Reflections on integrating bioinformatics into the undergraduate curriculum: The Lancaster experience

Derek Gatherer

Abstract
Bioinformatics is an essential discipline for biologists. It also has a reputation of being difficult for those without a strong quantitative and computer science background. At Lancaster University, we have developed modules for the integration of bioinformatics skills training into our undergraduate biology degree portfolio. This article describes those modules, situating them in the context of the accumulated quarter century of literature on bioinformatics education. The constant evolution of bioinformatics as a discipline is emphasized, drawing attention to the continual necessity to revise and upgrade those skills being taught, even at undergraduate level. Our overarching aim is to equip students both with a portfolio of skills in the currently most essential bioinformatics tools and with the confidence to continue their own bioinformatics skills development at postgraduate or professional level.

KEYWORDS
genomics, proteomics, bioinformatics, curriculum design, development and implementation, learning and curriculum design

1 | WHAT IS BIOINFORMATICS?

Most of the readers of this article will probably know the answer to the above question and, if they read further, may wonder why I feel it necessary to offer a potted history of the field. I do this because the main contention of this article is that bioinformatics teaching is in a greater state of flux than other branches of biological science education, and that we can only decide what we need to teach now in bioinformatics by considering what was taught in the past. In the light of these issues, I then present the new curriculum for undergraduate bioinformatics at Lancaster University, outlining how it has developed since 2013 and how I think it is likely to develop into the middle of the next decade.

As the name suggests, bioinformatics might be regarded as anything that can be done on a computer that is of relevance to biology. An occasionally undignified scramble for precedence as the inventor of the word "bioinformatics" was ended by the eventual collective acknowledgment that the first usage was by Hogeweg in 1978.1 In practice, however, bioinformatics does not have such a wide definition. The first papers to use the word in its modern sense appeared around 1993 or 1994, for instance those by Boguski2 and Harper,3 and since then there have been several narrower areas where labor in the field known as bioinformatics has been concentrated. These have varied over the years as funding priorities and intellectual fashions have waxed and waned but, despite this, bioinformatics has been accepted for at least the last two decades as an essential discipline within biology. Consequently, the lack of bioinformatics skills among biology graduates is regularly lamented by both the pharmaceutical industry, which has historically been
one of the major career destinations for those interested in bioinformatics, and by UK central government as part of a more general anxiety concerning lack of quantitative skills among British graduates. In the words of one report from 2017: "Data analytics, especially bioinformatics, appear to be particularly vulnerable".4 National initiatives in the United Kingdom to stimulate "Science, Technology, Engineering and Mathematics" have regularly included development of bioinformatics skills as one of their key goals.5

However, due to the rapidity of technological advancement in biology and transformation of the field into a “big data” science,6,7 it has not always been clear exactly what bioinformatics skills need to be developed among biology graduates. Prior to the launch of the Human Genome Project (HGP) in 1990, bioinformatics was seen very much as an eccentric alternative occupation for those whose careers as laboratory-based researchers had foundered. Despite roots going back to the 1950s, and a modestly thriving literature, bioinformatics was a backwater of science. Suddenly in the mid-1990s, it became hugely in vogue, and the rebranding of Oxford University Press’s journal Computer Applications in the Biosciences as Bioinformatics in 1998, marked a coming-of-age moment. The late 1990s saw the simultaneous mass desertion of academia by bioinformaticians for higher-paid jobs in the pharmaceutical industry—an industry eager to put the data of the HGP to its own use—and the rapid development of 1-year Masters-level courses in bioinformatics by those who remained. The bioinformatics “gold rush” had arrived. For a flavor of the time, see Brass.8 For a more detailed account, see Leendert den Besten.9

The turn of the millennium saw the peak of this first wave of bioinformatics. The bursting of the “dot.com” bubble on March 11, 2000 and the further stock market slump following September 11, 2001, confronted many biotech companies with a withdrawal of investor capital and consequent liquidation or hostile merger. These events occurred just as the HGP was drawing to a close and its results were becoming public domain. A bruised pharmaceutical industry began to move away from the analysis of the genome itself (“target discovery”) to specific drug design projects on what had been discovered (“target validation” and “lead discovery”).10 Those sequence analysts who survived the initial financial crash in industry now found themselves elbowed aside by other bioinformaticians specializing in the analysis of three-dimensional protein structures and how these interact with drug molecules—the subdiscipline of computer-assisted drug design, or “docking” (as the drugs “dock” into small crevices in the proteins). Crucially, dockers often had more of a background in chemistry than the molecular biology-trained sequence analysts.

Meanwhile, in academic bioinformatics, attention during the first 5 years of the new millennium turned away from genome sequencing and became oriented toward gene expression analysis using microarrays.11 Although the major genome projects of the late 90s had been massive undertakings by the standards of previous molecular biology, the advent of microarray genomics and other “-omics” technologies such as proteomics, brought bioinformatics for the first time into the territory of a “big data” science. Omics practitioners, confronted with the problem of making sense of all their data, reached out to the biochemical discipline of metabolic control theory, which had for many years been wrestling with the problems of how to model far smaller-scale biochemical networks. The result was the birth of “systems biology”,12 and an influx of statisticians, mathematicians, and computer scientists into biology. For a short while, it seemed as if most academic bioinformaticians were intent on rebranding themselves as systems biologists or systems bioinformaticians. Network analysis tools became the new center of attention. However, just as this new mainstream in bioinformatics was becoming established, it was once again undermined, not this time by market forces and international politics, but by technological developments.

In the late 1990s, while the HGP was still underway, novel sequencing technologies began to be developed, with an eye to faster and cheaper sequence analysis on a grand scale—“deep sequencing." Many of these technologies were highly innovative and initially beset with multiple technical and engineering problems. However, by the end of the first decade of the twenty-first century, these difficulties began to be solved and deep sequencing entered the research mainstream.13,14 Even microarray analysis, although barely a decade old, began to be edged out by deep sequencing-based transcriptomics as the preferred method for studying gene activity.15 As the third decade of the present century approaches, another technological shift is underway, as long-read sequencing technologies begin to edge out the short-read technologies of the first wave of deep sequencing.16 Table 1 summarizes the rapid development of bioinformatics during this time, identifying the main trends in molecular biology and how they have impacted bioinformatics. It is evident that anyone trained in bioinformatics in the 1990s or even in the 2000s will be seriously in need of a refresher course.

Table 1 also demonstrates how bioinformatics has always been both a discipline that creates new software and one in which that software is put to use. Those who wish to have a career as bioinformaticians need to learn how to write computer programs and, furthermore, to be prepared to learn new computer languages every few years as these are adopted into the field. Bioinformatics has benefitted over the years by influxes of computer science graduates, particularly at times of transition, for example, when microarrays, systems biology, or deep
sequencing made their first appearances each with a whole raft of new problems to be solved. Not all bioinformaticians, however, are full-time software developers. Many spend most of their time using existing software tools to analyze data produced in the lab, and need to know only enough programming to be able to organize their data workflows. This distinction between the “pure” bioinformaticians engaged in software development, and the “applied” bioinformaticians engaged in data analysis is often based on undergraduate degree background: computer scientists being the former and biologists the latter. Teaching bioinformatics in a mixed-background Masters-level course often feels like a struggle to explain biology to computer scientists while simultaneously explaining computing to biologists. The focus of this article, however, is on teaching bioinformatics to biology undergraduates. This is a narrower remit, but one which presents its own challenges.

##  2 | THE EMERGENCE OF BIOINFORMATICS CURRICULA

Table 1 may also be read as an exercise in the bioinformatics subdiscipline of “workbenching,” the heyday of which happened around the turn of the millennium. Workbenchers focused on defining a minimum toolkit for bioinformatics, a suite of “must-have” programs. For an example of this approach, see Baker et al.49 Workbenchers saw their contribution as helping other bioinformaticians to adopt common working methods and shared tool sets, to make starting out in the field easier and to encourage reproducibility and sharing of results. The peak of the field was achieved with the release of Bio-Linux,50 which provided in a single download an entire bioinformatics-oriented operating system preinstalled with hundreds of tools. After the appearance of Bio-Linux, workbenching evaporated as an area of research interest. However, since the last update of Bio-Linux was version 8 in 2014, the necessity for workbenching studies is beginning to arise once more. In applying a workbench ethos to bioinformatics curriculum development, I follow in the footsteps of Greene and Donovan.51 Before describing this in detail, I shall briefly review previous published bioinformatics curricula and discuss the philosophy behind them.

Although, as mentioned above, bioinformatics in its modern sense was well underway by the mid-90s, it took a while for articles on bioinformatics curriculum development to be written. Altman’s 1998 paper52 may be the
first. Many of these initial efforts were possibly responses to the ad hoc nature of the first bioinformatics Masters courses during the 90s “gold rush” era, and the need to inform universities where there were no actual bioinformaticians among the staff, about what was needed if their graduate product was to be fit for purpose in industry. One early influential paper by Hughey and Karplus53 reviewed the experience of the first 5 years of undergraduate bioinformatics teaching at University of California, Santa Cruz, culminating in a degree major in the subject. Dubay et al.54 were the first to describe a Masters curriculum. One of the most striking things in these pioneering papers is their description of the heavy mathematics and engineering pre-requisites for entry to the final year of the course, which would exclude most prospective bioinformatics students in the UK. Some curricula were specifically aimed at computer science students55,56 or emphasized the need for a strong computer science grounding.57 A second surprising feature is how theoretical the courses are, but it must be remembered that they were constructed in an era when far less bioinformatics software had been written, and the emphasis was on teaching students to be program new tools rather than master existing ones. The next few years after Hughey and Karplus’s seminal 2001 paper saw a huge surge in similar descriptive and discursive considerations of bioinformatics teaching (e.g., Zadeh58). Zatz59 produced something almost equivalent to a “which guide” to bioinformatics courses. A workbench perspective was represented by Green and Donovan,51 and Rustad60 explored if special tools are needed for bioinformatics education. Tusch et al.61 were the first to discuss the technical infrastructure needed to run such a course. Most papers were written from a U.S. perspective, but bioinformatics education became a global phenomenon and Shamsir et al.,62 Tastan Bishop et al.,63 and Richard et al.64 provided views from other continents.

The precursors of today’s mixed “Bioinformatics and ....” courses also began to appear in the 5 years after the turn of the millennium, and these also became subjects for discussion in the burgeoning bioinfocurricular literature. For instance, see LeBlanc and Dyer65 on the “Genomics” course at Wheaton College, and Pham et al.66 on the University of Wisconsin-Parkside’s “Molecular Biology and Bioinformatics” undergraduate course. Governmental bodies and professional societies also began to take an interest67,68 and as early as 2003, discussions began to appear of how to do it all online,69–72 and for those with no prior experience.73 One interesting trend74–77 is to choose to emphasize structural bioinformatics, perhaps with an eye to continued demand for drug development “dockers” within the pharmaceutical industry. At the other end of the spectrum, Wightman and Hark78 emphasize the positive impact bioinformatics education has on the mathematical skills of biologists otherwise disinclined to numeracy.

Debate concerning which methods really are the best has had to wait for more recent publications, where a variety of education research perspectives have been presented, such as the core competencies approach,79,80 case study-based learning,81 peer-assisted and team-based learning,82–84 and the use of the popular hobbyist 427sp hardware system.85 Now bioinformatics education has sufficient scholarly groundwork to be considered a field in its own right and reviews have begun to appear.86

3 | THE LANCASTER UNDERGRADUATE BIOINFORMATICS CURRICULUM

The scarcity of bioinformatics provision in the undergraduate curriculum was lamented in 2005 by Hack and Kendal.87 At Lancaster University, bioinformatics only began to appear in the undergraduate biology curriculum in academic year 2013–2014. In writing about the integration of bioinformatics into the undergraduate curriculum, I follow in the footsteps of various authors.55,57,58,74,76,78,81

My own efforts to stand on the shoulders of these giants began initially in a single module, BIOL273 DNA Technology. This module had been running for several years and was a techniques-based course focused on teaching second-year undergraduates the basic skills required in gene cloning, polymerase chain reaction and DNA sequencing. To introduce bioinformatics, two of the laboratory sessions were replaced with bioinformatics computer workshops. In the following academic year, bioinformatics content was added to BIOL113 Genetics and BIOL313 Protein Biochemistry, again by removing some of the existing material to make space for bioinformatics workshops. These module contributions constituted the undergraduate bioinformatics component for the academic years 2014–2015 to 2016–2017 inclusive. In academic year 2017–2018, two major changes were introduced: BIOL313 was redesigned and rebranded as Proteins: Structure, Function and Evolution, removing the remnants of classical protein biochemistry from the course to make way for greater bioinformatics content, and a 4-year course BIOL445 Bioinformatics was initiated. This latter course was the first module at Lancaster devoted entirely to bioinformatics. Lancaster University fourth-year modules have a very mixed group of students, divided approximately equally into undergraduates on 4-year extended undergraduate degrees (MSci), postgraduates on a taught Masters degrees (MSc) and
postgraduates in the first year of a 4-year joint PhD program with the Liverpool School of Tropical Medicine (LSTM). Many of the last category are medical or veterinary graduates with several years of professional experience. Those in the second category are divided fairly equally between overseas students, often from China, and our own undergraduates who have opted to stay for an MSc after graduation. BIOL445 is also unusual in that the entire content is delivered in a single week, rather than the 5- or 10-week courses normal at Lancaster. The compression is designed to minimize student travel between Lancaster and Liverpool for the joint LSTM PhD students.

Finally in the academic year 2018–2019, bioinformatics content was withdrawn from BIOL273 DNA Technology, replaced by material on CRISPR and synthetic biology. A new module BIOL275 Bioinformatics was introduced. Just as BIOL445 was the first Lancaster course dedicated entirely to bioinformatics, BIOL275 was the first offered at exclusively undergraduate level. Table 2 summarizes the bioinformatics content of the modules mentioned earlier.

Table 2 illustrates how the bulk of the bioinformatics delivery at Lancaster takes place in second and fourth years. For the majority of undergraduates who are only on 3-year degrees, bioinformatics is introduced in first year, studied intensively in second year, and then applied to the subject of protein evolution in third year. Those staying for the fourth year receive the same experience as the Masters students. The first 3 years are designed to develop progression from point-and-click internet-focused bioinformatics in first year, through advanced internet-focused bioinformatics and basic Windows stand-alone tool use in second year, to a more advanced command of the tools and their application to a specific problem in protein evolution in the third year. For biochemistry undergraduates, all levels are compulsory. Students from other degree programs are only compelled to enroll for BIOL113 Genetics. This can mean that occasionally students may appear in the third year class without the second year grounding. However, as the tools used within BIOL313 Proteins: Structure, Function, and Evolution are a subset focused on protein evolution, the time required to catch up with the rest of the class is limited. The fourth year partly sits within this learning arc insofar as, for the undergraduates on 4-year degrees, it represents a return from the narrower focus of the third year bioinformatics teaching to the general scope and emphasis on mastery of tools introduced in second year. However, since postgraduate students of various types must also be catered for in fourth year, some of whom will be complete beginners, a certain amount of crash

TABLE 2  Summary of the Lancaster bioinformatics curriculum

| Modules                  | BIOL113 Genetics | BIOL275 Bioinformatics | BIOL313 Proteins: Structure, Function and Evolution | BIOL445 Bioinformatics |
|--------------------------|-------------------|------------------------|-----------------------------------------------------|------------------------|
| Length of course (weeks) | 10                | 10                     | 5                                                   | 1                      |
| Hours of bioinformatics lectures | 1 (of 12 total)   | 2 (of 2 total)         | 6 (of 10 total)                                     | 10 (of 10 total)        |
| Hours of bioinformatics labs          | 1.5 (of 4.5 total) | 15 (of 15 total)       | 12 (of 15 total)                                    | 15 (of 15 total)        |
| Lecture content           | NCBI website and how to search for resources | Introductory and revision lectures, "book-ending" the practical bloc | Selection (dN/dS); phylogenetics; structural bioinformatics | Algorithmic foundations of a. cluster detection b. alignment c. phylogenetics d. motif detection |
| Bioinformatics lab content | Systematic literature search technique | Extensive suite of tools (Windows); Introduction to Bio-Linux | Tools (Windows) related to lectures | Extensive suite of tools (Windows and Bio-Linux) |
| Coursework                | None              | Report demonstrating competence in techniques | Report analyzing protein for selection, and its Bayesian phylogenetics | Report demonstrating competence in techniques |
| Exam                      | Multiple choice   | Multiple choice        | Essays (2 from 4 options)                           | Essays (2 from 3 options) |
course introduction must also be delivered in that module. Whether fourth year undergraduates find this a welcome refresher or an annoying distraction largely depends on the extent to which they absorbed their second year course.

We therefore deliver bioinformatics across our degree programs as an almost equal mixture of dedicated modules (second and fourth years) and integration (first and third years). Our general trajectory has been away from integration toward dedicated modules, with the removal of bioinformatics from BIOL273 DNA Technology in 2018–2019, and the transformation in 2017–2018 of BIOL313 Protein Biochemistry into a strongly bioinformatics-oriented Proteins: Structure, Function, and Evolution. We therefore do not follow the trend of integrating bioinformatics teaching as a minor component of several modules (e.g., Furge et al.,94 or for an extreme example the integration of bioinformatics into 10 courses at University of Wisconsin—La Crosse95).

Table 3 summarizes the software training in our two applications based modules

| Modules                     | BIOL275 Bioinformatics | BIOL445 Bioinformatics |
|-----------------------------|------------------------|------------------------|
| Genome structure viewing    | Artemis21              |                        |
| Sequence alignment          | Clustal17 EMBoss (needle, water),24 Muscle,96 MAFPr97 |                        |
| Sequence search             | BLAST,20 Pfam,31 Prosite98 |                        |
| General tools               | EMBoss (seqet, getorf),24 Primer-BLAST99 |                        |
| Protein structure           | Swiss-Model,37 GOR,100 Coils,101 FPROM102 | +Chimera36 |
| Phylogenetics/phylodynamics | MEGA22                 | +Simplot,28 BEAST,43 SPREAD44 |
| Evolution                   | Not covered            | DNASP,26 DataMonkey41  |
| Next-generation sequencing  | Not covered            | BWA,38 Velvet103       |

Table 3  Software training in the Lancaster bioinformatics curriculum, grouped by subdiscipline

4 | TECHNICAL DELIVERY OF TEACHING AND LEARNING

Likić91 emphasized the introduction of programming skills and the need to go beyond “internet bioinformatics.” My own experience at Lancaster (and in previous bioinformatics teaching in Glasgow) is that teaching biology students a programming language from scratch requires more time than is available. Within a dedicated Masters course on bioinformatics, programming is of course essential, and several languages need to be mastered (Table 1), even if only those currently in vogue are chosen. However, a decision not to include programming skills in undergraduate bioinformatics need not confine us to internet-focused techniques. The large quantity of open-source or closed-but-free tools in the field means that there is ample scope for developing expertise that goes beyond simple knowledge of the best bioinformatics websites (although that is important and is included in first- and second-year teaching). Lancaster University deploys AppsAnywhere (https://www.appsanywhere.com) as an interface to deliver a large range of software to all Windows PCs fully connected to the university network, including both computer lab PCs, staff offices and the personal devices of students. Lancaster University is a Windows-only desktop environment, which precludes the deployment of some popular classic Macintosh applications such as MacClade.104 We use VMware Horizon (https://www.vmware.com/uk/products/horizon.html) to deliver a virtual Bio-Linux server. The Bio-Linux file system is shared with Windows, allowing students to work on the same files within both Windows and Bio-Linux (cf. Floriano105).

5 | EVOLUTION OF LEARNING OBJECTIVES AND ASSESSMENT METHODS OVER TIME

The extensive changes to course content and delivery described earlier have also necessitated change in the learning objectives over the years. At Lancaster, a cascade system of learning objectives is used, starting with overarching objectives for degree programs, then developing more specific learning objectives to each module, with the bottom level consisting of detailed objectives for each teaching session. Approval of new teaching, or of changes to existing teaching, is governed at the module level. Consideration of learning objectives for bioinformatics teaching at Lancaster must therefore take account of the fact that first- and third-year teaching are embedded...
within modules—BIOL113 Genetics and BIOL313 Proteins: Structure, Function, and Evolution—where most or some of the content, respectively, is not bioinformatics, and therefore the learning objectives must be congruent with the broader aims of the module. With the modules entitled Bioinformatics—BIOL275 and BIOL445—there is considerably more room to specify relevant learning objectives in more detail.

See “Data Availability Statement” below for a link to the handouts for the various courses on which lists of learning objectives may be found. These have varied from year-to-year as the emphasis of teaching has evolved. To give one particular example, in BIOL113 Genetics the 2014–2015 bioinformatics content covered recognition of common sequence formats, retrieval of sequences from GenBank, BLAST searching, multiple alignment, and phylogenetic tree building in MEGA. These session-specific detailed learning objectives report upwards to the module learning objectives for BIOL113, among which are two bioinformatics-focused objectives: (1) to become aware of bioinformatics as a discipline and (2) to be able to perform a set of basic bioinformatics techniques. The specific bioinformatics workshop content in BIOL113 changed on two occasions since 2014–2015, requiring adjustments to the detailed sessional learning objectives but without any need to change the overarching module-level objective pertinent to the bioinformatics content. Similar adjustments have been made to BIOL313 over the years, changing sessional learning objectives while maintaining relevance to those of the module as a whole. In the dedicated bioinformatics modules, by contrast, module-level learning objectives often appear directly at sessional level, sharpened, or elaborated as appropriate.

Assessment is also governed at the module level (Table 2). BIOL275 Bioinformatics is part of a series of techniques-focused second-year modules, which includes BIOL273 DNA Technology in which bioinformatics was previously taught, that are all assessed via equally weighted multiple-choice test and practical report. BIOL313 Proteins: Structure, Function, and Evolution is assessed via an exam in which two out of the four essay choices will be on bioinformatics—and the students must write one bioinformatics essay—and a practical report, weighted 60:40, respectively. A similar 60:40 exam/report structure is used for BIOL445 Bioinformatics. In the first run of BIOL445, the exam was a mixture of problem-solving questions and essays, but in subsequent runs only essay questions have been used. This change resulted from an observation in the first run of BIOL445, that there was a very bipolar marks distribution for problem solving questions which skewed the overall exam marks distribution from the bell-curve ideal.

## 6 | THE FUTURE OF BIOINFORMATICS TEACHING

The future of bioinformatics teaching is difficult to predict. The only things that can confidently be said are that bioinformatics will continue to be of central importance to biology education in general, and that bioinformatics teaching a decade from now will look very different to that of today. Table 1 provides a guide to what would have been taught in each of what I conjecture to be the five eras of the discipline. Many of the earlier columns of Table 1 contain software of continued usefulness in the present day, whereas other mentioned software has reached obsolescence (compare Tables 1–3). A particularly rapid turnover is evident in the field of sequencing assembly. The decade spent developing tools for short read deep sequencing assembly, and the corresponding time spent teaching those tools, may soon seem an archaic epoch if the latest long read sequencing technologies fulfill their initial promise. A movement away from the recent years of intense focus on sequence assembly may produce a situation reminiscent of the early 2000s, with systems biology and the omics field beginning to figure once more as a main research orientation of bioinformatics. What is new now in 2020 that was not around in 2005 is the potential for bringing virtual reality, artificial intelligence and the internet-of-things approaches into bioinformatics. I speculate that the first of these, especially as applied to protein structure and electron microscopy, would seem to be the most likely to break through soon into the mainstream. Perhaps bioinformatics classes in the year 2030 will be delivered to students encased in headsets, spinning detailed simulations of proteins and cells before their virtual eyes.

In the meantime, students need to have certain fundamental skills, and they need to have skills that are in demand. Some of those skills are challenging to acquire, especially for those who have not had much previous experience of thinking abstractly, or of thinking quantitatively. There are several places where “threshold concepts,” as defined by Meyer and Land,106 need to be grasped. Given the fickle nature of the employment market in bioinformatics, students also need to have a foundation that will enable them to build new bioinformatics skills once graduated and in the workplace. As with so much in higher education, it is the ability to learn to the highest level, rather than what is actually learned, that is the key.

### ACKNOWLEDGMENTS

I thank all the students, undergraduate and postgraduate, who have participated in my bioinformatics classes at Lancaster University since 2013, and who have been so forthcoming with their feedback. I also thank the Lancaster University Organizational and Educational
Development (OED) team for indirectly prompting me to write this article.

DATA AVAILABILITY STATEMENT
Selected bioinformatics laboratory class protocols and instructional videos from the courses mentioned are available under CC-BY at https://doi.org/10.17635/lancaster/researchdata/308.

ORCID
Derek Gatherer © https://orcid.org/0000-0002-7385-5734

REFERENCES
1. Hogeweg P. Simulating the growth of cellular forms. Simulation. 1978;31:90–96.
2. Boguski MS. Bioinformatics. Curr Opin Genet Develop. 1994;4:383–388.
3. Harper R. Access to DNA and protein databases on the Internet. Curr Opin Biotechnol. 1994;5:4–18.
4. BBSRC and MRC (2017) BBSRC and MRC review of vulnerable skills and capabilities. Executive summary in http://www.bbsrc.ac.uk/documents/1501-vulnerable-capabilities-report-pdf/
5. Pain E. Report: U.K. postdocs need more skills. Science. 2015. http://www.sciencemag.org/careers/2015/05/report-uk-postdocs-need-more-skills. https://doi.org/10.1126/science.caredit.a1500138.
6. Leonelli S. The challenges of big data biology. Elife. 2019;8:e47381. https://doi.org/10.7554/eLife.47381.
7. Marx V. Biology: The big challenges of big data. Nature. 2013;498:255–260.
8. Brass A. Bioinformatics education—a UK perspective. Bioinformatics. 2000;16:77–78.
9. Leendert den Besten M. The rise of bioinformatics. An historical approach to the emergence of a new scientific discipline. Oxford: University of Oxford, 2003.
10. Hughes JP, Rees S, Kalindjian SB, Philpott KL. Principles of early drug discovery. Br J Pharmacol. 2011;162:1239–1249.
11. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science. 1995;270:467–470.
12. Wolkenhauer O. Systems biology: The reincarnation of systems theory applied in biology? Brief Bioinform. 2001;2:258–270.
13. Wicker T, Schlagenhaufl E, Graner A, Close TJ, Keller B, Stein N. 454 sequencing put to the test using the complex genome of barley. BMC Genomics. 2006;7:276.
14. Rougemont J, Amzallag A, Iseli C, Farinelli L, Xenarios I, Naef F. Probabilistic base calling of Solexa sequencing data. BMC Bioinformatics. 2008;9:431.
15. Tariq MA, Kim HJ, Jejelowo O, Pourmand N. Whole-transcriptome RNAseq analysis from minute amount of total RNA. Nucleic Acids Res. 2011;39:e120.
16. Jain M, Koren S, Miga KH, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat Biotechnol. 2018;36:338–345.
17. Higgins DG, Sharp PM. CLUSTAL: A package for performing multiple sequence alignment on a microcomputer. Gene. 1988;73:237–244.
18. Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. Proc Natl Acad Sci USA. 1988;85:2444–2448.
19. Fink WL. Microcomputers and phylogenetic analysis. Science. 1986;234:1135–1139.
20. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–410.
21. Mural RJ. ARTEMIS: A tool for displaying and annotating DNA sequence. Brief Bioinform. 2000;1:199–200.
22. Kumar S, Tamura K, Nei M. MEGA: Molecular evolutionary genetics analysis software for microcomputers. Comput Appl Biosci. 1994;10:189–191.
23. Xia X, Xie Z. DAMBE: Software package for data analysis in molecular biology and evolution. J Hered. 2001;92:371–373.
24. Rice P, Longden I, Bleasby A. EMBOSS: The European molecular biology open software suite. Trends Genet. 2000;16:276–277.
25. Walsh S, Anderson M, Cartinhour SW. ACEDB: A database for genome information. Methods Biochem Anal. 1998;39:299–318.
26. Rozas J, Rozas R. DnaSP, DNA sequence polymorphism: An interactive program for estimating population genetics parameters from DNA sequence data. Comput Appl Biosci. 1995;11:621–625.
27. Yang Z. PAML: A program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 1997;13:555–556.
28. Lole KS, Bollinger RC, Paranjape RS, et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol. 1999;73:152–160.
29. Sayle RA, Milner-White EJ. RASMOL: Biomolecular graphics for all. Trends Biochem Sci. 1995;20:374–376.
30. McClure MA, Smith C, Elton P. Parameterization studies for the SAM and HMMER methods of hidden Markov model generation. Proc Int Conf Intell Syst Mol Biol. 1996;4:155–164.
31. Bateman A, Birney E, Durbin R, Eddy SR, Finn RD, Sonnhammer EL. Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. Nucleic Acids Res. 1999;27:260–262.
32. Birney E, Durbin R. Using GeneWise in the Drosophila annotation experiment. Genome Res. 2000;10:547–548.
33. Clamp M, Cuff J, Searle SM, Barton GJ. The Jalview Java alignment editor. Brief Bioinform. 2001;2:258–270.
34. Dudoit S, Gentleman RC, Quackenbush J. Open source software for the analysis of microarray data. Biotechniques. 2003;34:155–167.
39. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10:R25.

40. Trapnell C, Pachter L, Salzberg SL. TopHat: Discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25:1105–1111.

41. Poon AF, Frost SD, Pond SL. Detecting signatures of selection from DNA sequences using Datamonkey. Methods Mol Biol. 2009;537:163–183.

42. Algan E, Baker D, van den Beek M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res. 2016;44:W3–W10.

43. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012;29:1969–1973.

44. Bielejec F, Rambaut A, Suchard MA, Lemey P. SPREAD: Spatial phylogenetic reconstruction of evolutionary dynamics. Bioinformatics. 2011;27:2910–2912.

45. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27:722–736.

46. Sovic I, Sikic M, Wilm A, Fenlon SN, Chen S, Nagarajan N. Fast and sensitive mapping of nanopore sequencing reads with GraphMap. Nat Commun. 2016;7:11307.

47. Li H. Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094–3100.

48. Hadfield J, Megill C, Bell SM, et al. Nextstrain: Real-time tracking of pathogen evolution. Bioinformatics. 2018;34:4121–4123.

49. Baker PG, Goble CA, Becherhofer S, Paton NW, Stevens R, Brass A. An ontology for bioinformatics applications. Bioinformatics. 1999;15:510–520.

50. Field D, Tiwari B, Booth T, et al. Open software for biologists: From famine to feast. Nature biotechnology. 2006;24:801–803.

51. Greene K, Donovan S. Ramping up to the biology workbench: A multi-stage approach to bioinformatics education. Bioscience J Coll Biol Teach. 2005;31:3–11.

52. Altman RB. A curriculum for bioinformatics: The time is ripe. Bioinformatics. 1998;14:549–550.

53. Hughey R, Karplus K. Bioinformatics: A new field in engineering education in 31st Annual Frontiers in Education Conference. Reno, NV: IEEE, 2001.

54. Dubay C, Brundege JM, Hersh W, Spackman K. Delivering bioinformatics training: Bridging the gaps between computer science and biomedicine. Proceedings of the AMIA Symposium 2002. American Medical Informatics Association, 2002;220–224.

55. Doom T, Raymer M, Krane D, Garcia O. A proposed undergraduate bioinformatics curriculum for computer scientists. SIGCSE Bull. 2002;34:78–81.

56. Morrow C, Wilkins D. A bioinformatics course in the computer science curriculum. Proceedings of the 2nd annual conference on Mid-South College computing. Mid-South College, 2004;192-199.

57. Burhans DT, Skuse GR. The role of computer science in undergraduate bioinformatics education. ACM SIGCSE Bull. 2004;36:417–421.

58. Zadeh J. An undergraduate program in bioinformatics, potentials. IEEE. 2006;25:43–46.

59. Zatz MM. Bioinformatics training in the USA. Brief Bioinform. 2002;3:353–360.

60. Rustad DL. Developing an interactive web-based learning environment for bioinformatics. Oslo, Norway: University of Oslo, 2005.

61. Tusch G, Leidig P, Wolffe G, Elrod D, Strebel C. Technology infrastructure supporting a medical and bioinformatics masters degree. ACM SIGCSE Bull. 2004;36:264.

62. Shamsir MDS, Hussein H, Hashim SZMD, Salim N. Educating the educators: Incorporating bioinformatics into biological science education in Malaysia. National Biology Conference. UTM, Universiti Teknologi Malaysia, Melaka, 2006.

63. Tastan Bishop O, Adebiyi EF, Alzohairy AM, et al. Bioinformatics education—Perspectives and challenges out of Africa. Brief Bioinform. 2015;16:355–364.

64. Richard RJA, Sirraam N. A feasibility study of challenges and opportunities in computational biology: A Malaysian perspective. Am J Appl Sci. 2005;2:1296–1300.

65. LeBlanc MD, Dyer BD. Teaching together: A three-year case study in genomics. J Comput Sci Coll. 2003;18:85–95.

66. Pham DQD, Higgs DC, Statham A, Schleitter MK. Implementation and assessment of a molecular biology and bioinformatics undergraduate degree program. Biochem Mol Biol Educat. 2008;36:106–115.

67. Canadian Genetic Diseases Network (2002) White Paper: Bioinformatics curriculum. Recommendations for undergraduate, graduate and professional programs. http://wikifuse.pbworks.com/f/bioinformatics_whitepaper_en.pdf

68. Welch LR, Schwartz R, Lewitter F. A report of the Curriculum Task Force of the ISCB Education Committee. PLoS Comput Biol. 2012;8:e1002570.

69. Searls DB. An online bioinformatics curriculum. PLoS Comput Biol. 2012;8:e1002632.

70. Searls DB. A new online computational biology curriculum. PLoS Comput Biol. 2014;10:e1003662.

71. Tolvanen M, Vihinen M. Virtual bioinformatics distance learning suite. Biochem Mol Biol Educat. 2004;32:156–160.

72. Lim YP, Höög JO, Gardner P, et al. The S-star trial bioinformatics course: An on-line learning success. Biochem Mol Biol Educat. 2003;31:20–23.

73. Vincent AT, Bourbonnais Y, Brouard JS, et al. Implementing a web-based introductory bioinformatics course for non-bioinformatics majors that incorporates practical exercises. Biochem Mol Biol Educat. 2018;46:31–38.

74. Centeno NB, Villa-Freixa J, Oliva B. Teaching structural bioinformatics at the undergraduate level. Biochem Mol Biol Educat. 2003;31:386–391.

75. Badotti F, Barbosa AS, Reis ALM, do Valle ÍF, Ambrósio L, Bitar M. Comparative modeling of proteins: A method for engaging students’ interest in bioinformatics tools. Biochem Mol Biol Educat. 2014;42:68–78.

76. Oke M, Aghalajobi R, Osifeso M, et al. Design and implementation of structural bioinformatics projects for biological sciences undergraduate students. Biochem Mol Biol Educat. 2018;46:547–554.
GATHERER

77. Inlow JK, Miller P, Pittman B. Introductory bioinformatics exercises utilizing hemoglobin and chymotrypsin to reinforce the protein sequence-structure-function relationship. Biochem Mol Biol Educat. 2007;35:119–124.

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

79. Welch L, Lewitter F, Schwartz R, et al. Bioinformatics curriculum guidelines: Toward a definition of core competencies. PLoS Comput Biol. 2014;10:e1003496.

80. Wu H, Palani A. Bioinformatics curriculum development and skill sets for bioinformaticians. Front Educ. El Paso, TX: IEEE; 2015; p. 1–7.

81. Serve KM, Clayton N, Thomas MA. Using an on-line case study to introduce undergraduate students to bioinformatics. J Idaho Acad Sci. 2013;49:35–36.

82. Brown JAL. Evaluating the effectiveness of a practical inquiry-based learning bioinformatics module on undergraduate student engagement and applied skills. Biochem Mol Biol Educ. 2016;44:304–313.

83. Smith JT, Harris JC, Lopez OJ, Valverde L, Borchert GM. ‘On the job’ learning: A bioinformatics course incorporating undergraduates in actual research projects and manuscript submissions. Biochem Mol Biol Educ. 2015;43:154–161.

84. Shapiro C, Ayon C, Moberg-Parker J, Levis-Fitzgerald M, Sanders ER. Strategies for using peer-assisted learning effectively in an undergraduate bioinformatics course. Biochem Mol Biol Educ. 2013;41:24–33.

85. Barker D, Ferrier DE, Holland PW, et al. 4273pi: Bioinformatics education on low cost ARM hardware. BMC Bioinform. 2013;14:243.

86. Magana AJ, Taleyarkhan M, Alvarado DR, Kane M, Springer J, Clase K. A survey of scholarly literature describing the field of bioinformatics education and bioinformatics educational research. CBE Life Sci Educ. 2014;13:607–623.

87. Hack C, Kendall G. Bioinformatics: Current practice and future challenges for life science education. Biochem Mol Biol Educ. 2005;33:82–85.

88. Zhang X. Exploring cystic fibrosis using bioinformatics tools: A module designed for the freshman biology course. Biochem Mol Biol Educ. 2011;39:17–20.

89. Weisman D. Incorporating a collaborative web-based virtual laboratory in an undergraduate bioinformatics course. 2010; 38:4–9.

90. Chapman BS, Christmann JL, Thatcher EF. Bioinformatics for undergraduates: Steps toward a quantitative bioscience curriculum. Biochem Mol Biol Educ. 2006;34:180–186.

91. Likić VA. Computer programming and biomolecular structure studies: A step beyond internet bioinformatics. Biochem Mol Biol Educ. 2006;34:1–4.

92. Boyle JA. Bioinformatics in undergraduate education: Practical examples. Biochem Mol Biol Educ. 2004;32:236–238.

93. Feig AL, Jabri E. Incorporation of bioinformatics exercises into the undergraduate biochemistry curriculum. Biochem Mol Biol Educ. 2002;30:224–231.

94. Furge LL, Stevens-Truss R, Moore DB, Langeland JA. Vertical and horizontal integration of bioinformatics education. Biochem Mol Biol Educ. 2009;37:26–36.

95. Miskowski JA, Howard DR, Ahler ML, Grunwald SK. Design and implementation of an interdepartmental bioinformatics program across life science curricula. Biochem Mol Biol Educ. 2007;35:9–15.

96. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–1797.

97. Katoh K, Asimenos G, Toh H. Multiple alignment of DNA sequences with MAFFT. Methods Mol Biol. 2009;537:39–64.

98. Sigrist CJ, Cerutti L, Hulo N, et al. PROSITE: A documented database using patterns and profiles as motif descriptors. Brief Bioinform. 2002;3:265–274.

99. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinform. 2012;13:134.

100. Garnier J, Ogut Thorpe DJ, Robson B. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J Mol Biol. 1978;120:97–120.

101. Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. Science. 1991;252:1162–1164.

102. Solovyev VV, Shahnuradov IA. PromH: Promoters identification using orthologous genomic sequences. Nucleic Acids Res. 2003;31:3540–3545.

103. Zerbino DR. Using the Velvet de novo assembler for short-read sequencing technologies. Curr Protocol Bioinf. 2010; Chapter 11, Unit 11 5. https://doi.org/10.1002/0471250953.bi1105s31.

104. Maddison WP, Maddison DR. Interactive analysis of phylogeny and character evolution using the computer program MacClade. Folia Primatol (Basel). 1989;53:190–202.

105. Floriano WB. A portable bioinformatics course for upper-division undergraduate curriculum in sciences. Biochem Mol Biol Educ. 2008;36:325–335.

106. Meyer J, Land R. Threshold concepts and troublesome knowledge: Linkages to ways of thinking and practising within the disciplines. Enhancing Teaching-Learning Environments in Undergraduate Courses. Edinburgh: School of Education, University of Edinburgh, 2003.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Gatherer D. Reflections on integrating bioinformatics into the undergraduate curriculum: The Lancaster experience. Biochem Mol Biol Educ. 2020;48: 118–127. https://doi.org/10.1002/bmb.21320