

Workflow and Its Navigation System Using Web API for Biology (WABI)

Yeondae Kwon¹
yekwon@lab.nig.ac.jp

Yasumasa Shigemoto¹
yshigemo@genes.nig.ac.jp

Yoshikazu Kuwana¹
ykuwana@genes.nig.ac.jp

Hideaki Sugawara¹
hsugawara@genes.nig.ac.jp

¹ Laboratory for Research and Development of Biological Databases, CIB-DDBJ Center, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan

Abstract

We present an overview of workflow and its navigation system using Web API for biology (WABI) developed by DNA Data Bank of Japan (DDBJ). The system helps non-programming biologists perform analysis tasks which use multiple software tools and databases only by following links on Web browsers. The services are available at <http://xml.nig.ac.jp/> and <http://cyclamen.ddbj.nig.ac.jp/>.

Keywords: Web service, SOAP, REST, workflow, workflow navigation system

1 Introduction

Biologists often use multiple software tools and databases for one analysis task via the Internet. One of the most time-consuming and cumbersome steps in such a task is to process the output of one data source and convert it into a proper input format to a next data source [1]. Especially when accessing heterogeneous Web-based systems, this conversion should be done by manually copying strings to another form on browsers. To ease this burden, DNA Data Bank of Japan (DDBJ) provides an extensive set of Web APIs for biology (WABI) based on SOAP and REST technologies [2]. Using WABI, users only need to define their analysis tasks with some programming language (Perl, Java, C, Ruby or Python), and thus, can avoid manual copy and paste for each analysis step. In addition, typical workflows, combination use of WABI, are provided so that fixed analysis procedures can be carried out without any programming. Non-programming users can execute even other non-predefined workflows by following links on Web browsers. These links are generated by our workflow navigation system which selects next candidate services according to the output of a previously used service. We present an overview of the workflow and its navigation system using WABI. WABI and workflow navigation system are available at <http://xml.nig.ac.jp/> and <http://cyclamen.ddbj.nig.ac.jp/>, respectively.

2 Workflows and Workflow Navigation System

DDBJ currently provides 131 Web APIs, such as keyword search, data retrieval, and homology search, with both SOAP and REST interfaces. A workflow is a predefined task of continuous applications of Web APIs. As an example, we show SNP workflow in Figure 1. SNP workflow first retrieves all human genes, RNA, and amino acid sequences using “NCBIGenomeAnnotation” service in WABI. Next, using “RefSeq” service, group the sequences by whether they have amino acid substitutions or not. Then, for those sequences that have amino acid substitutions, return their 3D structures using GTOP [3], a structure prediction tool provided by DDBJ. From the result of this workflow, users can predict whether positions where amino acid substitutions occur are important from the viewpoint of 3D structures.

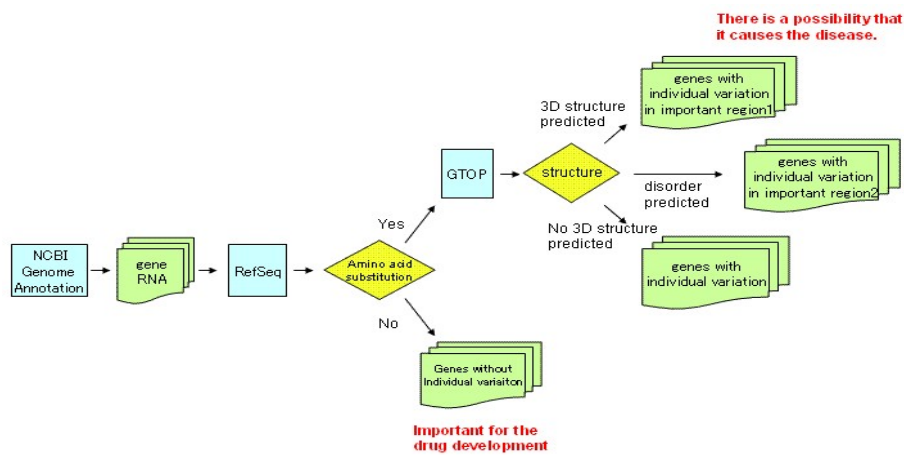


Figure 1: SNP workflow.

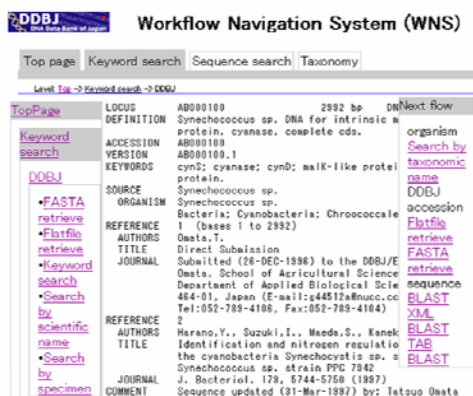


Figure 2: An example of workflow navigation.

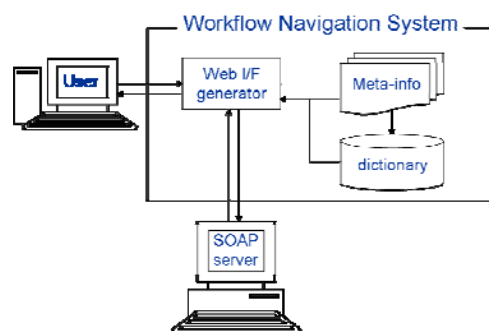


Figure 3: Workflow navigation system.

A workflow navigation system presents users a list of services that are executable as a next step according to results of a service. For example, when sequences are contained in output items, the system suggests all the services that contain sequences as inputs as next services. This function is useful when one does not ensure which services are applicable after performing a particular service or wants to perform analysis tasks without any programming. Figure 2 shows an example of a Web page generated by the workflow navigation system. A list of services are displayed in the rightmost column in Fig. 2 and ranked by service frequency at DDBJ. Users only need to choose one of these services as a next service. Figure 3 shows the system architecture. The system consists of three components: (i) meta information about services such as categories and WSDL locations, (ii) dictionaries on categories, the input/output type of each service (this is used to suggest next applicable services), and HTML tags, and (iii) Web interface generator that generates a Web page from meta information, dictionaries, and SOAP results.

3 Future Developments

Currently, we are developing a communication board for general users which contains tutorials, frequently asked questions, and a wiki tool for knowledge sharing to effectively use our services.

References

- [1] Stein, L., Creating a bioinformatics nation, *Nature* 417:119-120, 2002.
- [2] Sugawara, H. and Miyazaki, S., Biological SOAP servers and Web services provided by the public sequence data bank, *Nucleic Acids Res.* 31(13):3836-3839, 2003.
- [3] Kawabata, T., Fukuchi, S., Homma, K., Ota, M., Araki, J., Ito, T., Ichiyoshi, N., and Nishikawa, K., GTOP: a database of protein structures predicted from genome sequence, *Nucleic Acids Res.* 30(1):294-298, 2002.