

ACS Technology Project Report Genomics Tools Online

Laurie J. Heyer and A. Malcolm Campbell
Davidson College

Our ACS Technology grant funded three separate projects in computational genomics: microarray data analysis software (MAGIC Tool), online clustering, and online phylogenetic tree analysis. This is an interim report, describing the completion of one of the three projects, progress on the other two projects, and plans for completion of all three projects during this academic year. We have built a home page for our work on this grant (http://gcat.davidson.edu/Online_Genomics/Online.html), which we will continue to update as we make further progress.

MAGIC Tool

MicroArray Genome Imaging and Clustering (MAGIC) Tool software was completed during the summer of 2003. MAGIC Tool facilitates microarray data analysis, beginning with TIFF images, and continuing through several layers of exploratory analysis, including clustering. The program is written in Java, so it is easily used on all major computer platforms. Executable files and source code files are available for downloading from www.bio.davidson.edu/magic. Under the GNU Public License for open source software, which we have applied to MAGIC Tool, all users are welcome to make changes to the code. Therefore, other faculty who wish to investigate imaging or clustering algorithms with their students (or for their research) can easily plug their own algorithms into the MAGIC Tool platform. The MAGIC Tool web site will serve as a repository for users to share their additions and modifications to the code.

Also on the MAGIC Tool web site, we provide several sample files (TIFF images, corresponding gene lists, gene “info” database files, and publicly available gene expression data) that users can download and use for practice as they learn how to use MAGIC Tool. We also provide several levels of documentation, including a brief tutorial, a User’s Guide containing complete software instructions, and an Instructor’s Guide with detailed descriptions of algorithms.

Malcolm will be using MAGIC Tool in his Genomics, Proteomics and Bioinformatics class this fall, and Laurie will use it in her Computational Biology class in the spring. Both classes already have a microarray component; giving students hands-on experience with data analysis will significantly enrich these courses. We have disseminated the software through the following: a short course for biology faculty at the Association for Biology Laboratory Educators (ABLE; Las Vegas; June 2003), a two day workshop for California State University system faculty (Sacramento; June 2003), a two day workshop for Eastern Michigan University’s new Bioinformatics faculty (June, 2003), and a workshop for Genome Consortium for Active Teaching (GCAT) faculty (Seattle; August 2003 www.bio.davidson.edu/projects/GCAT/workshop.html). Another workshop is scheduled for October of this year in Chapel Hill, NC. A manuscript describing MAGIC

Tool will be submitted to *Bioinformatics* within the next few weeks, and we still anticipate writing a second paper on pedagogical uses of the software.

GCAT OSX Server

Using Davidson College funds, we purchased a Mac OSX Xserver with RAID and a total hard disc capacity of 1.25 terabytes. We have fully installed the server gcat.davidson.edu in a closet that is kept cooler than most rooms in the building. We continue to improve the functionality of this server and will increasingly rely upon it to provide flexibility and robust capacity for student projects and course work (see below).

Online Clustering

Computational Biology student projects for online clustering (see, for example, www.bio.davidson.edu/courses/compbio/jas/home.htm) are the basis for this project. We were having technical difficulties with these web pages because of their reliance on the graphing program Gnuplot. Although Gnuplot is written for Windows, the Windows 2000 server that hosts these pages occasionally fails to run the program properly, resulting in broken pages. Gnuplot failures have been intermittent and unpredictable, creating a frustrating situation for faculty who wish to use these pages for instructional purposes. Our proposal for this project was to find new ways to graph results on the Windows server that would work consistently. However, since the proposal was written, we were able to obtain funding for a Mac OS X Server (see above). We are now collaborating with a Davidson alumnus to install Gnuplot and host these pages on the new server. We believe that because Gnuplot is better suited to a Unix type environment, that the graphing will be stable under Mac OS X. Once the pages are working consistently, we plan to enhance them with tutorials and instructors guides, facilitating their use by faculty in their courses. We anticipate finishing this project early in the spring semester.

Phylip

This project was to develop a web page for phylogenetic tree analysis. We planned to have sample files that users can ask to be analyzed with Phylip, a publicly available package for phylogenetic tree construction, and a short tutorial for faculty and students new to this computational genomics technique. We have familiarized ourselves with Phylip, and worked with a student to develop a Perl script for running Phylip from the web. The web interface for this Perl script is still under development. It will be hosted on the GCAT server described above. We expect to complete this project later in the spring semester.

10/13/2003