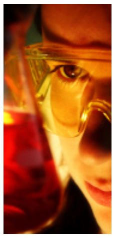
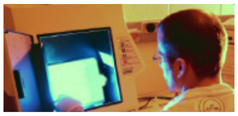
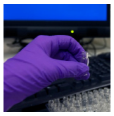






Vol. 1 No. 2
February 2008

Bioinformatics Organization

In this issue

Unveiling Pipet (part one)

J.W. Bizzaro

Pipet was one of the first software development projects to be hosted at Bioinformatics.Org. In fact, the open development model that we used helped us form the mission of the Bioinformatics Organization. And now, having undergone several name changes and a refinement of its technical objectives, Pipet is approaching its public release.

In this article, the first in a series, I'll introduce the system to you.

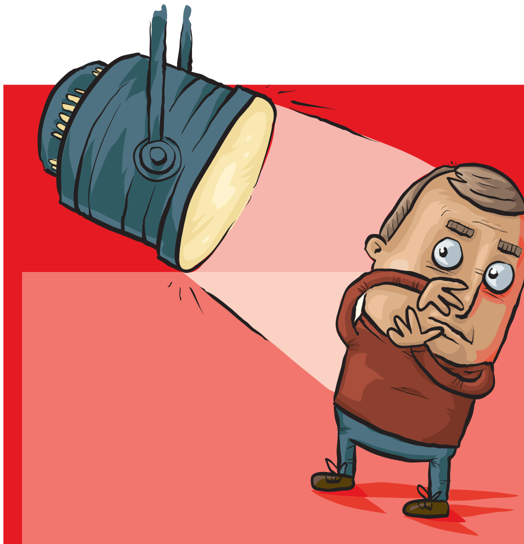
In the broadest sense, Pipet is a World Wide Web system that allows groups of researchers to collaborate on the construction of graphical user interface (GUI) skins for command-line interface (CLI) scripts. This is done by combining elements of visual programming with those as disparate as wiki development, and it has resulted in an analytical platform that can provide a simpler and more robust alternative to conventional website development and Common Gateway Interface (CGI) programming.

One of Pipet's main objectives is to allow the construction of these skins and scripts without the use of a text editor. In terms of implementation, Pipet begins with a generic container for various GUI and CLI components. This container is called a "cell," and it merges the GUI concept of "a container for widgets" with the CLI concept of "a block of code."

In Pipet, cells can be compiled (simply merged) into a single cell while exposing only the parts that should be seen. This process, aptly called "cell compilation" (CC), means that a compiled cell can be shared and used with additional cells, which can then be merged into yet another, single cell, etc., *ad infinitum*. And this lends itself to a very straightforward description via eXtensible

Continued on page 2.

Project spotlight



SAMMD

Staphylococcus aureus microarray meta-database

SAMMD is a relational database that uses MySQL as the back end and PHP/JavaScript/DHTML as the front end. The database is normalized and consists of five tables, which holds information about gene annotations, regulated gene lists, experimental details, references, and other details. SAMMD data is collected from the peer-reviewed published articles. Data extraction and conversion was done using perl scripts while data entry was done through phpMyAdmin tool. The database is accessible via a web interface that contains several features such as a simple search by ORF ID, gene name, gene product name, advanced search using gene lists, comparing among datasets, browsing, downloading, statistics, and help. The database is licensed under General Public License (GPL).

SAMMD is hosted and available at http://www.bifx.org/sammd/. Currently there are over 9500 entries for regulated genes, from 67 microarray experiments. SAMMD will help staphylococcal scientists to analyze their expression data and understand it at global level. It will also allow scientists to compare and contrast their transcriptome to that of the other published transcriptomes

Continued from page 1.

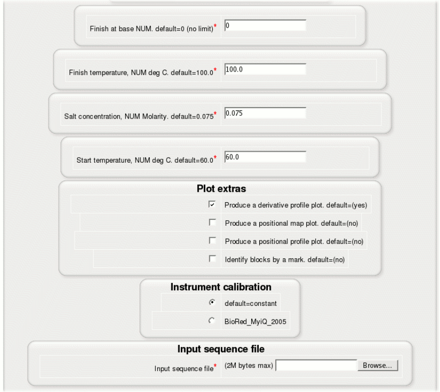
Markup Language (XML). (Next month's article will shed more light on this.)

Another objective for Pipet is to integrate with the existing Web resources at Bioinformatics.Org. This will allow life scientists to maintain accounts where Pipet cells, as well as the resulting data from analyses, can be stored and shared with other account holders—and everyone else in the world if they choose. Plus, cells are organized into "clusters," which can also serve as research groups.

Pipet doesn't come with its own CLI tools for bioinformatics. There are plenty to be found on the Internet, and we see no need to reinvent the wheel here (or thousands of wheels). And this gives us lots of options for expansion and customization. For one, cells for EMBOSS can be combined with those for statistics, physics, chemistry, etc.

Also, Pipet doesn't need to directly manage how analyses are queued and balanced. It can pass scripts along to third-party batch and workload management systems while simply providing information on the status of an analysis.

In next month's article, I'll detail cell creation and compilation. In the meantime, we invite people to help construct cells and test Pipet. If you'd like to help, please contact me at jeff@bioinformatics.org



Finish at base NUM. default=0 (no limit)* 0

Finish temperature, NUM deg C. default=100.0* 100.0

Salt concentration, NUM Molarity. default=0.075* 0.075

Start temperature, NUM deg C. default=60.0* 60.0

Plot extras

☑ Produce a derivative profile plot. default=(yes)
☐ Produce a positional map plot. default=(no)
☐ Produce a positional profile plot. default=(no)
☐ Identify blocks by a mark. default=(no)

Instrument calibration

● default=constant
○ BioRed_MyiQ_2005

Input sequence file

Input sequence file* (2M bytes max)

Screenshot of a cell created for the CLI version of MeltSim (a DNA denaturation simulator)

Job search

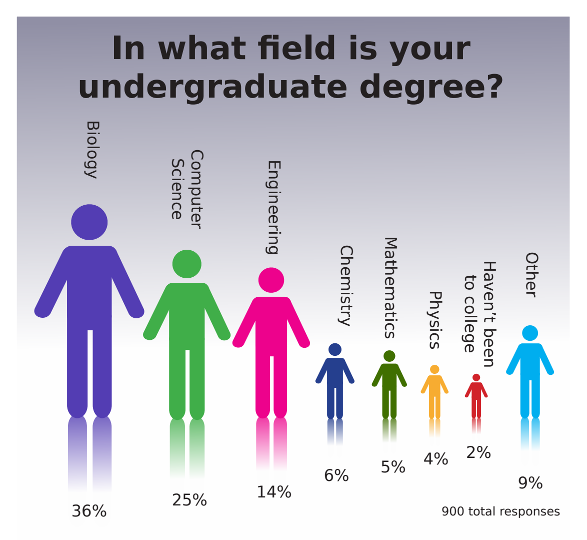
Looking for a job in the bioinformatics field?

We have an open position for a Software Engineer at dictyBase, the model organism database for Dictyostelium discoideum (http://www.dictybase.org). The database consists of a small group of bioinformatics engineers and scientific curators who work closely together in a stimulating environment to develop a leading bioinformatics resource to maintain and distribute the genome of Dictyostelium. We are located at Northwestern University in downtown Chicago, IL, USA.

We're looking for an experienced software engineer to build and maintain bioinformatics applications. The developer will work on a team with life science investigators and other engineers to develop applications that incorporate data from sequencing centers, stock centers, public biomedical databases, and analyses from bioinformatics toolsets. The responsibilities include analyzing requirements for the different projects, identifying technical and functional requirements, and designing and building the systems based on those specifications.

The software engineering team at dictyBase is highly collaborative and committed to Agile Software practices. We are especially interested in applicants who are team players with a demonstrated commitment of software practices such as unit testing, functional testing, build management and excellent documentation skills

Full announcement:
http://www.bifx.org/forums/forum.php?forum_id=6134



Robert Gentleman: the 2008 Benjamin Franklin Award laureate

The Bioinformatics Organization is proud announce that the members have chosen, by vote, Robert Gentleman of the Fred Hutchinson Cancer Research Center as the laureate for this year's Benjamin Franklin Award in the Life Sciences.


Photo by Todd McNaught

In the words of his nominators, Robert is one of the minds behind R (http://www.r-project.org/), a powerful suite of statistical tools with an ever increasing following. And, together with other core members, he founded and developed BioConductor (http://www.bioconductor.org/), an open-source and open-development software project for the analysis and comprehension of genomic data. BioConductor is one of the most popular open-source suites of tools for genomic data analysis. More importantly, Robert has a strong ethical view on the meaning of publishing data, with an emphasis on sharing data-transformation methods as well as the underlying data.

The ceremony for the presentation of the Award will be held on April 29 at the 2008 Bio-IT World Conference & Expo in Boston (http://bio-itworldexpo.com/). It involves a short introduction, the presentation of the certificate, and the laureate seminar

Upcoming events

MySQL for Biologists, Level 1, Monday, 31 March » Friday, 4 April
- Taught in the context of biological research, this course helps biologists learn how to use the database management system, MySQL to store, manage and query biological information.

http://wiki.bioinformatics.org/CS101C_MySQL_for_Biologists,_Level_1

Perl for Biologists, Level 1; Monday, 7 April » Friday, 11 April
- Taught in the context of biological research, this course helps biologists learn how to use the scripting language Perl to automate certain tasks.

http://wiki.bioinformatics.org/CS101A_Perl_for_Biologists,_Level_1

R for Biologists, Level 1; Monday, 14 April » Friday, 18 April
- Taught in the context of biological research, this course helps biologists learn how to use the statistical scripting language R for data analysis. Unlike Bioinformatics Organization course MA101A (Biostatistics with R, Part 1), this course focuses on the syntax and functionality of the language R and may be used in preparation for MA101A.

http://wiki.bioinformatics.org/CS101B_R_for_Biologists,_Level_1

Bio-IT World Conference and Expo; Boston, MA; Monday, 28 April » Wednesday, 30 April
- Location of the Bioinformatics.Org Annual Meeting and presentation of the Benjamin Franklin Award. This premier event demonstrates how technology is transforming the life sciences throughout all discovery and development processes, including over 100 speakers. In addition to affording you access to these high quality talks this event will include an exhibit hall comprised of over 50 booths providing another great way to maximize your participation.

http://www.bio-itworldexpo.com/

Algorithms in Molecular Biology (ALBIO'08); Technical University of Vienna, Austria; Monday, 7 July » Wednesday, 9 July
- Computational Molecular Biology has emerged from the Human Genome Project as an important discipline for academic research and industrial application. The exponential growth of the size of biological databases, the complexity of biological problems and the necessity to deal with errors in biological sequences, result in time efficiency and memory requirements. The development of fast, low memory requirements and high-performances algorithms is thus increasingly important in Computational Molecular Biology.

http://www.birdconf.org/

Bioinformatics™
Organization
17 Pope Street
Hudson, MA 01749
USA