**Essay**

**A Survey of Scholarly Literature Describing the Field of Bioinformatics Education and Bioinformatics Educational Research**

Alejandra J. Magana,*† Manaz Taleyarkhan,* Daniela Rivera Alvarado,* Michael Kane,* John Springer,* and Kari Clase‡

*Department of Computer and Information Technology, †School of Engineering Education, and ‡Department of Technology, Leadership and Innovation, Purdue University, West Lafayette, IN 47906

Submitted October 4, 2013; Revised August 20, 2014; Accepted September 5, 2014

Monitoring Editor: Mary Lee Ledbetter

Bioinformatics education can be broadly defined as the teaching and learning of the use of computer and information technology, along with mathematical and statistical analysis for gathering, storing, analyzing, interpreting, and integrating data to solve biological problems. The recent surge of genomics, proteomics, and structural biology in the potential advancement of research and development in complex biomedical systems has created a need for an educated workforce in bioinformatics. However, effectively integrating bioinformatics education through formal and informal educational settings has been a challenge due in part to its cross-disciplinary nature. In this article, we seek to provide an overview of the state of bioinformatics education. This article identifies: 1) current approaches of bioinformatics education at the undergraduate and graduate levels; 2) the most common concepts and skills being taught in bioinformatics education; 3) pedagogical approaches and methods of delivery for conveying bioinformatics concepts and skills; and 4) assessment results on the impact of these programs, approaches, and methods in students’ attitudes or learning. Based on these findings, it is our goal to describe the landscape of scholarly work in this area and, as a result, identify opportunities and challenges in bioinformatics education.

**MOTIVATION FOR THE STUDY**

Many complex problems in biomedical systems and public health research and development require cross-disciplinary approaches to integrate diverse perspectives into a collective whole (Adams et al., 2010). Cross-disciplinary approaches encompass “a set of practices associated with thinking and working across perspectives such as multidisciplinary, interdisciplinary, and transdisciplinary” (Adams et al., 2010, p. 1158). Bioinformatics is a cross-disciplinary field that resulted from the advancement of biological sciences due to the integration and application of information technology and computational science to solve biological problems in emerging fields such as genomics and systems biology. This advancement has resulted in a massive amount of biological data that has impacted the manner in which research and education has and will continue to be conducted in our pursuit of improved human health and prolongation of human life (Yang et al., 2008). Hence, it is vital to leverage education and training to fulfill the ongoing need for competent scientists and technicians in bioinformatics (Hersh, 2008; Ranganathan, 2005). Moreover, to ensure the efficacy of the education and training, we must have a better understanding of effective educational approaches and strategies for improving student learning in this field.

Bioinformatics education can be broadly defined as the teaching and learning of the use of computer and information technology to gather, store, analyze, interpret, and integrate data to solve biological problems (Counsell, 2003; Koch and Fuellen, 2008). This means that training and education in bioinformatics should encompass knowledge and skills...
from biology, mathematics, statistics, physics, chemistry, medicine, pharmacology, computer science, and information technology (Ranganathan, 2005). However, effectively integrating bioinformatics education into formal and informal educational settings has been a challenge due to: 1) its cross-disciplinary nature; 2) the disparate methods, outlooks, and cultures of its related disciplines (Zauhar, 2001); 3) the lack of an integrated training support structure (e.g., United Kingdom; Brass, 2000); and 4) the lack of collaboration between funding agencies (Brass, 2000). In this article, we seek to provide an overview of the state of research in bioinformatics education and bioinformatics educational research as described in the published literature. Our goal is to 1) help inform current education and training in the field, 2) motivate educators to share the effectiveness or challenges of their efforts through scholarly articles with the broader community, and 3) invite educational researchers to collaborate with educators in helping to shape and evaluate the effectiveness of specific implementation efforts.

Like many other fields, bioinformatics has emerged due to the development of new computing tools and environments that allow researchers to more easily share data. While traditional bioinformatics has been used to perform genome and proteome analysis, this field is growing or morphing rapidly to now encompass simulation-based approaches give rise to computational biology (Kitano, 2002). We argue that, as the technical advances in this area progress, we also need to simultaneously concentrate on how these advances can be effectively integrated into educational settings and on sharing the approaches with the broader community. Coupling advances in technical research and identifying how these advances can be transformed into specific training and educational experiences will allow the development of a future workforce that will be ready to take jobs and conduct research in these emerging fields.

To make steps toward this end, we aim in this survey study to identify: current approaches for integrating bioinformatics at the undergraduate and graduate levels; the most common concepts and skills being taught in bioinformatics education; pedagogical approaches and methods of delivery for conveying bioinformatics concepts and skills; and evaluation or assessment results (when available) on the impact of these programs, approaches, and methods in students’ learning. Specific research questions for this survey study were:

How does published literature describe curricular efforts aimed at integrating bioinformatics education?

How does published literature describe specific bioinformatics education content, delivery methods, and assessment?

Our findings identify both the opportunities and challenges that exist in bioinformatics education and reveal a need for further educational assessments in this emerging area.

METHODS

Owing to the recent emergence and continuously evolving nature of the field of bioinformatics, we chose to narrow our searches to key phrases such as “bioinformatics education,” “biomedical informatics,” or “bioinformatics” in some educational context. We limited our searches to these three key phrases to keep the focus on the field of bioinformatics and to limit the scope of the paper. We expected that, by using this approach, we would identify articles from better established efforts that at the same time could have had an impact in educational programs or classroom implementations. Therefore, findings from this study are limited to those articles with any of those phrases in at least the title, abstract, and keywords and with a publication date of 2013 or earlier.

The research team followed a systematic procedure of searching, categorizing, and analyzing the papers presented in this review. The interdisciplinary research team consisted of four faculty members and two students from backgrounds in educational research (A.M.), biology (M.K.), biotechnology (K.C.), computer science (J.S.), computer and information technology (D.R.), and engineering (M.T.). To account for reliability, A.M., D.R., and M.T. performed all the searches and the initial categorization of the papers. They also took the first step at constructing the tables. Then, the tables, the columns, and the categorization of the articles were reviewed by the rest of the team members to account for validity, with each team member focusing on his or her own discipline. Finally, A.M. and the lead team member from each discipline conducted discussions associated with each of the tables. At times, the discussion was performed by three team members (i.e., mathematics, statistics). Details of the specific procedures are detailed in the following paragraphs.

The publications selected for this analysis were identified by conducting searches in Google Scholar, Web of Science, ACM Digital Library, ERIC, and PubMed. When the searches were conducted, we used the default settings of those databases: searching on multiple fields and matching at least title, abstract, and indexing subject headings/descriptors. In Web of Science, the default “topic” search parameters include title, abstract, and descriptor fields. Default settings of PubMed include all fields, but PubMed also tries mapping the search terms to phrases in the MeSH (medical subject headings), journal, and author databases. For ACM Digital Library, the default search is “any field.” In Google Scholar, the default search setting is “anywhere in the article.” ERIC default parameters include searches in title, author, source, abstract, and descriptor. Therefore, we can argue that the searches conducted were in at least the title of the document, the abstract, and the keywords. Additionally, we performed a search on the National Science Foundation database to identify awards on the same topic. Using the key term “bioinformatics education,” we identified 25 awards. Each award was reviewed individually to identify related lists of publications.

This search returned 113 (out of a total of 140) documents ranging in their publication year from 1998 to 2013 and returned primarily three types of scholarly publications: journal papers, conference proceedings, and magazine articles. The 113 identified papers were analyzed as individual units (as opposed to groups of papers published by the same research groups), potentially resulting in overlapping themes. Multiple levels of analyses were conducted on the 113 documents identified. To provide an overview of our sample, we first started by identifying frequencies of types of articles found as well as the year in which those were published. Details of the specific procedures are detailed in the following paragraphs.

The 113 identified papers were analyzed as individual units (as opposed to groups of papers published by the same research groups), potentially resulting in overlapping themes. Multiple levels of analyses were conducted on the 113 documents identified. To provide an overview of our sample, we first started by identifying frequencies of types of articles found as well as the year in which those were published. Details of the specific procedures are detailed in the following paragraphs.
decided to use a qualitative approach that consisted of analyzing the content of the abstracts in each of the categories to identify the themes; the details of this process follow.

In the first level of analysis, we aimed to identify a general perspective on the kinds and number of publications related to bioinformatics education, including the number of papers published per year, the types of publications (i.e., journal or magazine articles or conference papers), and the journals or conferences wherein bioinformatics education is disseminating and finding a community. In the second level of analysis, we categorized the papers based on their primary theme. Four primary themes were identified: 1) description, 2) position, 3) pedagogy, and 4) educational research or evaluation. In the description category, we included papers that described some sort of program, degree, course, or series of courses. In the position category, we considered all papers that described opinions about the need for bioinformatics education, opportunities and challenges of bioinformatics education, and issues related to the interdisciplinary nature of bioinformatics education. In the pedagogy category, we included papers that described a teaching method, multimedia technology, or method for the delivery of bioinformatics education. In the educational research or evaluation category, we considered all of the papers that included some sort of evaluation of programs, materials, knowledge, skills, or attitudes in general. Some papers were assigned to two or more of these categories. Papers describing programs at the K–12 level were omitted.

In the third level of analysis, we qualitatively examined the abstracts of the 113 papers, using categorical analysis to identify themes. Categorical analysis refers to a systematic approach of data analysis in which findings are inductively derived from the data (Strauss and Corbin, 1990). The process of inductive analysis identifies similarities and differences within the data, resulting in a set of interrelated themes (Glaser and Strauss, 1967). Through this analysis, we manually identified common themes among the abstracts, clustered these themes together into similar topics, and used the most descriptive wording or phrase for the topics. Finally, all papers from the same topic were further reviewed to provide a more detailed and in-depth description of each of the newly identified topics. All papers presented in the tables in the Results section are ordered alphabetically based on the first author’s last name followed by year of publication.

RESULTS

In this section, we present the results of the three levels of analysis conducted on the articles identified using the search parameters and methodology described earlier. As shown in Figure 1, the publication year for the 113 identified articles ranged from 1998 to 2013.

Results in Figure 1 illustrate that, so far, the maximum number of articles appeared in 2007. Of the three types of publications, journal articles were the majority (81%), with conference papers (16%) and magazine articles (3%) forming much smaller contributions. On the other hand, Figure 2 shows that the published papers classified by discipline or field of study fell into four main categories. The biology education category (Figure 2, Biology Ed.) included 37 papers (34%) in education journals or conference proceedings focusing on microbiology, biology, pharmaceutical, biochemistry, or life sciences. The bioinformatics category included 38 papers (35%) discussing topics about bioinformatics, medical informatics, biomedical informatics, or computational biology. Computer or technology education (Figure 2, Computer Ed.) contained 14 papers (13%) published in journals or as conference papers related to technology education or computer science education. The biology and biotechnology category (Figure 2, Bio. and Biotech.) included 12 papers (11%) related to medical and biological engineering, biotechnology, science, and systems biology. The remaining seven papers (6%) were categorized as published in journals and as conference papers related to engineering education, education or science education, and information science or information science technology (Figure 2, Eng. Ed. & Inf. Science).
technology majors at both the undergraduate and graduate levels. This distribution supports the notion of bioinformatics as interdisciplinary.

When bioinformatics is integrated as a university degree or program, the course of study is mainly focused on degrees for graduate students or geared toward a combination of undergraduate and graduate students. A less common focus is in the area of degrees for undergraduate students only. This may exemplify a trend observed in other new fields that emerge and are taught first at a graduate level and then, as the discipline evolves and becomes more clearly defined, at the undergraduate level.

How Does the Published Bibliography Describe Specific Bioinformatics Education Content, Delivery Methods, and Assessment?

This discussion starts by describing the content related to bioinformatics education in terms of subject matter expertise; specific concepts, methods, and tools; and services. This section also describes specific pedagogical methods described in the literature and delivery methods such as online learning. Finally, the analysis concentrates on literature that describes evaluation and assessment components.

Concepts, Methods, Tools, and Services in Bioinformatics Education

We divide this section into four parts: concepts, methods, tools, and services. The term “concept” refers to the notion of a subject or topic that any student must know and understand in the area of study in order to solve certain problems. These topics include computer science, biology/genetics, and mathematics/statistics. As for “tools,” this refers to all the pre-existing instruments and applications available for students to use in order to facilitate the analysis and solution of a certain problem. Additionally, “methods” refers to the experimental, statistical, and computational processes. Finally, “services” refers to all the resources available for
Table 1. Summary of articles in category “course, series of courses or university degree or program”

| Reference            | Area                                                                 | Format                      | Level                        | Discipline                                      |
|----------------------|----------------------------------------------------------------------|-----------------------------|------------------------------|-------------------------------------------------|
| Brame et al., 2008   | Bioinformatics, molecular biology, and genetic approaches and research | Course                      | Undergraduate                | Genetics                                        |
| Brame and Brazeau, 2006 | Human genetics and genomics in drug therapy, optimization, and patient care and counseling | Course                      | Undergraduate                | Pharmacy students                               |
| Campbell, 2003       | Genomics, proteomics, and bioinformatics                             | Course                      | Undergraduate                | Biology                                         |
| Cattley, 2004        | Bioinformatics                                                       | University degree or program | Undergraduate                | Science                                         |
| Craddock et al., 2007 | Pathosystems biology                                                | Course                      | Undergraduate                | Interdisciplinary                               |
| Feig and Jabri, 2002  | Data mining                                                         | Integrated collection of exercises | Undergraduate              | Biochemistry                                    |
| Fetrow and John, 2006 | Bioinformatics                                                      | Course                      | Undergraduate                | Interdisciplinary                               |
| Floraino, 2008       | Bioinformatics                                                       | Course                      | Upper undergraduate          | Biological sciences, chemistry, and computer science |
| Furge et al., 2009   | Bioinformatics                                                       | Modules                     | Undergraduate                | Interdisciplinary                               |
| Goode and Trajkovski, 2007 | Molecular biology, biochemistry, and bioinformatics                   | University degree or program | Undergraduate                | Interdisciplinary                               |
| Hack and Kendall, 2005 | Bioinformatics                                                      | University degree or program | Graduate                    | Life sciences                                   |
| Haux, 2004           | Health, medical, and biomedical informatics                          | University degree or program | Graduate                    | Biomedicine and health sciences                 |
| Hersh, 2007          | Biomedical informatics                                              | University degree or program | Graduate                    | Health science                                  |
| Honts, 2003          | Biomedical informatics                                              | Three courses               | Undergraduate                | Biology                                         |
| Hughey and Karplus, 2003 | Bioinformatics                                                    | University degree or program | Graduate and undergraduate  | Engineering                                     |
| Kane and Brewer, 2007 | Biomedical informatics                                              | Three courses               | Undergraduate                | Information technology                          |
| Kane et al., 2006    | Biomedical informatics                                              | Courses                     | Undergraduate                | Information technology                          |
| Khuri, 2008          | Computer science                                                    | Track collection of courses | Undergraduate                | Health, medical, and life sciences              |
| Koch and Fuellen, 2008 | Bioinformatics                                                        | Courses                     | Gradate and undergraduate    | Interdisciplinary                               |
| Kriłowiec et al., 2007 | Bioinformatics                                                       | Summer program             | Gradate and undergraduate    | Interdisciplinary                               |
| Kulkarni-Kale et al., 2010 | Information technology and biotechnology                         | Course                     | Gradate                      | Biotechnology                                   |
| Lim et al., 2003     | Bioinformatics                                                       | Online course               | Graduate and undergraduate  | Interdisciplinary                               |
| Luo, 2013            | Online bioinformatics resources                                      | One semester course         | Graduate and undergraduate  | Biology                                         |
| Rainey et al., 2007  | Bioinformatics                                                       | Two course                  | Graduate and undergraduate  | Interdisciplinary                               |
| Ranganathan, 2005    | Bioinformatics                                                       | University degree or program | Graduate                   | Interdisciplinary                               |
| Reisdorph et al., 2013 | Genomics, proteomics, and bioinformatics                           | Hands-on workshops          | Graduate                     | Bioinformatics                                  |
| Sahinidis et al., 2005 | Bioinformatics                                                        | University degree or program | Graduate and graduate        | Interdisciplinary                               |
| Sczyrba et al., 2008 | Sequence analysis and programming in bioinformatics                  | Two courses                 | Undergraduate                | Bioinformatics                                  |
| Smith and Emmeluth, 2002 | Bioinformatics                                                    | Learning modules           | Undergraduate                | Biology                                         |
| Tolvanen and Vihrinen, 2004 | Bioinformatics                                                     | Two courses                 | Graduate                     | Interdisciplinary                               |
| Toth and Connelly, 2006 | Sequence analysis                                               | Course                     | Upper undergraduate          | Biology and computer science                    |
| Yang and Zhang, 2008 | Bioinformatics                                                       | Conferences, workshops, and tutorials | General                  | Interdisciplinary                               |
| Zatz, 2002           | Bioinformatics                                                       | University degree or program | Graduate                    | Medical                                         |
| Zauhar, 2001         | Bioinformatics                                                       | University degree or program | Graduate and undergraduate  | Interdisciplinary                               |
| Zhang, 2011          | Bioinformatics                                                       | Lab exercises               | Undergraduate                | Biology                                         |
| Zhang et al., 2007   | Bioinformatics                                                       | Track collection of courses | Graduate and undergraduate  | Computer science                                |
| Zhong et al., 2003   | Bioinformatics                                                       | Courses and programs        | Graduate and undergraduate  | Bioinformatics                                  |
users to access information, including reference and literary databases such as GenBank and PubMed. We recognize that the “database” term is burdened with multiple meanings, including a simple container for data (such as a file or even a written record), a means to manage data (i.e., a database management system), and even a Web-based interface providing access to an underlying database. For our purposes, we place “database” in the services category due to databases’ need for a retrieval mechanism to provide their fullest value; consider that, if one did not have the need to eventually retrieve the data, then the need to store it is highly dubious. These services can be paired with several tools to organize and analyze data.

These categories are not mutually exclusive, in that items may appear in two or more of the categories based on context and audience. For instance, the term “BLAST” subsumes concepts such as DNA sequencing and pattern-matching; methods involving algorithms, statistics, and experimental processes/workflows; tools implementing the BLAST algorithms, such as the software supplied by the National Center for Biotechnology Information (NCBI); and services such as the BLAST search capabilities provided by the NCBI on its website (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Given the richness of the field, some overloading of terms is natural, and when such cases occur in our analysis, we attempted to include such items in the most applicable category while recognizing others with different perspectives may place them elsewhere.

The most basic computer science concepts introduced within bioinformatics-related courses are databases, algorithms, and the Perl programming language, along with proper uses of the BLAST, Clustal, and FASTA tools and the NCBI, GenBank, PubMed, and MEDLINE services (see Table 2). Effective use of these tools requires data-mining skills and object-oriented programming skills. Additional skills are systems analysis and design, database design, software engineering, and human–computer interaction (HCI).

Table 3 shows that the application of bioinformatics methods to biological problems is the most common method for integrating bioinformatics and biology. Other less common applications relate to demonstration of evolution-related concepts and uses of information technology, such as conducting searches and mining data. The most frequent biological concepts include genomics, proteomics, and DNA sequence concepts, as well as biochemistry and molecular biology. Methods commonly taught include sequence analysis and alignment, DNA microarray data analysis, and the use of protein structure prediction and classification tools.

From results depicted in Table 4, we can identify probability and statistics as main concepts highlighted by different authors. No specific tools or statistical analysis or methods in this area were identified in any of the reviewed papers.

**Pedagogical Approaches and Methods for the Delivery of Bioinformatics Education**

During the content analysis, we attempted to identify the most common pedagogical methods used to convey bioinformatics-related concepts and procedures (see Table 5). In the process, we also determined that bioinformatics education has been delivered through both traditional face-to-face classroom formats and online distance-learning experiences, as described in Table 6.

Results from Table 5 indicate that bioinformatics has been introduced primarily through problem-to-be-solved methods (i.e., challenge-based learning) including student-centered approaches such as inquiry learning and collaborative learning. This also supports the trend discussed earlier, that within an educational context, bioinformatics is introduced as a tool to solve biological problems.

Diverse multimedia and delivery methods have been used in bioinformatics education. Primarily, we can identify the use and development of Web-based applications and learning environments to support inquiry learning and also the use of distance learning as a means of delivering content.

**Impact of Programs, Approaches, and Methods on Students’ Attitudes or Learning**

During the analysis stage we attempted to identify whether some educational research or program evaluations have been conducted in the area of bioinformatics education and the kinds of constructs that have been measured. In Table 7, we present the summary of articles that fall within this category.

We note that that most of these articles are focused on evaluation of learning materials in the form of final course grades and student self-assessments, perceptions of the materials, confidence in attaining specific learning outcomes, and attitudes toward the learning experience. When assessment of learning was provided, results reported positive outcomes on student learning and perceptions in most cases. In cases in which learning data were reported, gains were related to working knowledge of bioinformatics concepts and methods.

**DISCUSSION**

This analysis provides a broad perspective of the state of scholarly work in bioinformatics education on the following dimensions: 1) a spectrum of the literature related to bioinformatics education; 2) ways of integrating bioinformatics into educational settings; 3) concepts, methods, computational tools, and services used in bioinformatics education; 4) pedagogical approaches for integrating bioinformatics education; and 5) evidence of the effectiveness of bioinformatics education.

Looking at trends of research in bioinformatics education, we can identify interesting patterns. The data reveal that publishing on bioinformatics education has been increasing since 1998. After 2007, a relatively constant rate of fewer than 12 papers a year is observed. It is possible that this decline can be attributed to the bioinformatics education community using additional terms (i.e., such as terms described in U.S. Department of Energy, 2010) to describe fields that are using some of the tools and curricula developed by the bioinformatics field and that are now applying them to new emerging areas of need. Additional terms may include “genomics,” “proteomics,” “quantitative biology,” “systems biology,” “computational biology,” “biological systems,” and “computing for biological systems,” among others.

The most common method found for integrating bioinformatics into the undergraduate curriculum is either through a series of stand-alone courses or through learning modules integrated into existing courses. In contrast, university degrees
| Reference               | Concepts                                           | Tools and services                                                                 |
|------------------------|---------------------------------------------------|-------------------------------------------------------------------------------------|
| Bagga, 2012            | Data mining and programming                       | Perl                                                                                |
| Beck et al., 2007      | Data mining                                       | BLAST, GenBank, NSF, MAGI, and Inter-ProScan                                        |
| Bednarski et al., 2005 | Data mining                                       | BLAST, ClustalW, LocusLink, PSIPRED, DeepView, NCBI, OMIM, ExPASy’s, KEGG, and Swiss-Prot |
| Boyle, 2004            | Data mining                                       | BLAST, Biology Workbench of San Diego Super-computer Center, NCBI, OMIM, PubMed, and Google |
| Burhans and Skuse, 2004 | Object-oriented programming, information management, data mining, and HCI | Perl, algorithms and complexity, and human–computer interaction                      |
| Campbell, 2003         | Data mining and object-oriented programming        | PubMed, PubCrawler, Perl, Perl programming                                          |
| Cattley and Arthur, 2007 | Data mining and object-oriented programming       | BLAST, Clustal, BioManager, Phylip, Course DNA, GenBank, MEDLINE, Sydney Bioinformatics, AGIC, AGNIS, UniProt, Swiss-Prot, Uniform Resource Locator URL, and UNIX |
| Cooper, 2001           | Data mining                                       | BLAST, Biology Workbench of San Diego Super-computer Center, Protein Explorer, Chime, and RasMol |
| Craddock et al., 2007   | Data mining                                       | NCBI, GenBank, PATRIC, PathInfo, and MiNet                                           |
| Doom et al., 2003       | Data mining, database design, object-oriented programming, and HCI | Introductory programming, entity-relationship models, Perl, artificial intelligence algorithms, formal and comparative languages, pattern recognition, human–computer interaction, and evolutionary computation |
| Feig and Jabri, 2002    | Data mining                                       | Chime plug-in module                                                                 |
| Fetrow and John, 2006   | Data mining, systems analysis and design, and software engineering | Software engineering protocol, workflow model, dynamic programming algorithms, clustering methods, and artificial neural networks |
| Furge et al., 2009      | Data mining, data structures, and machine learning | BLAST, BLASTp, FASTA, ClustalW, ClustalX, CBS, Trident, GlobPlot, VAST, FoldIndex, Swiss Deep View, ConSurf, MSA, Protein Explorer, MAFFT, MapViewer, dbSNP, BLASTx, ENTREZ, CDD-CART, CN3D, NetPhos 2.0, Phi-Blast, UniGene, BioQUEST, GARLI, GCG, LAMARC, MrBayes, PAML, PAUP*, PHYLIP, NetPhos, SignalP, Spartan, FirstGlance, NCBI, OMIM, PubMed, KEGG, ExPASy, PDB, BRENDA, data structures, machine learning, Perl, GUI programs, and query |
| Gelbart and Yarden, 2006 | Data mining                                      | BLAST                                                                                |
| Gollery, 2006           | Data mining and object-oriented programming        | BLAST, PFAM, Linux, and Perl                                                          |
| Goode and Trajkovski, 2007 | Data mining and object-oriented programming       | Perl and database management                                                         |
| Haux, 2004              | Data mining, object-oriented programming, data structures, and software engineering | Algorithms, data structures, software engineering, programming, and information systems |
| Honts, 2003             | Data mining                                       | BLAST, FASTA, VAST, ClustalW, GraailEXP, RasMol, OpenRasMol, Chime, CN3D, Deep View, SwissPDBViewer, TreeView, BCM Search Launcher, COILS Server, NCBI, OMIM, GenBank, PubMed, CDART, PDB, Human Genome Project, UCSC, RCSB, and Perl |
| Howard et al., 2007     | Data mining                                       | BLAST, RPSBLAST, ClustalW, and PROSEARCH                                               |
| Hughey and Karplus, 2003 | Data mining and object-oriented programming        | Perl                                                                                |
| Kane and Brewer, 2007   | Data mining, systems analysis and design, database design, and HCI | BLAST, FASTA, Clustal, managing databases, information systems, information management, evolving systems modeling, microarray, development languages, client–server architectures, algorithms, queries, data structures, human–computer interaction, data modeling, data organization architecture, system architecture, and system integration |
| Kane et al., 2006       | Data mining                                       | BLAST and Clustal                                                                    |
| Kane and Springer, 2007 | Data mining and object-oriented programming        | mpiBLAST, FASTA, NCBI, GenBank, Visual Basic .NET, SDK, and hands-on training in blade server architecture |
| Khuri, 2008             | Data mining and object-oriented programming        | BLAST, FASTA, ClustalW, GenBank, Swiss-Prot, PDB, C, Perl, UNIX, algorithms, and assembly packages |
| Koch and Fuellen, 2008  | Data mining, object-oriented programming, and database design | Database design, algorithm design, pattern matching, programming paradigms, Perl, Phyton, C, C++, and Java |
| Krilowicz et al., 2007  | Data mining, object-oriented programming, systems analysis and design, and software engineering | C++, Phyton, Perl, and basic programming methods including but not limited to: data representations, data processing, file input/output, user interfaces, software engineering, algorithms, documentation, testing, debugging, and data structures |
| LeBlanc and Dyer, 2004  | Data mining and object-oriented programming        | Algorithms, and complexity programming fundamentals                                      |
| Luo, 2013               | Data mining                                       | Dot plots                                                                           |

(Continued)
or programs are mostly oriented toward graduate degrees. A more recent trend seems to be the integration of bioinformatics education via online educational resources and programs. Kampov-Polevoi and Hemminger (2011) conducted a systematic comparison of curricula among bioinformatics programs. They developed a categorization scheme that is grounded in the analysis of the content of existing academic programs. Categories of their proposed scheme align with our categorization of bioinformatics domains such as informatics and computer science, statistics and research methods, domain-specific information systems (e.g., tools and services), and domain-specific knowledge (e.g., biology). In addition, they identified management- and business-related concepts, ethical and societal issues, and communication skills.

As we observed earlier, we may demarcate bioinformatics education into four main areas: concepts, methods, computational tools, and services. The conceptual basis for bioinformatics is found in topical areas such as genomics and proteomics; as for analysis tools, these include approaches such as data mining and object-oriented programming. In addition to concepts and analysis tools, we naturally see the wide deployment of computational tools and methods such as BLAST and databases such as GenBank. We also note services as playing a significant role in bioinformatics research and, hence, in education; in fact, we may characterize BLAST as provided by the NCBI as a service. Moreover, the capacity to search databases such as GenBank—again provided by the NCBI—as a service available to researchers and students alike. Underlying all four of these areas are threads that suggest a more conceptual basis for considering and incorporating computational elements as warranted. For example, moving beyond the specifics of BLAST into reflections on and investigations of string pattern matching and, even more broadly, pattern recognition will enable students to move beyond the current state of the art computationally and into a future of computing that will complement and enhance the next generation of biotechnologies. It is also interesting to point out that, from patterns found in Tables 2–4, little emphasis has been placed on mathematical and statistical concepts and procedures. Computational thinking—including a stronger focus on mathematics and statistics curriculum—serves to provide the ties that bind the aforementioned threads into a stronger weave surrounding bioinformatics education. Two additional topics that were found in more recent literature

### Table 2. Continued

| Reference                  | Concepts                                                                 | Tools and services                                                                 |
|----------------------------|--------------------------------------------------------------------------|------------------------------------------------------------------------------------|
| Marceglia et al., 2007    | Data mining, object-oriented programming, and database design            | Database management systems, relational database theory, relational models, entity-relationship diagrams, database design, and SQL. |
| Moll et al., 2006         | Data mining and object-oriented programming                              | BALLView and Python                                                                 |
| Nehm and Budd, 2006       | Data mining                                                              | GenBank, NMITA, and analyzing the Human Genome Project                             |
| Nichols et al., 2003      | Data mining                                                              | Navigation through DNA data banks, NCBI, Sequence Manipulation Suite, Nucleotide Frequency Program, DHPLC Melt Program, Biology Workbench, Cold Spring Harbor Sequence Server, Codon Usage Database, T-COFFEE Sequence Alignment |
| Obom and Cummings, 2009   | Data mining, object-oriented programming, and software engineering       | Perl, microarrays, algorithms, and software engineering                            |
| Perez-Iratxeta et al., 2007| Data mining                                                             | MEDLINE                                                                            |
| Pezvner and Shamir, 2009  | Data mining                                                              | BLAST and principal component analysis                                             |
| Rainey et al., 2007       | Data mining                                                              | Course management system, CART, and BSC                                            |
| Rao et al., 2008          | Data mining and object-oriented programming                              | FASTA, EMBoss, PHYLiP, GCG Wisconsin, ArrayQuest, NCBI, GenBank, SRS, PubMed, UniProt, PDB, KEGG, MUSC, GEO, Human Genome Project, EMBL, database, SOAP, algorithm, Web services, Web server, BioPerl, C++, microarray, and information systems |
| Sahinidis et al., 2005    | Data mining and object-oriented programming                              | Databases, algorithms, microarrays, integer programming, and computational complexity |
| Sansom and Smith, 2000    | Data mining                                                              | BLAST, FASTA, Prints, ProDom, TREMBL, Kabat ENZYME, PSI-Blat server, HGMP-RC Dali, Pfam, PROSITE, Jpred, NCBI, OMIM, EMBL, GenBank, DDBJ, Human Genome Mapping Project Resource Centre, PBD, and Swiss-Prot |
| Sczyrba et al., 2008      | Data mining                                                              | BLAST, FASTA, GenBank, PubMed, EMBL, DDBJ, InterPro, PDB, UniProt, FlyBase, Wormbase, NCBI, and NEWT |
| Smith and Emmeluth, 2002  | Data mining                                                              | NCBI, PubMed, MEDLINE, NLM, Human Genome Project                                    |
| Toth and Connelly, 2006   | Data mining and object-oriented programming                              | BLAST, Folding@Home, Clustal, Perl, regular expressions, dynamic programming, call stack, call tree, and memoization |
| Umarji et al., 2009       | Data mining and object-oriented programming                              | Extreme programming and requirements engineering and documentation                  |
| Yang and Zhang, 2008      | Data mining                                                              | BLAST, Genome Browser, Ensembl, UCLC, NCBI, and PDB                               |
| Zauhar, 2001              | Data mining and object-oriented programming                              | Databases, algorithms, C, C++, Java, and Perl                                      |
| Zhang et al., 2007        | Data mining                                                              | BLAST, FASTA, OMIM, GenBank, and GEO                                              |
Table 3. Summary of biology and genetics concepts and methods

| Reference | Concepts | Method |
|-----------|----------|--------|
| Beck et al., 2007 | Global gene expression and shoot apical meristem | Microscopy |
| Butler et al., 2008 | Multiple sequence alignment, PCR primer design, restriction mapping, evolution, phylogeny, gene detection, microarray analysis, protein structure and function prediction, proteomics, protein identification and characterization, motif searching, and sequence assembly | Obtain pathogen information from Patho-Systems Resource Integration Center and Center for Pathogen Information |
| Cattley and Arthur, 2007 | Chemistry, biochemistry, molecular biology, genetics, DNA sequencing, gene expression, X-ray crystallography, protein structure and function, gene structure and density, introns and exons, transposition and repetitive elements, introduction to gene microarrays, and proteomics | Comparative model of protein structure, DNA isolation, gel electrophoresis, molecular visualization, structural modeling, ligand screening, inhibition, and drug design |
| Craddock et al., 2007 | Exploring concepts of information content of different biopolymers, the relationship between primary sequence and tertiary structure, and how sequence conservation can be used to find an enzyme active site | Pairwise sequence alignment, protein secondary structure prediction, gene expression, gene prediction, and gene sequencing |
| Doom et al., 2003 | Chemistry, biochemistry, molecular biology, genetics, DNA sequence, gene expression, X-ray crystallography, protein sequence and function, gene structure and density, introduction to gene microarrays, and proteomics | NCBI and BLASTN |
| Feig and Jabri, 2002 | Exploring concepts of information content of different biopolymers, the relationship between primary sequence and tertiary structure, and how sequence conservation can be used to find an enzyme active site | DNA research and sequencing |
| Fetrow and John, 2006 | Genomics and proteomics | Biochemistry laboratory, protein structure prediction, mRNA expression analysis, gene finding, RNA prediction and alignment, X-ray crystallography, and NMR spectroscopy |
| Fuselier et al., 2011 | Sequencing and PCR | Pairwise sequence alignments, multiple sequence alignments, and DNA microarray |
| Goode and Trajkovski, 2007 | Organic chemistry | Mass spectrometry and cellular imaging |
| Harmon et al., 2002 | Phylogenetic trees, molecular biology, cellular biology, DNA sequence, protein structure and function, gene structure and expression, and genome | BLAST |
| Honts, 2003 | Genomics and proteomics | Macromolecular structures and machines and DNA microarray technology |
| Howard et al., 2007 | Biochemistry, genetics, cell biology, molecular biology, bacterial diversity, microbial genetics, microbiology, genes, protein sequence, protein structure, nucleotide sequence, amino acid sequence, DNA sequence, genetic mutation, and proteomics | Sequence/structure analysis, microarray data, and phylogenetic tree inference |
| Hughey and Karplus, 2003 | Molecular biology, chemistry, biochemistry, organic chemistry, cell biology, basics of DNA, RNA, protein sequence and structure, enzymes, regulation, metabolism, amino acids, genomics, phylogeny, and proteomics | PubMed searches, UniProt database queries, sequence alignments, dot plots, BLAST, WebLab, Jemboss, MEGA, and SPDBV |
| Kane and Brever, 2007 | Amino acids, nucleotides, genes, proteins, and single-nucleotide polymorphisms | Blast, GenBank, Protein Data Bank, Science Direct, PubMed |
| Kane et al., 2006 | Single nucleotide polymorphisms, protein sequence and function, genomics, DNA microarrays, proteomics, and atomic force | Gene expression |
| Kerfeld and Scott, 2011 | Evolution and biological principles | | |
| Khuri, 2008 | Genetics, molecular biology, biochemistry, genomics, and proteomics | | |
| Koch and Fuellen, 2008 | Molecular biology, biochemistry, genetics, microarrays, molecular life science, and bioethics | | |
| Krilowicz et al., 2007 | Data mining, DNA and protein sequence analysis, motif identification, gene structure prediction, tree construction, and protein structure visualization and analysis | | |
| Luo, 2013 | Primers and data mining | | |
| Medin and Nolin, 2011 | Biology, transcription, translation, mutations, microbiology, genetics, evolutionary conservation, biochemistry, protein structure and function, enzyme kinetics, cell biology, phylogeny, protein sequence alignments, conserved protein domains, molecular biology, genomics, developmental biology, model organisms/comparative genomics, bacterial diversity, diversity of morphologies, physiologies and ecological niches throughout the microbial phylogenetic tree, DNA replication, structural RNA, and proteomics | | |
| Miskowski et al., 2007 | Gene expression, protein secondary structure prediction, gene expression, gene prediction, and gene sequencing | NCBI and BLASTN |
| MacMullen and Denn, 2005 | Microarray gene expression, gene function, and molecular biology | | |

(Continued)
also emphasized the importance of providing bioinformatics students with core skills in scientific communication (e.g., Willighagen, 2010) and knowledge about ethics in the field (Taneri, 2011).

Pedagogical approaches identified to convey bioinformatics-related concepts and procedures include diverse forms of challenge-based learning. Specifically, inquiry-based learning approaches such as problem-based learning and research-based projects are the most common pedagogical methods coupled with the use of learning management systems that provide learners with an entry point where all tools and materials can be organized cohesively. It was also identified that, because of the nature of bioinformatics tools and resources, distance learning another viable approach for delivery of bioinformatics education.

Considering the results on the evaluation and educational research in bioinformatics education, we identified that assessment of educational materials (and not learning in particular) has been the common thread across most of the articles found in this category. The most common assessment mechanisms include pretest and posttest assessments, laboratory reports with their corresponding rubrics, and course examinations. This finding is consistent with the idea that the educational community is at an early stage of the instructional design process, focusing strongly on deciding the content and “what to teach” and the effectiveness of such content. But less progress has been made in identifying most effective ways of how to teach bioinformatics, including a focus on how students learn bioinformatics. This observation can be derived from the descriptions in Table 7, wherein most of the findings relate to student perceptions of learning materials or student self-assessments. These more experimental designs, as well as the introduction of qualitative research methods to identify how learning happens, are needed. Usability studies would also be useful in identifying how the technology supports or limits learning. This finding can also resonate with the circumstance that most of the authors outlined in this research come from disciplines in biology, computer science, and the like, where disciplines such as the social sciences, specifically science

Table 3. Continued

| Reference                  | Concepts                                                                 | Method                                                                 |
|----------------------------|--------------------------------------------------------------------------|------------------------------------------------------------------------|
| Nehm and Budd, 2006        | Evolution, geochemistry, molecular biology, paleobiology, and genetics |                                                                       |
| Nichols et al., 2003       | Illustrate the different percentage of guanine-cytosine content present in the same gene across various organizations; variations in codon usage for each amino acid among organisms; relationships between nucleotide frequency, codon usage, and melting temperatures; and building phylogenetic trees based on a single gene from different organisms |                                                                       |
| Obom and Cummings, 2009    | Molecular biology, cell biology, human genetics, immunology, stem cell biology, proteomics, microbial genomics, molecular structure, and systems biology | Gene organization and expression, genome analysis, microarrays and analysis, and simulation of biological complex systems |
| Perez-Iratxeta et al., 2007| Genomics and protein sequences                                           | DNA microarray data                                                     |
| Rao et al., 2008           | Genomics                                                                 | DNA microarray technology and gene-specific measurements and detections |
| Robertson and Phillips, 2008| PCR                                                                     |                                                                       |
| Sahinidis et al., 2005     | Proteomics, metabolic networks, biochemistry, molecular biology, cell biology, organic chemistry, and genomics | DNA microarray data, complex feedback and control mechanisms          |
| Saier, 2003                | Bioinformatic and biosystematic approaches to address fundamental questions about transmembrane transport systems and to develop probable answers based on systematic phylogenetic analyses | Bioinformatic tools applied to macromolecular evolution                |
| Tolvanen and Vihinen, 2004 | Biochemistry, molecular biology, proteins, and DNA                      |                                                                       |
| Toth and Connelly, 2006    | Molecular biology                                                        |                                                                       |
| Wefer and Sheppard, 2008   | Evolution, nucleotide sequences, amino acid sequences, DNA sequences, and mutations and variations | Evolutionary models                                                     |
| Wightman and Hark, 2012    | Data mining and sequences                                                | PubMed, NCBI, OMIM, and Blast                                           |
| Yang et al., 2008          | Transmembrane proteins, amino acid, amino acid pair composition, RNA, and RNase digestion | Microarray data analysis, Bayesian biclustering model, Gibbs sampling procedure, protein structure prediction and classification, and protein disorder predictor |
| Yang and Zhang, 2008       | Protein structure and function and genomes                              | Sequence analysis, pairwise sequence alignment, multiple sequence alignment, protein structure comparison and classification, protein structure prediction, and gene expression |
| Zhang et al., 2007         | PCR, DNA, theory of molecular evolution, comparative genomics, phylogenetic trees, sequence alignment, biochemistry, and biology | Isolation of cell DNA and preparation of PCRs                          |
education, engineering education, and computing education, can play a larger role in contributing with pedagogical methods, instructional design theories, and assessment mechanisms.

Finally, through this analysis we also found that little research has focused on 1) how faculty members conceive bioinformatics and education in bioinformatics, 2) the identification of an integrated curriculum describing the required content and skills in bioinformatics education at different levels (K–12, undergraduate and graduate), and 3) how students learn bioinformatics.

The limitations of the study relate to the searching methodology and the categorization of the articles found. Specifically, we only considered articles that resulted from searches conducted in the above-mentioned databases and the use of very specific terminology. Also, the categorization was done manually, and although the categorizations were performed at least two times, there is still a possibility that some papers should have been categorized in additional themes. Also, the main source of the categorizations were title, abstract, and keywords, and while most of the papers were reviewed manually, and although the categorizations were performed very specific terminology. Also, the categorization was done in the above-mentioned databases and the use of symposia. Another limitation is the scope of the level of description for each of the resources. For instance, we do not have a way to identify that all courses reported in Table 1 had the same workload, credits, or content, nor do we have specific details to include on the tables, a thorough analysis was not performed on the bodies of the documents.

Another limitation is the scope of the level of description for each of the resources. For instance, we do not have a way to identify that all courses reported in Table 1 had the same workload, credits, or content, nor do we have specific details to include on the tables, a thorough analysis was not performed on the bodies of the documents.

### Table 4. Summary of math and statistics concepts

| Reference                  | Concepts                                      |
|----------------------------|-----------------------------------------------|
| Andersson et al., 2001     | Mathematics: factorial calculations           |
| Burhans and Skuse, 2004    | Mathematics: discrete mathematics             |
| Doom et al., 2003          | Mathematics: calculus and discrete mathematics |
| Fetrow and John, 2006      | Mathematics: abstract modeling and logical and quantitative problem solving |
| Furge et al., 2009         | Mathematics: specific concepts not specified  |
| Goode and Trajkovski, 2007 | Mathematics: discrete mathematics and calculus |
| Hack and Kendall, 2005     | Mathematics: mathematical modeling methods, specific concepts not specified |
| Haux, 2004                 | Mathematics: specific concepts not specified  |
| Kane and Brewer, 2007      | Mathematics: calculus                         |
| Koch and Fuellen, 2008     | Statistics: “probability and statistics,” specific concepts not specified |
| Khuri, 2008                | Mathematics: specific concepts not specified  |
| Krilowicz et al., 2007     | Mathematics: college-level mathematics skills, specific concepts not specified |
| Sahinidis et al., 2005     | Statistics: “statistics and probability,” specific concepts not specified |
| Wightman and Hark, 2012    | Statistics: ratios, probabilities, logarithms |

### Table 5. Summary of articles in category “pedagogical methods”

| Reference                  | Pedagogical method                                  |
|----------------------------|-----------------------------------------------------|
| Andersson et al., 2001     | Theory-anchored evaluation approach and reciprocal evaluation-based collaborative teaching and learning, and design of online learning materials' virtual teacher |
| Beck et al., 2007          | Collaborative research project                      |
| Bednarski et al., 2005     | Inquiry-based labs                                 |
| Boyle, 2004                | Problem-based learning                             |
| Burhans et al., 2004       | Laboratory practices                               |
| Burhans and Skuse, 2004    | Laboratory practices                               |
| Butler et al., 2008        | Close-ended research experience integrating student-centered research projects |
| Cooper, 2001               | Inquiry-based exercises                            |
| Craddock et al., 2007      | Hands-on skills, project-based learning             |
| Fetrow and John, 2006      | In-class exercises and a research-based course project |
| Floraino, 2008             | Hands-on experience                                |
| Furge et al., 2009         | Active learning                                    |
| Hersberger, 1999           | Interactive website for student research           |
| Honts, 2003                | Computer laboratory problems and group research projects |
| Jungck and Donovan, 2000   | Use of the theme of “evolution” to convey bioinformatics |
| Jungck et al., 2010        | Phylogenetic thinking and problem solving           |
| Kane and Springer, 2007    | Training modules and applied scientific computing   |
| Lim et al., 2009           | Problem-based learning                             |
| Ranganathan, 2009          | Tutorials and symposia                             |
| Robertson and Phillips, 2008 | Interactive Primer Design Exercise using the principles of scientific teaching |
| Shapiro et al., 2013       | Problem-based learning (PBL), process-oriented, guided inquiry learning, and peer-led team learning |
| Toth and Connelly, 2006    | Research project                                   |
| Williams et al., 2010      | Informal resources in bioinformatics education     |
| Yang et al., 2008          | Application-oriented approaches and student-centered instructional strategies |
major skills needed to use bioinformatics tools other than identifying data-mining skills and object-oriented programming skills (see Table 2). For instance, specific definitions and examples of how those skills must be developed to become a skilled bioinformatics student are missing. These descriptions were not included mainly because they were outside the scope of the study, but also because not enough details were provided in the original sources. We welcome educators to help the broader community to define these skills and report specific examples of how those can be integrated into working classrooms. This limitation extends to the case of the evaluation data reported in Table 7. Specifically, we do not report specific or quantitative measures of learning (e.g., means, SDs, \(p\) values or effect sizes), making the evaluation of the effectiveness of the educational initiatives difficult to compare. Finally, for the case of Table 4, several papers identified a need for mathematics and statistics content knowledge, yet authors listed only broad terms such as “college-level statistics,” and no specific topics or skills were able to be identified for these fields.

### IMPLICATIONS FOR BIOINFORMATICS EDUCATION

The results of this study have implications for both the teaching and learning of bioinformatics at the undergraduate and graduate levels and the development of a more scholarly based body of knowledge in bioinformatics educational research.

The implications of this study, as related to teaching and learning bioinformatics, focus on the design, validation, and implementation of curricular materials and learning resources. The first step toward this goal is to appropriately orchestrate bioinformatics-related learning outcomes, including concepts, skills, and procedures; to identify the evidence of the learning; and to use appropriate pedagogical approaches (Wiggins and McTighe, 1997). The process for coordinating these three main steps toward the design of instructional curriculum has been denominated “understanding by design” (Wiggins and McTighe, 1997). This framework is a simple but complete model, accessible to all audiences. This model has also been suggested to be particularly effective for the design of outcome-based curricula (Streveler et al., 2012) and for the practice of scientific teaching (Handelsman et al., 2004). Its core consists of a set of tools composed of three main steps: 1) identifying the desired learning outcomes, 2) determining the acceptable evidence of learning, and 3) planning the experiences and instructional approach. Understanding by design is then an educational tool that focuses on the processes essential to the act of teaching and learning and is centered on the design of curriculum and learning experiences to accomplish specified purposes (Wiggins and McTighe, 1997). According to Wiggins and McTighe (1997), effective curricular designs can be accomplished by starting with the desired results, then deriving the curriculum from the evidence of learning and subsequently focusing on the educational methods that will move the students to the desired performance. Through this review of the literature, we have identified a need to clearly define the curriculum and the content that should be taught together with appropriate pedagogical approaches (i.e., inquiry-based learning) and evaluation and assessment mechanisms that go beyond perceptions and motivational aspects and move toward assessing learning outcomes and rigorous research in bioinformatics education.

Along the same lines, educators should also be aware that, as important as the identification of concepts, skills and procedures (i.e., the learning outcomes), is the assessment of the attainment of learning outcomes. Wiggins and McTighe (1997), called for us “to operationalize our goals or standards in terms of assessment evidence as we begin to plan a unit or course” (p. 8). In this work, we have identified that little has been reported in this area. However, a good example is provided by Robertson and Phillips (2008), who utilized this framework to design and implement an active learning activity aimed at designing DNA parameters for PCR. In their implementation, they carefully aligned learning objectives with assessment and activities. They also developed a rubric delineating different levels of performance.

### Table 6. Summary of articles in category “delivery method”

| Reference | Delivery method |
|-----------|-----------------|
| Brazas and Ouellette, 2013 | Online videos, discussion forums |
| Butler et al., 2008 | Active learning and enhanced student–faculty interaction |
| Buttigieg, 2010 | Multimedia presentation and visual communication |
| Campbell, 2003 | Student-based discoveries |
| Cattley, 2004 | Web-based bioinformatics application integrating a variety of common bioinformatics tools for teaching BioManager |
| Cooper, 2001 | Open-ended, inquiry-based exercises |
| Craddock et al., 2007 | Problem-based approach, course management service |
| Crawford, 2007 | Inquiry-based strategies |
| Floraino, 2008 | Lecture and computer practice topics, free for academic use, with software and Web links required for the laboratory exercises |
| Gelbart and Yarden, 2006 | Web-based learning environment and inquiry-based processes |
| Hersh, 2007 | Distance-learning program |
| Jungck et al., 2010 | Academic community of BioQUEST Curriculum Consortium |
| Lim et al., 2003 | Online course distance education |
| Lim et al., 2009 | Learning Activity Management System e-learning tool |
| Machluf and Yarden, 2013 | Learning environment |
| Moll et al., 2006 | Molecular viewer and modeling tool BALLView. |
| Obom and Cummings, 2009 | Online master of science in bioinformatics program |
| Perry et al., 2013 | Games |
| Ranganathan, 2009 | e-Learning tools |
| Searls, 2012 | Online videos virtual course catalogue |
| Shapiro et al., 2013 | Hybrid delivery including peer-assisted learning approaches incorporated into a bioinformatics tutorial for a genome annotation research project |
| Tolvanen and Vihinen, 2004 | Distance-learning program |
| Williams et al., 2010 | Informal sources of bioinformatics education |
### Table 7. Summary of articles in category “educational research or evaluation”

| Reference                  | Focus of evaluation                              | Learning assessment                                                                 | Result                                                                 |
|---------------------------|-------------------------------------------------|--------------------------------------------------------------------------------------|----------------------------------------------------------------------|
| Bednarski et al., 2005    | Student learning and attitudes                   | Group quiz and a pretest and posttest assessment                                     | Students gained understanding of the Web-based databases and tools and enjoyed of the investigatory nature of the lab |
| Brame et al., 2008        | Student knowledge and skills and interest in research | Pretest and posttest assessments and student laboratory reports scored with a rubric | Student response to the project was positive, both in terms of knowledge and skills increases and interest in research |
| Brazeau and Brazeau, 2006 | Student perceptions                            |                                                                                      | Significant improvement on student course ratings on the pedagogical format of the course and the relevance of course material to professional practice |
| Campbell, 2003            | Student academic achievement and perceptions     | Course examinations                                                                   | Students gained the ability to utilize online information to achieve the educational goals of the course and perceived this as a positive experience with respect to how they might contribute to biology |
| Furge et al., 2009        | Student learning strategies and learning        | Student laboratory reports, solutions to problem sets, and in-class presentations      | Largely promotion of active learning in the classrooms and enhanced student understanding of course materials |
| Honts, 2003               | Student learning                                | A take-home final examination                                                         | Students developed working knowledge of bioinformatics concepts and methods |
| Howard et al., 2007       | Student confidence and performance              |                                                                                      | Students gained confidence in solving and ability to solve bioinformatics-related problems. Increased student performance on bioinformatics-related problems |
| Howard et al., 2007       | Faculty perceptions of students’ increased awareness |                                                                                     | Faculty members perceived an increased awareness of the applications of bioinformatics among the students in their courses |
| Krilowicz et al., 2007    | Instructor and student self-reported required prior knowledge and skills |                                                                                     | Identified skills and knowledge from the fields of computer science, biology, and mathematics that are critical for students considering bioinformatics research |
| Lim et al., 2003          | Student perceptions                            |                                                                                      | The course was rated as informative, interactive, and effective for distance learning. Participants expressed that the course content was useful and well presented with good technical support |
| Lim et al., 2009          | Student perceptions                            |                                                                                      | Identified a positive response regarding the usefulness of an e-learning tool in guiding the learning and discussion process involved in problem-based learning and enhanced the learning experience by breaking down PBL activities into a sequential workflow |
| Machluf et al., 2013      | Teachers’ design of an assessment tool as a means of probing their knowledge and beliefs in adopting contemporary scientific research into their classroom |                                                                                     | The analysis of the assessment tool revealed that teachers perceived research as combining laboratory experiments and bioinformatics approaches. Thus, the assessment tool represented characteristics of authentic modern scientific research and the teachers’ appropriation of the new bioinformatics curriculum by extending its roots into the traditional curriculum |
| Medin and Nolin, 2011     | Student self-assessment of their learning gains |                                                                                      | Students reflected that the design aspect of the experiments increased their understanding and retention of molecular biology |
| Obom and Cummings, 2009   | Compared student onsite and online learning and satisfaction | Course examinations                                                                  | Perceived similar levels of satisfaction between most online and on-site student responses, obtained similar performance in grades earned by students in online and on-site courses, and perceived more rigorous course load and more opportunities for participation in online environment |
| Robertson and Phillips, 2008 | Learning goals and assessments of student performance and perceptions | Pretest and posttest assessments, with instructor rubric to report perceived student learning | Students were more poised to troubleshoot problems that arose in real experiments. Students were receptive to the new materials and the majority achieved the learning goals |

(Continued)
As for assessment, findings from Table 7 show that the main purpose of these has been to assess curricular change. Their primary focus has been on identifying student perceptions and attitudes, with the secondary focus being on learning gains. These findings are consistent with the focus of assessment in genomics and bioinformatics reported by Campbell and Nehm (2013). Campbell and Nehm (2013) also performed a critical analysis of the quality of these assessments in which they raised concerns about the validity and reliability of these instruments and thus of the evidence derived from them.

Some of the assessment reported in this literature review is fragmented and superficial. It is fragmented in the sense that is not thorough and in depth. It is superficial in the sense that it is focused on perceptions and motivation and, in very few instances, on learning. This trend is understandable, because biology faculty members specifically, or science, technology, engineering, and mathematics faculty members in general, usually do not have formal training in educational research methods. However, in the same way faculty members seek interdisciplinary collaborations to complement their technical research agendas (e.g., biologists working with computer scientists), they can seek similar interdisciplinary collaborations with educational researchers to support their scholarship of teaching and learning. Furthermore, the National Science Foundation has called for discipline-based education research enterprises combining expertise of scientists providing disciplinary priorities, worldviews, knowledge, and practices, with expertise of educational researchers providing knowledge, theories, and methods that explain learning and cognition (Singer et al., 2012). Having identified the appropriate learning outcomes and evidence of the learning, the last component of the “understanding by design” is to determine the pedagogical method to introduce the concepts and skills. Through this review, it was determined that bioinformatics education has a tremendous potential to be integrated through inquiry-based learning due to the scientific nature of the field. It can also be integrated through face-to-face and online delivery mechanisms such as computing tools and services that are open to the public.

### CONCLUSIONS

In this study, we have investigated the state of undergraduate and graduate bioinformatics education as described in the published literature. In particular, this study emphasizes the opportunities for and challenges to the integration of computing and biology earlier in the curriculum. Opportunities include the enormous potential to integrate both scientific thinking and computational thinking (Wing, 2006) through bioinformatics. Bioinformatics offers the practical and technological infrastructure to introduce computing principles and practices in an applied and scientific way. In the same way, learners can have the opportunity to obtain computing knowledge and skills by using or creating meaningful applications in biology. Challenges, on the other hand, include the need to move beyond a basic application of the tools of bioinformatics to a deeper conceptual understanding of the field. It also suggests the need for training the future workforce not only as consumers of tools and services but also as future professionals who will be producers of tools and resources. This study also revealed the increased importance of preparing professionals in this field at earlier stages. That is, we need to develop the pipeline of workforce at earlier stages. Starting at or even before the high school level. As a result, existing biology education communities of practice (Wenger, 2000) need to be expanded to support teachers and faculty as they adopt and adapt potential learning modules.

Bioinformatics education is an emerging field that requires attention from educators at all levels as well as from educational researchers. The results from this study also reveal the need for a better identification of learning outcomes and a better integration of assessment and pedagogical methods. Furthermore, the appropriate integration of these three components can be disseminated and validated through a more scholarly based integration and development of learning experiences ranging from K–12 to graduate levels. Such integration also calls for involvement from investigators in the social sciences who can help integrate educational research into bioinformatics education, representing an opportunity for educational researchers to study an emerging interdisciplinary field that integrates scientific thinking with computational thinking.

---

**Table 7. Continued**

| Reference                      | Focus of evaluation                   | Learning assessment                                                                 | Result                                                                 |
|-------------------------------|---------------------------------------|-------------------------------------------------------------------------------------|------------------------------------------------------------------------|
| Shachak et al., 2005          | Faculty curriculum design activities   | Identified that Gagne’s conditions of learning instructional design theory provides a useful framework for developing bioinformatics training, but may not be optimal as a method for teaching it |
| Van Mulligen et al., 2008     | Faculty curriculum design activities   | Participants indicated that the training challenge experience had contributed to their understanding and appreciation of multidisciplinary teamwork |
| Wefer and Sheppard, 2008      | Science standards as related to bioinformatics | Identified a generally low representation of bioinformatics-related content in science standards |
| Yang et al., 2008             | Student stimulation to learn           | Increased stimulation on students’ activities in bioinformatics learning based on proper application-oriented bioinformatics curriculum and student-centered instructional strategy |
Jungck JR, Donovan S (2000). Evolution as a basis for bioinformatics education. Mol Biol Cell 11, 26A.

Jungck JR, Donovan SS, Weisstein AE, Khiripet N, Everse SJ (2010). Bioinformatics education dissemination with an evolutionary problem solving perspective. Brief Bioinform 11, 570–581.

Kampov-Polevoy I, Hemminger BM (2011). A curricula-based comparison of biomedical and health informatics programs in the USA. J Am Med Inform Assoc 18, 195–202.

Kane M, Brewer J (2007). An information technology emphasis in biomedical informatics education. J Biomed Inform 40, 67–72.

Kane M, Brewer J, Goldman J, Moidu K (2006). Integrating bioinformatics, clinical informatics, and information technology in support of interdisciplinary curriculum development. Paper presented at the 7th Conference on Information Technology Education, Minneapolis MN.

Kane M, Springer J (2007). Integrating bioinformatics, distributed data management, and distributed computing for applied training in high performance computing. Paper presented at the 8th ACM SIGITE Conference on Information Technology Education, Destin, FL.

Kerfeld CA, Scott KM (2011). Using BLAST to teach “e-value-tion-ary” concepts. PLoS Biol 9, 3389–3402.

Khuri S (2008). A bioinformatics track in computer science. ACM SIGCSE Bull 40, 508–512.

Kitano H (2002). Computational systems biology. Nature 420, 206–210.

Koch I, Fuellen G (2008). A review of bioinformatics education in Germany. Brief Bioinform 9, 232–242.

Krilovicz B, Johnston W, Sharp S, Warter-Perez N, Momand J (2007). A summer program designed to educate college students for careers in bioinformatics. CBE Life Sci Educ 6, 74–83.

Kulkarni-Kale U, Savant S, Chavan V (2010). Bioinformatics education in India. Brief Bioinform 11, 616–625.

LeBlanc M, Dyer B (2004). Bioinformatics and computing curricula 2001: why computer science is well positioned in a post-genomic world. ACM SIGCSE Bull 36, 64–68.

Lim S, Khan A, De Silva M, Lim K, Hu Y, Tan C, Tan T (2009). The implementation of e-learning tools to enhance undergraduate bioinformatics teaching and learning: a case study in the National University of Singapore. BMC Bioinformatics 10(suppl 15), S12.

Lim YP, Hőög JO, Gardner P, Ranganathan S, Andersson S, Subbiah S, Tan T, Hide W, Weiss A (2003). The S star trial bioinformatics course: an online learning success. Biochem Mol Biol Educ 31, 20–23.

Luo J (2013). Teaching the ABCs of bioinformatics: a brief introduction to the applied bioinformatics course. Brief Bioinform, bbt065.

MacLulch Y, Gelhart H, Yarden A (2013). High-school teachers’ appropriation of an innovative curriculum in bioinformatics. Paper presented at the 9th Conference of European Researchers in Didactics of Biology, Berlin, Germany, September 18–22.

MacLulch Y, Yarden A (2013). Integrating bioinformatics into senior high school: design principles and implications. Brief Bioinform 14, 648–660.

MacMullen W, Denn S (2005). Information problems in molecular biology and bioinformatics. J Am Soc Inform Sci Technol 56, 447–456.

Marceglia S, Bonacina S, Mazzola L, Pinciroli F (2007). Education in biomedical informatics: learning by doing bioimage archiving. Paper presented at the 29th Annual International Conference of the IEEE, Engineering in Medicine and Biology Society, Lyon, France, August 23–26.

Medin CL, Nolin KL (2011). A linked series of laboratory exercises in molecular biology utilizing bioinformatics and GFP. Biochem Mol Biol Educ 39, 448–456.

Miskowski J, Howard D, Abler M, Grunwald S (2007). Design and implementation of an interdepartmental bioinformatics program across life science curricula. Biochem Mol Biol Educ 35, 9–15.

Moll A, Hildebrandt A, Lenhof H, Kohlbacher O (2006). BALLView: a tool for research and education in molecular modeling. Bioinformatics 22, 365–366.

Nehm R, Budd A (2006). “Missing links” in bioinformatics education: expanding students’ conceptions of bioinformatics using a biodiversity database of living & fossil reef corals. Am Biol Teach 68, 91–97.

Nichols AJ, Coonrod EM, Elwess NL (2003). Incorporating bioinformatics into the biology classroom through DNA sequence analysis. Bioscene 29, 9–15.

Obom K, Cummings P (2009). Comparison of online and onsite bioinformatics instruction for a fully online bioinformatics master’s program. J Microbiol Biol Educ 8, 22–27.

Perez-Iturzeta C, Andrade-Navarro M, Wren J (2007). Evolving research trends in bioinformatics. Brief Bioinform 8, 88–95.

Perry D, Lynch A, Joshi A, Hellman K, Robinson JJ, Alyssa-Cyze O, Richtarik M, Aragon CR (2013). Diverse player experiences in the design of science games for bioinformatics learning. In: Proceedings of the 2013 Chilean Conference on Human-Computer Interaction, 104–109.

Pevzner P, Shamir R (2009). Computing has changed biology—biology education must catch up. Science 325, 541–542.

Rainey D, Mutter B, Craddock L, Faulkner S, Hart F, Eborall M, Foster L, Cammer S, Tretola B, Sobral B, Crasta O (2007). A project-centric approach for cyberinfrastructure in bioinformatics. Paper presented at the Annual Conference of the American Society for Engineering Education, Honolulu, HI, June 24–27.

Ranganathan S (2005). Bioinformatics education—perspectives and challenges. PLoS Comput Biol 1, e52.

Ranganathan S (2009). Towards a career in bioinformatics. BMC Bioinformatics 10 (suppl 15), S1.

Rao V, Das S, Rao V, Srinubabu G (2008). Recent developments in life sciences research: role of bioinformatics. Afr J Biotechnol 7, 495–503.

Reisdorph N, Stearman R, Kechriss K, Phang TL, Reisdorph R, Prenni J, Erle D, Holden C, Schey K, Nesvizhskii A, Geraci M (2013). Hands-on workshops as an effective means of learning advanced technologies including genomics, proteomics and bioinformatics. Genomics Proteomics Bioinformatics 11, 368–377.

Robertson AL, Phillips AR (2008). Integrating PCR theory and bioinformatics into the biology classroom through DNA sequence analysis. Bioscene 29, 9–15.

Sahinidis N, Harandi M, Heath M, Murphy L, Snir M, Wheeler R, Zukoski CF (2005). Establishing a master’s degree programme in bioinformatics: challenges and opportunities. Syst Biol (Stevenage) 11, 127–131.

Sivasubramanian M, Sivasubramanian P, Sahinidis N, Harandi M, Heath M, Murphy L, Snir M, Wheeler R, Zukoski CF (2005). Establishing a master’s degree programme in bioinformatics: challenges and opportunities. Syst Biol (Stevenage) 11, 243–249.

Sears DB (2012). An online Bioinformatics curriculum. PLoS Comput Biol 8, e1002632.

Shachak A, Ophir R, Rubin E (2005). Applying instructional design theories to bioinformatics education in microarray analysis and primer design workshops. Cell Biol Educ 4, 199–206.
Shapiro C, Ayon C, Moberg-Parker J, Levis-Fitzgerald M, Sanders ER (2013). Strategies for using peer-assisted learning effectively in an undergraduate bioinformatics course. Biochem Mol Biol Educ 41, 24–33.

Singer SR, Nielsen NR, Schweingruber HA (2012). Discipline-based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering, Washington, DC: National Academies Press.

Smith T, Emmeluth D (2002). Introducing bioinformatics into the biology curriculum: exploring the National Center for Biotechnology Information. Am Biol Teach 64, 93–99.

Strauss A, Corbin J (1990). Basics of Qualitative Research: Grounded Theory, Procedures, and Techniques, Newbury Park, CA: Sage.

Streveler RA, Smith KA, Pilotte M (2012). Aligning course content, assessment, and delivery: creating a context for outcome-based education. In: Outcome-Based Education and Engineering Curriculum: Evaluation, Assessment and Accreditation, ed. K Mohd Yusof, S Mohammad, N Ahmad Azli, M Noor Hassan, A Kosnin, and SK Syed Yusof, Hershey, PA: IGI Global.

Taneri B (2011). Is there room for ethics within bioinformatics education? J Comput Biol 18, 907–916.

Tolvanen M, Vihinen M (2004). Virtual bioinformatics distance learning suite. Biochem Mol Biol Educ 32, 156–160.

Toth C, Connelly R (2006). A bioinformatics experience course. J Comput Sci Coll 21, 100–107.

Umarji M, Seaman C, Koru AG, Liu HF (2009). Software Engineering Education for Bioinformatics 22nd Conference on Software Engineering Education and Training, Proceedings, IEEE Xplore, 216–223.

U.S. Department of Energy (2010). Grand Challenges for Biological and Environmental Research: A Long-Term Vision, Washington, DC.

Van Mulligen E, Cases M, Hettne K, Molero E, Weeber M, Robertson K, Baldanero O, de la Calle G, Maojo V (2008). Training multidisciplinary biomedical informatics students: three years of experience. J Am Med Inform Assoc 15, 246–254.

Wefer S, Sheppard K (2008). Bioinformatics in high school biology curricula: a study of state science standards. CBE Life Sci Educ 7, 155–162.

Wenger E (2000). Communities of practice and social learning systems. Organization 7, 225–246.

Wiggins G, McTighe J (1997). Understanding by Design, Alexandria, VA: Association for Supervision and Curriculum Development.

Wightman B, Hark AT (2012). Integration of bioinformatics into an undergraduate biology curriculum and the impact on development of mathematical skills. Biochem Mol Biol Educ 40, 310–319.

Williams J, Mangan M, Perreault-Micale C, Lathe S, Sirohi N, Lathe W (2010). OpenHelix: bioinformatics education outside of a different box. Brief Bioinform 11, 598–609.

Willighagen E (2010). Teaching scientific communication in pharmaceutical bioinformatics education. Abstr Pap Am Chem S 240.

Wing JM (2006). Computational thinking. Commun ACM 49, 33–35.

Yang JY, Yang MQ, Zhu MM, Arabnia H, Deng Y (2008). Promoting synergistic research and education in genomics and bioinformatics. BMC Genomics 9 (suppl 1), 25–28.

Yang X, Zhang Z (2008). Bioinformatics Education in Medical College. Paper presented at the 7th Asian-Pacific Conference on Medical and Biological Engineering, Beijing, China, April 22–25.

Zatz M (2002). Bioinformatics training in the USA. Brief Bioinform 3, 353–360.

Zauhar R (2001). University bioinformatics programs on the rise. Nat Biotechnol 19, 285–286.

Zhang M, Lin C, Olsen G, Beck B (2007). A bioinformatics track with outreach components. Paper presented at the 12th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education, held 23–27 June, in Dundee, Scotland, UK.

Zhang X (2011). Exploring cystic fibrosis using bioinformatics tools: a module designed for the freshman biology course. Biochem Mol Biol Educ 39, 17–20.

Zhong Y, Zhang XY, Ma J, Zhang L (2003). Rapid development of bioinformatics education in China. J Biol Educ 37, 75–78.