Broader incorporation of bioinformatics in education: opportunities and challenges

Michael P. Cummings and Glena G. Temple

Abstract

The major opportunities for broader incorporation of bioinformatics in education can be placed into three general categories: general applicability of bioinformatics in life science and related curricula; inherent fit of bioinformatics for promoting student learning in most biology programs; and the general experience and associated comfort students have with computers and technology. Conversely, the major challenges for broader incorporation of bioinformatics in education can be placed into three general categories: required infrastructure and logistics; instructor knowledge of bioinformatics and continuing education; and the breadth of bioinformatics, and the diversity of students and educational objectives. Broader incorporation of bioinformatics at all education levels requires overcoming the challenges to using transformative computer-requiring learning activities, assisting faculty in collecting assessment data on mastery of student learning outcomes, as well as creating more faculty development opportunities that span diverse skill levels, with an emphasis placed on providing resource materials that are kept up-to-date as the field and tools change.

Keywords: bioinformatics; computational biology; education

INTRODUCTION

The life sciences and related disciplines have become very computationally intensive in both research and educational techniques. This trend is increasing and is reflected in the development of newer disciplines, such as bioinformatics and computational biology that were not part of undergraduate or graduate education, in whole or in part, a short time ago. Bioinformatics and computational biology mean different things depending on the audience, so for the purpose of this manuscript, the following definitions [1] will be used:

- Bioinformatics: Research, development or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze or visualize such data.

- Computational biology: The development and application of data analytical and theoretical methods, mathematical modeling and computational simulation techniques to the study of biological, behavioral and social systems.

Although the distinctions conveyed in these definitions are not always considered in many contexts, they are useful in understanding the opportunities and challenges for broader incorporation of bioinformatics in education, because most opportunities are

Corresponding author. Michael P. Cummings, Center for Bioinformatics and Computational Biology, Biomolecular Sciences Building, University of Maryland, College Park, MD 20742 USA. Tel: +1 301 405 9903; Fax: +1 301 314 1341; E-mail: mike@umiacs.umd.edu

Michael P. Cummings is Associate Professor in the Center for Bioinformatics and Computational Biology, with appointments in the Department of Biology and the Institute for Advanced Computer Research at the University of Maryland. He is Director for the Workshop on Molecular Evolution, Woods Hole, and the Co-Director for the Workshop on Molecular Evolution, Europe. His research focuses on molecular evolution, phylogenetics, genotype–phenotype relationships, and grid, public and General Purpose Graphics Processing Unit (GPGPU) computing in phylogenetics and molecular evolution.

Glena G. Temple is Professor of Biology, and Dean of Letters and Sciences at Viterbo University. Her research is in the area of plant biochemistry with a focus on biochemical and molecular adaptations in plants in response to environmental stress. She has taught biochemistry, biochemical techniques, genetics, general biology, plant biology and comparative bioinformatics.

those to incorporate computational biology into curricula.

Although there has been tremendous growth in bioinformatics and computational biology research, an equal amount of attention to education in these fields has not been realized. For example, there has been extensive growth in bioinformatics-related web sites over the period 2006–09 with sites dedicated to proteins increasing 82% and those for expression increasing 56% [2]. However, the number of web sites devoted to bioinformatics education increased by just three web sites, the second lowest growth rate over the same period at 4% (Figure 1). Donovan identified this disparity in attention to education in the field of bioinformatics in 2008, stating,

'failure to challenge the status quo in science education will lead to an even greater divergence between scientific practice and education, which could potentially undermine the continued rapid evolution of science research' [3]. We echo this concern, and argue that to maintain the growth of the field in the next decade, the nature of education at all levels needs to change to include bioinformatics and computational biology and mirror the research trends occurring.

THE OPPORTUNITIES

The major opportunities for broader incorporation of bioinformatics in education can be placed into
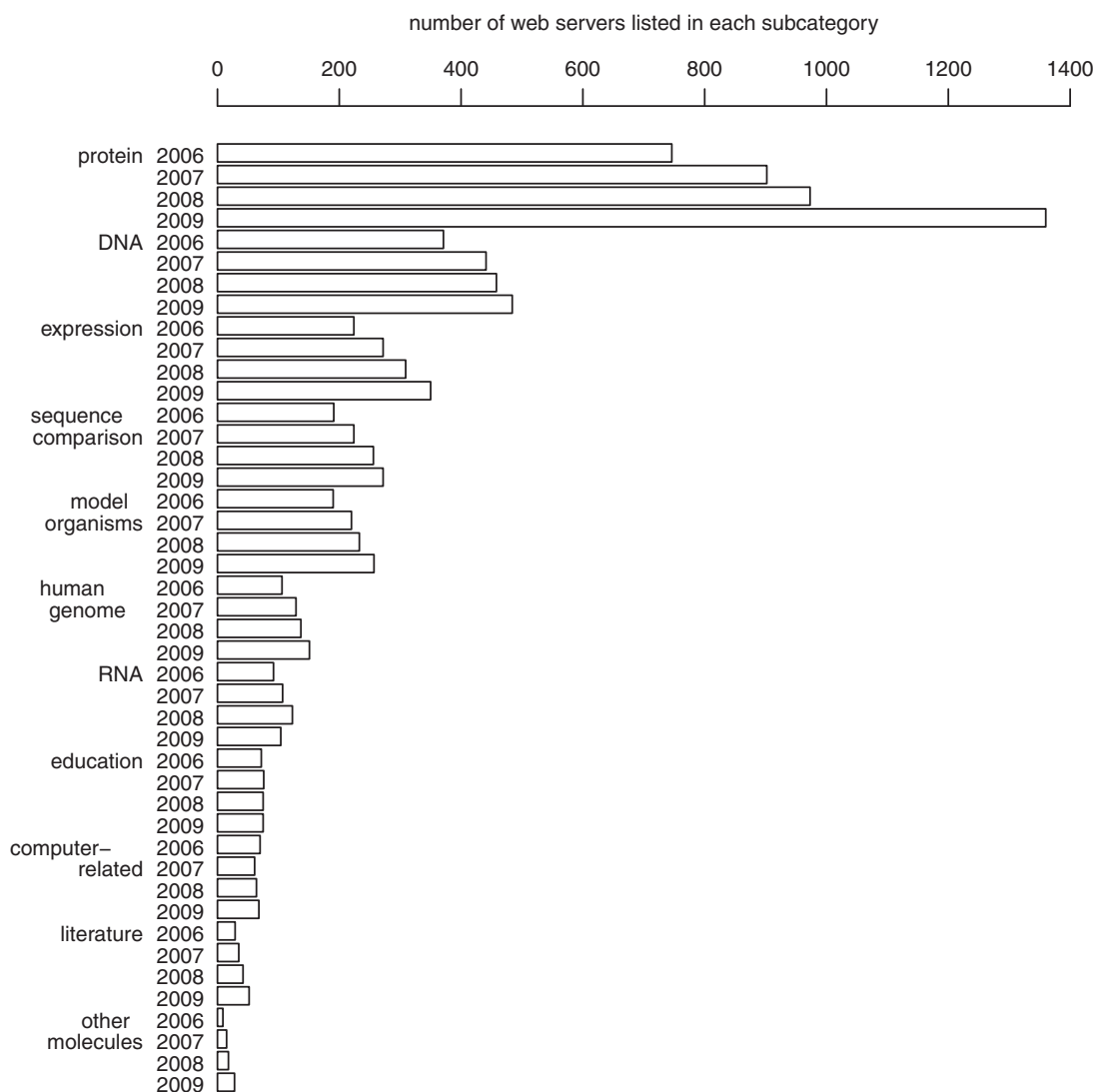

Figure 1: Plot of the number of web servers listed in each subcategory from the Bioinformatics Links Directory 2006–09 [2]. For the full period there has been an increase in web servers for all subcategories with the exception of RNA, and education has the second lowest growth rate.

three general categories: general applicability of bioinformatics in life science and related curricula; inherent fit of bioinformatics for promoting student learning in most biology programs; and the general experience and associated comfort students have with computers and technology. Here, we briefly examine each of these in turn.

General applicability of bioinformatics in life science and related curricula

The life sciences and related areas are getting more quantitative, deal with larger data sets and employ increasingly sophisticated statistics and models, all of which makes computer-based activities beneficial if not necessary for high-quality education. This is no longer just applicable to a genetics course or a bioinformatics course, but is impacting the entire field. Consequently opportunities for broader incorporation of bioinformatics occur throughout the curriculum in most life science programs. The changing dynamic of science education and particularly biology education was outlined in *Bio 2010* [4], which challenged educators to make biology education more interdisciplinary, inquiry based and with an emphasis on quantitative skills. The recommendations for *Bio 2010* [4] outlines several key themes essential for preparing future biomedical researchers that undergraduate biology programs should highlight. Many of these stated themes would be strengthened by the incorporation of inquiry based bioinformatics activities including:

- All living things have evolved from a common ancestor, through processes that include natural selection and genetic drift acting on heritable genetic variation.
- Understanding biological systems requires both reductionist and holistic thinking because novel properties emerge as simpler units assemble into more complex structures.
- Although fundamental molecular and cellular processes are conserved, biological systems and organisms are extraordinarily diverse. Unlike atoms and simple molecules studied in chemistry and physics, no two cells are identical.
- Information encoded in DNA is organized into genes. These heritable units use RNA as informational intermediates to encode protein sequences, which become functional on folding into distinctive three-dimensional structures. In some situations RNA itself has catalytic activity.

Bioinformatics and computational biology provide useful mechanisms for engaging the students in the mastery of learning outcomes in most biology programs. The literature demonstrates that active learning strategies including bioinformatics and computational biology tools have been used successfully in a range of courses, from lower division general biology to upper division genetics and molecular biology courses [5–8]. Several static web sites exist with data sets, resources and learning modules for educators that have been developed in the last 15 years, primarily with funding from the National Science Foundation. Examples include BioQuest [9] and the Molecular Science Student Work Bench [10], and others with more specific focuses [11].

For many educational activities computers may be a useful tool, but the pedagogic objectives can be met without the use of computers. In contrast, some learning activities like the research science on which they are based, require computers and this is true for a substantial amount of material in the life sciences. Interest in computer-required learning modules has increased in recent years in a variety of courses. Contributing factors for this increase include the advances in technology and the high cost of laboratory activities [12]. In many aspects, bioinformatics and computational biology are natural fits for computer-required learning activities in biology and related subjects, and allow for active engagement by the student with data sets and current technologies in the field.

Inherent fit of bioinformatics for promoting student learning

In several respects, bioinformatics education has particular advantages for effective education when properly integrated into life science and biochemistry curricula. For example, computer-requiring learning activities in bioinformatics education have great potential to promote learning because they well fit prescriptive principles common to various instructional design theories [13]. Two principles for promoting learning are particularly relevant to bioinformatics: 'learning is promoted when learners are engaged in solving real-world problems' and 'learning is promoted when new knowledge is applied by the learner'. There are more than ample opportunities to incorporate real-world examples, some extremely timely, into lectures and computer-requiring laboratory activities. For example, the developments in medical genetics [14], pathogen outbreaks [15] and

ancient DNA research [16–18] can readily be used as the basis for lectures and computer activities. Doing so provides a direct linkage between learning through course work, and learning through news media and other venues. Furthermore, the opportunities for the application of new knowledge is perhaps unparalleled within the life sciences because of the inherent complementary reciprocity of conceptual information conveyed in readings and lectures on one hand, and computer activities on the other.

General experience and associated comfort students have with computers and technology

The pervasiveness of computers has instilled in most students at least a semblance of technological competence that can be leveraged for more specialized education in bioinformatics. Previous experience searching for and accessing remote information with web browsers, mobile devices and a variety of user interfaces means the introduction of computing for bioinformatics education builds upon previous knowledge. So here too bioinformatics fits another prescriptive principle common to various instructional design theories, 'learning is promoted when existing knowledge is activated as a foundation for new knowledge' [13].

In a 2008 report, over 63% of survey respondents in the public and private sector indicated that technological innovation would have major influences on pedagogy and higher education in the next 5 years [19]. Students today call for technology to be included in their courses, and prefer to be engaged through technology in learning activities instead of traditional lecture. In our experience, students have the technology skills and interest to readily become adept in navigating web sites and performing analyses in bioinformatics. A challenge in many cases is to raise the comfort of the faculty member integrating the new technology, not the student, through interfaces that are simple and have solid support materials provided.

THE CHALLENGES

The major challenges for broader incorporation of bioinformatics in education can be placed into three general categories: required infrastructure and logistics; instructor knowledge of bioinformatics and continuing education; and the breadth of bioinformatics,

and the diversity of students and educational objectives. Here, we briefly examine each of these in turn.

Required infrastructure and logistics

Like research in the subject itself, teaching bioinformatics requires some varying degree of computational infrastructure depending on the material covered and tools employed. The operational logistics for both faculty and students constitutes a substantial challenge to the broader incorporation of bioinformatics in education. The specifics of the computing environment available for instruction is usually determined by what is provided by institution or course-specific policies. These environments may include relatively uniform computers (i.e. similar hardware and operating system) in a teaching lab to diverse computers (i.e. disparate hardware and operating systems) composed of students' own computers. Instructional computing support also varies greatly, with some institutions providing knowledgeable, flexible and responsive support policies and personnel, some institutions providing much narrower support, and others providing none at all. Lack of adequate support often represents a substantial barrier for faculty even in those cases where they simply wish to adopt instructional activities developed and employed elsewhere. Ultimately, someone, often the instructor, needs to choose, install and maintain the software, databases and data sets necessary for the activities to further the learning objectives.

Setting up and maintaining the appropriate computing infrastructure and developing appropriate learning activities represent major challenges to using computing resources for instruction, especially for underserved faculty populations (e.g. some community colleges or primarily undergraduate institutions) and faculty not trained in this technology. For example, many existing online learning activities currently:

- are static and cannot be easily tailored to meet individual faculty learning outcomes and objectives;
- do not provide significant faculty development opportunities and support materials without travel to workshops;
- provide limited and static additional materials to assist the faculty member when introducing the material to the students;

- do not demonstrate to the student the algorithms and computer resources required to analyze the data sets;
- do not interact with the students (e.g. for results, feedback on student understanding);
- do not interact with the faculty member to provide data on mastery of student learning;
- are not constructed for sustainability—leading to 'dead links' and outdated modules; and
- are not designed for online courses (lacking sufficiently clear instructions and background information for the students to work on independently).

Instructor knowledge and continuing education in bioinformatics

Neumann and Provart (2006) described the benefits and problems faced by instructors when incorporating customized tools and databases related to bioinformatics into biology courses. They reported that the complexity of tools and the interpretation of output can discourage many students, and when teaching diverse students in large lab or lecture sections the majority became overwhelmed and lost sight of the biological problem they were using the tools to address. Also, the faculty and teaching assistants had unequal training, previous experience and comfort with the activities, and students were requesting additional support in completing the activities. They also reported that existing static web pages were not ideal, and customizable tools and data sets were recommended for these courses to allow instructors to optimize the tools for their unique introductory biology courses [8].

The rapidly changing and dynamic nature of bioinformatics makes it specially challenging for instructors, particularly those who do not regularly use bioinformatics in their research programs. Much of the material to be covered in lectures and software to be used in computer laboratory activities are not likely part of the instructor's own formal training experience. Addressing this challenge of instructor knowledge will likely require a multi-fronted approach including: (i) expanding the availability of short (e.g. several days to several weeks) courses focusing on the conceptual background and educational strategies for implementing specific aspects of bioinformatics into courses and curricula; and (ii) the development of lecture material and faculty-customizable web-based student learning activities and associated background materials for deepening faculty understanding. Some training

opportunities for educators do exist through Bioquest, American Society for Microbiology/DOE Joint Genome Institute (ASM/JGI), National Computational Science Institute (NCSI) and DNA Subway [9, 20–22], and these are primarily focused on introductory skills in bioinformatics to be incorporated into courses. However, these educator-training opportunities are neither sufficient for increasing incorporation of bioinformatics more broadly into undergraduate curricula, nor do they provide the range of training options (from beginner to advanced) to develop the skill sets required for broad incorporation into the curriculum.

Breadth of bioinformatics, and the diversity of students and educational objectives

One challenge is the lack of user-friendly interfaces for bioinformatics tools appropriate for upper-level undergraduate courses, as the majority of resources are aimed at the introductory level, and do not fit the diversity of courses offered. Even these resources, however, are often out-of-date because of frequently changing web sites and interfaces. These weaknesses impact both student and instructor, and are a challenge that must be overcome for training future generation of scientists. As a field, we cannot rely solely on separate courses in bioinformatics to train our current students. To incorporate bioinformatics into diverse classes within the curriculum, we must recognize the challenges of static learning activities available on web sites serving diverse needs. Courses and programs have very specific learning outcomes, populations of students and strength areas. These characteristics limit the potential of a predetermined activity developed for broad application to fit the needs of the diverse audiences. At a minimum an increased number of activities needs to be provided.

FUTURE DIRECTIONS FOR IMPROVING BIOINFORMATICS EDUCATION
Customizable web-based activities for bioinformatics

Broader incorporation of bioinformatics in education requires overcoming the challenges to using transformative computer-requiring learning activities, and to assist faculty in collecting assessment data on mastery of student learning outcomes. One possible approach to overcoming infrastructure-related

challenges is through use of software and cyberinfrastructure via the web. This would reduce the complexity of the computing environment and expertise required to a web browser. Rather than categorize courses, there should be recognition that each course is different, because of dissimilarities in student populations, curricula and faculty. Therefore an appropriate model might be a set of customizable web-based bioinformatics activities where faculty can choose features, questions, complexity/sophistication of background information and other characteristics to fit their students and teaching style. Such flexibility would overcome a principal problem with many 'canned' modules, web sites, activities, textbooks and other teaching aids that are currently available: they are fixed. Assessment methods that document the mastery of student learning outcomes, and provide feedback to students and faculty members should be included. This will create the interactive characteristics that students are demanding in education today.

Expanding training opportunities for bioinformatics and computational biology

One training opportunity for faculty teaching bioinformatics, albeit rather specialized, is the Workshop on Molecular Evolution [23]. An intensive 2-week course, the Workshop is designed for established investigators, faculty, postdoctoral scholars and advanced graduate students with prior experience in molecular evolution and related fields. Potential to improve teaching and incorporate Workshop-related materials into courses are among the criteria for admission. Workshop faculty members are chosen exclusively for their effectiveness in teaching theory and practice in molecular evolution. Included among the faculty are developers and other experts in the use of computer programs and packages who provide demonstrations and consultations. There are no handouts or other printed materials distributed at the course, and the web site [23] helps prepare participants in advance of their participation in the workshop, is a resource for participants while in attendance at the workshop, provide a means to maintain and increase learning after participating in the workshop, and provides a resource for those who do not attend the workshop.

Achieving computing environment consistency

The Workshop also provides some examples about overcoming other challenges to the broader incorporation of bioinformatics in education. Recognizing that people are more comfortable using their own computers, and this is especially true for the Workshop, which typically has participants from 24 to 33 countries spanning a range of native languages and keyboard layouts, we work with whatever computers people bring to the course. We achieve a sufficient degree of computing environment consistency by having everyone use UNIX or Linux, which is a target environment for all the software used in the course. We provide a flash drive to each participant with all the necessary programs and files for Mac OS X, Linux and for Windows users, a bootable customized Linux OS. During the first computer laboratory session we ensure the software (including Linux where appropriate) is properly installed. In a subsequent session we give an introduction to UNIX/Linux, shell commands and Emacs, and these sessions are all supported by web pages. Although several participants each year express reservations about learning UNIX/Linux, shell commands and Emacs, none of the more than a thousand participants expressed concerns about these issues in course evaluations. Indeed, subsequent feedback has all been positive as people feel they have overcome a major hurdle to using many bioinformatics programs. This approach provides a successful example of establishing a common base of technical competence in a diverse student population that allows for independent work in bioinformatics. For broad improvement of bioinformatics education at all levels, more faculty development opportunities that span diverse skill levels need to be developed, with an emphasis placed on providing resource materials that are kept up-to-date as the field and tools change.

Key Points

- More emphasis needs to be placed on education as bioinformatics as a field develops and advances.
- Educational opportunities need to span the range from beginner to advanced, with an emphasis on incorporating material into classes to correspond to learning outcomes.
- Barriers for educators need to be removed so that bioinformatics is more readily applied to a wide range of students and courses.

References

1. Huerta M, Downing G, Haseltine F, *et al*. *NIH Biomedical Information Science and Technology Initiative Consortium Definition Committee*. http://bisti.nih.gov/docs/CompuBio Def.pdf (17 July 2000, date last accessed).

2. Brazas MD, Yamada JT, Ouellette BFF. Evolution in bioinformatic resources: 2009 update on the Bioinformatics Links Directory. *Nucleic Acids Res* 2009;**37**(**Suppl 2**):W3–5.

3. Donovan S. Big data: teaching must evolve to keep up with advances. Correspondence. *Nature* 2008;**455**:25.

4. National Research Council, Committee on Undergraduate Education to Prepare Research Scientists for the 21st Century. *Bio2010: Transforming Undergraduate Education for Future Research Biologists*. Washington, DC: NRC, 2003.

5. Bednarski AE, Elgin SCR, Pakrasi HB. An inquiry into protein structure and genetic disease: introducing undergraduates to bioinformatics in a large introductory course. *CBE Life Sci Educ* 2005;**4**(3):207–20.

6. Holtzclaw JD, Eisen A, Whitney EM, *et al*. Incorporating a new bioinformatics component into genetics at a historically black college: outcomes and lessons. *CBE Life Sci Educ* 2006;**5**(1):52–64.

7. Honts JE. Evolving strategies for the incorporation of bioinformatics within the undergraduate cell biology curriculum. *Cell Biol Educ* 2003;**2**:233–47.

8. Neumann M, Provart N. Using customized tools and databases for teaching bioinformatics in introductory biology courses. In: O'Donnell MA (ed). *Tested Studies for Laboratory Teaching. Proceedings of the 27th Workshop/Conference of the Association for Biology Laboratory Education [ABLE]*, Vol. 27. ABLE, 2006, 321–8.

9. Bioquest. http://bioquest.org/ (23 April 2010, date last accessed).

10. Molecular Science Student Work Bench. http://medsocnet.ncsa.illinois.edu/MSSW/ (23 April 2010, date last accessed).

11. Statistics in Bioinformatics. http://gmein.uib.es/bioinformatica/statistics/ (23 April 2010, date last accessed).

12. Ma J, Nickerson JV. Hands-on, simulated, and remote laboratories: a comparative literature review. *ACM Comput Surveys* 2006;**38**:(article 7), 19pp.

13. Merrill MD. First principles of instruction. *Educ Technol Res Dev* 2002;**50**:43–59.

14. Genes and Disease. http://www.ncbi.nlm.nih.gov/book shelf/br.fcgi?book=gnd (1 May 2010, date last accessed).

15. Smith GJD, Vijaykrishna D, Bahl J, *et al*. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* 2009;**459**:1122–5.

16. Green RE, Krause J, Ptak SE, *et al*. Analysis of one million base pairs of Neanderthal DNA. *Nature* 2006;**444**:330–6.

17. Miller M, Drautz DI, Ratan A, *et al*. Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* 2008;**456**: 387–90.

18. Noonan JP, Coop G, Kudaravalli S, *et al*. Sequencing and analysis of Neanderthal genomic DNA. *Science* 2006;**314**: 1113–8.

19. Glenn M. The future of higher education: how technology will shape learning. *A Report from the Economist Intelligence Unit Sponsored by the New Media Consortium*. London: Economist Intelligence Unit, 2008;28pp.

20. ASM/JGI Bioinformatics Institute. http://www.facultypro grams.org/page02a.shtml (23 April 2010, date last accessed).

21. NCSI. http://www.computationalscience.org/workshops. 2010 (6 May 2010, date last accessed).

22. DNA Subway. http://dnasubway.iplantcollaborative.org/ (25 April 2010, date last accessed).

23. Workshop on Molecular Evolution. http://www.molecular evolution.org/workshops/ (1 May 2010, date last accessed).