

FOCUS: EDUCATING YOURSELF IN BIOINFORMATICS

Development of an Undergraduate Bioinformatics Degree Program at a Liberal Arts College

Paramjeet S. Bagga

Bioinformatics, School of Theoretical and Applied Science, Ramapo College of New Jersey, Mahwah, New Jersey

The highly interdisciplinary field of bioinformatics has emerged as a powerful modern science. There has been a great demand for undergraduate- and graduate-level trained bioinformaticists in the industry as well in the academia. In order to address the needs for trained bioinformaticists, its curriculum must be offered at the undergraduate level, especially at four-year colleges, where a majority of the United States gets its education. There are many challenges in developing an undergraduate-level bioinformatics program that needs to be carefully designed as a well-integrated and cohesive interdisciplinary curriculum that prepares the students for a wide variety of career options. This article describes the challenges of establishing a highly interdisciplinary undergraduate major, the development of an undergraduate bioinformatics degree program at Ramapo College of New Jersey, and lessons learned in the last 10 years during its management.

INTRODUCTION

Bioinformatics is a highly interdisciplinary modern science that combines life sciences and computer science with information technology. The National Institutes of Health (NIH[†]) defines bioinformatics as “research, development, or application of computational tools and approaches for ex-

panding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data.” According to the National Center for Biotechnology Information (NCBI), “the ultimate goal of the field is to enable the discovery of new biological insights as well as to create a global

To whom all correspondence should be addressed: Paramjeet S Bagga, 505 Ramapo Valley Road, Mahwah, NJ 07430; Tele: 201-684-7722; Fax: 201-684-7637; Email: pbagga@ramapo.edu.

[†]Abbreviations: NIH, National Institutes of Health; NCBI, National Center for Biotechnology Information; IBD, inflammatory bowel disease; CAGR, Compound Annual Growth Rate; RCNJ, Ramapo College of New Jersey; TAS, school of Theoretical and Applied Science; BLAST, Basic Local Alignment Search Tool; SCP, School Core Requirements.

Keywords: bioinformatics, genomics, undergraduate education, undergraduate curriculum

perspective from which unifying principles in biology can be discerned.” It is hard to define interdisciplinary sciences because of their inherently dynamic scope. However, via integration of multiple disciplines, the science of bioinformatics offers powerful research tools that would be difficult for individual disciplines to provide on their own. Bioinformatics is already practiced in the fields of molecular biology, personalized and preventive medicine, biotechnology, and nanobiotechnology [1-5]. The information obtained by the application of this powerful technology can potentially be applied to prevent, diagnose, and treat inherited and other diseases, design new drugs and vaccines, and help shorten the development cycle of new medicines [1,6-10]. Bioinformatics-based genome studies have completely transformed cancer research in recent years [11]. Application of bioinformatics has the potential to revolutionize the way scientific research and medicine is done in the 21st century.

The post-genomic era has brought significant changes in the life sciences and associated fields. Owing to lower costs and technological advances, large amounts of data are coming out of genomic sequencing projects [12,13]. Consequently, there has been an explosion of biological information being generated [14]. A variety of many other efforts such as high throughput gene expression profiling, miRNA profiling, and biological data-mining also result in large datasets [15-17].

A wide variety of databases store well-organized information that is continuously being generated by scientific research and by further computations of the data coming out of research projects [18-21]. There are several databases that hold information about genomes [22] and genes [23], gene expression data [24], or a variety of information on proteins [25] and biochemical pathways [26]. Databases are very useful for analyzing genomic variations and regulated gene expression, analyzing structure and function of proteins, or studying complex relationships among biochemical reactions in the cell. New databases are regularly being

developed to accommodate biological data from novel efforts.

Bioinformatics-based mathematical and computational methods are required to manage, analyze, interpret, and draw conclusions from experiments that result in large and compound datasets with multifaceted relationships such as those observed in biochemical pathways. Many new methods are being regularly developed to address the complex needs of the rapidly evolving field of bioinformatics and associated sciences. Comparative genomics, proteomics, biochemical pathway modeling associated with systems biology, and many other bioinformatics approaches have been applied in recent times to address biological problems that were beyond reach with traditional techniques [27-29].

Bioinformatics methods are already proving very useful for scientific research, improved diagnosis, and treatment of diseases. For instance, bioinformatics researchers have used innovative strategies such as video gaming to resolve the structure of a retroviral protease, opening up new avenues for AIDS drug design [30]. Microarray analysis has already begun to help correctly identify carcinomas and help decide streamlined approaches for personalized treatments. Large datasets that result from microarray functional genomics experiments are also used for extraction of useful information that could be directly applied in medicine. For example, based on computational analysis of the gene expression profiles, researchers were able to successfully identify new therapeutic uses for existing medicines. The “drug repositioning” strategy resulted in prediction and experimental validation of cimetidine, an antilucer drug for use in lung adenocarcinoma [31]. Using a similar approach, anticonvulsant drug topiramate has been indicated for treatment of inflammatory bowel disease (IBD) [32].

The importance of modern interdisciplinary bioinformatics science cannot be overemphasized. As a result, there has been a great demand for undergraduate- and graduate-level trained bioinformaticists in the industry as well in the academia. The global

bioinformatics market, estimated at USD 3 billion in 2010, is expected to grow at Compound Annual Growth Rate (CAGR) of 25 percent to reach USD 8.3 billion within the next 3 years [11,33]. Bioinformatics and related fields of proteomics and genomics have been the top contributors to advances in drug discovery and are projected to be the major contributors for the next decade [34]. Therefore, the demand for individuals with training in bioinformatics seems justified. A search of simplyhired.com found that roughly 10 percent of all science jobs require bioinformatics skills. According to this web service, the average salary of a bioinformaticist in New York area was \$98,000, and for a senior bioinformaticist, it was \$154,000. These salaries were much higher than those of comparable positions in other sciences. The jobs-salary.com posted 45 positions requiring bioinformatics skills with salaries approaching \$112,000. Similarly, glassdoor.com, another job posting website, posted 66 positions requiring bioinformatics skills in 21 companies with salaries in the range of \$90,000.

Several graduate programs have been developed in bioinformatics in recent years to satisfy the needs of the job market. However, it is difficult to find suitable candidates for graduate programs. Most prospective graduate students either have a biology background or computer science skills, but not both. Availability of undergraduate level instruction in bioinformatics cannot only start building up a highly desirable trained workforce, it can also better prepare the students for graduate programs. Several elite research institutions already have an undergraduate degree program related to the bioinformatics field or include courses in bioinformatics, genomics, computer programming, mathematical biology, and systems biology in their undergraduate biology curricula. Very few undergraduate degree colleges have been able to develop instructional programs in bioinformatics. (Please see [35] for a list of bioinformatics and related educational programs in the United States and rest of the world). Albeit, a majority of students in the United States gets their education at either associate-degree

colleges or four-year bachelor's degree granting institutions. Of the 1,650 institutions offering baccalaureate and higher degrees in the country, 1,388 are primarily non-PhD-granting educational institutions [36]. If we are to address the needs for trained bioinformaticists, a bioinformatics curriculum must also be offered in the numerous four-year colleges and not just at the universities with doctoral programs.

This article at first describes the challenges of establishing a highly interdisciplinary undergraduate major in a typical four-year college and the development of an undergraduate bioinformatics degree program at a liberal arts college in New Jersey, as well as lessons learned in the last 10 years during its management.

CHALLENGES OF DEVELOPING AN UNDERGRADUATE BIOINFORMATICS CURRICULUM

Interdisciplinary programs face many challenges, especially at four-year educational institutions. It is hard to find professionals with bioinformatics background who are willing to teach in undergraduate programs that offer modest remunerations as compared to industry compensations. According to a recent genomics-based salary survey [37], median salary range for a senior scientist in academia is \$50,000 to \$75,000, whereas this amount for an equivalent position in the pharmaceutical is \$100,000 to \$125,000. Nevertheless, the undergraduate institutions have generally recognized the importance of establishing such programs and have had some success attracting dedicated educators or encouraging existing faculty members to obtain appropriate training. The next challenge is in designing an interdisciplinary science curriculum that will be able to accommodate additional bioinformatics courses without compromising education in the fundamentals of biology, chemistry, and physics. In fact, bioinformatics courses require biology, chemistry, mathematics, and computer science as foundations on which the higher level of interdisciplinary instructions need to be built. Therefore, interdisci-

iplinary programs like bioinformatics are inherently larger than other disciplines. Most undergraduate institutions have an upper limit for the total number of credits (generally in the range of 120 to 130) for graduation. Students at many institutions have to satisfy college core and then their specific school core requirements in addition to general education course credits. These requirements pose a limit on the number of credits that can be assigned to the science courses. Whereas liberal arts institutions provide a well-rounded educational experience in science that include humanities and other general education courses, they have limited room for curricular expansion.

In the absence of established guidelines or accreditations in this modern field, it is hard to decide which courses to include in the bioinformatics curriculum. Bioinformatics is a young science, and there have been rapid developments in this field. Since the scope of this science and the jobs that require its expertise is constantly changing, the undergraduate curricula need to be dynamic. The rigorous process associated with curricular changes makes it difficult for the undergraduate programs to adapt in a timely fashion.

The biggest challenge perhaps comes from a general lack of interdisciplinary collaboration among colleagues at undergraduate institutions. The school of science at these institutions is usually divided into individual departments, such as chemistry, biology, mathematics, and computer science. Establishment of a major in bioinformatics requires cross-department partnerships, mutual cooperation in sharing resources, and collaborative development of new interdisciplinary courses. Boundaries of the departments are generally inhibitive of interdisciplinary alliances.

Attracting students to the undergraduate bioinformatics program can also be a challenge. The concept of bioinformatics is generally not discussed at the high school level. Therefore, high school seniors are usually not aware of the significance or even existence of this interdisciplinary science. Most incoming undergraduate freshmen

tend to pick the traditional majors in biology, biochemistry, chemistry, mathematics, physics, or computer science. Consequently, the small number of undergraduate bioinformatics programs that do exist have had difficulty enrolling students despite a great demand for trained bioinformaticists in academia and industry. Many students who do get attracted to the programs due to impressive placement rates may not be prepared to handle a broad spectrum of courses native to interdisciplinary curricula, resulting in dropouts and low retention rates.

INTEGRATION OF BIOINFORMATICS INTO EXISTING UNDERGRADUATE COURSES AND CURRICULA

Despite the inherent challenges in establishing highly interdisciplinary undergraduate curricula in bioinformatics, there have been many successful efforts in the right direction. One way to get started is to “infuse” concepts of bioinformatics into existing courses. This concept originated at Wheaton College 10 years ago [38]. Bioinformatics has become an essential component of modern life sciences. Therefore, it makes sense to discuss its principles and applications in existing courses. For example, courses in genetics, molecular biology, and biochemistry, which are taken by almost all life science majors, essentially discuss gene mapping, mutations, gene expression, genetic diseases, biochemical pathways, and structure and functions of nucleic acids and proteins. Bioinformatics methods and databases are intimately associated with all of these concepts and can easily be integrated into these courses to provide up-to-date knowledge and skills to undergraduate students. There have been several successful efforts using a variety of strategies for incorporating bioinformatics into the existing courses [39-41].

The concept of “linking courses,” which also originated at Wheaton College [42], is a great way to demonstrate cross-disciplinary connections between concepts taught in two independent courses. For example, a biology course that covers se-

quence comparisons can easily be linked to a computer programming course. The educators can give a number of reciprocal guest lectures and combine selected labs of the courses. The idea is for students of both the courses to understand the biological problem and collaboratively come up with computational solutions. This could be a class project for students working in teams consisting of a biology and a computer science student each. Besides enhancing learning through project-based instruction, this exercise will help encourage interdisciplinary partnerships among students as well as faculty, thereby opening up new avenues in pursuing undergraduate bioinformatics education. "Linked-courses" strategy has been successful at Wheaton and is being tried at other institutions [43].

Many undergraduate degree programs, such as the Genetics program at Rutgers University in New Jersey, Biology programs at Spelman College in Georgia, and Wheaton College in Massachusetts, have developed individual courses in bioinformatics and related disciplines. Some models of bioinformatics courses are available [44-47]. These courses are commonly used as electives in the existing curricula. Some of the undergraduate programs in life sciences have gone a step further and developed tracks in bioinformatics and genomics. These tracks provide a great alternative to students who were not aware of the significance of the bioinformatics field initially and later decided to include them in their curricula.

DEVELOPMENT OF UNDERGRADUATE CURRICULA IN BIOINFORMATICS

In recent years, several new undergraduate degree programs have been developed by overcoming many challenges. Since there are no nationally or internationally accepted guidelines, the early undergraduate bioinformatics curricula come in a wide variety of flavors. A comprehensive list of undergraduate and graduate programs in bioinformatics is available [35].

It is tempting to quickly develop undergraduate programs in bioinformatics by putting together existing courses from different disciplines. However, there is a need to carefully design a well-integrated and cohesive bioinformatics curriculum that prepares the students for applying their problem-solving skills in research and addresses the needs of academia and industry. The concept of "infusing" principles of bioinformatics, genomics, and proteomics, as appropriate, into the contents of existing courses before adopting them for a program in bioinformatics has much merit. Development of several carefully designed new courses in the core curriculum to build cohesiveness and integrate knowledge and skills learned from basic science courses is instrumental to the success of the program. Availability of a variety of elective courses helps enrich the program and helps develop specialized skills of students' interest.

Bioinformatics Undergraduate Degree Program at RCNJ

In 2002, Ramapo College of New Jersey (RCNJ) became the first in the state and one of the first primarily undergraduate colleges in the entire United States to offer a baccalaureate degree in bioinformatics. It all started with a bioinformatics course offering in 2000 as an upper level elective for the existing biology and biochemistry curricula. The success of the bioinformatics course and a tremendous student response led to the idea of establishing an entire program of study in this field. After many months of deliberations among faculty members from the biology, chemistry, mathematics, and computer science programs, a curriculum was agreed upon. The proposed bachelor of science degree in bioinformatics program went through a series of administrative and peer reviews before being officially approved by the state of New Jersey in 2002.

Fortunately, Ramapo College has a unique governance structure that provides a fertile ground for interdisciplinary efforts. Instead of division into individual departments, the RCNJ school of Theoretical and Applied Science (TAS) is comprised of

Table 1. BS in Bioinformatics Curriculum at Ramapo College of New Jersey.

| Bioinformatics Major Requirements | As of Fall 2012 |
|---|------------------------|
| Bioinformatics CORE | Credits |
| Bioinformatics courses (8 credits): | |
| BIIN 430 Bioinformatics Lec and Lab Integrated | 4 |
| BIIN 450 Advanced Bioinformatics | 4 |
| Biology courses (16 credits): | |
| BIOL 110/112 Fundamentals of Biology I & II Lec and Lab | 8 |
| BIOL 331 Genetics Lec and Lab | 4 |
| BIOL 406 Cell & Molecular Biology Lec and Lab | 4 |
| Computer Science courses (20 credits): | |
| CMPS 147/148 Computer Science I & II | 8 |
| CMPS 231 Data Structures | 4 |
| CMPS 364 Database Design | 4 |
| CMPS 345 Analysis of Algorithms | 4 |
| Chemistry/Biochemistry courses (12-16 credits): | |
| CHEM 110/111 Fundamentals of Chemistry I Lec & Lab | 4 |
| CHEM 112/113 Fundamentals of Chemistry II Lec & Lab | 4 |
| CHEM 210/211 Organic Chemistry I Lec and Lab and | 4 |
| CHEM 212/213 Organic Chemistry II Lec and Lab or | 4 |
| CHEM 205 Bio-organic Chemistry Lec and Lab | 4 |
| Math courses (12 credits): | |
| MATH 121 Calculus I | 4 |
| MATH 237 Discrete Structures | 4 |
| PSYC 242 Statistics | 4 |
| Electives (minimum 3 courses required) | |
| Group 1: Bioinformatics/Biology/Biochemistry (one course minimum) | |
| BIIN 350 Molecular Genetics | 4 |
| Two semesters of Research Honors (SRSH 301/302)/Independent Studies/CO-OP | 2-4 |
| BIIN 351 Protein Structure and Modeling | 4 |
| CHEM 425 Biochemistry | 4 |
| CHEM 445 Medicinal Chemistry | 4 |
| Group II: Computer Science / Mathematics (one course minimum) | |
| BIIN 210 Introduction to PERL Scripting | 4 |
| CMPS 331 Artificial Intelligence CMPS 342 Computer Graphics | 4 |
| CMPS 361 Software Design | 4 |
| CMPS 369 Web Application Development | 4 |
| MATH 122 Calculus II | 4 |
| MATH 122 Calculus II | 4 |
| Credit Calculations | |
| General Education | 32 |
| School Core Requirements (SCP) | 4 |
| Bioinformatics Core Requirements | 78-84 |
| Required Core courses | 68-72 |
| Required Electives | 10-12 |
| Free Electives | 8-14 |
| Total Credits for Graduation | 128 |

“convening groups” to manage individual majors like biology, chemistry, physics, mathematics, etc. There is significant resource sharing among the convening groups. Almost all science faculty members actively participate in multiple convening groups. For example, the bioinformatics convening group consists of faculty members drawn from biology, mathematics, chemistry, and computer science groups. The governance structure at Ramapo College helped circumvent many problems commonly faced in establishing an interdisciplinary curriculum.

The undergraduate bioinformatics program at RCNJ has evolved in the last 10 years and is now well established. Following sections describe the current state of the project.

RCNJ Undergraduate Bioinformatics Curriculum

The curriculum for the undergraduate bioinformatics program at RCNJ has been designed to lay a solid multidisciplinary foundation in modern life sciences, chemistry, and mathematics, as well as computer science using informatics as an interdisciplinary cohesive tool (Table 1). In advanced courses, students are trained to manage biological data, develop computational methods to analyze and interpret data, solve scientific problems, and make new discoveries.

The first year of study in this program is generally dedicated to introductory courses in sciences and mathematics. Students take two semesters each of fundamentals of biology (BIOL 110 and 112) and chemistry (CHEM 110-113). They also take at least one course in mathematics (Calculus I: MATH 121) and computer programming (CMPS 147). All the science courses also have associated lab sections that expose students to the concept of experimentation, data analysis and computation, data interpretation, and reporting. In their second year, students typically take two semesters of organic chemistry (CHEM 210-213) along with associated labs, statistics (PSYC 242) and discrete mathematics (MATH 237) courses, a genetics course (BIOL 331) with lab, and a

mid level computer programming course (CMPS 148). All along, students are assigned professional advisors to ensure that they stay on course with the curriculum. They are also encouraged, through advisement, to take courses in physics. This year essentially prepares the students for higher level junior and senior courses in which they will be able to integrate the concepts learned from chemistry, biology, and mathematics and start applying computer programming strategies to address biological problems.

In their junior year, students will need to take an upper level cell and molecular biology course (BIOL 406) with lab, covering the current state of knowledge about the cells and their biology at molecular level, and a bioinformatics course (BIIN 430), which covers current topics in bioinformatics, including sequence analysis and alignments, phylogenetic analysis, structural bioinformatics, genomics, and proteomics. By this point, each student has been assigned a bioinformatics faculty member as an adviser. In this year, students also take upper level courses like data structures (CMPS 231) and database design (CMPS 364) and are encouraged by the faculty advisors to take a course in Perl programming (BIIN 210). The junior year trains them well to integrating their knowledge and skills gained from diverse scientific disciplines of biochemical, mathematical, computational, and life sciences. By this time, the students are aware of latest advances in bioinformatics, understand key biological problems, and know how to apply their technical skills to seek solutions and manage and analyze a variety of biological data, including large data sets.

The courses in the senior year of the RCNJ undergraduate bioinformatics program help prepare the students for graduate studies and jobs in the industry. The majors will take a high level bioinformatics algorithms course (CMPS 345), followed by the advanced bioinformatics course (BIIN 450), which is designed to provide a capstone experience. Besides covering current topics in bioinformatics research, the advanced bioinformatics course involves working on major

research projects for designing computational tools and databases. Students also take several electives from two groups of courses (Table 1) to enhance their knowledge and skills in the areas of their choice and to complement the core curriculum. They are highly encouraged by their faculty advisors to take a course in protein structure and modeling (BIIN 351) and an upper level course in software development (CMPS 361 and/or CMPS 369).

Interdisciplinary curricula tend to be generally larger than the traditional majors. The RCNJ bioinformatics core curriculum ranges from 78 to 84 credits based on the choices students make (Table 1). The comprehensive program helps students to be prepared for a wide variety of challenges in scientific careers.

In addition to the scientific and technical concepts, RCNJ bioinformatics majors are also exposed to the social, business, and ethical aspects of science as part of the school of science core and college general education courses. There is a key advantage in having an interdisciplinary major like bioinformatics at a public liberal arts institution like Ramapo College. It offers a well-rounded curriculum leading to a comprehensive experience of undergraduate education, which, in addition to equipping them with highly technical bioinformatics skills, also helps to create coherence and integrity in the students and preparation for life ahead.

Project-Based Instructions

Project-based laboratories, especially those associated with interdisciplinary courses, provide opportunities for adequately comprehending cross-discipline integrated knowledge inherent to such courses and to develop invaluable skills for solving realistic biological problems encountered in research. Several mid- to upper-level courses in the undergraduate bioinformatics curriculum at RCNJ require independent research projects. These projects involve application of knowledge and skills learned during the semester to solve biological problems. Provided below are descriptions of three such

projects associated with courses of the bioinformatics curriculum.

Students of the junior level cell and molecular biology course (BIOL 406) spend the last three weeks of the lab section working on their projects. Every student is assigned a unique partial amino acid sequence of an apparently purified protein. Searching appropriate databases with the help of the BLAST (Basic Local Alignment Search Tool) [48,49] program, the students investigate if that protein has already been identified and previously studied. Once the protein is identified, they explore a wide variety of Internet-based molecular databases and other resources to compile information about the structure and function of their protein and its gene in significant details. The project reports are submitted in the form of well-designed and published websites dedicated to their gene and protein. They provide detailed information about the domain composition, three dimensional and quaternary structure of the protein, known functions of the protein, interactions with other proteins, and biochemical pathway involvements, as well as sub-cellular localization. They also provide information about alternative and tissue specific gene expression, involvement of the gene in diseases, and any treatments that may be available. The website is heavily hyper-linked to the Internet-based resources in a context-specific way. This project is quite popular, since it provides a real-life active learning opportunity to the students. A plus for the students is that they can showcase the published websites in their resumes.

Students of the junior level bioinformatics course (BIIN 430) are assigned at the end of the semester to individual research projects for the lab component of this course. They spend 3 weeks solving a “real-life”-like problem using bioinformatics skills learned during the semester. Each student is assigned a “novel” protein, which has “never” been studied before. (The instructor personally “designs” the proteins by “stitching” together amino acid sequences corresponding to different functional domains obtained from a variety of proteins. A sub-

cellular localization signal is also added.) Every student is assigned a unique protein. The goal of the project is to predict the structure and function of the protein using the instructor-provided partial amino acid sequence. Students use a variety of bioinformatics skills to predict the domain structure, secondary and tertiary structure, sub-cellular localization, and other properties of the protein using existing databases and web-based applications that are freely available. The project requires students to figure out a molecular function of the protein and any diseases it may cause when mutated. They get to name the protein themselves based on their findings. Although it is a lot of work, the students have fun playing detective.

The instructor of the advanced bioinformatics (BIIN 450), which is a capstone course for the BS in bioinformatics program, works closely with the students to assign independent research projects in the beginning of the semester. The semester projects are designed to answer a question and consist of original research. The research uses a bioinformatics/computational approach and usually includes development of new programs and databases. For example, one of the students worked on developing an application to computationally footprint genes using Bioperl and ClustalW. Another student worked on a method to correlate physical properties of binding sites of proteins that bind to similar ligands. This is considered a capstone or senior research project of the undergraduate bioinformatics curriculum that every student of the program needs to accomplish.

Undergraduate Research Opportunities and Internships

Learning to apply the knowledge and skills outside the classroom are pivotal to student success in the graduate school and for industry jobs. Research opportunities provide invaluable means of active learning. In the process, the students benefit by getting trained in the latest techniques, learn to design experiments, and thus become more marketable for jobs and higher studies. Con-

sequently, research is rapidly becoming an essential component of undergraduate education. Undergraduate level research experiences are particularly important for the students of interdisciplinary curricula because they help integrate cross-disciplinary knowledge effectively via direct application of interdisciplinary skills to solving real problems. Therefore, in addition to incorporating research into the curriculum, there is a need for the bioinformatics programs to provide sufficient extra-curricular research opportunities.

RCNJ bioinformatics faculty members are actively engaged in research and encourage involving undergraduate students in their research projects. Several of these research efforts represent cross-disciplinary collaborations, for example, between computer scientists and biologists. Interdisciplinary research efforts have led to greater collaborations for developing interdisciplinary courses and eventually the bioinformatics curriculum at RCNJ.

Owing to the growing interest in undergraduate research, the school of Theoretical and Applied Science (TAS) at RCNJ, in which the undergraduate bioinformatics program resides, has established a Research Honors program. It provides a structured system for students to regularly engage in independent research under the mentorship of faculty members for two to four semesters. This experience can count toward elective credits in the bioinformatics curriculum. Establishment of this research program has significantly increased the number of faculty-student research projects. Typically, bioinformatics undergraduates get involved in the research honors program after their sophomore year and continue working on their research projects until they graduate.

Faculty-student research endeavors have resulted in several top quality research projects in the bioinformatics program at RCNJ. For example, one of the projects involves studying G-quadruplex structural motifs as *cis*-regulatory elements of post-transcriptional gene expression. This project has resulted in the development of many bioinformatics web servers and databases

(<http://bioinformatics.ramapo.edu/GQRS/>) designed by collaborative efforts of faculty and undergraduate bioinformatics students. These tools help predict *cis*-regulatory motifs in the genomic and expressed nucleotide sequences and have been used to perform genome wide bioinformatics studies [50]. Other research projects in the bioinformatics program have involved, for example, exploration of ligand sites on the malarial parasite proteins using structural bioinformatics approaches, identification and functional characterization of genes involved in the regulation of bone mass, and evolution of proteins involved in cystic fibrosis.

The bioinformatics faculty-student research projects have resulted in more than 35 conference presentations, including several international conferences, by faculty and undergraduate students in the past 5 years alone. RCNJ bioinformatics undergraduates have, in fact, won multiple awards for their research accomplishments at regional conferences. In recent years, bioinformatics faculty and students have co-authored many publications in international peer-reviewed journals [50-53].

The school of Theoretical and Applied Science organizes its own student research conference every year. Bioinformatics undergraduates make good use of the local event to get their first conference experience at home and hone their presentation skills. In addition, the weekly journal club has proven a great way for students to learn about recent developments in bioinformatics and related fields.

In addition to the research opportunities on campus, the students are highly encouraged to seek off-campus training opportunities in the form of research internships and co-ops in academia and industry. These experiences are important for them to broaden their skills, apply the knowledge they have obtained in the classroom, and gain invaluable hands-on experience. Many RCNJ bioinformatics majors opt for paid summer internships at top-ranked research universities. For example, students have been able to intern in Cornell University's plant genome research project. Students have also been successfully

able to work in the local pharmaceutical industry to gain on-the-job work experiences.

Undergraduate research experiences provide a glimpse into the expectations of the graduate school and provide essential skills and the ability to succeed in doctoral programs. Research and internship experiences are also essential for marketability and success in post-baccalaureate jobs in the pharmaceutical and biotech industry. Prospective employers look for trainable candidates. Past experiences working on independent research projects are instrumental in determining the suitability of job candidates with desired skills and abilities.

Career Placements

The Ramapo College bioinformatics program prepares students to obtain employment in the industry and academia, pursue graduate education, careers in medicine and related disciplines, or engage in research in this technological field of great demand. The breadth as well as depth of knowledge obtained from a well-rounded, comprehensive interdisciplinary curriculum, combined with a variety of research experiences, help prepare the students for doctoral programs as well as industrial jobs. That is why perhaps the bioinformatics program has had impressive student placement record.

The majority of the RCNJ-trained bioinformaticists have been accepted into elite doctoral programs like those at the University of Pennsylvania, Cornell University, Albert Einstein College, Columbia University, and Yale University. RCNJ bioinformaticists have also been very successful in finding jobs in the pharmaceuticals and biotech industry.

Although the RCNJ bioinformatics curriculum is well balanced in terms of providing training in life sciences and computer programming, trained bioinformaticists generally show a dichotomy in their preferences. Some students are a bit more interested in the molecular biology as compared to the computer science. They tend to pick more elective courses related to their interest. On the other hand, some students prefer computer programming. RCNJ bioin-

formatics program has evolved in the last 10 years to provide sufficient background for every graduate to develop a balance of skills or be successful in either direction. RCNJ-trained bioinformaticists with biology preference have sufficient training in “wet” bench experimentation techniques to work as molecular biologists, and at the same time to communicate aptly with computer scientists. Computer science-oriented students are trained well enough to be successful as programmers as well as capable of working closely with “wet” bench research scientists. Consequently, RCNJ-trained bioinformaticists have been successful in working as bioinformaticists, computer programmers, and molecular biologists.

The undergraduate bioinformatics program prepares students for a variety of career choices, including medical school. The courses in the RCNJ bioinformatics curriculum are sufficient for students to prepare for careers in medicine. The curriculum contains most of the courses, such as biology, chemistry, organic chemistry, genetics, and molecular biology, needed for the pre-med track. Students can choose appropriate courses from the list of electives, such as biochemistry and higher level calculus, and take additional courses like physics as free electives within the credit limits for graduation from RCNJ with a BS in the bioinformatics program to prepare for medical school. Since bioinformatics research is directly applied to modern medicine, it only makes sense for bioinformaticists to pursue careers in this field.

CONCLUSIONS

Bioinformatics can be introduced into the undergraduate curricula by infusing its concepts into existing courses and developing tracks of bioinformatics in existing majors. However, there are many challenges in developing an undergraduate level bioinformatics program that needs to be carefully designed as a well-integrated and cohesive interdisciplinary curriculum that prepares the students for a variety of career options.

Several new undergraduate programs have been developed in recent years by

overcoming many challenges. Ten years ago, Ramapo College of New Jersey (RCNJ) became the first in the state and one of the first primarily undergraduate colleges in the United States to offer a baccalaureate degree in bioinformatics. Being a liberal arts institution, RCNJ offers a well-rounded and balanced curriculum leading to a comprehensive experience of undergraduate education.

The RCNJ undergraduate bioinformatics program has evolved over 10 years and has had many successes. It provides many opportunities for undergraduate research experiences inside and outside the classroom. As apparent from their successful placements, the graduates of this program are well prepared for doctoral programs as well as jobs in the industry. The undergraduate bioinformatics program successfully prepares students for a wide variety of career options.

REFERENCES

1. Real world applications of bioinformatics [Internet]. [cited 2012 June 29]. Available from: http://www.ebi.ac.uk/2can/bioinformatics/bioinf_realworld_1.html#1.4.
2. Adnan A. Importance and Applications of Bioinformatics in Molecular Medicine 2010 [Internet]. [cited 2012 June 30]. Available from: <http://www.biotecharticles.com/Bioinformatics-Article/Importance-and-Applications-of-Bioinformatics-in-Molecular-Medicine-214.html>.
3. Safdar M. Application of Bioinformatics in Medicine 2010 [Internet]. [cited 2012 June 29]. Available from: <http://www.biotecharticles.com/Bioinformatics-Article/Application-of-Bioinformatics-in-Medicine-222.html>.
4. Safdar M. Bioinformatics Application in Nanotechnology 2010 [Internet]. [cited 2012 June 29]. Available from: <http://www.biotecharticles.com/Bioinformatics-Article/Bioinformatics-Application-in-Nanotechnology-217.html>.
5. Gonzalez-Nilo F, Perez-Acle T, Guinez-Molinos S, et al. Nanoinformatics: an emerging area of information technology at the intersection of bioinformatics, computational chemistry and nanobiotechnology. *Biol Res.* 2011;44(1):43-51.
6. He Y, Rappuoli R, De Groot AS, et al. Emerging vaccine informatics. *J Biomed Biotechnol.* 2011;2010:218590.
7. Blundell TL, Sibanda BL, Montalvao RW, et al. Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery. *Phi-*

- los Trans R Soc Lond B Biol Sci. 2006;361(1467):413-23.
8. Poland GA, Ovsyannikova IG, Jacobson RM. Application of pharmacogenomics to vaccines. *Pharmacogenomics*. 2010;10(5):837-52.
 9. Altman RB. Translational Bioinformatics: Linking the Molecular World to the Clinical World. *Clin Pharmacol Ther*. 2012 May 2. [Epub ahead of print]
 10. Khoury MJ, Feero WG, Reyes M, et al. The genomic applications in practice and prevention network. *Genet Med*. 2009;11(7):488-94.
 11. Bioinformatics Market Outlook to 2015 [Internet]. 2012 [cited 2012 June 27]. Available from: <http://www.prnewswire.com/news-releases/bioinformatics-market-outlook-to-2015-142750225.html>.
 12. Batley J, Edwards D. Genome sequence data: management, storage, and visualization. *Biotechniques*. 2009;46(5):333-4, 336.
 13. Yandell M, Ence D. A beginner's guide to eukaryotic genome annotation. *Nat Rev Genet*. 2010;13(5):329-42.
 14. Center DM, Schwartz DA, Solway J, et al. Genomic Medicine and Lung Diseases: NHLBI Workshop. *Am J Respir Crit Care Med*. 2012 May 31. [Epub ahead of print]
 15. Zhang Y, Szustakowski J, Schinke M. Bioinformatics analysis of microarray data. *Methods Mol Biol*. 2009;573:259-84.
 16. Pritchard CC, Cheng HH, Tewari M. MicroRNA profiling: approaches and considerations. *Nat Rev Genet*. 2012;13(5):358-69.
 17. Faro A, Giordano D, Spampinato C. Combining literature text mining with microarray data: advances for system biology modeling. *Brief Bioinform*. 2012;13(1):61-82.
 18. Sayers EW, Barrett T, Benson DA, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2012;40(Database issue):D13-25.
 19. Wellcome Trust Sanger Institute. Databases [Internet]. [cited 2012 June 30]. Available from: <http://www.sanger.ac.uk/resources/databases/>.
 20. Databases at EBI [Internet]. [cited 2012 June 29]. Available from: <http://www.ebi.ac.uk/Databases/>.
 21. Kersey PJ, Staines DM, Lawson D, et al. Ensembl Genomes: an integrative resource for genome-scale data from non-vertebrate species. *Nucleic Acids Res*. 2012;40(Database issue):D91-7.
 22. e!Ensembl. Browse a Genome [Internet]. [cited 2012 June 28]. Available from: <http://useast.ensembl.org/index.html>.
 23. Maglott D, Ostell J, Pruitt KD, et al. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res*. 2011;39(Database issue):D52-7.
 24. Barrett T, Edgar R. Gene expression omnibus: microarray data storage, submission, retrieval, and analysis. *Methods Enzymol*. 2006;411:352-69.
 25. Magrane M, Consortium U. UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)*. 2011;2011:bar009.
 26. Haw RA, Croft D, Yung CK, et al. The Reactome BioMart. *Database (Oxford)*. 2011;2011:bar031.
 27. Brown JB, Okuno Y. Systems biology and systems chemistry: new directions for drug discovery. *Chem Biol*. 2012;19(1):23-8.
 28. Moore JB, Weeks ME. Proteomics and systems biology: current and future applications in the nutritional sciences. *Adv Nutr*. 2011;2(4):355-64.
 29. Rotilio D, Della Corte A, D'Imperio M, et al. Proteomics: bases for protein complexity understanding. *Thromb Res*. 2012;129(3):257-62.
 30. Khatib F, DiMaio F, Cooper S, et al. Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nat Struct Mol Biol*. 2011;18(10):1175-7.
 31. Sirota M, Dudley JT, Kim J, et al. Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci Transl Med*. 2011;3(96):96ra77.
 32. Dudley JT, Sirota M, Shenoy M, et al. Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. *Sci Transl Med*. 2011;3(96):96ra76.
 33. Markets and Markets: Global bioinformatics market worth \$8.3 billion by 2014 [Internet]. 2010. [cited 2012 June 26]. Available from: <http://www.marketsandmarkets.com/Press-Releases/bioinformatics-market.asp>.
 34. Peters RC. What lies ahead? Drug Discovery & Development [Internet]. 2008 Nov 10. [cited 2012 June 15]. Available from: <http://www.dddmag.com/articles/2008/11/wh-at-lies-ahead>.
 35. ISCB Listing of Degree/Certificate Programs Worldwide [Internet]. [cited 2012 April 29]. Available from: <http://www.iscb.org/iscb-degree-certificate-programs>.
 36. Carnegie-Foundation. Carnegie Classification of Institutions of Higher Education [Internet]. 2010. [cited 2012 June 30]. Available from: http://classifications.carnegiefoundation.org/summary/ugrad_prog.php.
 37. Rizk C. 10th Annual salary survey 2012. *Genome Web* [Internet]. [cited 2012 May 30]. Available from: <http://www.genomeweb.com/10th-annual-salary-survey>.
 38. Dyer BD, LeBlanc MD. Meeting report: Incorporating genomics research into undergraduate curricula. *Cell Biol Educ*. 2002;1(4):101-4.
 39. Honts JE. Evolving strategies for the incorporation of bioinformatics within the undergraduate cell biology curriculum. *Cell Biol Educ*. 2003;2(4):233-47.
 40. 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.

41. Weisman D. Incorporating a collaborative web-based virtual laboratory in an undergraduate bioinformatics course. *Biochem Mol Biol Educ.* 2010;38(1):4-9.
42. Leblanc MD, Dyer BD. Bioinformatics and computing curricula 2001: why computer science is well positioned in a post-genomic world. *ACM SIGCSE Bull.* 2004;36:64-7.
43. Maloney M, Parker J, Leblanc M, et al. Bioinformatics and the undergraduate curriculum essay. *CBE Life Sci Educ.* 2010;9(3):172-4.
44. Floriano WB. A portable bioinformatics course for upper-division undergraduate curriculum in sciences. *Biochem Mol Biol Educ.* 2008;36(5):325-35.
45. Brame CJ, Pruitt WM, Robinson LC. A molecular genetics laboratory course applying bioinformatics and cell biology in the context of original research. *CBE Life Sci Educ.* 2008;7(4):410-21.
46. Chapman BS, Christmann JL, Thatcher EF. Bioinformatics for undergraduates: Steps toward a quantitative bioscience curriculum. *Biochem Mol Biol Educ.* 2006;34(3):180-6.
47. Krawetz SA, Womble DD. Design and implementation of an introductory course for computer applications in molecular genetics. A case study. *Mol Biotechnol.* 2001;17(1):27-41.
48. Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25(17):3389-402.
49. Altschul SF, Wootton JC, Gertz EM, et al. Protein database searches using compositionally adjusted substitution matrices. *Febs J.* 2005;272(20):5101-9.
50. Kikin O, Zappala Z, D'Antonio L, et al. GRSDB2 and GRS_UTRdb: databases of quadruplex forming G-rich sequences in pre-mRNAs and mRNAs. *Nucleic Acids Res.* 2008;36(Database issue):D141-8.
51. Kikin O, D'Antonio L, Bagga PS. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.* 2006;34(Web Server issue):W676-82.
52. Kostadinov R, Malhotra N, Viotti M, et al. GRSDB: a database of quadruplex forming G-rich sequences in alternatively processed mammalian pre-mRNA sequences. *Nucleic Acids Res.* 2006;34(Database issue):D119-24.
53. Menendez C, Frees S, Bagga PS. QGRS-H Predictor: a web server for predicting homologous quadruplex forming G-rich sequence motifs in nucleotide sequences. *Nucleic Acids Res.* 2012;40(W1):W96-W103.