Genomics Tools Online
ACS Mellon Technology Fellowship Application

1. Background (Rationale for overall project)

Genomics is the hottest new field in biology. Furthermore, genomics is an interdisciplinary field that includes math and computer science as well as chemistry, physics and engineering. In short, there is a growing demand by teachers and students to learn more about genomics.

Laurie Heyer (Math Dept., Davidson College) and I have recently written the first textbook in genomics (*Discovering Genomics, Proteomics and Bioinformatics)* which was published jointly by Cold Spring Harbor Laboratory Press and Benjamin Cummings. Our book uses many well-documented teaching strategies: case-based learning for context, hands-on activities, and metacognitive questions for higher-order learning (McNeal and D'Avanzo, 1997; Allen and Duch, 1998; Uno, 1999; Novak et al., 1999; Astin and Astin, 2000). Therefore, we have provided a text-based approach of learning many aspects of genomics. Part of our methodology is to use "Discovery Questions" that lead students to databases designed for genomics researchers. However, databases designed for researchers are not ideal tools for beginning students. Therefore, we have identified a pedagogical need which is the focus of this proposal.

2. Description: Clustering Data – from data organization to analysis

Laurie and I propose to develop three computer tools for students and teachers to use when learning genomics. One tool will be an online resource for phylogenetic analysis of protein or DNA sequences. The second tool will be an online tutorial explaining how DNA microarray data are analyzed and clustered. The third tool will be free software for use for real DNA microarray data analysis. We will briefly describe all three projects below.

Phylogenetic Clusters

There are many programs available for downloading that allow users to compare sequences and produce a phylogenetic tree. However, each software package has its own platform requirements and generally are not amenable to teaching situations. Our first project will be to develop a web site that will allow students and teachers to submit sequences of their choice (with an upper limit such as 20,000 characters), align the sequences and produce either a rooted or unrooted phylogenetic tree. The program we will employ, PHYLIP, is free to academic users. In addition, we will produce online tutorial and instructions for new users. Laurie will be primarily responsible for managing the software side while Malcolm will handle the front end web page and tutorial/instructions. This work will be an extension of an independent study project being conducted during the fall 2002 semester.

DNA Microarray Tutorial

Laurie's Computational Biology class, and a subsequent independent study student, have established a prototype online tutorial web site for microarray tutorials. However, the site is buggy in that it often fails due to programming weaknesses. In addition, the site is not "lean and mean" to allow users to get at the heart of the matter quickly. Laurie will improve the software side while Malcolm will improve the user interface.

DNA Microarray Software

The largest project will be MAGIC Tool, *M*icro*A*rray *G*enome *I*maging and *C*lustering Tool. This

alpha version software was written by several of Laurie's summer students. It is written in Java and works on all computer platforms with a Java VM (e.g. Linux, Mac OSX, Mac OS 9.x, Windows). The final version of MAGIC Tool will begin with TIFF files and allow users to find spots, measure signal, determine ratios, cluster multiple datasets and interact with the dynamic displays. MAGIC Tool will require substantial modifications from its current version as well as a user manual. In addition, we will collect a number of pedagogically ideal datasets for instructors who want to cover the topic but cannot produce their own data.

3. Timeline (Deliverables/Milestones)

All three projects will have working drafts by May 2003. Laurie and Malcolm intend to spend a large portion of their summers working out kinks and producing user manuals and course modules. We will begin with the Phylogenetic tree because that one should be the easiest. By June 5, we should have a version ready for posting online. The microarray tutorial will require a lot of manipulations, and some reworking of the code. We will target June 30 as our date for completion of this project. Finally, we will complete the MAGIC Tool project by August 1.

4. Technology Required

The only technological issue will be the Microarray Tutorial. The program used to generate the graphs was written for Unix and does not work well on the Windows web server. Ideally, we would be able to host the program on a Mac OSX machine, but currently we do not have one on campus that is available to us. Therefore, we will investigate new ways to generate the graphs on the PC server. The other two projects require computers that are already up and running on campus and the software is already purchased or downloaded. Laurie and Malcolm have the necessary computer skills to perform all the necessary tasks.

5. Other support

MAGIC Tool was supported in part by an NSF grant awarded to Laurie and Malcolm. This paid for the summer salaries of several students.

6. Learning outcomes: How it will enhance teaching/learning?

Each project will have its own impact on student learning.

The Phylogenetic Tree Project will provide two substantial improvements in our educational offerings. First, students will have an interactive way to learn how phylogenetic trees are produced. Currently, all students can do is examine trees built by others – they cannot manipulate datasets to see what effects their manipulations have on the phylogenetic tree output. With our proposal, students will be able to submit data for analysis. By carefully designing pedagogically valuable data, students will be guided through various features in order to understand the software better. Finally, students will be able to submit any sequences of their own choosing for analysis.

The Microarray Tutorial will significantly enhance what is currently available. To our knowledge, we have the only publicly available web site where users can cluster data in real time. The buggy output needs to be fixed and the accompanying text needs reworking. However, once these aspects are improved, students and teachers will be able to learn about microarray data analysis through hands-on activities. We

will create ideal course modules that include datasets and problems that illustrate key points in this process. Currently, there is nothing like it available on the web. DNA microarrays are especially popular tools and many teachers in a wide range of classes will find this project beneficial.
MAGIC Tool software will provide high powered software to any academic user. This integrated program is unique for many reasons. 1) It is written in Java so it works on all platforms. 2) It combines in one program all the tools necessary to being with raw data (TIFF files of scanned microarrays) and finish with clustered data from complex datasets. 3) MAGIC Tool was written with students in mind. MAGIC Tool has fewer bells and whistles than commercial software which tend to overwhelm and confuse novices. 4) MAGIC Tool will be distributed to academic users free of charge. Although a few other analysis tools are freely available, they do not do as much as MAGIC Tool and they do not work on all platforms. Commercial products start at about $5,000 and go as high as $500,000 – no kidding! With MAGIC Tool, students and teachers will be able to analyze data for course work or research which is alter the terrain of undergraduate access to microarray technology. Finally, we will be very responsive to feedback and intend to update MAGIC Tool to meet the needs of the users.

7. Curriculum: How it will be integrated into the curriculum?

Biology courses such as Genetics, Cell Biology, Development, Evolution, Ecology, Molecular Biology and Genomics could incorporate some or all of the tools we propose to develop. In addition, these tools could be used by mathematics/ computer science courses such as the Computational Biology course taught by Laurie. Statistics, Java programming, and Mathematical Modeling are potential sources for further interdisciplinary use of our three proposed projects.

8: Assessment

As co-Editor-in-Chief of the new online journal *Cell Biology Education*, Malcolm is well aware of the importance of assessment in educational projects. To this end, we propose several levels of assessment. First, we will require all users who download MAGIC Tool to register with us so we can track its popularity and contact users for evaluation. Second, students will post online portfolios of their work to reveal their intellectual growth. In addition, we will compile emails where students submit their self-evaluation and perceptions of genomics; these will not be posted online. Third, students will be given assignments that encourage creative uses of lessons learned in order to demonstrate degree of learning and skill acquisition that is extends well beyond simple rote memorization. Finally, we will measure particular learning goals by giving students pre- and post-surveys that are not graded but will reveal changes in knowledge. All our assessment tools will be available from ACS faculty upon request.

9. Dissemination

We have several means for dissemination. First, we will present MAGIC Tool to a number of workshops over the following years. For example, we have an NSF grant proposal due in October to fund two summers of workshops coordinated through the Genome Consortium for Active Teaching (GCAT; www.bio.davidson.edu/GCAT) which Malcolm directs. GCAT members will be contacted by email and encouraged to use the resources we propose. We will write at least two papers; one will describe MAGIC Tool and be submitted to an appropriate journal (e.g. *Genomics* or *Bioinformatics*). The second paper will focus on the pedagogical issues outlined in this proposal and will be submitted to *Cell Biology Education*.

As Malcolm has done for past ACS projects, we will email all biology and math departmental chairs in ACS schools to notify them of our projects and encourage these chairs to forward the email to all departmental members. Finally, Laurie and Malcolm are interested in offering a summer workshop for ACS faculty to learn how to use the tools we develop. If ACS is interested, we would like to offer a three day course at the ACS facility at Southwestern University during the summer of 2004. Ideally, participants would come in pairs from their home institutions with at least one biologist and one mathematician/computer scientist.

References

1. Allen, Deborah E. and Barbara J. Duch. 1998. Thinking Toward Solutions: Problem-based Learning Activities for General Biology. Saunders College Publishing.
2. Astin, Alexander W. and Helen S. Astin. 2000. Leadership Reconsidered: Engaging Higher Education in Social Change. W.K. Kellogg Foundation.
3. McNeal, Ann P. and Charlene D'Avanzo. 1997. Student-Active Science: Models of Innovation in College Science Teaching. Saunders College Publishing.
4. Novak, Gregor M., Evelyn T. Patterson, Andrew D. Gavrin, and Wolfgang Christian. 1999. Just-in-time Teaching: Blending Active Learning with Web Technology. Prentice Hall.
5. Uno, G. 1999. Handbook on Teaching Undergraduate Science Courses: A Survival Training Manual. Thomson Custom Publishing.