. These transporters were classified into orthologs and paralogs from the sequence similarity of the ATP-binding domain and the domain structure, i.e., the ordering of the two domains. The resulting set of clusters (families) is shown in Table 1.

Furthermore, hidden Markov models were built using individual clusters, and they were used to search for similar sequences in other genomes in the KEGG/ GENES database.

By the HMM search in bacteria, archaea and eukarya, we identified a specific ATP-binding domain group, whose homologs are found in only plants and fungi. The HMM for this ATP-binding domain group was constructed by the members of *S.cerevisiae* only, and they belong to multidrug resistance family (family 4 in table 1). Each sequence of this family has two ATP-binding domains, and the homologs of C-terminal side ATP-binding domain are found in bacteria and other eukarya, but the homologs of N-terminal side ATP-binding domain are found in only *S.pombe*, *C.albicans* and *A.thaliana*. Therefore, it is possible that N-terminal side of the sequences may have a function special to the medicine tolerance of fungi and plant cells.

Table1: The number of eukrayotic ABC transporters in *S.cerevisiae*, *C.elegans* and *D.melanogaster*

	family	S.cer	C.ele	D.mel
1	p-glycoprotein	4	21	9
2	multidrug resistance associated protein	7	7	14
3	peroxisomal membrane transporter	2	5	4
4	multidrug resistance (white protein homolog)	10	7	14
5	ABC-2 type?	0	5	8
6	miscellaneous	1	2	1
		24	46	50

References

1. Tomii K. and Kanehisa M., *Genome Research*, **8**, 1048 - 1059 (1998)
2. Saurin W. and Dassa E., *Protein Sci*, **3**, 325 - 344 (1994).
3. Tam R. and Saier M.H. Jr., *Microbiol Rev*, **57**, 320 - 346 (1993)
4. Hirokawa T, Boon-Chieng S, Mitaku S, *Bioinformatics*, **14**, 378-379 (1998)