Development and Implementation of the G-Language Genome Analysis Environment Version 2

Kazuharu Arakawa\textsuperscript{1,2}  
\texttt{gaou@g-language.org}  
Ryo Hattori\textsuperscript{1,3}  
\texttt{hanzo@g-language.org}  
Koya Mori\textsuperscript{1,2}  
\texttt{mory@g-language.org}  
Yohei Yamada\textsuperscript{1,3}  
\texttt{skipper@g-language.org}  
Tatekimi Matsuzaki\textsuperscript{1,4}  
\texttt{tero@g-language.org}  
Masaru Tomita\textsuperscript{1,3}  
\texttt{mt@sfc.keio.ac.jp}  
\textsuperscript{1} Institute for Advanced Biosciences, Keio University, Tsuruoka 997-0035, Japan  
\textsuperscript{2} Bioinformatics Program, Graduate School of Media and Governance  
\textsuperscript{3} Department of Environmental Information  
\textsuperscript{4} Department of Policy Management  

Keywords: analysis software, G-language Genome Analysis Environment, workbench

1 Introduction

G-language Project of Institute for Advanced Biosciences aims to introduce higher efficiency in genome analyses by:

1. Constructing an integrated environment for the development of analysis software.
2. Systematically accumulating existing analysis software, methodologies for analysis and their results.
3. Constructing generic analysis packages that allow users to avoid redundancy in the process of analysis.

Development of version 1 of the G-language Genome Analysis Environment (G-language GAE) \cite{1} has reached a stable stage in July 2002, and we are currently implementing version 2 of the software with greater efficiency and flexibility.

2 System Overview

The version 2 core is rewritten to gain further speed and integrity through more flexible object-oriented architecture. Database access drivers, IO controllers, and calculation engines are easily pluggable, and by defining a specific super class, the internal data structures can also be altered. Supported genome database formats are extended to GenBank, Fasta, EMBL, Swiss, PIR, SCF, GCG, Ace, raw, ptt, eri, Qual, Phd, and the XML genome database formats such as GAME, BSML, and EML. Most common relational database systems such as Oracle, MySQL, Postgres, Ingres, Sybase, and Informix can be accessed under unified procedures. The whole system is perfectly compatible with version 1 of the software and Perl \cite{3}, and it can be easily linked to bioperl \cite{2}. An instance of G-language GAE is interchangeable with that of bioperl, and all analysis methods provided by the system is directly accessible with a bioperl instance.

Large scale caching and storage of data internally with relational database reduces the cost of computation and memory, and a structured and persistent data realizes a "session" of analysis that allows continuous analysis through cascades of analysis programs. An instance of G-language GAE can be instantly converted to and stored as a relational database.
The new front-end "Inspire" achieves platform independence and freedom in presentation by making HTML the basic framework of the output. With HTML, anyone can create multi-media presentation with little experience, and an expert will be able to create interactive and dynamic presentation with enhancements with Java, Flash, SVG, and etc. G-language GAE version 2 provides a set of API for the development of SVG, the Scalable Vector Graphics. SVG presentation can be clicked, zoomed in and out, dragged, and animated, and it can of course gain all the merits of Hyper Link to connect to the useful resources online. Inspire front-end runs both locally and on the Web; therefore, an application created for Inspire can be directly converted into a web application. This achieves complete operation system independence. Figure 1 shows the new genome map output in SVG.

Analysis systems concurrently developed with G-language GAE version 2 are the new version of cDNA Analysis System (CASYS: Matsuzaki et al.), Comparative Genome Analysis System (COMGA: Nakamura et al.), Chi Sequence Analysis System (Arakawa and Kyuma et al.), and Bacteria Analysis System (BAS: Mori et al.).

Figure 1: Genome map enhanced with the new front-end, Inspire.

Acknowledgments
This work is supported by the members of G-language Project. URL: http://www.g-language.org/. G-language Project official Web site.

References