FlyGlycoDB: a database of glycan-related genes and proteins in Drosophila

Kiyoshi Tadahisa e0656112@soka.ac.jp
Shoko Nishihara shoko@soka.ac.jp
Kiyoko F. Aoki-Kinoshita kkiyoko@soka.ac.jp

Dept. of Bioinformatics, Faculty of Engineering, Soka University, 1-236 Tangi-machi, Hachioji, Tokyo, 192-8577, Japan

Keywords: drosophila, glyco-gene, database, glycobiology

1 Introduction

Glycan structure databases such as GlycomeDB [1] and KEGG GLYCAN [2] have been developed in recent years. Furthermore, databases containing information about genes and proteins in Drosophila such as FlyBase have been established. However, databases pertaining to glycan-related genes in Drosophila do not exist. The glycobiology of Drosophila has been extensively studied in recent years, and much information has accumulated regarding glycan-related genes and proteins in Drosophila. In order to make this data more accessible to the public for bioinformatics analysis, we have developed a web resource of glycan-related genes in Drosophila, called FlyGlycoDB [3]. This web resource is still in the beginning stages, but we hope that it will become useful for glycome informatics research in the future.

2 Method and Results

Because the glycobiology research in Drosophila continues to progress, as a first step, we decided to utilize the data summarized in [4] as the base data set in FlyGlycoDB. After confirmation of this data, we created tables in a mySQL database to organize the information. The tables that we created include information such as gene name, type (lectin, glycosyltransferase, etc.), CG number, mutants, and activities. Many of these data included citations, for which we collected the PubMed IDs individually. For each CG number, we also retrieved the corresponding FlyBase ID [5].

For web access to this data, we have developed web services using REST as well as web interfaces using perl-CGI. These have been organized through a middleware layer for abstraction. The REST interface allows users to search FlyGlycoDB using the tags listed in Table 1.

Table 1: REST query tags for FlyGlycoDB.

Name	Tag	REST URL
Full name	fullname	/search/query+fullname[/offset[,limit]][/ret]
Abbreviated name	abbrevname	/search/query+abbrevname[/offset[,limit]][/ret]
Biomolecule type	type	/search/query+type[/offset[,limit]][/ret]
CG number	CGnumber	/search/query+CGnumber[/offset[,limit]][/ret]
Flybase ID	fbgn	/search/query+fbgn[/offset[,limit]][/ret]
PubMed ID	pmid	/search/query+pmid[/offset[,limit]][/ret]

`query`, `offset`, `limit` and `ret` represent parameters that can be specified by the user. These parameters are explained in Table 2.

Table 2: Variables used in the REST query tags for FlyGlycoDB.

Parameter Name	Description
query	keyword or ID to search for
ret	either `cnt` to return the number of matching entries, or `list` to return the query results as a list
offset	a number indicating to return the entries starting from number `offset`
limit	the number of entries to return

Specific database entries can be retrieved as text or xml using the CG number. The format is as follows: `/entry/CGxxxx.[xml/txt]`.

In addition to the REST interfaces described previously, we have developed various web interfaces such that biologists can query and view entries via the web. A simple query interface lists those entries that match the keyword entered by the user. Furthermore, information for individual entries can be viewed in detail.

3 Discussion

FlyGlycoDB is the first resource to provide easy access to glycan-related gene and protein information in Drosophila. Access to FlyGlycoDB is available both through REST as well as the world wide web. We are currently developing new interfaces such that the data in FlyGlycoDB can be updated with newly-discovered glycan-related information from Drosophila. This will entail checking of the data by internal experts such that the database remains consistent. We hope that this new resource will help push the research in fly glycobiology forward.

References

[1] Ranzinger, R., Herget, S., Wetter, T., von der Lieth, C.-W., GlycomeDB - integration of open-access carbohydrate structure databases, *BMC Bioinformatics*, 9:384, 2008.

[2] Hashimoto, K., Goto, S., Kawano, S., Aoki-Kinoshita, K.F., Ueda, N., Hamajima, M., Kawasaki, T., and Kanehisa, M., KEGG as a glycome informatics resource, *Glycobiology*, 16:63R-70R, 2006.

[3] `http://flyglycodb.t.soka.ac.jp/index.html`

[4] Nishihara, S., *Drosophila Development, RNAi, and Glycobiology* in Comprehensive Glycosciences (Kamerling, J.P., ed.), Elsevier, 2007.

[5] Grumbling, G., Strelets, V., and the FlyBase Consortium, FlyBase: anatomical data, images and queries, *Nucleic Acids Res.*, 34(Database issue):D484-D488, 2006.