

## Web Service Interface for G-language Genome Analysis Environment

Kazuharu Arakawa<sup>1</sup> (gaou@sfc.keio.ac.jp), Nobuhiro Kido<sup>1</sup>, Kazuki Oshita<sup>1</sup>, and Masaru Tomita<sup>1</sup>

<sup>1</sup>Institute for Advanced Biosciences, Keio University, Fujisawa, 252-8520, Japan

G-language Genome Analysis Environment (G-language GAE) is a set of Perl libraries for genome sequence analysis that is compatible with BioPerl, equipped with several software interfaces (interactive Perl/UNIX shell with persistent data, AJAX Web GUI, Perl API). The software package contains more than 100 original analysis programs especially focusing on bacterial genome analysis, including those for the identification of binding sites with information theory, analysis of nucleotide composition bias, analysis of the distribution of characteristic oligonucleotides, analysis of codons and prediction of expression levels, and visualization of genomic information. First version of G-language GAE was released in 2001, and the latest release is currently 1.8.8 (14<sup>th</sup> March, 2009).

In this lightning talk, we highlight the following recent implementations of web-service interfaces for G-language GAE, mainly developed during and after the BioHackathon 2009 to provide higher interoperability:

### **REST Services** (<http://rest.g-language.org/>):

This interface provides RESTful URL-based access to all functions of G-language GAE, which is highly interoperable to be accessed from other online resources. For example, graphical result of the GC skew analysis of *Escherichia coli* K12 genome is given by [http://rest.g-language.org/NC\\_000913/gc skew](http://rest.g-language.org/NC_000913/gc skew).

### **SOAP Services** (<http://soap.g-language.org/g-language.wsdl>):

This interface provides language-independent access to more than 100 analysis programs. The WSDL file contains descriptions for all available programs in a single file, and can be readily loaded in Taverna 2 workbench to integrate with other services to construct workflows.

### **Lightweight Module** (Bio::Glite):

This module is a wrapper around the REST services available at CPAN, with minimal number of external modules for easy installation, and with minimal computational resource requirement.

### References:

1. Arakawa K, Mori K, Ikeda K, Matsuzaki T, Kobayashi Y, Tomita M, "G-language Genome Analysis Environment: a workbench for nucleotide sequence data mining", *Bioinformatics*, 2003, **19**(2):305-306.
2. Arakawa K, Tomita M, "G-language System as a platform for large-scale analysis of high-throughput omics data", *Journal of Pesticide Science*, 2006, **31**(3):282-288.
3. Arakawa K\*, Suzuki H, Tomita M, "Computational Genome Analysis Using The G-language System", *Genes, Genomes and Genomics*, 2008, **2**(1): 1-13.

URL(project): <http://www.g-language.org/>

URL(code): <http://sourceforge.jp/projects/glang/releases/>

License: GNU General Public License