A study of glycan evolution by comprehensive analysis of human glycosyltransferases using phylogenetic profiling

Takayoshi Tomono [1]
sj004081@ed.ritsumei.ac.jp

Hisao Kojima [1]
hkojima@fc.ritsumei.ac.jp

Yukako Tohsato [1, 2]
yukako.tohsato@riken.jp

Masahiro Ito [1]
maito@sk.ritsumei.ac.jp

[1] Department of Bioinformatics, College of Life Sciences, Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577, Japan
[2] Laboratory for Developmental Dynamics, RIKEN Quantitative Biology Center, 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe 650-0047, Japan

Keywords: genome-wide analysis, lineage-specific glycan, enzyme localization

1 Introduction

Glycosylation of proteins and lipids is observed in various eukaryotes. An interspecies comparison of glycan structures suggests that related species have analogous glycan structure. It has been predicted that the conserved glycan structures in multiple lineages are involved in important biological functions. Glycans play an important role in cell-cell interactions like signaling and adhesion. They influence pathogenic infections, cancers, and neurological diseases. Glycans are biosynthesized by multiple glycosyltransferases, which function sequentially.

Glycosyltransferases are membrane proteins that are generally localized to the endoplasmic reticulum (ER) and the Golgi apparatus. The non-reducing terminal of glycans is biosynthesized in the Golgi apparatus after the reducing terminal is biosynthesized in the ER. The glycosyltransferases exist as widely conserved or lineage-specific conserved enzymes. These widely conserved glycosyltransferases are usually located in the ER and contain sequence motifs that are conserved across several families.

Ten years have passed since the sequencing of the complete human genome. Moreover, the complete genomes of various organisms have been sequenced and lineage relationships have been revealed by molecular phylogenetic analysis. For example, when sequencing the amphioxus (*Branchiostoma floridae*) genome in 2008, the lineage relationship between cephalochordates (amphioxus) and urochordates (ascidian) was revealed, which indicated that the cephalochordates diverged from a common ancestor of mammals before urochordates' diverging [1]. With the increase in the number of complete genome sequences, genome-wide approaches to analyze protein function are being invented. Of them the phylogenetic profiling [2] predicted genetic co-occurrence among genomes. Phylogenetic profiling compiled data on orthologous genes from other genomes, and converted these data into a binary sequence (phylogenetic profile) based on whether the orthologous genes were present (1) or not (0) in the genomes. In this method, similarity among the profile have been utilized to predict protein-protein interaction, subcellular localization of proteins, metabolic pathways and lineage relationship among organisms.

In this study, we aim to perform genome-wide analyses of human glycosyltransferases, and to elucidate the lineage relationship between cephalochordates and urochordates. For this reason, 186 human glycosyltransferases were analyzed using phylogenetic profiling and clustering method. Our analyses indicated that the non-reducing terminal of glycans was biosynthesized by the newly evolved glycosyltransferases. In addition, it suggested that the urochordates lost some genes like sialyltransferases, which were conserved in metazoa.

2 Methods and Results

2.1 Dataset

The dataset of 186 glycosyltransferases was obtained as described. First, we extracted glycosyltransferase data from UniProt database [3] by keyword "glycosyltransferase AND organism:human AND reviewed:yes". In addition, we obtained data on the glycosyltransferases of the complex glycoconjugate (*O*-glycan, *N*-glycan, Glycosphingolipid, Proteoglycan, and GPI anchor) biosynthesizing pathway from the KEGG PATHWAY database [4]. We excluded the proteins not biosynthesizing complex glycoconjugates from the glycosyltransferase data by using the UniProt and KEGG databases.

2.2 Phylogenetic profiling analysis

The phylogenetic profile of human glycosyltransferases was compiled using the data from 1,356 organisms (Archaea: 68, Bacteria: 915, Eukaryota: 198, Viruses: 175) whose complete genome have been sequenced. First, the data on the presence of orthologous (E-value < e^{-10}) genes were obtained from the GTOP database [5]. Next, the data was converted

into a bit string, containing the characters 1 or 0, which indicated whether orthologous genes are present or not, respectively.

Clustering of 186 human glycosyltransferases was performed using Manhattan distance and Ward's method of phylogenetic profiling. Four clusters were observed, and categorized as Classes 1–4 (Figure 1): Class1 contained 44 glycosyltransferases mostly conserved in deuterostomes, Class2 contained 103 glycosyltransferases mostly conserved in metazoans, Class3 contained 35 glycosyltransferases mostly conserved in eukaryotes, and Class4 contained 4 other glycosyltransferases. The glycosyltransferases evolved from class 1 to 4. Class1 and Class2 glycosyltransferases were mostly localized to the Golgi apparatus, whereas Class 3 glycosyltransferases were mostly localized to the ER. We mapped each class of glycosyltransferases to the glycan biosynthesing pathway (Figure 2). Glycosyltransferase function was in the order of Class 3 to Class 1. Notably, 20 sialyltransferases, which could catalyze the non-reducing terminal sialic acid, were clustered into Class 1. According to the above results, the widely conserved glycosyltransferases (Class 3) catalyzed transfers to the reducing terminal, and only the deuterostomian glycosyltransferases (Class 1) catalyzed transfers to the non-reducing terminal.

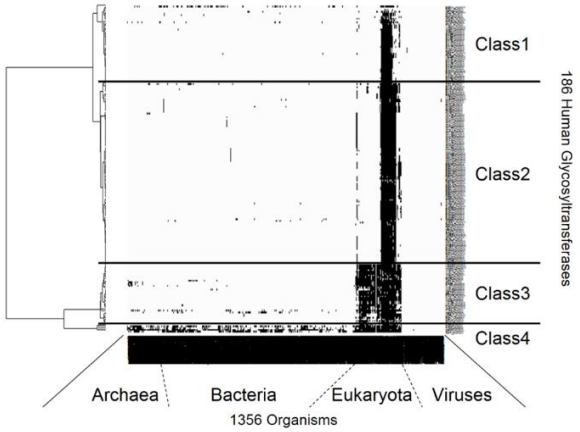

Figure 1: The clustering analysis of 186 human glycosyltransferases. The plotted black point indicates that orthologous proteins were coded in genomes. The branch length of dendrogram tree indicates the distances between glycosyltransferase pairs.

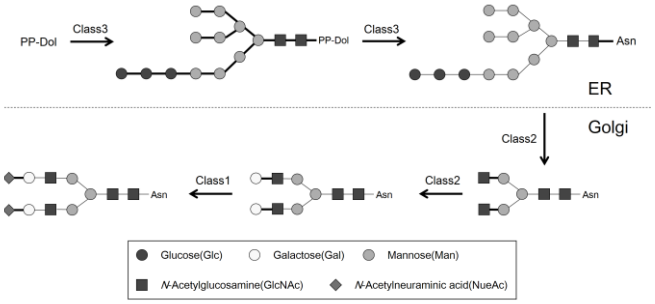

Figure 2: Synthesis of *N*-glycans. Arrows indicate that the pathway was correlated with the specific classes.

The phylogenetic profile of 65 metazoans was extracted from 1,356 organisms. Clustering of data from 65 metazoans was performed using this phylogenetic profile. As a result, deuterostomes and protostomes were roughly clustered. However, ascidians were clustered in the latter group. Class1 glycosyltransferases were hardly conserved in ascidians.

About half of the glycosyltransferases that were clustered into Class1 were sialyltransferases. Therefore, we focused on the evolution of sialyltransferases in metazoans and analyzed its evolution. Sialyltransferases were divided into 4 families: ST3Gal, ST6Gal, ST6GalNAc, and ST8Sia. These were conserved not only in vertebrates but also in some invertebrates. However, ST6Gal, ST6GalNAc, and ST8Sia were not conserved in the ascidians, despite their close relationship to the vertebrates. This result indicated that ascidians lost these sialyltransferases.

3 Discussion

This study compiled the phylogenetic profile of 186 human glycosyltransferases in 1,356 organisms, and performed clustering analysis using this profile. Several lineage-specific glycans were found with class-specific glycosyltransferases. These glycans had important functions, which developed in specific lineages. For example, all sialyltransferases were clustered into Class 1, which was conserved in deuterostomes. Sialic acid containing glycans are related to cell differentiation, signaling, immunity, etc. In the evolution of multicellular animals, these glycans might be important for cell differentiation and tissue development.

In the glycan biosynthesis pathway, glycosyltransferases functioned in the order of Class 3 to Class 1. Thus, it appears that glycosyltransferases act in order from the evolutionarily-old to the new. It suggests that the Golgi apparatus has acquired ability of synthesizing glycan after the ER's doing, because Class 3 is mainly occupied by the ER-localized glycoslytransferases, and Class 1/2 by the Golgi-localized glycosyltransferases.

References

[1] Putnam, N.H. Butts, T. Ferrier, D.E. Furlong, R.F. Hellsten, U. et al, The amphioxus genome and the evolution of the chordate karyotype, *Nature*, 453:1064-1071, 2008.

[2] Pellegrini, M. Marcotte, E.M. Thompson, M.J. Eisenberg, D. and Yeates, T.O., Assigning protein functions by comparative genome analysis: protein phylogenetic profiles, *Proc Natl Acad Sci USA*, 96:4285-4288, 1999.

[3] http://www.uniprot.org/

[4] http://www.genome.jp/kegg/

[5] http://spock.genes.nig.ac.jp/~genome/