Classification and Analysis of Eukaryotic ABC Transporters in Complete Eukarya Genomes

Yoshinobu Igarashi[1]
igarashi@kuicr.kyoto-u.ac.jp

Daisuke Kihara[2]
kihara@danforthcenter.org

Minoru Kanehisa[1]
kanehisa@kuicr.kyoto-u.ac.jp

[1] Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan
[2] Donald Danforth Plant Science Center, St.Louis, Missouri 63141, USA

Keywords: ABC, ATP-binding cassette, transporter, complete genome

1 Introduction

We performed classification and analysis of the eukaryotic ABC (ATP-Binding Cassette) transporters in *S. cerevisiae*, *C. elegans*, and *D. melanogaster* which are the three eukarya whose genomes were sequenced completely. The ABC transporters form a major class of active transporters which are widespread in archaea, bacteria, and eukarya. These transporters convey a wide variety of substrates, such as peptides, amino acids, simple sugars, sulfate, phosphate and various drugs.

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

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 (Tam and Saier 1993; Saurin and Dassa 1994). 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*. Next, they were classified into orthologs and paralogs from sequence similarity and domain structure according to the hierarchical cluster analysis. 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.

2 Method and Results

First, we collected the sequences annotated as ABC transporters from the eukarya entry in the SWISS-PROT protein sequence database release36.0+.

Next, we searched for candidate sequences of eukaryotic ABC transporters in *S. cerevisiae*, *C. elegans* and *D. melanogaster* in the KEGG/GENES database by SSEARCH using the sequences from SWISS-PROT as queries.

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 (Hirokawa et al 1998). 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 25 in *S. cerevisiae*, 44 in *C.elegans*, 50 in *D. melanogaster* were identified (Table 1). We counted alternatively spliced proteins as one entry.

Finally, 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.

Table 1: The number of eukaryotic ABC transporters in *S. cerevisiae*, *C. elegans* and *D. melanogaster*.

	Family	*S. cer*	*C. ele*	*D. mel*
1	p-glycoprotein	4	19	8
2	multidrug resistance	7	6	14
3	peroxisomal membrane transporter	2	5	4
4	multidrug resistance (and white protein homolog)	10	7	14
5	ABC-2 type?	0	5	8
6	miscellaneous	2	2	1

3 Discussions

By the HMM search in bacteria, archaea and eukarya, we identified a specific ATP-binding domain group, whose homologs are found in S 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.

4 Acknowledgements

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

References

[1] Saurin, W. and Dassa, E., Sequence relationships between integral inner membrane proteins of binding protein dependent transport systems: evolution by recurrent gene duplications, *Protein Sci.*, Feb;3(2):325–44, 1994.

[2] Tam, R. and Saier, M.H. Jr., Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria, *Microbiol Rev.*, Jun;57(2):320–46, 1993.

[3] Tomii, K. and Kanehisa, M., ABC transporters in complete microbial genomes, *Genome Research*, Oct;8(10):1048–1059, 1998.