

Essay

Bioinformatics and the Undergraduate Curriculum Essay

Mark Maloney,^{*} Jeffrey Parker,[†] Mark LeBlanc,[‡] Craig T. Woodard,[§]
Mary Glackin,^{||} and Michael Hanrahan^{**}

^{*}Department of Biology, Spelman College, Atlanta, GA 30314-4399; [†]Department of Computer Science, Merrimack College, North Andover, MA 01845; [‡]Department of Computer Science, Wheaton College, Norton, MA 02766; [§]Department of Biological Sciences, Mount Holyoke College, South Hadley, MA 01075; [¶]Library, Information & Technology Services, Mount Holyoke College, South Hadley, MA 01075; ^{||}Information and Library Services, Bates College, Lewiston, ME 04240

Submitted March 16, 2010; Revised June 17, 2010; Accepted June 20, 2010
Monitoring Editor: John Jungck

Recent advances involving high-throughput techniques for data generation and analysis have made familiarity with basic bioinformatics concepts and programs a necessity in the biological sciences. Undergraduate students increasingly need training in methods related to finding and retrieving information stored in vast databases. The rapid rise of bioinformatics as a new discipline has challenged many colleges and universities to keep current with their curricula, often in the face of static or dwindling resources. On the plus side, many bioinformatics modules and related databases and software programs are free and accessible online, and interdisciplinary partnerships between existing faculty members and their support staff have proved advantageous in such efforts. We present examples of strategies and methods that have been successfully used to incorporate bioinformatics content into undergraduate curricula.

INTRODUCTION

In recent years, great progress has been made in our ability to sequence entire genomes of a variety of organisms and understand how cells access and use the information stored in DNA. Mathematics and the languages of computer science can be used to describe changes in the DNA-related codes of biology. Scientists can use the many programs developed with applied computer science and mathematics to access DNA, RNA, and protein sequence data as well as other biologically relevant information generated by biologists and stored in databases.

Operations to compare genes or gene products by matching nucleotide or related sequences use a variety of bioinformatics algorithms and computer software. Clearly, a working under-

standing of bioinformatics requires a synthesis of principles from biology and computer science as well as applied mathematics and chemistry. Due to recent technological advances, data accumulation is currently far outpacing analysis. There are many sequences and even complete genomes being entered into databases that require annotation. These can be used as sources of data for research as well as for class projects. In addition to their usefulness as teaching tools, some important research will undoubtedly emerge from such projects. Students need a robust introduction to bioinformatics tools and a solid understanding of related principles and technologies. Circumstances are favorable for computer scientists to participate in and support bioinformatics programs (LeBlanc and Dyer, 2004), and many interdisciplinary programs have been established at the undergraduate and graduate levels (Goode and Trajkovski, 2007; Hemminger *et al.*, 2005). For now, the fruit is still low because the possibilities for interdisciplinary collaboration are far from exhausted, the ever-increasing store of information is largely held in databases with public access, and many software programs are freely available that can be used to access and analyze the data. Valuable research can be performed on a computer (in

DOI: 10.1187/cbe.10-03-0038

Address correspondence to: Mark Maloney (mmaloney@spelman.edu).

© 2010 M. Maloney *et al.* CBE—Life Sciences Education © 2010 The American Society for Cell Biology under license from the author(s). It is available to the public under Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

silico) with no further investment in equipment or laboratory supplies.

BRINGING BIOINFORMATICS INTO THE CLASSROOM

No science curriculum can remain current without a bioinformatics component. Projects such as the sequencing of the human genome have changed the nature of instruction. A modern biology course must address new techniques in gene mapping. Students need to understand what bioinformatics-related computer software programs do and how they do it. How do programs align nucleic acid sequences or generate structural models? To evaluate the results obtained using standard bioinformatics tools such as Basic Local Alignment Search Tool (BLAST), students must understand the strengths and limitations of the programs and data. This is basic knowledge for today's biologist.

At the undergraduate level, a solid introduction to a computer programming language is very important. Although many of the bioinformatics exercises available use the computer language Perl, many computer scientists have preferences for other languages. Computer science enrollment is declining just as bioinformatics has made the need for computer programming experience in biology undeniable. These trends give computer science faculty a strong incentive to invest in learning relevant biological concepts and collaborate with biology faculty. Computer science graduates with a basic knowledge of biological principles will be especially in demand, as will biologists with a computer science background.

In addition to the necessary participation of biologists and computer scientists, faculty from other disciplines also play a vital role in bioinformatics instruction. Mathematicians provide instruction in applied statistics and algorithm development, and biochemists and chemists provide instruction in molecular modeling, proteomics, and metabolomics.

Biologists will continue to use so-called wet labs to give students a chance to experience experimental techniques involving organisms, tissues, and cellular components first hand. But in silico dry labs involving bioinformatics techniques and virtual lab exercises can be very effective, especially in genetics, cell biology, and molecular biology. Release time for existing faculty to gain workshop or short course experience may be necessary for related curriculum development purposes. Many excellent workshops and short courses are available from various institutions and organizations. In addition to the National Institute for Technology in Liberal Education-funded workshops, these include the National Science Foundation (NSF) funded Bioquest/BEDROCK project, the Teagle-funded Teaching Big Science at Small Colleges: a Genomics Collaboration, the NSF- and Howard Hughes Medical Institute-funded project The Genome Consortium for Active Teaching, and the workshops offered by the National Center for Biotechnology Information (NCBI).

Mount Holyoke College (MHC) and Bates College obtained a National Institute for Technology in Liberal Education grant (www.nitle.org) to support collaboration for curricular planning among small liberal arts colleges. These bioinformatics retreats convened 46 participants in teams of

biologists, mathematicians, and computer scientists as well as faculty from related disciplines, instructional technologists, and librarians from 18 colleges. Topics included novel ways to initiate and foster interdisciplinary dialogue between Science, Technology, Engineering, and Mathematics departments; to use advances in bioinformatics to enhance curricular and research opportunities for faculty and undergraduate students; and to engage technical support staff and administrators in these endeavors. Participants discussed ways to use bioinformatics to excite millennial generation students who are very comfortable with the use of computers but whose learning styles often are not well suited to a traditional lecture-based format. We concluded that hands-on computer exercises using software freely available on the Web would engage students in using bioinformatics tools. See sample exercises developed and tested by participants in the Supplemental Material. These workshops have inspired innovative interdisciplinary instructional and research-oriented projects. Some examples follow.

At MHC, faculty from biology, biochemistry, computer science, and mathematics and staff from information technology (IT) explored strategies for bringing bioinformatics into the curriculum. The group decided to start by integrating bioinformatics and cheminformatics tools throughout the existing curriculum, from introductory-level biology through advanced courses. They plan to add Bioinformatics and Proteomics courses as appropriate. MHC hosted an NCBI training workshop on searching tools and participated in 2 yr of summer workshops to share curricula and strategies. MHC faculty have introduced NCBI tools, including BLAST and Cn3D, into their Introductory Genetics and Molecular Biology courses. They use the Protein Data Bank and visualization tools in Genetics, Molecular Biology, and Cell Biology courses and have developed a new Computational Chemistry course.

The Department of Biology at Spelman College has undertaken a major revision of its undergraduate curriculum (Maloney *et al.*, 2007). The first Bioinformatics course was offered as an elective in 2004 with assistance from guest lecturers funded through a science seminar series. Basic bioinformatics content is now incorporated into the required four-course introductory biology sequence with curriculum development support from the Advancing Spelman's Participation in Informatics Research and Education (www.spelman.edu/academics/research/aspire/index.shtml) program funded by Historically Black Colleges and Universities Undergraduate Program. Modules currently in use or under development use software available online at the NCBI or the San Diego Super Computer Center's Biology Workbench sites and include protein modeling using Cn3D, sequence similarity searches using various BLAST programs, ORF Finder, CLUSTALW for multiple sequence comparisons, and Phylip for researching evolutionary relationships and generating dendrograms.

Advanced genomics, proteomics, and related bioinformatics topics are now offered in upper-level electives enhanced by recent collaborative efforts and new hires in biology and computer science. The faculty has experimented with teaching computing skills to Spelman students by using the open-source R-language for statistical computing. In addition, a revised version of the Introductory Computer Science course is being developed that will involve enhanced training with a programming language and various examples of compu-

tational software applications that form the basis of informatics in the new millennium.

Since 1998, Wheaton College (Norton, MA) has actively engaged with various modes of interdisciplinary teaching. In fact, collaborative pedagogy in the new curriculum is centered on connections, or pairs of linked courses that connect significantly different disciplines (LeBlanc *et al.*, 2010). In short, the faculty voted to teach together in various combinations. This provides an exciting way to explore different areas of knowledge and different approaches to problems. With NSF funding, Wheaton faculty have developed, tested, and disseminated course materials for linking courses in biology and computer science, such as biology's Genetics with Computer Science's Algorithm course. Other curricula changes include a new bioinformatics major and a team-taught course called DNA, an elective for both the biology and computer science majors in which students learn to program and design experiments to data mine entire genomes. Bioinformatics courses in biology and computer science also may be connected with Philosophy's Ethics course. In both the ethics and computing courses, particular attention is paid to the increasing use and challenges of sequenced genomes as applied to personalized medicine.

WORKSHOP RECOMMENDATIONS

Multiple teaching models are available for bioinformatics-related courses and diversity is encouraged. No one model fits all curriculum and personnel situations. We see three distinct models in the spectrum of teaching opportunities:

1. Infusion of bioinformatics content into existing courses
2. Linking of Biology and Computer Science courses (or Bio-Math, Chem-CS, etc.) by sharing lab exercises or interdisciplinary final projects
3. Team-teaching of Bioinformatics involving instructors from multiple departments

Also see Supplemental Material and <http://genomics.wheatoncollege.edu> for examples of course materials.

Alternatively, targeting candidates with bioinformatics experience when hiring new faculty may be possible in biology, computer science, chemistry, mathematics, or related searches. Such coordination of hiring above the department level in order to assemble appropriate interdisciplinary teams is not typical practice and may require novel administrative initiatives.

Institutions may not need to implement new majors or minors in bioinformatics. The majority of schools can be well served by introducing the concepts and technologies into existing majors. However, teaching across disciplines is often difficult to coordinate and administer, and innovations in team teaching must be properly valued and evaluated. Faculty members need administrative encouragement and support to experiment with such courses. Junior faculty involved in interdisciplinary efforts are often concerned that their efforts will not be valued as much as instruction within

their department. Representatives of both the administration and faculty governance, including those involved in tenure and promotion decisions, should move to allay these fears. These members of the shared governance structure should make it clear that such efforts to keep current in increasingly interdisciplinary practices in higher education will be valued and rewarded.

Use the computer savvy and technical inclinations of the millennial generation of students to advantage in recruiting students for classes featuring hands-on software program experiences in bioinformatics. The most capable students then can serve as the next wave of teaching assistants.

Support for bioinformatics, or informatics in general, goes beyond the need for informed classroom instructors. IT, library science, and laboratory support staff may need to work together to assemble a bioinformatics computer lab in a manner that is novel to the institution. Such critical technology support may require new duties for existing personnel or new staff hires.

Use of external expertise is often desirable, if not necessary, and may come in different forms and through different venues. Visiting scientists from established bioinformatics programs at either teaching or research-intensive institutions are especially helpful. The duration and form of expertise sharing is dependent on time and budgetary considerations. For institutions with divisional or departmental seminar series, now may be the optimal time to choose bioinformatics as the series theme. Speakers can perform double duties as guest lecturers in nascent Bioinformatics courses or in related Genetics, Cell and Molecular Biology, or Biochemistry courses. The large-scale application of computer science-based technology to biological problems is an exciting new development in science. The need to introduce bioinformatics to the undergraduate curriculum is irrefutable. Done properly, this will excite the new generation of students as well as prepare them for their future careers.

REFERENCES

- Goode, E., and Trajkovski, G. (2007). Developing a truly interdisciplinary bioinformatics track: work in progress, *J. Comput. Sci. Coll.* 22, 73–79.
- Hemminger, B., Losi, T., and Bauers, A. (2005). Survey of bioinformatics programs in the United States. *J. Am. Soc. Inf. Sci. Technol.* 56, 529–537.
- LeBlanc, M. D., and Dyer, B. D. (2004). Bioinformatics and computing curricula 2001—why computer science is well positioned in a post-genomic world. *ACM SIGCSE Bull.* 36, 64–67.
- LeBlanc, M. D., Gousie, M. and Armstrong, T. (2010). Connecting across campus. In: Proceedings of the 41st SIGCSE Technical Symposium on Computer Science Education, 10–13 March 2010, Milwaukee, WI. ACM Special Interest Group on Computer Science Education. www.sigcse.org/sigcse2010 (accessed 16 July 2010).
- Maloney, M., Imumorin, I., and Bauerle, C. (2007) Teaching millennial science students in the (bio)informatics age. *Network: A Journal of Faculty Development.* www.nyu.edu/fm/publications/millennial_student/Bioinformatic.html (accessed 16 July 2010).