

Comparative genome analysis of glycan transporters

Yoshinobu Igarashi Shin Kawano Susumu Goto
{igarashi, kawano, goto}@kuicr.kyoto-u.ac.jp

Minoru Kanehisa
kanehisa@kuicr.kyoto-u.ac.jp

Bioinformatics Center, ICR, Kyoto University, Uji, Kyoto, 611-0011, Japan

Keywords: sugar, transporter, channel, post-translation modification, glycosylation

1 Introduction

Genome projects revealed the sequence more than a hundred genomes, including a draft of human genome in the past ten years. Now, as well as a large number of bacterial and archaeal genomes, many higher eukaryote genomes are being deciphered. In order to elucidate the blue print of life from bacteria to higher eukaryotes, one of the major challenges is to identify individual functions of the proteins that are encoded in genomes. In general, proteins are assumed to work only by forming 3D structures after a folding process from linear polypeptides. In fact, real proteins obtain their functions after undergoing a variety of modifications on them. These modifications are called “post-translation modification” which include phosphorylation, methylation and glycosylation, among others. In these post-translation modification systems, glycosylation plays an important role. It is believed that over 50% of all proteins are glycosylated after translation according to SWISS-PROT data. The term “glycosylation” includes modifications by monosaccharides, disaccharides, trisaccharides, oligosaccharides and polysaccharides, and produce glycolipids and proteoglycans, among others as well as glycoproteins. Most of glycans are imported to the cell across the plasma membranes. When cells transport these glycans, they use specific transporters or channels for importing and exporting. This means that transporters and channels are important factors for controlling the kinds and amounts of glycans within cells. Thus, these glycan transporters are potential targets for pharmacogenomics, and classification and evolutionary analysis of these glycan transporters are essential to a fundamental data set for various studies, such as predicting and finding out specific variation of glycans in individual cells and organisms.

In this study, we identified the glycan transporter genes in all genomes in KEGG/GENES database [1] using a data set of known glycan transporters in TCDB (transporter classification database) [2]. We also analyzed the distribution among the three kingdoms and individual evolutionary relationships of these transporters with their substrates.

2 Methods

We extracted glycan transporters data with their substrates from TCDB (<http://tcdb.ucsd.edu/tcdb/>). TCDB is a transporter database based on published literature. The gapped-blast [3] with default parameters was used to detect homologous sequences in the KEGG/GENES sequence database. The precompiled homology database, KEGG/SSDB [4], based on KEGG/GENES was used to identify orthologous groups.

3 Results

We extracted 300 protein sequences that are involved in glycan transporting. These protein sets cover four of five types of transporters in TCDB. These four types of transporters are comprised of beta-barrel type

porins, secondary transporters which are driven by proton/sodium motive force, ABC (ATP-binding cassette) primary transporters which are driven by ATP hydrolysis energy and PTS (phosphotransferase system) group translocators. The repertoires of substrates of these transporter and channels are monosaccharides, disaccharides, trisaccharides, oligosaccharides, polysaccharides and other kinds of sugars. In order to detect the orthologous sequences to the source dataset from TCDB, we detected the top five similar sequences in KEGG/GENES using gapped-blast, and we used these homologous sequences as queries for detecting orthologous sequences, which are defined by bidirectional best hit between different organisms (Table 1). In this table, we can refer to several kinds of previously described facts. For example, many orthologous sequences were found in categories 3.A.1 and 2.A.1 that correspond to ABC transporter and major facilitator superfamily, respectively. Both of them are known as superfamilies which are comprised of large number of genes in bacterial genomes [5]. The category 2.A.7 corresponds to the transporters that are located in subcellular organelle for uptaking sugar-nucleotide compounds. Thus, this category transporter is observed only in eukaryotes. The categories from 4.A.1 to 4.A.6 correspond to PTS families, at the present time, they are found mainly in gram-negative bacteria.

Table1 The number of bidirectional best hit glycan transporter genes in the KEGG/GENES sequence database.

Category of organisms and the number of organisms		TCDB category																											
		1.B.3	1.B.13	1.B.14	1.B.15	1.B.17	1.B.18	1.B.19	1.B.26	1.B.35	2.A.1	2.A.2	2.A.6	2.A.7	2.A.8	2.A.10	2.A.21	2.A.66	2.A.84	3.A.1	4.A.1	4.A.2	4.A.3	4.A.4	4.A.5	4.A.6	8.A.1	8.A.3	
eukaryotes	animals	6	-	-	-	-	-	-	-	-	47	4	-	22	-	-	24	-	-	-	7	-	-	-	-	-	-	-	
	plants	2	-	-	-	-	-	-	-	-	17	-	-	7	-	-	-	-	-	-	2	-	-	-	-	-	-	-	
	fungi	4	-	-	-	-	-	-	-	-	83	-	-	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	protists	5	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
bacteria	gamma	35	6	12	6	-	8	22	4	57	10	115	11	2	-	2	-	14	39	-	355	66	-	4	14	-	16	6	-
	proteo- bacteria	beta	8	-	2	-	-	9	-	-	10	-	10	-	2	-	-	-	-	4	-	133	2	-	-	-	-	-	-
	epsilon/delta	7	-	-	-	-	-	-	-	7	-	4	-	-	-	-	-	2	4	-	8	-	-	-	-	-	-	4	-
	alpha	12	-	-	3	-	2	10	-	10	-	32	2	-	-	-	-	43	2	488	-	-	-	2	-	-	-	8	
	bacillales	13	-	-	-	-	-	-	-	2	-	38	-	-	-	6	-	2	37	-	171	36	-	52	6	-	12	-	16
	firmicutes	lactobacillales	13	-	-	-	-	-	-	-	2	-	16	6	-	-	-	-	28	-	83	6	-	58	8	-	2	-	2
	clostridia/mollicutes	12	-	-	-	-	-	-	-	-	2	-	14	2	-	-	-	-	17	-	128	10	-	6	-	-	2	-	-
	actinobacteria	13	-	-	-	-	-	-	-	-	-	-	14	-	2	-	-	-	-	-	173	19	-	-	-	-	-	-	-
	fusobacteria	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	7	-	-	-	-	-	-	-
	planctomyces	1	-	-	-	-	-	2	-	2	-	-	-	-	-	-	-	-	-	-	4	-	-	-	-	-	-	-	-
	chlamydia	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	8	-	-	-	-	-	-	-	-
	spirochete	5	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	11	2	-	-	-	-	-	-	-
	bacteroid	2	-	-	5	-	-	3	-	2	-	19	-	-	-	-	-	-	9	-	5	-	-	-	-	-	-	2	2
	cyanobacteria	9	-	-	-	-	-	3	-	-	-	-	-	2	-	-	-	-	2	10	-	66	-	-	-	-	-	-	2
green_sulfur_bacteria	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6	-	-	-	-	-	-	-	-	
deinococcusthermus	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	7	-	-	-	-	-	-	-	-	
hyperthermophilic_bacteria	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	38	-	-	-	-	-	-	-	-	
archaea	euryarchaeota	13	-	-	-	-	-	-	-	-	-	18	-	-	-	-	-	8	59	-	112	-	-	-	-	-	-	-	2
	crenarchaeota	4	-	-	-	-	-	-	-	-	-	23	-	-	-	-	-	-	2	-	74	-	-	-	-	-	-	-	-
	nanoarchaeota	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

References

- [1] Kanehisa, M., Goto, S., Kawashima S., Okuno, Y., Hattori M., The KEGG resource for deciphering the genome, *Nucleic. Acid. Res.*, 32:D277-80, 2004.
- [2] Busch, W., Saier, M.H. jr, The IUBMB-endorsed transporter classification system, *Methods Mol. Biol.*, 227:21-36, 2003.
- [3] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* 25:3389-3402, 1997.
- [4] Sato, Y., Nakaya, A., Shiraishi, K., Kawashima, S., Goto, S., and Kanehisa, M.; SSDB: sequence similarity database in KEGG, *Genome Informatics* 12, 230-231, 2001.
- [5] Paulsen, I.T., Sliwinski, M.K., and Saier, M.H.Jr., Microbial genome analyses: Global comparisons of transport capabilities based on phylogenies, bioenergetics and substrate specificities, *J. Mol. Biol.*, 277:573-592, 1998.