

# TEACHING TOGETHER: A THREE-YEAR CASE STUDY IN GENOMICS\*

*Mark D. LeBlanc and Betsey D. Dyer*  
*Computer Science and Biology*  
*Wheaton College*  
*Norton, MA 02766*  
*mleblanc@wheatoncollege.edu*  
*bdyer@wheatoncollege.edu*

## ABSTRACT

The emerging field of "Genomics" (the analysis of DNA sequences) requires interdisciplinary collaborations between computer scientists and biologists. Whereas most colleges and universities do not and may not ever have full programs in bioinformatics/genomics, new models of teaching and associated supporting materials are needed if we are to provide undergraduate science majors with experiences in the art of cooperation necessary to solve tomorrow's problems in genomics. We report here on our recent three-year experiences with "linking" computer science and biology courses. By infusing genomics into existing computer science and biology courses ("Algorithms," "Software Engineering," "Genetics," and "Cell Evolution"), interdisciplinary classes and labs in genomics are provided to all computer science majors and over half of all majors in the biological sciences (biology, biochemistry, environmental science, and psychobiology).

## 1. BIO MEETS BIG OH

It is a wonderful time to learn to be a computer scientist. The paradigm shift of genomes as data is the newest science to appreciate that very little of the work can be done without computers. Bioinformatics, including the subset focused on DNA or genomics, has in the last five years emerged as a new discipline, at a faster rate than for example the emergence of

---

\* Copyright © 2003 by the Consortium for Computing in Small Colleges. Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the CCSC copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Consortium for Computing in Small Colleges. To copy otherwise, or to republish, requires a fee and/or specific permission.

electrical engineering. For computer science faculty, biology is the new(est) client and the solution of problems in bioinformatics demand a serious understanding of computing. Biology needs Big Oh and more. Can we deliver?

## **2. SOFTWARE TEAMS**

The next generation of genomic analyses will require the continual development of new interdisciplinary approaches. Whereas today's software is intended for today's problems, it is necessary but not sufficient for biologists to learn to use existing pieces of software such as BLAST, the sequence alignment tool. There is an urgent need for interdisciplinary work on cataloguing and annotating genomic data (Brown, 1999; Colwell, 2002). Biology students need opportunities to ask original questions and participate in algorithm and software design in order to learn how to set up new genomics projects with programmers. Likewise, programmers entering the field of genomics must have a richer facility with the types of analyses and hypotheses that are useful in genomics. To address these needs, our goals and objectives are to:

1. infuse genomics into the biology and computer science curricula via strategic, yet flexible scheduling of "linked" courses ("Algorithms" or "Software Engineering") with ("Genetics" or "Cell Evolution");
2. increase students' awareness of and appreciation for the challenges and rewards of interdisciplinary research;
3. increase students' knowledge of bioinformatic/genomic data, applications, and potential areas of research;
4. expose students to areas of research through these course experiences in order to create a pipeline for future research interest and work;
5. design, evaluate, and disseminate course materials for an alternative teaching/learning model (not just biologists using software; not just computer scientists writing software) where the art of collaboration is demonstrated and practiced.

## **3. IMPLEMENTATION**

Our goal of infusion is intended to both complement and extend full programs in bioinformatics that are beginning to emerge (Berkowitz, 2002) at some colleges and universities (e.g., Rensselaer Polytechnic Institute, Wright State University, Canisius College, Ramapo College). For most colleges, especially small liberal arts colleges like our own, the infusion of genomics will reach far more students than would a new major. The suite of courses for a bioinformatics major would attract a very small number of students, some of whom are likely to switch out of already small and rigorous programs, e.g., biochemistry. In the pilot iterations of our model ("Algorithms" and "Genetics" [twice]; "Algorithms" and "Cell Evolution" [twice]), we have reached 100% of the computer science majors of a given class year ("Algorithms" is

a required course) and up to 70% of all majors in the biological sciences (biology, biochemistry, environmental science, and psychobiology).

We define "linked" courses as two independently run courses that share genomics as a common thread in the respective syllabi and that share time in the form of guest lectures, some common lab sessions (e.g., four out of 12 labs over the semester), collaborative programming assignments outside of lab time, and final interdisciplinary team projects and presentations. Linked courses offer faculty a flexible way to infuse genomics content into appropriate courses, gain the benefits of interdisciplinary experiences, but still maintain control of most topics in the syllabus. For example, both "Genetics" and "Algorithms" are core courses in small departments with a considerable amount of "traditional" material that must be covered. Like Goodrich (2001) when focusing on examples from the internet, foundational topics in Algorithms remain in the course, as a part of the link when possible.

Of particular importance was our goal to develop course materials that focus faculty and student attention on genomics research, facilitate various types of collaborative work between students, and are easy to integrate into various combinations of course linkages. To meet our objective of reaching a significant number of majors in computer science and across the biological sciences, we have been careful to include a core course from each discipline ("Algorithms" for computer science and "Genetics" for biology). However, our plan was to develop course materials that facilitate collaborations around genomics content that are not necessarily specific to a particular course. We consider this to be one of the more significant challenges of our current effort. In preliminary efforts, we linked "Algorithms" with both "Cell Evolution" and with "Genetics". Other potential computer science courses include: "Artificial Intelligence", "Software Engineering", and "Database Systems".

Table 1 summarizes a typical plan for two linked courses, in this case "Genetics" and "Algorithms." An underlying assumption in our plan is that new and exciting research questions will be answered by new software (or at least modifications to existing software). Our personal research collaboration and classroom experiences to date have convinced us that successful collaborations emerge over time, not, for example, in one or two isolated laboratory sessions. Thus, we are developing materials to bring students from both disciplines together in class and lab and then later in follow-ups outside of lab, for example, in a joint homework assignment, in modifications to a programming assignment, or preparations for final project presentations.

Approximate timeline during the semester (14-week semester)	Type of Collaboration		Course materials
	Biology	Computer Science	

Week 2		<b>Guest lecture</b> by biology professor	<b>"DNA 101"</b> - an introductory lecture for computer science students along with a demonstration of DNA extraction "in the manner of Julia Childs"
Week 2 or 3	<b>Genomics Lab1</b> - Biology and computer science students share initial joint lab; work in BIO-CS teams ..... followed-up with homework		<b>BLASTing the Flagellar Genes</b> - an intro to NCBI, BLAST and PubMed database
	<b>Joint homework</b>		<b>BLAST homework</b>
Week 3		initial genomics programming assignment	<b>Motif Finder</b> - an intro to DNA (*.fna) files and protein table (*.ptt) files; software to find all locations of a user-entered motif
Week 3	<b>Guest lecture</b> by computer science professor		Explanation of <b>DNA data files</b> , the <b>algorithm</b> in the Motif Finder software, and <b>potential tweaks</b> to suggest to the programmer
Week 4	<b>Genomics Lab2</b>		Collaborative model of a complete experimental design, methods, tests, and results
Week 6	<b>Genomics Lab3</b>		<b>Tweaking Motif Finder</b> - demonstrations of software; biologists suggest changes to programmers for (new) added functionality

Week 8		second genomics programming assignment	<b>IsPal</b> - finding "palindromes" or inverted repeats with potential mismatched base pairs: comparing algorithms $O(2N)$ vs. $O(N^2)$
Week 10	<b>Genomics Lab4</b> ... followed-up by informal out-of-class work on projects		<b>Planning for final projects</b> - review of general specifications (e.g., triplet repeat diseases), <b>suggested timeline</b> , and <b>coordination</b> of team day-timers
Near end of the semester	<b>Final oral presentations</b>		Biology-Computer Science teams give talks in conference format

Table 1. Summary of a linked course plan as viewed within an example of points of contact between "Genetics" and "Algorithms"

### 3.1 Infusing genomics within each of the syllabi

The range of four courses allows us room to experiment with both general, course-independent materials as well as course-specific content. Table 2 summarizes some ways that genomics has been and will be infused in the respective syllabi.

Course	Topic	Infusing Genomics	Materials
Algorithms	Growth rate of algorithms (Big Oh)	Volumes of data and the perils of programs that never finish	Lecture Notes and Homework
	Recurrence Relations	Finding Inverted Repeats	Lecture Notes and Homework exercise
	Dynamic Programming		Programming Assignment
Software Engineering	Design of a new project	Collaborations in software design and specification	Lecture notes and homework
	Writing Specifications	Agreeing on an Application Programmers Interface (API)	Writing exercises (three iterations)

Cell Evolution	Phylogenetic trees	Collaborations in programming search and analysis software	Lecture notes and project specifications
	The origin of specific structures and pathways	Collaborations in programming search and analysis software	Lecture notes, help with online tools, and project specifications
Genetics	Gene regulation	Understanding the role of repeat sequences	Lecture notes, lab exercise, and project specifications
	Genetic diseases	Collaborating on projects to search for signature sequences	Lecture notes, lab exercise, suggestions when reading PubMed, and project specs
Table 2. Sample of topic areas where genomics materials will be infused into each course			

### 3.2 Exchanging Guest Lectures

We exchange guest lectures in each other's course, for example, see Table 1. In each case, an exposure to the presentation and language of a professor from "the other" discipline was as important as the content. Our lecture notes include annotations that highlight unique aspects of each discipline.

### 3.3 Sharing Interdisciplinary Laboratory Sessions

Three or four of the twelve laboratory sessions for "Cell Evolution" (or "Genetics") and "Algorithms" were shared so students could work together and benefit from the presence of both professors. In semesters where Algorithms did not have a two-hour lab in addition to three hours of lecture, lectures were scheduled in the same time slot to facilitate the shared meeting times. It was in such shared sessions that the dynamics for future out-of-class collaborations were established. The educational materials developed and being developed for shared labs include:

- Lab 1** - An introduction to the public databases of sequence data available at NCBI (National Center for Biotechnology Information) and associated tools such as BLAST.
- Lab 2** - The instructors model a complete experimental design in genomics.
- Lab 3** - In interdisciplinary teams, computer scientists demonstrate their designs and/or software from a previous programming assignment. Together the team agrees on and documents a set of "tweaks" to the design/software for increased functionality.
- Lab 4** - Beginning collaborations for research projects to be conducted out-of-class.

### 3.4 Collaborating on homework exercises

As follow-ups to interdisciplinary labs, we are developing homework exercises where biology and computer science students need to meet outside of scheduled classes/labs. The after-session homework guarantees another time, an informal time, where students from both disciplines can learn to work together.

### 3.5 Pairing biologists and computer scientists on independent projects

The interdisciplinary teams for final projects serve as a capstone experience for the infusion of genomics. Students are paired as collaborators to address a problem in genomics and to write both a final report and a piece of software. In initial pilots, we have used two models: (i) open-ended projects of the students' choice, and (ii) all groups focus on a single topic (e.g., triplet-repeat diseases). The associated materials to be developed to help faculty lead successful interdisciplinary projects are summarized in Table 3.

A list of <b>suggestions</b> to students who are unsure of the types of "open-ended" projects that might be available or appropriate. In particular, our own research in genomics has generated a number of hypotheses that might be tested by students working at an introductory level.
A <b>timeline</b> of action items, milestones, and critical paths to ensure that the project stays on track and both partners feel secure with preliminary and final results.
<b>Specifications</b> for the project, one for the biology partner that focuses on literature search and one for the computer science partner that focuses on the software; both specifications will highlight the need to establish a working hypothesis and how to work towards tangible results.
Sample <b>oral presentation</b> that includes hypothesis, methods, results, and future work; these slides can be used by the professors as they model a professional presentation to the students.
Table 3. Summary of course materials to help faculty coordinate interdisciplinary capstone projects in genomics

## 4. EVALUATION PLAN

Our approach to the evaluation of this project is multi-faceted. We are assessing not only some of the academic and affective outcomes of the linked courses in graded homeworks, labs, exams, and student pre- and post-evaluations, but also continue to evaluate the reach and implementation of this model at Wheaton and other institutions. Table 4 summarizes our evaluation plan. Pre- and post-evaluation instruments are being used to determine changes in students' attitudes and confidence levels.

In particular, we learned from initial pre/post assessments that some students initially believe that what is "challenging" or "difficult" is "bad." A primary goal of linking courses was to give students opportunities to work together on and through conceptual, organizational, and interpersonal challenges. We examined changes in these through students' ratings of expected personal contribution, and satisfaction from and importance of interdisciplinary collaborations.

<b>Evaluation Purpose</b>	<b>Objective</b>	<b>Data</b>
Measure efforts to change students' awareness of and appreciation for the challenges and rewards of interdisciplinary work	Reduce "concerns" over collaborations; Reduce the perception that collaboration is "challenging" and therefore "bad"	Pre- and Post-evaluation answers to questions
Measure efforts to increase students' knowledge of bioinformatic and genomic applications and potential areas of research	Infuse genomics techniques, research areas, and potential use of tools into "traditional" materials to instill pervasiveness and importance of this new area	homework, exams, and dual-grading of final project presentations and Pre- and Post-evaluation
Measure extent of experiments at other colleges with "linked courses" in genomics	Disseminate and encourage "linking" as an alternative teaching/learning model via NSF DUE #0126643	Workshop for Faculty from other institutions: # in workshop # who try model # who return June '03 Evaluations
Measure number of various science majors reached and percentage of majors	Ensure that we reach a significant percentage of students in computer science and the biological sciences	Number of different majors and percentages of CS and biological science majors (Registrar records)
Measure "pipeline" effect of research in class leading to research for work or internships	Experiences in collaborative research in "linked" courses leads to future research experiences.	Post-evaluation survey of students' expected research experiences; Tracking graduates

Table 4. Evaluation Plan

In particular, computer science students begin with fewer concerns about collaborating than biology students, perhaps because of the culture, where programmers are almost always "in service" to someone. However, in our pilot post-evaluations, neither the biologists nor computer scientists experienced a significant change in level of concern. We intend to focus more attention on this area as we continue to develop our materials and assessment tools.

Other outcomes that were evaluated include the extent to which we could encourage other faculty, both on and off campus, to experiment with linked courses, for example through our NSF-funded workshops and by the number of students who move on to research positions in bioinformatics over January breaks and summers or who enter graduate school or the workforce in bioinformatics. Participation of students by major shows the current impact of infusion in these courses, e.g., the number of different majors and percentages of each major. Early results show that two strategically chosen linked courses can bring genomics to 100% of the computer science majors of a given class year and as many as 70% of all majors in the biological sciences.

## 5. CONCLUSIONS

To our knowledge, resources supporting a linked course collaboration of this type (between undergraduate biologists and computer scientists) have not been published yet or made accessible online. However, training in and for interdisciplinary collaborations is a nationally recognized need. The National Computational Science Institute (NCSI) leads a national effort in establishing interdisciplinary training, fostering collaborations, and initiating new modes of undergraduate research in computational science (Panoff *et al.*, 2001). Although not originally part of the computational science scope, bioinformatics is now rightly included within "internetics" the interdisciplinary field between computer science and both simulation and information-based applications (Fox 2000).

Most other initiatives to introduce genomics (or bioinformatics) into curricula center around one or the other of two models:

1. developing full programs to train students in both disciplines (biologists who are competent programmers or programmers who have a considerable knowledge base in biology) (e.g., Doom *et al.*, 2002).
2. developing software packages and training tutorials for biologists to use in searching and analyzing sequences (e.g., Benson and Bruce, 2001).

In the case of the former, we would point out that small colleges are unlikely to be able to devote sufficient staff and resources to new programs that will reach only a small subset of especially versatile students. An infusion model reaches more students, using a preexisting infrastructure. Furthermore, it models realistic collaborative research in which partners are not necessarily expected to know each other's fields.

In the case of the latter, while we acknowledge the importance of encouraging biologists to use preexisting software and simulations for genomic analysis, we note that

neither of these models do what makes our own project unique. That is, we are developing course materials to help faculty make opportunities for collaborations of computer science and biology students such that new analytical software can be written by undergraduates to address specific, and sometimes novel, hypotheses. We believe our "linked" teaching model to be unique and especially useful, as it reaches a broad number of students and models on-going collaborations and programmer/client relationships in research and industry needed specifically for progress in the analyses of genomes.

#### **ACKNOWLEDGEMENTS:**

This material is based upon work supported by the National Science Foundation under NSF Grant DUE #0126643 and funding from Wheaton College including the Mars Foundation, the Davis Foundation, and the Provost's Office.

#### **6. REFERENCES CITED**

- Benson, A., & Bruce, B. (2001). Using the web to promote inquiry and collaboration: A snapshot of the Inquiry Page's development. *Teaching Education*, v12(2): 153-163.
- Berkowitz, M. (2002). How to develop and implement a bioinformatics curriculum: a hands-on workshop. *The Journal of Computing in Small Colleges*, v17(3): 183-186.
- Brown, S.M. (1999). "Dealing with genome project data." *Biotechniques*, v26: 266-268.
- Colwell, R. (2002). "NSF Director: 'New Bioscientist' will be future of genomics." *An interview in Genome Technology*, April 2002, No. 20: p72.
- Doom, T., Raymer, M. Krane, D. and Garcia, O. (2002). A proposed undergraduate bioinformatics curriculum for computer scientists. Published in the *Proceedings of the 33rd SIGCSE Technical Symposium on Computer Science Education*: 78-81.
- Goodrich, M. (2001). Teaching Internet Algorithmics. Abstract appears in the *The Journal of Computing in Small Colleges*, May, 2001, v16(4): xvii.
- Fox, G.C. (2000). From Computational Science to Internetics: Integration of Science with Computer Science. *Mathematics and Computers in Simulation*, v54: 295-306.
- Panoff, R., Hirst, H., Jakobsson, E., and Stevenson, D. (2001). "National Computational Science Institute." NSF CCLI-National Dissemination grant #0127488.

#### **Relevant URLs**

"The Computer Science Teaching Center" (CSTC)  
<http://www.cstc.org>

"Genome Consortium for Active Teaching" (GCAT)  
<http://www.bio.davidson.edu/Biology/GCAT/GCAT.html>

"National Center for Biotechnology Information" (NCBI)  
<http://www.ncbi.nlm.nih.gov/>

"National Science, Mathematics, Engineering, and Technology Education Digital Library"  
(NSDL)  
<http://www.nsdlnet.org/>

San Diego SuperComputer Center's "Biology Workbench"  
<http://workbench.sdsc.edu/>

"National Computational Science Institute" (NCSI) -- The Shodor Education Foundation,  
Inc.  
<http://www.computationalscience.net/>

"Wheaton College Genomics Group"  
<http://genomics.wheatoncollege.edu>