

## Poster B-38

### Novel "shim" Service to Develop Flexible Workflows with Web Services



#### Authors:

Toshiaki Katayama (*University of Tokyo*)

Minoru Kanehisa (*Kyoto University*)

**Short Abstract:** It is difficult to develop flexible pipelines for biological analysis only by the methods currently available in the public web services. We developed a new web service containing a collection of "shim" services to process various data types especially for the integration with the KEGG API.

#### Long Abstract:

It is difficult to develop flexible pipelines for biological analysis only by the methods currently available in the public web services. We started to develop a new web service containing a collection of "shim" services to process various data types especially for the integration with the KEGG API.

The KEGG API, which is a SOAP/WSDL based web service, has become an increasingly popular mode of access to the KEGG database. We are continuously adding new features in the KEGG API, however, we also wanted to release experimental version of the KEGG API for obtaining feedbacks from users. For this purpose, we developed the KEGG API-beta, which contains various enhanced methods to utilize KEGG PATHWAY database and other databases in KEGG such as COMPOUND, GLYCAN, REACTION and DRUG.

On the other hand, there are many other web services publicly available such as SOAP interface for NCBI Entrez-utilities, various web services at EBI, and DDBJ XML. Users can integrate these services to their program by writing small amount of codes for building a specific workflow of their biological analysis. However, for making use of them, users need to transform their data to the format the service accepts, parse the result of their methods for the next step, and also need to care about the conversion of the corresponding IDs of the biological objects between the services. This circumstance limits the potential usage of each web service and also prevents the integration of new service that will be needed to obtain complex biological knowledge such as pathway information in the KEGG database. To solve this problem, some applications dedicated for creating a workflow by web services, such as Taverna, are designed to connect various services seamlessly by adding modules to format the previous output for the next input, however, this approach limits others to develop their own program without the application.

For this purpose, we developed a collection of "shim" services to process various data formats and application results in addition to the KEGG API-beta service. This service is hosted at Human Genome Center in the University of Tokyo and utilizes the open source bioinformatics library, BioRuby, which is also developed by us. The BioRuby library is similar to the BioPerl but has strength in handling the KEGG data. In this service, user can retrieve various database entries, parse major database formats and extract information from outputs

of major bioinformatics applications to make it available as the input of the other service. Besides, we also integrated various services provided at Human Genome Center, such as a full-text database search engine HiGet, sequence similarity search system SSS, genome annotation database KEGG DAS and others. This service will be freely available and makes researchers to create their own workflow very easily, especially when combining their analysis with the KEGG system.