Comprehensive Survey of Intracellular Transport System-Related Proteins in Complete Genomes and Draft Genomes

Akiyasu C. Yoshizawa acyshzw@kuicr.kyoto-u.ac.jp
Masumi Itoh itoh@kuicr.kyoto-u.ac.jp
Shujiro Okuda okuda@kuicr.kyoto-u.ac.jp
Vachiranee Limviphuvadh limvip@kuicr.kyoto-u.ac.jp
Tadahiko Sakiyama sakiyama@kuicr.kyoto-u.ac.jp
Shuichi Kawashima shuichi@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

Keywords: eukaryotic genome, intracellular vesicular transport, cluster analysis

1 Introduction

The intracellular vesicular transport is an indispensable function for eukaryotic cells to maintain the intracellular membrane compartments. This transport is classified into three processes as (a) budding, (b) moving, and (c) fusion, of the vesicles. It is thought that the processes (a) and (c) are specific to the vesicular transport, and many proteins involved in each of the three processes are already known. Especially in the process (c), the destination of transport vesicles are determined by unique mutual recognition events in order to fuse the vesicle membrane with the correct target membrane, making it possible to transfer macromolecules selectively.

We previously collected SNARE proteins [3] and Rab small GTPases, both of which function in the process (c), from the KEGG/GENES database, and extracted sequence features of those proteins. We also surveyed the numbers of paralogous genes for each of the intracellular positions across the species involved in eukaryotic evolution [2]. As yet, these transport molecules in multiple eukaryotic genomes have been surveyed comparatively and comprehensively only in few studies and in a small number of genomes. For example, only four draft genomes of human, fly, worm and yeast were surveyed comparatively yet [1].

Now we have collected 18 eukaryotic whole genomes, and performed an exhaustive survey of the vesicular transport-related proteins that belong to the process (a) or (c).

2 Methods

2.1 Genomic Data

(1) For the survey of the following seven organisms, we have used the sequences in KEGG/GENES because the whole genomes were already stored in:

Arabidopsis thaliana (ath), *Caenorhabditis elegans* (cel), *Drosophila melanogaster* (dme), *Encephalitozoon cuniculi* (ecu), *Plasmodium falciparum* (pfa), *Saccharomyces cerevisiae* (sce), *Schizosaccharomyces pombe* (spo).

(2) For the following ten organisms, we have used the complete set of predicted ORF amino acid sequences of the whole draft genomes that are publicly available from Ensembl, TIGR, and JGI:

Anopheles gambiae (aga), *Ciona intestinalis* (cin), *Danio rerio* (dre), *Homo sapiens* (hsa), *Mus musculus* (mmu), *Neurospora crassa* (ncr), *Oryza sativa* japonica *Nipponbare* (osj), *Plasmodium yoelii* (pyo), *Takifugu rubripes* (tru).

(3) For the following two organisms, we have used all of the predicted ORF amino acid sequences available in public from Ensembl:

Caenorhabditis briggsae (cbr), *Rattus norvegicus* (rno).

2.2 Searching for Homologs

(1) We searched the homologs of the following molecules involved in the vesicular transport system: (a:budding) AP1/2/3/4, Clathrin, COP-I, COP-II, Retromer, small GTPase (Arf, Arl, Sar); (c:fusion) Sec1/munc18, SNARE, small GTPase (Rab).

(2) The homologs were extracted with the methods as described in the reference [3]. The initial PSI-BLAST search (blastp2.2.4, E-value < 0.01) against 18 genomes was performed using yeast orthologs [1] as quries of the above molecules. Then, the sequence similarity of all PSI-BLAST hits was examined with the SSEARCH 3.4t06 program and the complete-linkage hierarchical cluster analysis was performed using the SSEARCH raw scores.

3 Results and Discussion

Generally, when the intracellular structure is complicated, the sorting process is also complicated, and the number of the molecules involved in the process (c) is likely to grow. The numbers of SNAREs, Rabs, and Sec1s, which are involved in the process (c), grow in this tendency when multiple genomes are compared. Especially in vertebrates, the number grows drastically. On the contrary, there is no apparent difference between the numbers in the unicellular organisms and the multicellular organisms except for Rab GTPases.

On the other hand, because the purpose of the process (a) is just the budding of the membrane of organelles, it is thought that many paralogs are not required in this process. However, the number of the molecules involved in the process (a) also grows in vertebrates.

These results may represent the complexity of the associated control system for the intracellular transport and indicate characteristics of vertebrates in the eukaryotic evolution.

Table 1: The numbers of the paralogs involved in the process (a)(budding) or (c)(fusion).

	pyo	pfa	ecu	ncr	spo	sce	cel	cbr	dme	aga	cin	dre	tru	rno	mmu	hsa	ath	osj
(a)	36	34	18	39	36	42	46	45	46	41	52	78	95	91	96	92	80	109
(c)	30	29	13	34	31	37	65	58	75	66	78	131	158	155	138	150	128	138

Acknowledgments

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

References

[1] Bock, J.B., Matern, H.T., Peden, A.A., and Scheller, R.H., A genomic perspective on membrane compartment organization, *Nature*, 409:839–841, 2001.

[2] Yoshizawa, A.C., Kawashima, S., and Kanehisa, M., Comparative analysis of Rab GTPases and SNAREs in eukaryotic genomes, *Genome Informatics*, 13:404–405, 2002.

[3] Yoshizawa, A.C., Kawashima, S., and Kanehisa, M., The construction of a database on the intracellular vesicular transport, *Genome Informatics*, 11:276–277, 2000.