University bioinformatics programs on the rise

Fueled by strong demand from students and industry's need for trained bioinformaticists, universities are increasing their offerings in this fast-growing field.

Randy J. Zauhar

Information is found in the arrangement of things with respect to each other. Whether we observe the arrangement of ink on the page to form words, or the order of nucleotide bases to form a gene sequence, we recognize that information is the key to understanding the world around us. Bioinformatics studies how information is stored, reproduced, and used by living systems. It is not an overstatement to say that bioinformatics is what biology is evolving to become in the 21st century.

The subject matter of bioinformatics has existed as long as there has been molecular biology, but it has emerged as a distinct discipline only in the last two decades. This was prompted by a veritable explosion of information triggered by the innovations of recombinant DNA and DNA sequencing, which have culminated in the human genome project (HGP). Of course, the HGP is only the most visible of a host of sequencing projects aimed at a variety of species, ranging from microorganisms to plants and animals, which will produce a torrent of data for the foreseeable future. As the time needed to store, annotate, and analyze this stream of information has grown from a laboratory pastime to a full-time job, bioinformatics has, by necessity, emerged as a separate and vital field of study. There is expected to be a dramatic increase in demand for practitioners in this field, who are already in short supply.

Today, the typical bioinformaticist is by necessity largely self-taught. Although it has always been true that a savvy molecular biologist could take computer science courses (and the computer scientist attend a biochemistry class), organized programs of study that combine life and information sciences have arisen only in the last decade, and bioinformatics is still an oddity in the university curriculum. The central difficulty in establishing a bioinformatics curriculum is not only that the subject is cross-disciplinary, but that the disciplines it unites are so disparate in method, outlook, and culture. The ideal situation, establishing an independent bioinformatics department with its own specialized faculty, is not yet possible at most institutions; instead, it is necessary to identify and bring together an interested core of faculty in life and information sciences, hopefully with genuine experience and interest in bioinformatics.

Despite the barriers between the life and computing sciences, successful programs have risen, and are growing steadily. Rensselaer Polytechnic Institute is recognized as offering the first undergraduate degree in bioinformatics, and also has a strong graduate program; George Mason University has an outstanding graduate curriculum, with diverse and fascinating course offerings; and Georgia Tech has recently entered the fray with a graduate program that has a strong start and promises to expand quickly. Many other institutions are now contemplating programs in bioinformatics (or equivalently, in computational biology), and many more will soon offer degrees in computer or information science with a focus in bioinformatics. At each of these institutions there is one or more bioinformaticists, "legitimate" by reason of experience or research area, who have served as a center for organizing faculty around the new program.

The success of any of these graduate-level programs will depend on two important factors. First, what potential student population are they targeting: biologists who want to learn computers, or computer scientists/programmers who want to learn biology? Are the incoming students fresh from their undergraduate education, or are they seasoned professionals, perhaps already holding a PhD? Second, at what level are the students to be educated: as independent research scientists, or as skilled programmers with a focus in analyzing biological information? Both of these considerations are important in designing the curriculum, and in helping the students find rewarding and productive employment after graduation.

The target audience

Bioinformatics is an attractive field for professionals in the life sciences seeking a career transition, and these individuals have the makings of outstanding graduate students, given their hands-on experience and proven work ethic. Despite this, there is anecdotal evidence that the typical computer science department does not offer a warm welcome to the humble biologist who wants to learn about computing. Their focus is usually on research in algorithm development and computing architectures, and they have little patience with a PhD molecular biologist who "once had a course in Pascal". This is unfortunate, because the biologist would bring a different perspective and a wealth of practical experience. At the same time it is understandable; the computer scientist naturally wants a student who already speaks the language of his discipline.

The reverse migration of computer scientists to biology is almost unknown, mainly because of the intrinsic value of the programming skills of the computer scientist, and the expectation (or hope) that the he or she can quickly learn enough biology to "get by". At the laboratory bench, this inevitably leads to a communication gap, a difference in vocabulary that requires painful efforts to bridge. Clearly, although the computer scientists have a distinct advantage over the biologists simply by having facility with computers, they are still "stunted" as bioinformaticists. They also need formal, cross-disciplinary education if they are to reach their full potential in the field.

A successful bioinformatics program would ideally target both of these potential student populations, but in practice the strengths of a program will initially lie in the skills and experience of the faculty members who have formed it. One program may be dominated by computer scientists, and another by bioinformaticists whose roots lie in biology. But there is little doubt that it is the returning biologist in search of a career transition, or the newly graduated life science major who is fascinated by computers, who represents the future growth in graduate bioinformatics education. There is also little doubt that it is precisely this group of life scientists who are being underserved by the existing programs.

The curriculum

The successful bioinformatics curriculum must fulfill the difficult requirement of making both the incoming biologist and computer scientist comfortable in a territory that will initially be foreign and threatening for both. The biologist must quickly gain a foothold in the world of computing, learn about algo-

Randy J. Zauhar is associate professor of biochemistry and director of the graduate program in bioinformatics, Department of Chemistry and Biochemistry, University of the Sciences in Philadelphia, 600 S. 43rd Street, Philadelphia, PA 19104-4495 (r.zauhar@usip.edu).

Table 1. Some degree programs in bioinformatics/computational biology

School	Degree	URL
Boston University (Boston, MA)	MS/PhD	http://bioinformatics.bu.edu/
George Mason University (Fairfax, VA)	MS/PhD	http://www.science.gmu.edu/~michaels/Bioinformatics
Georgia Tech (Atlanta, GA)	MS	http://www.biology.gatech.edu/bioinformatics/
Iowa State University (Ames, IA)	PhD	http://www.bcb.iastate.edu/
Medical College of Wisconsin/Marquette Univ. (Milwaukee, WI)	MS	http://goliath.ifrc.mcw.edu/AP/announcements/initial.html
Rensselaer Polytechnic Institute (Troy, NY)	BS	http://www.rpi.edu/dept/bio/info/bioinformatics.html
Rutgers University (New Brunswick, NJ)	PhD	http://cmb.rutgers.edu/
University of the Sciences in Philadelphia (Philadelphia, PA)	MS	http://www.usip.edu/bioinformatics/

rithms and how to implement them in C/C++, Java, and PERL, and gain experience in constructing big databases. This can be intimidating for someone whose computing experience has been limited to Netscape and PowerPoint. On the other hand, the computer scientist must enter the squishy and complex world of living matter and experimental uncertainty, perhaps a frightening neighborhood for someone accustomed to tackling well-defined problems with rigorous logic. Both of our budding bioinformaticists must learn the computing techniques that form the foundation of current practice, including BLAST and other alignment algorithms, techniques for gene identification, and tracking profiles of protein expression—everything from the arcane (hidden Markov models) to the pedestrian (reformat that file with a PERL script!).

Whether the goal is to produce an independent research scientist or an efficient player in an industrial team, the fundamental education should be much the same, as it usually is when we compare MS and PhD scientists in the same discipline. The difference lies in the nature of the practical experience that follows the book learning. For a PhD, the emphasis is on innovation and creative thinking, whereas the expectation for the MS-level scientist is to be highly competent in applying a base of knowledge, and to be able to keep that base of knowledge current. For those interested in a career in independent research, the obvious road is a department that offers a PhD, and the expectation should be for a curriculum that is relatively heavy on computation and light on biology.

Our experience so far

At the University of the Sciences in Philadelphia (USP), we have just initiated an MS-level graduate program in bioinformatics; our undergraduate program will begin in Fall 2001. Our core faculty consists of this author, bringing the experience of having managed a bioinformatics computing facility along with background in commercial scientific software development, and Jim Pierce, a molecular biologist with much experience in experimental genomics and the practical application of bioinformatics software. We are currently recruiting an additional faculty member in the department of mathematics, physics, and computer science who will assist in the program.

It seems we will not have difficulty in finding students: at the end of our first academic year, we will have seven full- and part-time graduate students—all achieved with essentially no advertising (a formal marketing campaign will begin soon). As one might guess from the tone of my previous remarks, we are indeed targeting the somewhat ignored professional life scientist interested in a career transition, although we are beginning to receive inquiries from recent graduates with bachelor's degrees who want to begin graduate studies next year.

We believe that our curriculum is very well balanced, even for an incoming computer scientist. All of our students must complete (or show equivalency to) a two-semester "capstone sequence", the first semester covering genomics, the second being a crash course in bioinformatics algorithms and software development. Students must also complete an introductory course on biotechnology, which discusses modern experimental techniques, followed by an intensive hands-on laboratory course. In the second year, students will take courses in advance algorithms for bioinformatics, and systems analysis and design (focusing on client-server and web-based development).

An important feature of our program is the completion of a practical project over two semesters. Although the project can be kept in-house at USP, ideally it will be done in collaboration with an outside industrial or academic partner. The project must involve development of a software application to solve a significant real-world problem. For example, one of our students plans to develop custom software for displaying and analyzing microarray expression data. This project has evolved naturally as part of his job at a local research hospital, and his supervisor at the hospital will jointly oversee his MS project. This is exactly the sort of scenario we are aiming for; it provides the student with a solid educational background and significant practical experience to reinforce his or her résumé.

Conclusions

Bioinformatics is one of the key sciences of the new century, and demand for specialists in the area is strong and growing. The opportunity to train this new generation of scientists is one that we dare not ignore if we are to remain relevant to both our potential students and their future employers. Universities should view their participation in the molding of the New Biology not only from the narrow standpoint of filling a "market niche", but as a privilege and a chance to make a concrete contribution to the new era.

Table 2. Degree programs that offer specialization/emphasis in bioinformatics

School	Degree	URL
Baylor University (Waco, TX)	BS/Informatics	http://ecswww.baylor.edu/ecs/computer_science/csundergrad.htm
North Carolina State Univ. (Raleigh, NC)	MS, PhD/Genomics	http://genomics.ncsu.edu/
University of California, San Diego	PhD/Comp. Sci., Eng., Math., Chemistry, Physics, etc.	http://www.ogsr.ucsd.edu/bioinformatics/program.htm
University of Pennsylvania (Philadelphia, PA)	BS/MS/PhD, Computer Sci., Biotech., Biology, Math.	http://www.cbil.upenn.edu/UPCB/
University of Pittsburgh/ Carnegie Mellon (Pittsburgh, PA)	PhD/Computer Sci., Biol. Sci., Chemistry, Physics, etc.	http://www.cs.pitt.edu/keck/program.html
Washington University (St. Louis, MO)	PhD/Medicine, D.Sc./Biomed. Eng.	http://www.ibc.wustl.edu/CMB/