Genome-wide analysis of whole human glycoside hydrolases by data-driven analysis *in silico*

Takahiro Nakamura1†, Muhamad Fahmi1†, Jun Tanaka1, Kaito Seki1, Yukihiro Kubota2 and Masahiro Ito1,2\*

1 Advanced Life Sciences Program, Graduate School of Life Sciences, Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577, Japan

2 Department of Bioinformatics, College of Life Sciences, Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577, Japan

**\*** Correspondence: maito@sk.ritsumei.ac.jp; Tel.: +81-77-561-5301 (M.I.)

**†** These authors have equal contribution to this work.

**Supplementary Data 1. Glycoside hydrolase dataset used in this study from Uniprot and its overlapping status with dataset in GO**. The tables show the comparison of dataset retrieved from Uniprot and GO.

**Supplementary Data 2.** **Accession number of the glycoside hydrolase dataset used in this study (UniProt December, 2018).** The table includes the data information of entry name, protein name, and status in the UniProt database, the result of clustering based on the information of the phylogenetic profiling, and the ordered-disordered structure information including percentage of ordered-disordered residue, the absence and presence of continuous stretch of disordered peptide, structure category, and the longest continuous stretch of disordered region (Fig. 1).

**Supplementary Data 3. Organisms list (KEGG June, 2016).** The table shows the list of organisms used for generating phylogenetic profiling.

**Supplementary Data 4. The result of phylogenetic profiling of human glycoside hydrolases.** The column headers indicate the KEGG organism ID (Table S3), and the row headers indicate the UniProt ID of human glycoside hydrolases (Table S2). The 1 and 0 values indicate the presence and absence of organism ortholog glycoside hydrolase to human, respectively.

**Supplementary Data 5. Accession number of the glycoside hydrolase and glycosyltransferase dataset (UniProt December, 2018).** The table shows the information of human glycosyltransferase used in the previous study3 and human glycoside hydrolase used in this study. The table includes the data information of entry name, protein name, and status in the UniProt database, and the result of clustering based on the information of the phylogenetic profiling (Fig. 5a).

**Supplementary Data 6. The result of phylogenetic profiling of human glycoside hydrolases and glycosyltransferase.** The column headers indicate the KEGG organism ID (Table S3), and the row headers indicate UniProt ID of human glycoside hydrolases and glycosyltransferase (Table S5). The 1 and 0 values indicate the presence and absence of organism ortholog glycoside hydrolase to human, respectively.