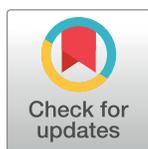# Bioinformatics core competencies for undergraduate life sciences education

Melissa A. Wilson Sayres[1], Charles Hauser[2], Michael Sierk[3], Srebrenka Robic[4], Anne G. Rosenwald[5], Todd M. Smith[6], Eric W. Triplett[7], Jason J. Williams[8], Elizabeth Dinsdale[9], William R. Morgan[10], James M. Burnette, III[11], Samuel S. Donovan[12], Jennifer C. Drew[7], Sarah C. R. Elgin[13], Edison R. Fowlks[14], Sebastian Galindo-Gonzalez[15], Anya L. Goodman[16], Nealy F. Grandgenett[17], Carlos C. Goller[18], John R. Jungck[19], Jeffrey D. Newman[20], William Pearson[21], Elizabeth F. Ryder[22], Rafael Tosado-Acevedo[23¤a], William Tapprich[24], Tammy C. Tobin[25], Arlín Toro-Martínez[26], Lonnie R. Welch[27], Robin Wright[28¤b], Lindsay Barone[8], David Ebenbach[29], Mindy McWilliams[29], Kimberly C. Olney[1], Mark A. Pauley[30¤b]*

1 School of Life Sciences, Arizona State University, Tempe, Arizona, United States of America,
2 Department of Biological Sciences, St. Edward's University, Austin, Texas, United States of America,
3 Bioinformatics Program, Saint Vincent College, Latrobe, Pennsylvania, United States of America,
4 Department of Biology, Agnes Scott College, Decatur, Georgia, United States of America, 5 Department of Biology, Georgetown University, Washington, D.C., United States of America, 6 Digital World Biology, Seattle, Washington, United States of America, 7 Microbiology and Cell Science Department, University of Florida, Gainesville, Florida, United States of America, 8 Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, United States of America, 9 Department of Biology, San Diego State University, San Diego, California, United States of America, 10 Department of Biology, College of Wooster, Wooster, Ohio, United States of America, 11 College of Natural & Agricultural Sciences, University of California, Riverside, Riverside, California, United States of America, 12 Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America, 13 Department of Biology, Washington University in St. Louis, St. Louis, Missouri, United States of America, 14 Department of Biological Sciences, Hampton University, Hampton, Virginia, United States of America, 15 Department of Agricultural Education and Communication, University of Florida, Gainesville, Florida, United States of America, 16 Department of Chemistry and Biochemistry, California Polytechnic State University, San Luis Obispo, California, United States of America, 17 Department of Teacher Education, University of Nebraska at Omaha, Omaha, Nebraska, United States of America, 18 Department of Biological Sciences, North Carolina State University, Raleigh, North Carolina, United States of America, 19 Department of Biological Sciences, University of Delaware, Newark, Delaware, United States of America, 20 Department of Biology, Lycoming College, Williamsport, Pennsylvania, United States of America, 21 Department of Biochemistry and Molecular Genetics, University of Virginia School of Medicine, Charlottesville, Virginia, United States of America, 22 Department of Biology and Biotechnology, Worcester Polytechnic Institute, Worcester, Massachusetts, United States of America, 23 Department of Natural Sciences, Inter American University of Puerto Rico, Metropolitan Campus, San Juan, Puerto Rico, United States of America, 24 Department of Biology, University of Nebraska at Omaha, Omaha, Nebraska, United States of America, 25 Department of Biology, Susquehanna University, Selinsgrove, Pennsylvania, United States of America, 26 Department of Biology, Chemistry, and Environmental Sciences, Inter American University of Puerto Rico, San Germán Campus, San Germán, Puerto Rico, United States of America, 27 Department of Computer Science, Ohio University, Athens, Ohio, United States of America, 28 Department of Biology Teaching and Learning, University of Minnesota, Saint Paul, Minnesota, United States of America, 29 Center for New Designs in Learning and Scholarship, Georgetown University, Washington, D.C., United States of America, 30 School of Interdisciplinary Informatics, University of Nebraska at Omaha, Omaha, Nebraska, United States of America

¤a Current address: National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, San Juan, Puerto Rico, United States of America
¤b Current address: Division of Undergraduate Education, Directorate for Education and Human Resources, National Science Foundation, Alexandria, Virginia, United States of America
* mpauley@unomaha.edu

Abstract

Although bioinformatics is becoming increasingly central to research in the life sciences, bioinformatics skills and knowledge are not well integrated into undergraduate biology education. This curricular gap prevents biology students from harnessing the full potential of their education, limiting their career opportunities and slowing research innovation. To advance the integration of bioinformatics into life sciences education, a framework of core bioinformatics competencies is needed. To that end, we here report the results of a survey of biology faculty in the United States about teaching bioinformatics to undergraduate life scientists. Responses were received from 1,260 faculty representing institutions in all fifty states with a combined capacity to educate hundreds of thousands of students every year. Results indicate strong, widespread agreement that bioinformatics knowledge and skills are critical for undergraduate life scientists as well as considerable agreement about which skills are necessary. Perceptions of the importance of some skills varied with the respondent's degree of training, time since degree earned, and/or the Carnegie Classification of the respondent's institution. To assess which skills are currently being taught, we analyzed syllabi of courses with bioinformatics content submitted by survey respondents. Finally, we used the survey results, the analysis of the syllabi, and our collective research and teaching expertise to develop a set of bioinformatics core competencies for undergraduate biology students. These core competencies are intended to serve as a guide for institutions as they work to integrate bioinformatics into their life sciences curricula.

Introduction

Over the past two decades, the rapid development of high-throughput technologies, data storage capacity, and sophisticated algorithms has produced substantial changes in research practices in the life sciences and medicine. In order for researchers and practitioners in these areas to take advantage of the changes in data availability, they need to have computational and quantitative skills—such as those encompassed by bioinformatics—beyond what was required of them in the past. For more than a decade, authoritative calls from a variety of professional organizations to update undergraduate life sciences curricula have stressed the importance of increasing quantitative and computational education to prepare life sciences students for 21st-century careers [1–14]. Examples include *BIO2010: Transforming Undergraduate Education for Future Research Biologists*, a report by the National Academy of Sciences [1]; *Math & Bio 2010: Linking Undergraduate Disciplines*, a joint project of the American Association for the Advancement of Sciences, the American Society for Mathematical Biology, and the Mathematical Association of America [11]; *Scientific Foundations for Future Physicians*, a publication from the Association of American Medical Colleges and the Howard Hughes Medical Institute [12]; and *Vision and Change in Undergraduate Biology Education: A Call to Action* [13] and *Chronicling Change, Inspiring the Future* [14]. These reports are consistent in their recommendations that life sciences majors should receive better training in chemistry, physics, mathematics, and computation because such knowledge and skills are necessary to address questions at all levels of biology. In addition, several publications have called for life scientists to develop more sophisticated data analysis and programming skills so that they can benefit from the bioinformatics revolution [7, 15–20]. These include a report from the RECOMB Bioinformatics Education Conference [7] and a recent essay in *Nature* [17]. Along these lines, it is important

to note that in its 2015 employment predictions, the United States Bureau of Labor Statistics projected that between now and 2024, 75% of new science, technology, engineering, and mathematics (STEM) jobs will involve computation [21, 22].

On a national scale, a number of innovative programs have been created to address the need for increased computational and quantitative training in biology. These include the Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) program [23], Genomics Education Partnership (GEP) [3, 24], Genome Consortium for Active Teaching (GCAT) [4], GCAT NextGen Sequencing Group (GCAT-SEEK) [25], and Genome Solver [10]. On a local scale, we are aware that the number of courses and meetings focused on undergraduate bioinformatics is increasing and that many existing courses have been modified to include bioinformatics. However, despite these efforts, the adoption of bioinformatics is often limited either to a small number of institutions or to particular courses within a given curriculum. To take an example, SEA-PHAGES engages first-year undergraduates in a genuine research experience and includes substantial bioinformatics analysis as part of its year-long curriculum. As reported in 2014, the program had been used at seventy-three institutions by 4,800 students over five years [23]. Although the reach of SEA-PHAGES is impressive, it impacts only a small fraction of the thousands of U.S. institutions that offer a biology degree and the approximately 110,000 students who now graduate annually with biological and biomedical sciences degrees (not to mention the 220,000 more who graduate from health and related professional programs) [26]. In addition, because the program is designed for first-year students, it does not necessarily impact the rest of the curriculum at a particular institution. The reaches of the other mentioned programs are similar to, or not as extensive as, that of SEA-PHAGES. Despite these large-scale efforts and those of independent faculty to build resources for bioinformatics education, bioinformatics is not yet a standard component of life sciences education [15].

Several groups have made independent efforts to develop or describe bioinformatics curricula, but they have primarily focused on training undergraduate and graduate bioinformaticians, not undergraduate life scientists. For example, the Curriculum Task Force of the Education Committee of the International Society for Computational Biology (ISCB) surveyed directors of bioinformatics core facilities in Europe, Israel, the United States, and Canada about the skills needed for success in the field of bioinformatics and what skills were lacking in the bioinformaticians they recently hired. Based on the results, the Task Force developed a set of core competencies for bioinformatics professionals [9, 27]. They also described three professional roles that require bioinformatics training and the different but overlapping competencies required for individuals in those roles. Since their publication, the ISCB competencies have been used to establish and refine bioinformatics degree and certificate programs, tracks, and specializations in the U.S., Africa [28], Australia, and the United Kingdom. As another example, Koch and Fuellen described the bioinformatics curricula at various German universities at both the graduate and undergraduate level [29]. However, a few authors have considered the specific bioinformatics needs of general life scientists. For example, Maloney et al. discussed the importance of incorporating bioinformatics into undergraduate biology curricula and presented examples of where this has been done successfully [30]. The East Asia Bioinformation Network developed a bioinformatics skill set for biology students in the developing countries that make up the Association for Southeast Asian Nations (https://eabn.apbionet.org/3eabn08/docs.shtml). Finally, Tan et al. proposed a minimum bioinformatics skill set for life sciences curricula (graduate and undergraduate) in developed nations, formulated from discussions at a conference focused on education in the Asia-Pacific region [8]. Thanks to developments such as these, the state of integration of bioinformatics into life sciences education is maturing.

The Network for Integrating Bioinformatics into Life Sciences Education (NIBLSE, pronounced "nibbles") is using an evidence-based approach to expand and promote the integration of bioinformatics skills and knowledge into undergraduate life sciences education [31]. NIBLSE is a National Science Foundation (NSF) Undergraduate Biology Education Research Coordination Network (RCN-UBE) that was formed in 2014 to build on and expand the curricular developments mentioned above. Since further integration of bioinformatics into undergraduate life sciences education will require the participation of non-expert faculty, one of the goals of NIBLSE is to provide resources (e.g., curricular materials and assessments) and support to faculty interested in expanding the integration of bioinformatics within their departments and programs. Note that NIBLSE is focused on the integration of bioinformatics into undergraduate life sciences curricula, which is related to, but distinct from, the curricular requirements of a "pure" bioinformatics degree. The goals of NIBLSE and the ISCB Task Force mentioned above are thus distinct.

In order for NIBLSE to provide curricular resources and assessment tools that align with the needs of undergraduate life sciences students, a set of core bioinformatics competencies for the training of these students is needed as a framework. To that end, NIBLSE recently conducted a survey (hereafter the "NIBLSE survey" or just the "survey") that targeted U.S. biologists. The survey was designed to assess the importance of bioinformatics and bioinformatics skills for undergraduate life scientists, and it included questions about teaching bioinformatics to this group of students and the challenges involved in doing so. In addition, to get a sense of which bioinformatics skills faculty are covering now, respondents were asked to submit syllabi from their courses that incorporate bioinformatics. The analyses of the survey results and syllabi were the subjects of a national NIBLSE conference in August 2016. Outcomes of the analyses and conference are described below and include a proposed set of core bioinformatics competencies for undergraduate life sciences majors. A separate paper addresses the barriers to integrating bioinformatics into life sciences education and NIBLSE's recommendations for overcoming them [32].

The NIBLSE Bioinformatics Core Competencies for Undergraduate Life Scientists (hereafter the "NIBLSE Core Competencies" or just the "Core Competencies") proposed here are intended to serve as a guide for institutions as they integrate bioinformatics into their own life sciences curricula. To facilitate this necessary and important change, NIBLSE is collecting and helping to develop curricular resources for faculty and will soon be developing assessment tools and faculty development resources that are aligned with the Core Competencies.

Materials and methods

Survey development and dissemination

The NIBLSE Core Competencies Working Group (CCWG), composed of biologists and bioinformaticians from a range of educational institutions and industry, developed the survey using an iterative process over the course of several months. Feedback from other NIBLSE members and two evaluation specialists, one with expertise in STEM education, was used to improve the questions and layout of the survey. The survey was implemented and distributed using Qualtrics (https://www.qualtrics.com) with assistance from the Center for New Designs in Learning and Scholarship at Georgetown University. Approval for the study was obtained from the University of Nebraska at Omaha Institutional Review Board (IRB #161-16-EX) before the survey was distributed.

The survey was branched, with some questions or sections presented or skipped based on responses to filtering questions. For example, the survey branched depending on whether the respondent taught at a four-year institution, a two-year institution, or provided not-for-credit

training (e.g., for a company or organization). The branched structure allowed us to formulate targeted questions. The survey in its entirety is provided as supporting information in two versions–one in which the branching structure of the survey can be followed (S1 Survey) and one that provides just the survey questions (S2 Survey).

The survey was divided into three sections. One section asked respondents to provide demographic information, such as gender, race, and highest degree earned (see Results for a complete list). Another questioned respondents about real and perceived barriers to the integration of bioinformatics into life sciences education. The third section asked respondents to rate the importance of fifteen bioinformatics skills (hereafter the "survey skills") in undergraduate life sciences education using a five-level Likert scale. A free-response question allowed respondents to specify skills they thought were missing. The survey skills, S1 to S15, are listed below; the text in parentheses is the abbreviation of the given skill.

- S1 (*Role*)—Understand the role of computation and data mining in hypothesis-driven processes within the life sciences

- S2 (*Concepts*)—Understand computational concepts used in bioinformatics, e.g., meaning of algorithm, bioinformatics file formats

- S3 (*Statistics*)—Know statistical concepts used in bioinformatics, e.g., E-value, z-scores, t test, type-1 error, type-2 error, employ R

- S4 (*Access genomic*)—Know how to access genomic data, e.g., in NCBI nucleotide databases

- S5 (*Tools genomic*)—Be able to use bioinformatics tools to analyze genomic data, e.g., BLASTN, genome browser

- S6 (*Access expression*)—Know how to access gene expression data, e.g., in UniGene, GEO, SRA

- S7 (*Tools expression*)—Be able to use bioinformatics tools to analyze gene expression data, e.g., GeneSifter, David, ORF Finder

- S8 (*Access proteomic*)—Know how to access proteomic data, e.g., in NCBI protein databases

- S9 (*Tools proteomic*)—Be able to use bioinformatics tools to examine protein structure and function, e.g., BLASTP, Cn3D, PyMol

- S10 (*Access metabolomic*)—Know how to access metabolomic and systems biology data, e.g., in the Human Metabolome Database

- S11 (*Pathways*)—Be able to use bioinformatics tools to examine the flow of molecules within pathways/networks, e.g., Gene Ontology, KEGG

- S12 (*Metagenomics*)—Be able to use bioinformatics tools to examine metagenomics data, e.g., MEGA, MUSCLE

- S13 (*Scripting*)—Know how to write short computer programs as part of the scientific discovery process, e.g., write a script to analyze sequence data

- S14 (*Software*)—Be able to use software packages to manipulate and analyze bioinformatics data, e.g., Geneious, Vector NTI Express, spreadsheets

- S15 (*Computational environment*)—Operate in a variety of computational environments to manipulate and analyze bioinformatics data, e.g., Mac OS, Windows, web- or cloud-based, Unix/Linux command line

S1 to S15 were based on the *CourseSource* Bioinformatics Learning Framework developed in 2014 by members of NIBLSE and refined using feedback from a number of groups including GEP and GCAT-SEEK [33]. (*CourseSource* is an open-access journal of peer-reviewed teaching resources for undergraduate biological sciences.) In turn, the Learning Framework was informed by the core competencies for bioinformaticians developed by the ISCB Curriculum Task Force [9, 27]. Two members of the CCWG serve on the ISCB Task Force, one of whom serves as co-chair. The results reported here focus on the first and third sections of the survey (demographics and bioinformatics skills); a separate paper addresses the second (barriers to integration) [32].

A list of more than 11,000 randomly-selected email addresses of biologists at U.S. institutions of higher education was purchased from MDR, an education marketing company (http://schooldata.com). The list included faculty at both four-year and two-year institutions at an approximately 70:30 ratio. Using Qualtrics, unique links to the survey were emailed to the addresses in this list; unique links allowed us to tie a given response to an email address and therefore to the home institution of the respondent. A generic link to the survey was sent to the members of scientific organizations, including the Society for the Advancement of Biology Education Research, GEP, and Biology Scholars alumni (American Society for Microbiology). (Responses to the generic link could not be tied to an institution.) In addition, the survey was advertised on Twitter and was announced in the CyVerse, Digital World Biology, Bio-Link, American Society for Cell Biology, and National Association of Biology Teachers newsletters as well as on the ScienceBlogs website. Finally, potential respondents were asked to forward the generic link to colleagues.

Survey results and analysis

A total of 1,260 responses to the survey were received (n = 1,260), 82% of which came from unique links (Fig 1). Survey results were analyzed by a subset of the NIBLSE CCWG, the interdisciplinary Core Competencies Team (CCT). Differences in Likert responses by self-reported demographics were determined by applying a two-sample Kolmogorov-Smirnov (KS) test to pairwise comparisons of groups. The KS tests were implemented in scripts written in R; these scripts are available at the NIBLSE repository on GitHub [34].

Using public sources [35–37] and information from MDR, we estimate that there are currently between 75,000 and 100,000 biological sciences faculty in the U.S. Given our overall sample size (1% to 2%) and the number of responses per question (between 970 and 1,009), we estimate that the mean margin of error for the survey questions described in this paper is ±3% at the 95% confidence interval.

Syllabi assessment

Ninety syllabi of courses with bioinformatics content were uploaded by survey respondents. Most (sixty-nine) were known or inferred to be from departments of biology, with the balance being from computer science/engineering (nine), biochemistry (five), math (three), and chemistry (two) departments; the home department of two syllabi could not be determined. Each syllabus was assessed independently by two CCT members, who determined which of the fifteen survey skills (S1 to S15) were covered in the course. When two assessments did not agree (approximately 25% of the time) the discordance was resolved by a third CCT member. Additionally, each member kept track of additional skills (i.e., not one of S1 to S15) covered in each syllabus.

Development of the Core Competencies

Using an iterative process, the survey skills were refined into the Core Competencies based on results from the survey, comments from survey respondents, assessment of syllabi, and the

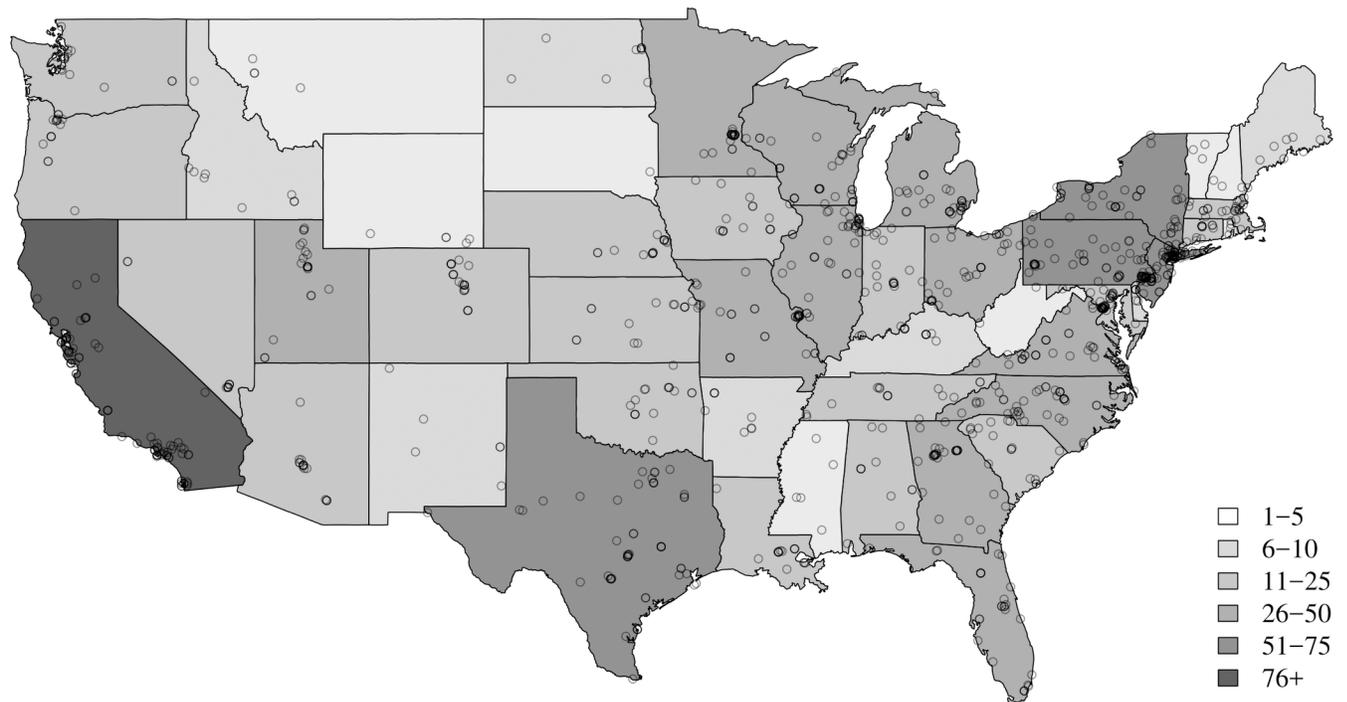## Geographic Distribution of NIBLSE Survey Respondents

Fig 1. Geographic distribution of NIBLSE survey respondents. The location (city/state) of each response to the survey was obtained using e-mail and/or IP addresses. The distribution of responses for the contiguous U.S. is shown ($n = 1,081$). A light circle represents one response at a particular location; a darker circle represents multiple responses at the same location (the darker the circle, the more responses). Note that the legend applies to the states themselves—e.g., there were more than seventy-five responses from California—and that there are no states with no responses. Responses (not shown) were also received from Alaska, Hawaii, Argentina, Australia, Canada, Denmark, France, Italy, Korea, New Zealand, Norway, Puerto Rico, the Republic of Poland, Switzerland, and the United Kingdom.

https://doi.org/10.1371/journal.pone.0196878.g001

collective expertise and experience of the authors in both teaching bioinformatics and applying it to research. As mentioned previously, the CCT analyzed the survey results and syllabi with respect to the survey skills. Based on this work, the CCT developed a set of tentative competencies. Included in this set were skills that showed up repeatedly in the syllabi and responses to the free-response question (see above) but were not in S1 to S15. On the other hand, some of the survey skills were revised or dropped from the tentative competencies set due to weak support, including S10 (*Access metabolomic*) and S15 (*Computational environment*). Prior to the August 2016 NIBLSE conference, the tentative set of competencies was distributed to the conference participants for their review and consideration.

At the NIBLSE conference, the CCT formally presented the tentative competencies to the conference participants and summarized the evidence on which they were based (survey results, syllabi assessment, etc.). The conference attendees then broke into small groups, each moderated by a member of the CCT. The groups discussed the evidence and revised the competencies based on both their interpretation of the evidence and their own expertise and experience. The groups then reconvened into one large group to discuss the resulting lists and develop a consensus set. During this discussion, considerations were made to balance specificity with generalizability. The result of this discussion was a set of nine competencies agreed to by all. Finally, the conference attendees voted to allow the NIBLSE leadership team (PI and Co-PIs of the NSF RCN-UBE) to decide on the specific final wordings of the competencies.

The resulting Bioinformatics Core Competencies for undergraduate life sciences education are presented below (see Results).

Results

Demographics

The group of survey respondents was distributed fairly evenly between males and females as well as across institutions by Carnegie classification, by measurement of institution size (total students, total undergraduates, majors, and faculty), and by self-reported minority-serving status versus not minority-serving (Fig 2). Respondents were from all fifty states with the distribution of responses roughly matching the population distribution of the country (Fig 1). The respondents were predominantly white and non-Latino/Hispanic; for the vast majority, a PhD was the highest earned degree (Fig 2). The level of bioinformatics training varied widely among respondents, including no training, self-taught, short courses, undergraduate and graduate classes, and graduate degrees (Fig 2).

Ratings of bioinformatics skills

As discussed in Materials and Methods, the survey asked respondents to provide Likert-scale responses about the importance of fifteen bioinformatics skills, S1 to S15, in undergraduate life sciences education (see that section for a description of each). Analysis of this data focused on determining, using a two-sample Kolmogorov-Smirnov (KS) test (see Materials and Methods), whether the Likert responses differed significantly across demographic variables. Results are summarized in Fig 3 and Table 1. Skills that received the highest mean responses overall included S1 (*Role*), S3 (*Statistics*), S4 (*Access genomic*), and S5 (*Tools genomic*) (Table 1). As detailed below, the perceived importance of some skills varied based on the demographics of the respondents.

No difference by gender or minority-serving institution status. Analyzing the results for the fifteen survey skills, responses did not vary based on gender or whether the respondent was at a minority-serving institution (see Materials and Methods and [34]). Respondents were overwhelmingly white and non-Hispanic/non-Latino (Fig 2). Individuals who identified as people of color (POC) and those who identified as Hispanic/Latino consistently rated the importance of every skill higher (and often significantly higher, i.e., indicated that it was more important) than those who did not identify as POC or Hispanic/Latino [34]. However, given the tremendous disparity in sample sizes (Fig 2), and known cultural differences in responses to surveys [38, 39], we hesitate to assign much weight to these particular differences.

Higher ratings of some skills at larger institutions. Respondents gave some survey skills different scores depending on the size of their institution, whether that is determined by total number of students, total number of undergraduates, or number of faculty in the department. Respondents at institutions with fewer than 5,000 total students or 5,000 undergraduates rated the importance of S1 (*Role*), S2 (*Concepts*), S13 (*Scripting*), and S15 (*Computational environment*) significantly lower (i.e., indicated that they were less important) than those at institutions with more than 15,000 total students or 15,000 undergraduates [34]. Similarly, faculty in larger departments gave significantly higher ratings to S1 (*Role*) and S2 (*Concepts*) than those in smaller departments (S1: $P = 6.73 \times 10^{-3}$ for a two-tailed KS test; S2: $P = 1.49 \times 10^{-3}$).

Among the four Carnegie Classifications, there were no significant differences in Likert ratings between respondents at Baccalaureate and Master's institutions (S1 Table). In contrast, respondents at Associate's institutions routinely rated every skill lower than did those at other institution types, whereas those at Doctoral institutions rated every skill higher than those at
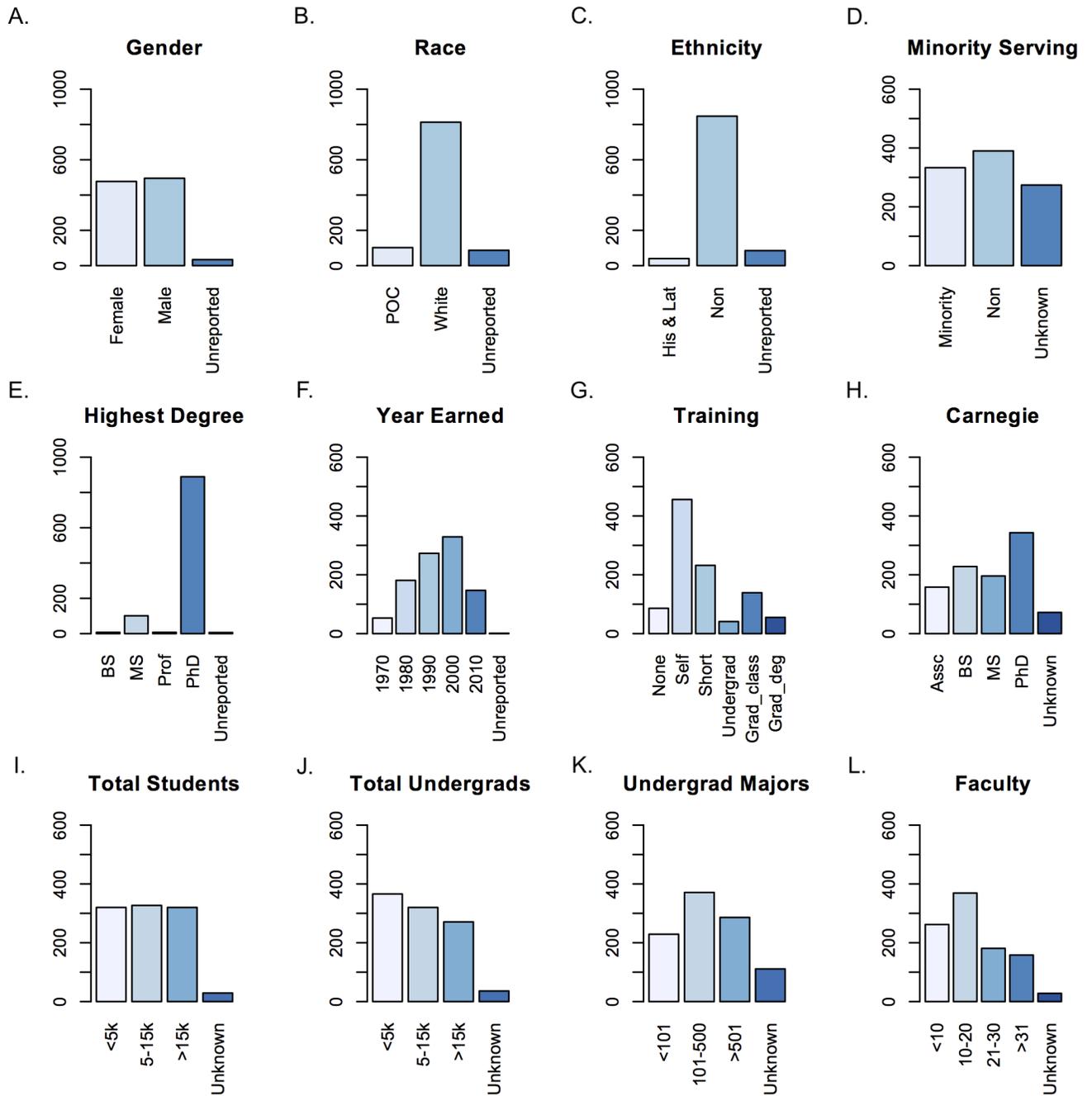
Fig 2. Demographics of survey respondents. The number of responses (y-axes) for each of the demographic variables (x-axes) on the survey, as follows: (A) Gender. (B) Race (People of Color and White); four categories in Race—American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or other Pacific Islander—were combined into People of Color (POC) due to very small sample numbers for each category. (C) Ethnicity (Hispanic-Latino, non-Hispanic/Latino). (D) Minority Serving (whether or not the respondent's home institution is classified as minority-serving). (E) Highest Degree (highest degree earned: Bachelor's, Master's, Professional Degree, PhD). (F) Year Earned (year that the highest degree was earned; responses were grouped in the following bins: Before 1980, 1980 to 1989, 1990 to 1999, 2000 to 2009, and After 2009). (G) Training (level of bioinformatics training: None, Self-taught, Short workshop, Undergraduate/PostBacc training, Graduate class, and Graduate degree); four categories in Training—Undergraduate course, Undergraduate certificate, Undergraduate degree, and Post-baccalaureate certificate—were grouped together into "Undergrad" (undergraduate/post-baccalaureate training) due to small sample numbers in these categories. (H) Carnegie (Carnegie classification of the respondent's home institution: Associate's, Baccalaureate, Master's, Doctoral). (I) Total Students (total number of students at the respondent's home institution). (J) Total Undergraduates (number of undergraduates at the respondent's home institution). (K) Undergraduate Majors (number of undergraduate majors in the respondent's home department). (L) Faculty (number of faculty in the respondent's home department).
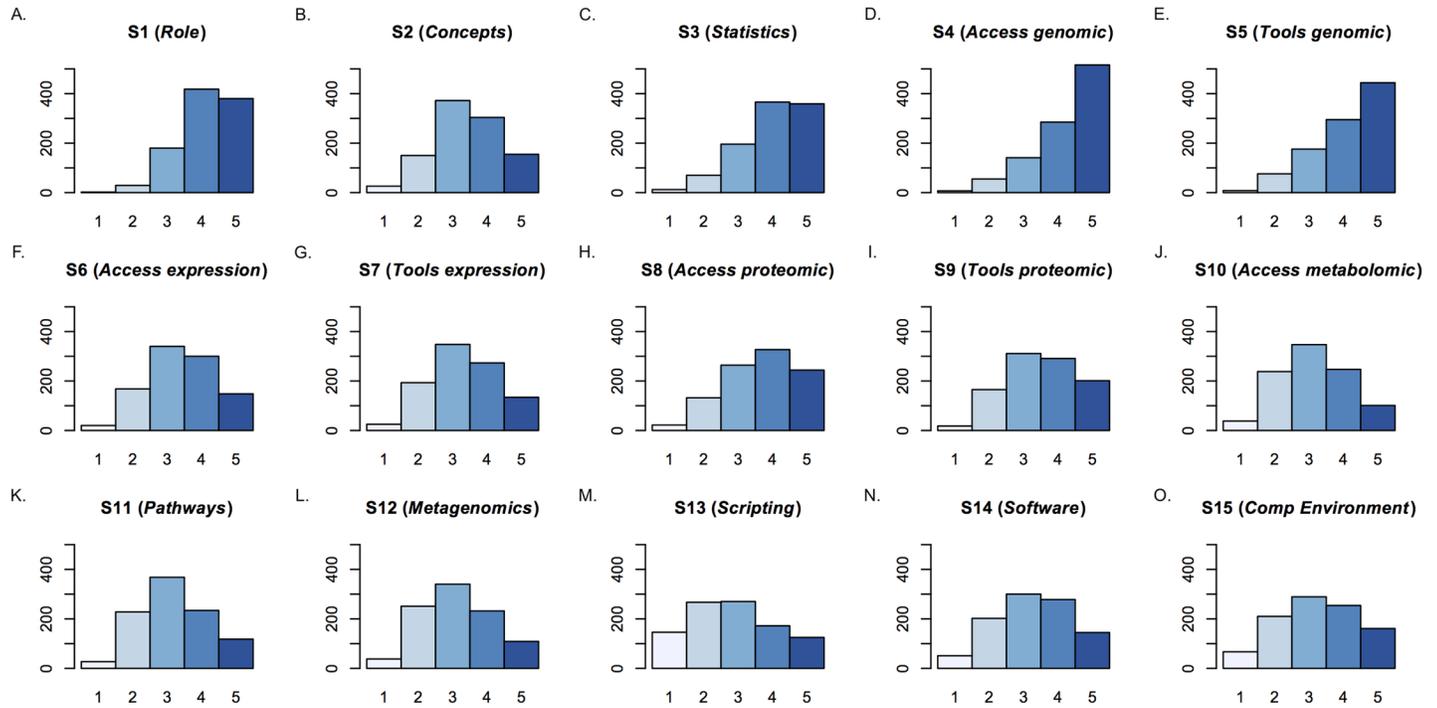
https://doi.org/10.1371/journal.pone.0196878.g002

Fig 3. Summary of bioinformatics skills ratings. The total number of responses (y-axes) by Likert-scale rating from 1 to 5 (x-axes)—1 being "Not at all important" to 5 being "Extremely important"—for each of the fifteen survey skills, S1 to S15, labeled in sequence from (A) to (O). As discussed in Results, these skills were divided into two broad categories: skills that just required familiarity ("knowing" skills: S1 to S4, S6, S8, S10), and those that required direct engagement ("practicing" skills: S5, S7, S9, S11 to S15).

https://doi.org/10.1371/journal.pone.0196878.g003

Table 1. Responses by degree.

Skill	MS	PhD	P value
S1 (*Role*)	3.90	4.16	0.188
S2 (*Concepts*)	3.12	3.44	0.180
S3 (*Statistics*)	3.77	4.02	0.286
S4 (*Access genomic*)	3.74	4.32	**1.13×10^{-4}**
S5 (*Tools genomic*)	3.51	4.17	**3.00×10^{-6}**
S6 (*Access expression*)	3.12	3.43	0.130
S7 (*Tools expression*)	2.96	3.34	0.0549
S8 (*Access proteomic*)	3.19	3.70	**0.0126**
S9 (*Tools proteomic*)	3.09	3.54	0.0789
S10 (*Access metabolomic*)	2.90	3.16	0.0863
S11 (*Pathways*)	2.88	3.22	0.0677
S12 (*Metagenomics*)	2.80	3.16	**0.0329**
S13 (*Scripting*)	2.44	2.89	**3.94×10^{-4}**
S14 (*Software*)	2.82	3.31	**8.10×10^{-3}**
S15 (*Computational environment*)	2.88	3.26	**0.0122**

Means of the Likert-scale responses for each survey skill for respondents with a Master's degree (MS) and those with a PhD. Two-sided *P* value from a Kolmogorov-Smirnov test of the Likert-scale responses is shown. In all cases, those with a PhD rated desired skills higher than those with an MS; significant differences are in bold.
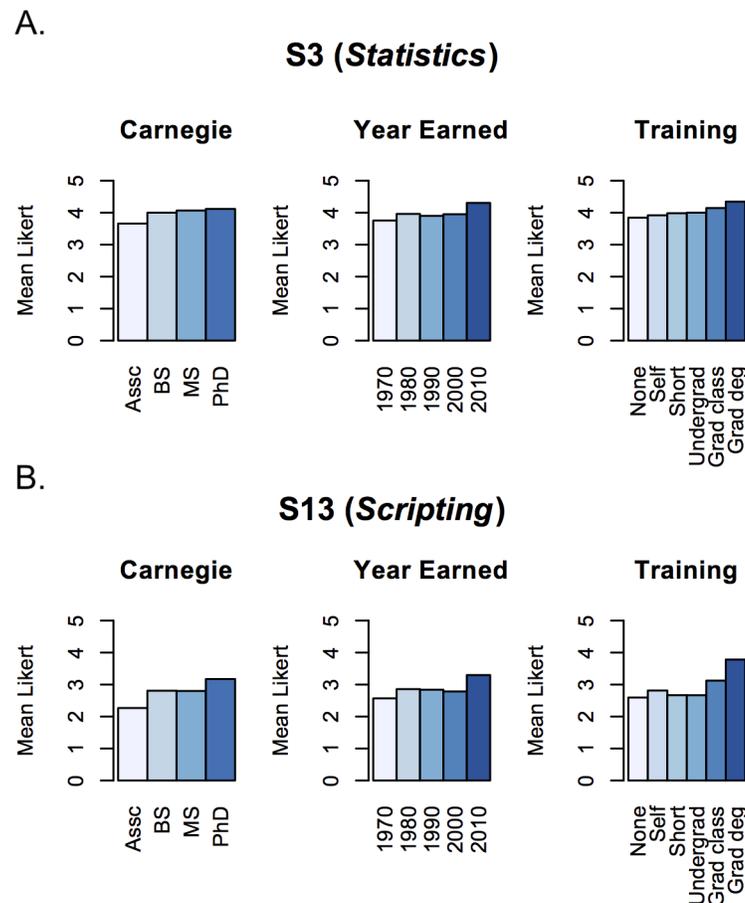
https://doi.org/10.1371/journal.pone.0196878.t001

A.

S3 (*Statistics*)

B.

S13 (*Scripting*)

Fig 4. Mean Likert responses for S3 (*Statistics*) and S13 (*Scripting*). Mean Likert responses are shown for (A) S3 (*Statistics*) and (B) S13 (*Scripting*) for three categories: Carnegie (Carnegie Classification of the respondent's home institution: Associate's, Baccalaureate, Master's, Doctoral), Year Earned (year that the highest degree was earned; responses were grouped in the following bins: Before 1980, 1980 to 1989, 1990 to 1999, 2000 to 2009, and After 2009), and Training (level of bioinformatics training: None, Self-taught, Short workshop, Undergraduate/PostBacc training, Graduate class, and Graduate degree). Means and *P* values from pairwise KS tests are reported in [34].

institutions with other classifications. In general, the rating of S13 (*Scripting*) increased with Carnegie Classification (Fig 4 and S1 Table).

Higher ratings of categories by year and level of bioinformatics training. The ratings of most of the fifteen survey skills did not depend significantly on the year in which the respondents earned their degree. However, skills S3 (*Statistics*) and S13 (*Scripting*) were rated significantly higher by respondents who earned their degree after 2009 (i.e., in 2010 or after) than by those who earned their degree prior to 2010 (Fig 4 and [34]). Furthermore, the majority of respondents had a PhD, and those with a PhD consistently rated every skill higher than those with a master's degree, often significantly higher (Table 1). Those with no training in bioinformatics routinely rated every skill significantly lower than respondents with training, formal or not (Fig 4 and [34]). Notably, S13 (*Scripting*) was rated significantly higher by respondents with a graduate degree in bioinformatics than by those with any other type of training (Fig 4 and [34]).

Coverage of skills varies across syllabi

To analyze the syllabi submitted by survey respondents (see Materials and Methods), the fifteen survey skills were divided into two groups—those that required students to be familiar
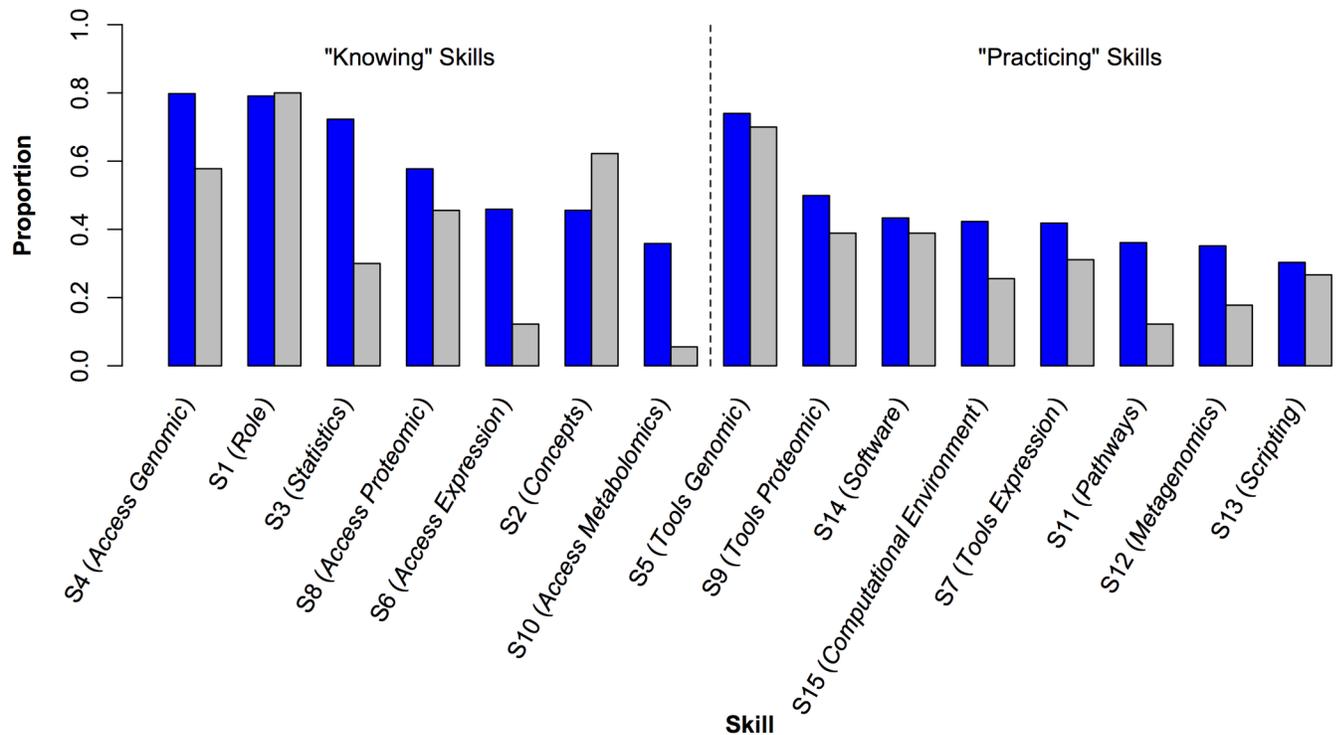
Fig 5. Importance of the fifteen bioinformatics skills as rated by survey respondents compared to coverage of the skills in the syllabi. Skills are shown with the proportion of survey responses rating the skill as either "Very Important" or "Extremely Important" (blue bars) and the proportion of submitted syllabi that exhibited evidence of the skill (grey bars). Skills requiring familiarity with a concept ("knowing" skills) are to the left of the vertical dashed line; skills requiring direct engagement ("practice" skills) are to the right. In their respective categories, skills are presented in order of decreasing proportion of survey responses rating the skill as Very or Extremely Important.

with a concept (i.e., to "know" about it; survey skills S1 to S4, S6, S8, and S10), and those that required direct engagement (i.e., to be able to use the skill in "practice"; skills S5, S7, S9, and S11 to S15). More of the "knowing" skills were rated as being either "extremely important" or "very important" than the "practice" skills and were more frequently covered in the syllabi; conversely, the "practice" skills were less likely to be covered (Fig 5). There were two exceptions to these trends. Few survey respondents thought students should be expected to be familiar with metabolomic and systems biology data (S10, a "knowing" skill), nor was it frequently covered in the submitted syllabi. (As described below, this skill was dropped from the Core Competencies.) On the other hand, a majority of respondents indicated that undergraduate life scientists should "be able to use bioinformatics tools" (S5). In addition, evidence of this was present in approximately 70% of the submitted syllabi (Fig 5, Fig 6).

Syllabi also varied substantially in the number of survey skills covered. On average, a syllabus covered 5.5 skills, with a median of six skills addressed. Although the submitted syllabi, in aggregate, covered all the survey skills, no single syllabus covered more than thirteen out of the fifteen skills, suggesting the difficulty of covering all the skills in one class (Fig 6).

Core competencies

The nine NIBLSE Core Competencies, C1 to C9, developed using the iterative process described in Materials and Methods, are given in Table 2. During the development process, several survey skills that received lower ratings on the survey (S10, S11, S12) were dropped or combined into a more general competency such as C4 ("Use bioinformatics tools to examine
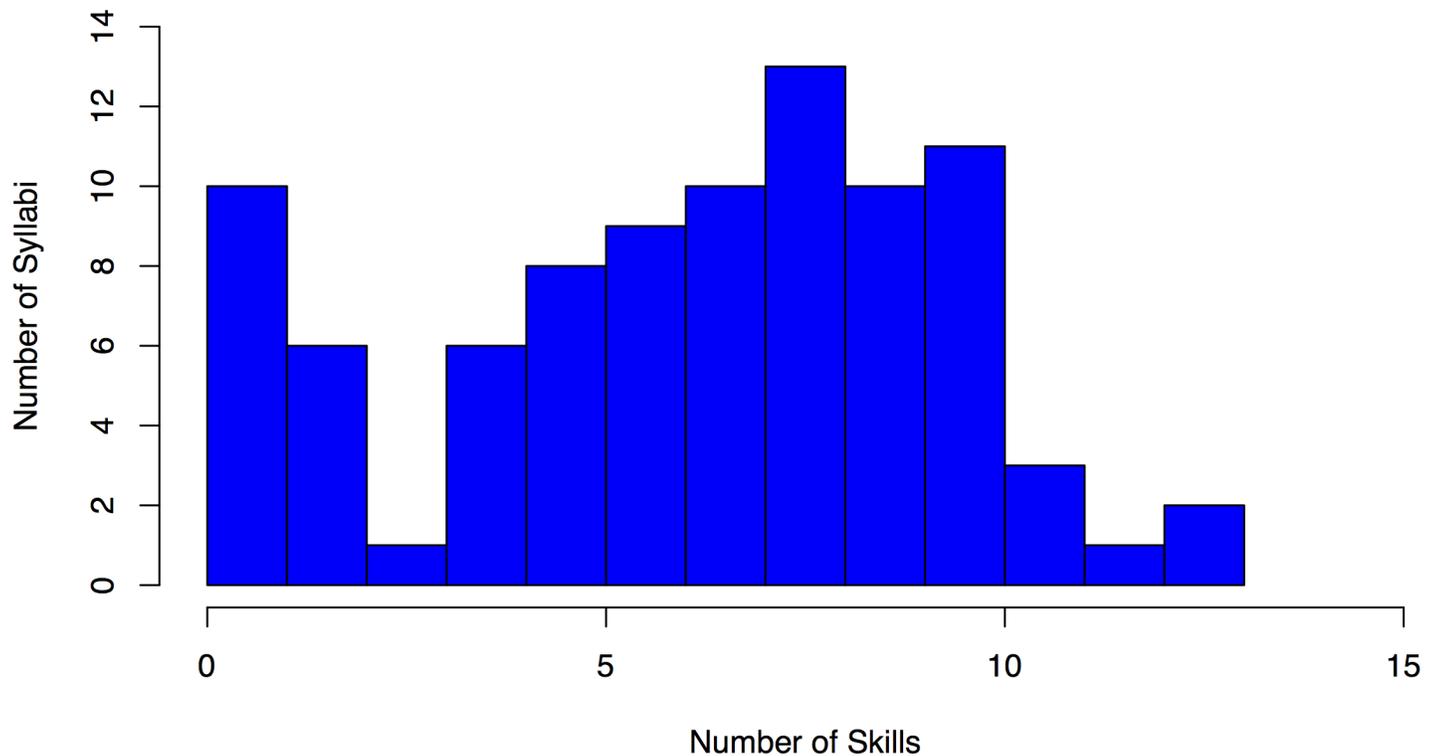
Fig 6. Histogram of number of skills covered per syllabus. The number of syllabi that addressed the specified number of survey skills (S1 to S15). For example, ten syllabi addressed only one of the fifteen skills. On average, a syllabus covered 5.5 skills, with a median of 6 skills addressed. In aggregate, the submitted syllabi covered all fifteen skills, but no single syllabus covered more than thirteen out of the fifteen skills.

https://doi.org/10.1371/journal.pone.0196878.g006

complex biological problems") and C6 ("Explore and/or model biological interactions, networks, and data integration using bioinformatics"). While scripting and use of the command line (S13) received relatively lower scores from survey respondents, the strong support for these emerging skills at Doctoral institutions and from the most recently trained respondents led to their inclusion in the final list. As mentioned above, the NIBLSE Core Competencies are intended to serve as a guide for institutions as they integrate bioinformatics into their own life sciences curricula.

Discussion

Faculty from a wide range of institutions showed strong support for greater integration of bioinformatics into undergraduate life sciences curricula: 95% of the 1,260 respondents agreed with the statement "I think bioinformatics should be integrated into undergraduate life sciences education." However, there are some differences among faculty perspectives at different types of institutions. While most respondents value the ability to retrieve information from public databases and use existing software tools to analyze data, respondents at Doctoral institutions place a higher priority on computational skills, such as being able to operate in multiple computational environments and being able to write short programs. A possible explanation for this difference is that respondents at research-based institutions are more directly exposed to the necessity of using these computational skills on a day-to-day basis, whether by themselves, their students, or their colleagues. These findings provide insight into different educational perspectives and the barriers institutions may face as they integrate bioinformatics into their own life sciences curricula.

Table 2. The NIBLSE bioinformatics core competencies for undergraduate life scientists.

C1. **Explain the role of computation and data mining in addressing hypothesis-driven and hypothesis-generating questions within the life sciences.** Life sciences students should have a clear understanding of the role computing and data mining play in modern biology. Given a traditional hypothesis-driven research question, students should have ideas about what types of data and software exist that could help them answer the question quickly and efficiently. They should also appreciate that mining large datasets can generate novel hypotheses to be tested in the lab or field.
• Compare and contrast computer-based research with wet-lab research.
• Explain the role of computation in finding genes, detecting the function of protein domains, and inferring protein function.
• Describe the role of various databases in identifying potential gene targets for drug development.

C2. **Summarize key computational concepts, such as algorithms and relational databases, and their applications in the life sciences.** To make use of sophisticated software and database tools, students should have a basic understanding of the principles upon which these tools are based and should be exposed to how these tools work.
• Explain the underlying algorithm(s) employed in sequence alignment (e.g., BLAST).
• Modify software parameters to achieve biologically meaningful results.
• Explain how data are organized in relational databases (e.g., NCBI and model organisms databases).

C3. **Apply statistical concepts used in bioinformatics.** In addition to the basic statistics found in many biology curricula, modern life scientists should have an understanding of the statistics of large datasets and multiple comparisons.
• Understand that there is a probability of finding a given sequence similarity score by chance (the *P* value) and that the size of the target database affects the probability that you will see a particular score in a particular search (the E-value).
• Explain the statistical modeling used to identify differentially expressed genes.
• Interpret data from a well-designed drug trial.

C4. **Use bioinformatics tools to examine complex biological problems in evolution, information flow, and other important areas of biology.** This competency is written broadly so as to encompass a variety of problems that can be addressed using bioinformatics tools, such as understanding the evolutionary underpinnings of sequence comparison and homology detection; distinguishing between genomic sequences, RNA sequences, and protein sequences; and interpreting phylogenetic trees. "Complex" biological problems require that students should be able to work through a problem with multiple steps, not just perform isolated tasks.
• Using multiple lines of evidence, annotate a gene.
• Develop and interpret a "tree of life" based on a BLAST search, multiple alignment, and phylogenetic-tree building.
• Explain how a mutation in a gene causes cancer, using a genome browser to identify the gene, transcript, and affected protein, and tools such as OMIM, GO, and KEGG to place it in the context of a function and pathway important to the disease

C5. **Find, retrieve, and organize various types of biological data.** Given the numerous and varied datasets currently being generated from all of the 'omics fields, students should develop the facility to identify appropriate data repositories, navigate and retrieve data from these repositories, and organize data relevant to their area of study in flat files or small local stand-alone databases.
• Navigate and retrieve data from a genome browser.
• Retrieve data from protein and genome databases (e.g., PDB, UniProt, NCBI).
• Store and interrogate small datasets using spreadsheets or delimited text files.

C6. **Explore and/or model biological interactions, networks, and data integration using bioinformatics.** Modeling of biological systems at all levels, from cellular to ecological, is being facilitated by technological and algorithmic advances. These models provide novel insights into the perturbations in systems that can cause disease, interactions of microbes with various eukaryotic systems, how metabolic networks respond to environmental stresses, etc. Students should be familiar with the techniques used to generate these analyses and should be able to interpret the outputs and use the data to generate novel hypotheses.
• Predict impact of a gene knockout on a cell-signaling pathway.
• Analyze gene expression data to build an expression network.
• Analyze metagenomics data from microbial samples obtained from environmental sources.

C7. **Use command-line bioinformatics tools and write simple computer scripts.** Most biological datasets (e.g., genomic and proteomic sequences, BLAST results, RNASeq and resulting differential expression data) are available as text files; the most powerful and dynamic way to interact with these datasets is through the command line or shell scripting. Students should be able to manipulate their own data and to create and modify complex data processing and analysis workflows.
• Run BLAST using command line options.
• Build and run statistical analyses using environments such as R or Python scripts.
• Write simple shell scripts to manipulate files.

(Continued)

Table 2. (Continued)

C8.	**Describe and manage biological data types, structure, and reproducibility.** This competency addresses two distinct concerns: 1) each of the varied 'omics fields produces data in formats particular to its needs, and these formats evolve with changes in technologies and refinements in downstream software; and 2) all experimental data is subject to error and the user must be cognizant of the need to verify the reproducibility of their data. Students need to develop an awareness of, and ability to, manipulate different data types given the versioning of formats. They also need to exercise caution, to carry out appropriate statistical analyses on their data as part of normal operating procedures and report the uncertainty of their results, and to provide the relevant information to enable reproduction of their results. • Describe the various sequence formats used to store DNA and protein sequences (e.g., FASTA, FASTQ). • Understand the representation of gene features using Gene Feature Format (GFF) files. • Compare reproducibility of biological and technical replicate data (e.g., transcriptomic data) using statistical tests (Spearman rank test and false discovery calculations).
C9.	**Interpret the ethical, legal, medical, and social implications of biological data.** The increasing scale and penetration of human genetic and genomic data has greatly enhanced our ability to identify disease-related loci, druggable targets, etc. and to identify potential genes for replacement with developing techniques. However, with this information also comes many ethical, legal, and social questions; suggested resolutions are often outpaced by the technological advances. As part of their scientific training, students should debate the medicinal, societal, and ethical implications of these information sets and techniques. • Explain the implications, good and bad, of being able to walk into a doctor's office and have your genome sequenced and analyzed or of being able to obtain information from direct-to-consumer testing services. • Be able to discuss different perspectives about who should have access to this data and how it should be protected. • Describe how the scientific community protects against the falsification or manipulation of large datasets.

The NIBLSE Bioinformatics Core Competencies. Following each compentency is a short description and a list of three representative examples that illustrate it.

https://doi.org/10.1371/journal.pone.0196878.t002

It is important to keep in mind the distinctions between the educational needs of bioinformaticians and life scientists as well as the differences between the goals of undergraduate education and those of graduate or professional education. Up to this point, much of the discussion of bioinformatics education in the literature has focused on the education of bioinformaticians or on graduate or professional development. However, some authors have addressed the question of bioinformatics education for undergraduate life sciences students. In particular, attendees of the first and second Workshop on Education in Bioinformatics and Computational Biology in 2008 (Taipei, Taiwan) and 2009 (Singapore), held as part of the International Conference in Bioinformatics, attempted to identify a minimum skill set for the training of bioinformaticians and life scientists with informatics capabilities [8]. A consensus list of five essential bioinformatics skills was reported. This skill set overlaps considerably with the NIBLSE Core Competencies, indicating that there is agreement about the skills that are necessary for modern life sciences students. However, the sample size of this effort was small (n = 56, including students) and the authors were attempting to find a consensus that could be applied internationally in widely differing contexts.

The NIBLSE Core Competencies include a computational competency, consistent with other reports [7, 15, 18, 19]. Although many life sciences programs are not currently equipped to provide training in basic programming and operating in a command-line environment, these are important skills that enable students to manipulate and analyze modern biological data. (As an added benefit, creating classes to teach these skills provides an opportunity for interdepartmental course development.) This does not mean that we recommend that life sciences students be able to write complex software applications or be trained to develop graphical user interfaces, but being able to write short programs and run command-line programs gives them flexibility in analyzing data and, perhaps more importantly, provides them with a better understanding of the data itself. The pedagogical literature from a variety of fields is clear that students learn more when they engage with data more deeply, as opposed to entering

data into a "black box" and reporting the results [16, 40–44]. Thus, as with any laboratory technique—e.g., PCR, dissection, or microscopy—bioinformatically-literate undergraduates don't need to be experts but should be expected to have basic skills in these areas when they graduate.

In addition to the survey data, the analysis of syllabi submitted by survey respondents shows that a variety of bioinformatics topics are already being covered in life science curricula nationwide. The fact that bioinformatics can be performed relatively inexpensively with freely available data and software makes it an attractive way for students to engage in research experiences and inquiry-based learning [3, 40, 41]. We would argue that this training would ideally occur in an integrated manner throughout a life sciences curriculum, as opposed to being isolated in a single course. Thus, there is no need to remove particular topics and replace them with "bioinformatics" units. Instead, we encourage faculty to find ways to incorporate bioinformatics techniques and applications as a way of exploring existing concepts in their curricula.

Among the syllabi analyzed, there appears to be a greater emphasis on "knowing" rather than "doing." This gap may reflect the degree of instructor training: if an instructor has little training or experience with bioinformatics, it is understandably easier to introduce a technique or concept in a lecture than it is to develop and implement an in-depth exercise. This idea is supported by the fact that survey respondents frequently commented on the lack of available teaching resources in this area. One effort to address this problem was the inclusion of bioinformatics as a course on *CourseSource*, which publishes evidence-based learning resources [33]. The issue of training was also raised repeatedly in responses to survey questions regarding real and perceived barriers to the integration of bioinformatics into life sciences education, with many respondents indicating that they wanted to integrate more bioinformatics into their courses but felt they lacked the necessary training to do so. This training deficit is a long-term problem that will be difficult to address. On the one hand, a variety of resources are currently available for faculty to receive training in bioinformatics or to educate themselves to address the training need. On the other, many faculty in the survey indicated they needed more training. Thus, it's not clear if the lack of training is due to faculty not being aware of the training resources available or if the existing resources are not useful to them because they cost too much, require too much time, are too advanced, or require too much work to adapt to their specific courses. Barriers to integrating bioinformatics into life sciences education, and our suggestions for overcoming them, are the topics of a separate paper [32]. Note that NIBLSE was formed in large part to provide resources to faculty to help overcome these kinds of barriers.

To conclude, the analysis presented here provides evidence of strong, widespread agreement that undergraduate life sciences students need to be trained in bioinformatics as well as considerable agreement about which skills are necessary. Furthermore, it resulted in development of the NIBLSE Core Competencies, which provide a framework of topics for this training. Although the results presented here could potentially be skewed—those interested in integrating bioinformatics into biology education could have been more motivated to respond to the survey—given the large number of responses [45] and the small margin of error in them (see Materials and Methods), the effects of such a skew, if present, would be minor. The minimal effects of a potential skew are further supported by the close alignment of our results and those in foundational works cited earlier [cf. 8, 9, 27, 28]. As such, we contend that our results accurately represent the opinions of the U.S. life sciences education community as a whole.

Supporting information

S1 Survey. The NIBLSE survey with branches. As explained in the narrative, the survey was branched, with some questions or sections presented or skipped depending on the responses

to previous, filtering questions. In this view, the branching structure of the survey can be followed.
(PDF)

S2 Survey. The NIBLSE survey without branches. In this view, only the survey questions are given—the branching structure of the survey is not evident.
(PDF)

S1 Table. Responses by Carnegie classification. Means of the Likert-scale responses for respondents whose home institution is classified as Associate's (Assc), Baccalaureate (BS), Master's (MS), or Doctoral (PhD). Two-sided P values from a Kolmogorov-Smirnov test of the Likert-scale responses for each pairwise test are shown. The pairs are indicated by the heading of the column; e.g., P_Assc_BS is the P value for the Associate's (Assc)/Baccalaureate (BS) pair. Significant differences are in bold.
(DOCX)

Acknowledgments

Author Contributions

Conceptualization: Melissa A. Wilson Sayres, Charles Hauser, Michael Sierk, Anne G. Rosenwald, Eric W. Triplett, Jason J. Williams, Elizabeth Dinsdale, William R. Morgan, Lonnie R. Welch, Mark A. Pauley.

Data curation: Melissa A. Wilson Sayres, Charles Hauser, Michael Sierk, Sebastian Galindo-Gonzalez, Nealy F. Grandgenett, Kimberly C. Olney.

Formal analysis: Melissa A. Wilson Sayres, Charles Hauser, Michael Sierk, Anne G. Rosenwald, Eric W. Triplett, Elizabeth Dinsdale, William R. Morgan, Lindsay Barone, Mark A. Pauley.

Funding acquisition: Anne G. Rosenwald, Eric W. Triplett, Elizabeth Dinsdale, William R. Morgan, Mark A. Pauley.

Investigation: Melissa A. Wilson Sayres, Charles Hauser, Michael Sierk, Srebrenka Robic, Anne G. Rosenwald, Todd M. Smith, Eric W. Triplett, Jason J. Williams, Elizabeth Dinsdale, William R. Morgan, James M. Burnette, III, Samuel S. Donovan, Jennifer C. Drew, Sarah C. R. Elgin, Edison R. Fowlks, Sebastian Galindo-Gonzalez, Anya L. Goodman, Nealy F. Grandgenett, Carlos C. Goller, John R. Jungck, Jeffrey D. Newman, William Pearson, Elizabeth F. Ryder, Rafael Tosado-Acevedo, William Tapprich, Tammy C. Tobin, Arlín

Toro-Martínez, Lonnie R. Welch, Robin Wright, David Ebenbach, Mindy McWilliams, Mark A. Pauley.

Methodology: Melissa A. Wilson Sayres, Charles Hauser, Michael Sierk, Anne G. Rosenwald, Eric W. Triplett, Jason J. Williams, Elizabeth Dinsdale, William R. Morgan, Sebastian Galindo-Gonzalez, Nealy F. Grandgenett, David Ebenbach, Mindy McWilliams, Mark A. Pauley.

Project administration: Mark A. Pauley.

Software: David Ebenbach, Mindy McWilliams.

Visualization: Melissa A. Wilson Sayres, Charles Hauser, Michael Sierk.

Writing – original draft: Melissa A. Wilson Sayres, Charles Hauser, Michael Sierk, Mark A. Pauley.

Writing – review & editing: Melissa A. Wilson Sayres, Charles Hauser, Michael Sierk, Srebrenka Robic, Anne G. Rosenwald, Todd M. Smith, Eric W. Triplett, Jason J. Williams, Elizabeth Dinsdale, William R. Morgan, James M. Burnette, III, Samuel S. Donovan, Jennifer C. Drew, Sarah C. R. Elgin, Edison R. Fowlks, Sebastian Galindo-Gonzalez, Anya L. Goodman, Nealy F. Grandgenett, Carlos C. Goller, John R. Jungck, Jeffrey D. Newman, William Pearson, Elizabeth F. Ryder, Rafael Tosado-Acevedo, William Tapprich, Tammy C. Tobin, Arlín Toro-Martínez, Lonnie R. Welch, Robin Wright, Lindsay Barone, David Ebenbach, Mindy McWilliams, Kimberly C. Olney, Mark A. Pauley.

References

1. National Research Council. BIO2010: transforming undergraduate education for future research biologists. Washington: National Academies Press; 2003. Available from: https://doi.org/10.17226/10497

2. Bialek W, Botstein D. Introductory science and mathematics education for 21st-century biologists. Science. 2004; 303:788–90. https://doi.org/10.1126/science.1095480 PMID: 14764865

3. Lopatto D, Alvarez C, Barnard D, Chandrasekaran C, Chung H-M, Du C, et al. Undergraduate research: Genomics Education Partnership. Science. 2008; 322:684–85. https://doi.org/10.1126/science.1165351 PMID: 18974335

4. Campbell AM, Ledbetter MLS, Hoopes LLM, Eckdahl TT, Heyer LJ, Rosenwald A, et al. Genome Consortium for Active Teaching: meeting the goals of BIO2010. CBE Life Sci Educ. 2007; 6:109–18. https://doi.org/10.1187/cbe.06-10-0196 PMID: 17548873

5. Drew JC, Triplett EW. Whole genome sequencing in the undergraduate classroom: outcomes and lessons from a pilot course. J Microbiol Biol Educ. 2008; 9:3–11. PMID: 23653818

6. Honts JE. Evolving strategies for the incorporation of bioinformatics within the undergraduate cell biology curriculum. Cell Biol Educ. 2003; 2:233–47. https://doi.org/10.1187/cbe.03-06-0026 PMID: 14673489

7. Pevzner P, Shamir R. Computing has changed biology—biology education must catch up. Science. 2009; 325:541–42. https://doi.org/10.1126/science.1173876 PMID: 19644094

8. Tan TW, Lim SJ, Khan AM, Ranganathan S. A proposed minimum skill set for university graduates to meet the informatics needs and challenges of the "-omics" era. BMC Genomics. 2009; 10(Suppl 3):S36.

9. Welch L, Lewitter F, Schwartz R, Brooksbank C, Radivojac P, Gaeta B, et al. Bioinformatics curriculum guidelines: toward a definition of core competencies. PLoS Comput Biol. 2014; 10(3):e1003496. https://doi.org/10.1371/journal.pcbi.1003496 PMID: 24603430

10. Rosenwald AG, Russell JS, Arora G. The genome solver website: a virtual space fostering high impact practices for undergraduate biology. J Microbiol Biol Educ. 2013; 13:188–90. https://doi.org/10.1128/jmbe.v13i2.444 PMID: 23653812

11. Steen LA. Math and Bio 2010: linking undergraduate disciplines. Washington: Mathematical Association of America; 2005.

12. Association of American Medical College and the Howard Hughes Medical Institute. Scientific foundations for future physicians. Washington: Association of American Medical Colleges; 2009 [cited 2017 Nov 1]. Available from: http://www.hhmi.org/sites/default/files/Programs/aamc-hhmi-2009-report.pdf

13. American Association for the Advancement of Science. Vision and change in undergraduate biology education: a call to action. Washington: American Association for the Advancement of Science; 2011. [cited 2017 Nov 1] Available from: http://visionandchange.org/files/2013/11/aaas-VISchange-web1113.pdf

14. American Association for the Advancement of Science. Vision and change in undergraduate biology education: chronicling change, inspiring the future. Washington: American Association for the Advancement of Science; 2015 [cited 2017 Nov 1]. Available from: http://visionandchange.org/files/2015/07/VISchange2015_webFin.pdf

15. Atwood TK, Bongcam-Rudloff E, Brazas ME, Corpas M, Gaudet P, Lewitter F, et al. GOBLET: the Global Organisation for Bioinformatics Learning, Education and Training. PLoS Comput Biol. 2015; 11 (4):e1004143. https://doi.org/10.1371/journal.pcbi.1004143 PMID: 25856076

16. Jungck JR, Donovan SS, Weisstein AE. Bioinformatics education dissemination with an evolutionary problem solving perspective. Brief Bioinform. 2010; 11:570–81. https://doi.org/10.1093/bib/bbq028 PMID: 21036947

17. Baker M. Scientific computing: code alert. Nature. 2017; 541:563–65.

18. Barone L, Williams J, Micklos D. Unmet needs for analyzing biological big data: A survey of 704 NSF principal investigators. PLoS Comput Biol. 2017; 13(10):e1005755. https://doi.org/10.1371/journal.pcbi.1005755 PMID: 29049281

19. Teal TK, Cranston KA, Lapp H, White E, Wilson G, Ram K, et al. Data Carpentry: workshops to increase data literacy for researchers. IJDC. 2015; 10:135–43.

20. Crowe M, Cameron G. EMBL Australia Bioinformatics Resource (EMBL-ABR) community survey report 2013 [Internet]. Melbourne: EMBL-ABR; 2013 [cited 2017 Nov 1]. Available from: https://www.embl-abr.org.au/news/braembl-community-survey-report-2013/

21. Bureau of Labor Statistics (US). Employment Projections [Internet]. Washington: US Department of Labor. c2015 [cited 2017 Aug 5]. Available from: https://www.bls.gov/emp/

22. Adams JC. US-BLS: Computing employment outlook remains bright. 2016 Apr 30 [cited 2017 Aug 5]. In: Blog@CACM [Internet]. New York: Communications of the ACM. c2016 -. 4p. Available from: http://cacm.acm.org/blogs/blog-cacm/201784-us-bls-computing-employment-outlook-remains-bright/

23. Jordan TC, Burnett SH, Carson S, Caruso SM, Clase K, DeJong RJ, et al. A broadly implementable research course in phage discovery and genomics for first-year undergraduate students. MBio. 2014; 5 (1):e01051–13. https://doi.org/10.1128/mBio.01051-13 PMID: 24496795

24. Shaffer CD, Alvarez CJ, Bednarski AE, Dunbar D, Goodman AL, Reinke C, et al. A course-based research experience: how benefits change with increased investment in instructional time. CBE Life Sci Educ. 2014; 13:111–30. https://doi.org/10.1187/cbe-13-08-0152 PMID: 24591510

25. Buonaccorsi V, Peterson M, Lamendella G, Newman J, Trun N, Tobin T, et al. Vision and change through the Genome Consortium for Active Teaching using Next-Generation Sequencing (GCAT-SEEK). CBE Life Sci Educ. 2014; 13:1–2. https://doi.org/10.1187/cbe.13-10-0195 PMID: 24591495

26. Snyder TD, de Brey C, Dillow SA. Digest of Education Statistics 2015. Washington: National Center for Education Statistics, Institute of Education Sciences, US Department of Education; 2016. Table 322.10; p. 634. Report No.: NCES 2016–014. Available from: https://nces.ed.gov/pubs2016/2016014.pdf

27. Welch L, Brooksbank C, Schwartz R, Morgan SL, Gaeta B, Kilpatrick AM, et al. Applying, evaluating and refining bioinformatics core competencies (an update from the curriculum task force of ISCB's education committee). PLoS Comput Biol. 2016; 12(5):e1004943. https://doi.org/10.1371/journal.pcbi.1004943 PMID: 27175996

28. Mulder NJ, Adebiyi E, Alami R, Benkahla A, Brandful J, Doumbia S, et al. H3ABioNet, a sustainable pan-African bioinformatics network for human heredity and health in Africa. Genome Res. 2016; 26 (2):271–77. https://doi.org/10.1101/gr.196295.115 PMID: 26627985

29. Koch I, Fuellen G. A review of bioinformatics education in Germany. Brief Bioinform. 2008; 9:232–42. https://doi.org/10.1093/bib/bbn006 PMID: 18310676

30. Maloney M, Parker J, LeBlanc M, Woodard CT, Glackin M, Hanrahan M. Bioinformatics and the undergraduate curriculum. CBE Life Sci Educ. 2010; 9:172–74. https://doi.org/10.1187/cbe.10-03-0038 PMID: 20810947

31. Dinsdale E, Elgin SCR, Grandgenett N, Morgan W, Rosenwald A, Tapprich W, et al. NIBLSE: a Network for Integrating Bioinformatics into Life Sciences Education. CBE Life Sci Educ. 2015; 14(4):le3. https://doi.org/10.1187/cbe.15-06-0123 PMID: 26466989

32. Williams JJ, Drew JC, Galindo-Gonzalez S, Robic S, Dinsdale E, Morgan W. Barriers to Integration of Bioinformatics into Undergraduate Life Sciences Education; 2017. Preprint. Available from bioRxiv https://doi.org/10.1101/204420

33. Rosenwald AG, Pauley MA, Welch L, Elgin SCR, Wright R, Blum J. The CourseSource bioinformatics learning framework. CBE Life Sci Educ. 2016; 15(1):le2. https://doi.org/10.1187/cbe.15-10-0217 PMID: 27290739

34. Network for Integrating Bioinformatics into Life Sciences Education (NIBLSE). Core Competencies Working Group analysis. GitHub repository [Internet]. c2017 –[cited 2017 Nov1]. Available from: https://github.com/niblse/

35. Snyder TD, de Brey C, Dillow SA. Digest of Education Statistics 2015. Washington: National Center for Education Statistics, Institute of Education Sciences, US Department of Education; 2016. Table 315.80; p. 583. Report No.: NCES 2016–014. Available from: https://nces.ed.gov/pubs2016/2016014.pdf

36. Bureau of Labor Statistics (US). Occupational Outlook Handbook, Postsecondary Teachers Summary [Internet]. Washington: US Department of Labor. c2016 - [cited 2017 Aug 5]. Available from: https://www.bls.gov/ooh/education-training-and-library/postsecondary-teachers.htm

37. Science and Engineering Indicators 2014 [Internet]. Appendix table 5–14: SEH doctorate holders employed in academia, by type of position and degree field: 1973–2010. Arlington (VA): National Science Foundation; 2014 [cited 2017 August 5]. Report No.: National Science Board 14–01. Available from: https://www.nsf.gov/statistics/seind14/content/chapter-5/at05-14.pdf

38. Bransford J, Brown A, Cocking R. How people learn: brain, mind, experience, and school. Washington: National Academies Press; 2000.

39. Eberlein T, Kampmeier J, Minderhout V, Moog RS, Platt T, Varma-Nelson P, et al. Pedagogies of engagement in science: a comparison of PBL, POGIL, and PLTL. Biochem Mol Biol Educ. 2008; 36:262–73. https://doi.org/10.1002/bmb.20204 PMID: 19381266

40. Saul JM, Redish EF. Final evaluation report for FIPSE Grant# P116P50026: evaluation of the workshop physics dissemination project. College Park (MD): University of Maryland; 1997.

41. Ditty JL, Kvaal CA, Goodner B, Freyermuth SK, Bailey C, Britton RA, et al. Incorporating genomics and bioinformatics across the life sciences curriculum. PLoS Biol. 2010; 8(8):e1000448. https://doi.org/10.1371/journal.pbio.1000448 PMID: 20711478

42. Overton TL, Randles CA. Beyond problem-based learning: using dynamic PBL in chemistry. Chem Educ Res Pract. 2015; 16(2):251–59.

43. Woods TL, Reed S, His S, Woods JA, Woods MR. Pilot study using the augmented reality sandbox to teach topographic maps and surficial processes in introductory geology labs. J Geosci Educ. 2016; 64 (3):199–214.

44. Theo L. Simplifying central place theory using GIS and GPS. J Geogr. 2011; 110(1):16–26.

45. Johnson TP, Wislar JS. Response rates and nonresponse errors in surveys. JAMA. 2012; 307 (17):1805–6. https://doi.org/10.1001/jama.2012.3532 PMID: 22550194