Check for updates

RESEARCH ARTICLE

Design and implementation of semester long project and problem based bioinformatics course [version 1; peer review: 3 approved with reservations]

Geetha Saarunya , Bert Ely

Biological Sciences, University of South Carolina, Columbia, South Carolina, 29208, USA

V1 First published: 25 Sep 2018, 7(ISCB Comm J):1547
<https://doi.org/10.12688/f1000research.16310.1>

Latest published: 25 Sep 2018, 7(ISCB Comm J):1547
<https://doi.org/10.12688/f1000research.16310.1>

Abstract

Background: Advancements in 'high-throughput technologies' have inundated us with data across disciplines. As a result, there is a bottleneck in addressing the demand for analyzing data and training of 'next generation data scientists'.

Methods: In response to this need, the authors designed a single semester "Bioinformatics" course that introduced a small cohort of students at the University of South Carolina to methods for analyzing data generated through different 'omic' platforms using variety of model systems. The course was divided into seven modules with each module ending with a problem.

Results: Towards the end of the course, the students each designed a project that allowed them to pursue their individual interests. These completed projects were presented as talks and posters at ISCB-RSG-SEUSA symposium held at University of South Carolina.

Conclusions: An important outcome of this course design was that the students acquired the basic skills to critically evaluate the reporting and interpretation of data of a problem or a project during the symposium.

Keywords

bioinformatics education, problem-based learning, project-based learning, hands-on course

This article is included in the **Bioinformatics** gateway.

Open Peer Review

Approval Status

	1	2	3
version 1			
25 Sep 2018	view	view	view

1. **Russell Schwartz**, Carnegie Mellon University, Pittsburgh, USA

2. **Mark A. Pauley** , National Science Foundation, Alexandria, USA

3. **Allegra Via** , Sapienza University of Rome, Rome, Italy

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the **Bioinformatics Education and Training Collection** collection.

Corresponding author: Geetha Saarunya (sreeramc@email.sc.edu)

Author roles: **Saarunya G:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Ely B:** Conceptualization, Data Curation, Formal Analysis, Methodology, Project Administration, Resources, Supervision, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2018 Saarunya G and Ely B. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

How to cite this article: Saarunya G and Ely B. **Design and implementation of semester long project and problem based bioinformatics course [version 1; peer review: 3 approved with reservations]** F1000Research 2018, 7(ISCB Comm J):1547 <https://doi.org/10.12688/f1000research.16310.1>

First published: 25 Sep 2018, 7(ISCB Comm J):1547 <https://doi.org/10.12688/f1000research.16310.1>

Introduction

Bioinformatics is a rapidly growing interdisciplinary field because of advances in both computer science and the life sciences. Rapid advances in sequencing technologies have led to a deluge of biological data, creating a need for expeditious, efficient, and effective analyses. Practitioners of bioinformatics now add techniques from statistics, information science and engineering to develop algorithms and build predictive models to understand the dynamics within a biological system. This paradigm shift in how bioinformatics is perceived has resulted in an evolutionary model of growth across both of its root disciplines¹. Bioinformatics as a field also enjoys a degree of duality: “episteme” (scientific knowledge) and “techné” (technical know-how), leading to the idea of ‘Science informing the tools and the tools enabling science’¹. In a 2017 survey of 704 NSF principal investigators, more than 90% of respondents replied that they were soon to be working with data sets that required high-performance computing, and they also identified bioinformatics data analyses to be the most urgent and unmet need required for successful completion of their projects². Increased exposure of students at an undergraduate level will help address the need for specialists working in this field and also make the students attractive for opportunities in industry or in graduate school^{3–5}. The Global Organization for Bioinformatics Learning, Education and Training (GOBLET) identified through surveys that the skills required for ‘basic data stewardship’ are taught only in ~ 25% of education programs creating a gulf between theory and practice^{6–8}.

Many courses have been designed and implemented to address the gaps faced in the field. They are project based, problem based or a combination of both to study one or more ‘next-generation’ datasets^{9–12}. The courses have been designed as workshops⁹ or as semester long courses using analyses from a single next-generation technology¹⁰. The authors haven’t come across a course that incorporates multi-omics data analyses in a single semester. There have been studies that address a single problem using multi-omics approaches¹¹ and there have been pipeline designs that help integrate these data under a single platform¹².

In response to this need, we designed a single semester course on bioinformatics in the Department of Biological Sciences at University of South Carolina that was targeted towards undergraduate seniors and graduate students who were mainly bench scientists working on experiments which generated data across different ‘omic technologies’ using different living systems.

Challenges in design of bioinformatics curriculum

The curriculum task force of the ‘International Society of Computational Biology’, a scholarly society for both bioinformatics and computational biology research scientists across the world, identified a set of 16 core competencies established through surveys and an iterative process of inputs from people associated with the fields of bioinformatics and computational biology¹³.

However, one of the biggest challenges is the heterogeneity of the backgrounds of the course participants. There is ‘no one size fits all’ while designing a bioinformatics course. In fact, there are three different types of user groups that employ bioinformatics in their research (Table 1), and each of these user groups requires different competencies^{14,15}.

Thus, there was considerable diversity in the backgrounds of the students registered for our course. In response, we chose to follow a ‘learner adaptable’ style of design of the curriculum. This approach allowed us to design the course based on the students’ knowledge of the subject and their expectations of the course.

Methods

Course design

Course conception. This course was designed to provide a structured Bioinformatics course that is geared towards the needs of students working on different “omics” experiments. The general premise of the course was to critically examine and analyze published or in-preparation datasets across different biological systems in a hands-on fashion. In addition, we wanted to introduce the students to the R programming language.

Course Participants. We had nine participants registered for the course. Four of the students were undergraduate seniors, four were first or second year graduate students and one of them was an emergency medical technician (EMT) with a Bachelor of Science degree who was taking additional classes for credit and is now in medical school.

Learning objectives and outcomes of the course. We sent a three-question survey (Table 2) to all the participants to understand their reasoning for registering in the course.

The primary learning objective of the course was to introduce the students to the breadth and depth of the field of Bioinformatics for ‘omics’ data analyses. We also identified the following three course outcomes for the students.

- I. At the end of the course, students should be able to identify and implement alternate strategies to answer genomics-based research questions.
- II. Students should be comfortable with the use open-source genomic software and command line programming, and be able to use R statistical packages.
- III. Students should be able to design and troubleshoot analyses of nucleotide sequence data and elicit biological information from the data.

Course structure

The course was divided into seven modules spread across the semester: Genome assembly and annotation, Comparative genomics, Introduction to Statistics, Metagenomics, Transcriptomics,

Table 1. Characteristics of user groups.

User groups	Characteristics
Bioinformatics Tool Users (BTU)	These users access bioinformatics resources, packages and software to perform analyses specific to their research domains. e.g. bench scientists, medical professionals
Bioinformatics Data Scientists (BDS)	These users utilize computational methods to analyze data and advance the scientific understanding of living systems.
Bioinformatics Engineers (BE)	These users, create, develop and manage novel computational methods needed for novel scientific discoveries.

Table 2. Survey questions sent out to the students.

Question premise	Reasons for the question	Responses
Q1) Previous Programming experience?	We wanted to gauge the level of expertise of the students and identify the level of programming to be introduced in class.	(i) 4 participants had taken a course on R.* (ii) 5 participants had no previous experience using any bioinformatics software or programming languages.
Q2) Motivation for registering in the course?	We wanted to understand the rationale of the students participating in the course	Unanimous response of the participants was that they were working on some type of benchwork that would generate "omic" data.
Q3 Take away from the course?	We wanted to ensure our learning outcomes matched the expectations of the course participants.	-Understand types of sequencing technologies -Learn how to analyze data -Learn better practices of biological data management

*Since we did not have this information in the pre-class survey answers, we asked students their experience with programming languages in class. We got 7 responses in total to the pre-lab survey.

Proteomics and Cancer data analysis. Each module ended with a graded research problem either in a prokaryotic system or a eukaryotic system (Table 3 and Supplementary File 1).

Results

Based on the responses of the students, we assigned potential user groups as explained in Table 1 at the start of the class with their expected competency levels at the end of the class. Seven students replied and two students did not reply to the pre-course survey. We were able to obtain permission from six of the seven students who replied to the survey to have their answers published online anonymously. Any identifying information in terms of names or project details have been edited from the responses (Table 4).

Successful completion of the project assigned to every student by the end of a course module determined their competency of the course. In lieu of a final exam, each student designed a research project, conducted appropriate analyses, and summarized their results in the form of a poster or a talk at the end of the semester as part of the ISCB-RSG-SE USA (International society of Computational biology-Regional student group-Southeast USA) conference held on campus on Dec 8/9 of 2017. They also had the opportunity to listen to talks from professors working on bioinformatics projects and interacted with their peers from University of South Florida and University of Alabama. In addition, two graduate students wrote papers on their projects with input from their respective research advisors.

Dataset 1. Pre-class survey

<https://dx.doi.org/10.5256/f1000research.16310.d218863>

Dataset 2. Post-class survey

<https://dx.doi.org/10.5256/f1000research.16310.d218864>

Discussion

This course covered a lot of topics in 13 weeks and some degree of mastery was required for each topic. In addition, half of the students had no familiarity with programming. As a result, many of the students were stretched beyond their comfort zone. However, since this was a small class, we were able to work with the students individually to help them be successful, and also tailor projects to the students' backgrounds and expectations. An important outcome of this course design was that the students acquired the basic skills to critically evaluate the reporting and interpretation of data of a problem or a project during the symposium.

Our leading goal was to develop a course that was responsive to the needs and background abilities of the participating students. It is important to recognize that every course will have students at different levels of learning with different goals. Hence when designing a course that caters to the needs of the students, it may be a good idea to have a small class.

Table 3. Summaries of course modules *.

Module	Topics covered	Software	Project
Genome assembly and annotation	(i) DNA sequencing and its advances over the years. (ii) Assembly of a bacterial genome from nucleotide sequencing data, and submission to NCBI GenBank	Artemis : A free genome browser and annotation tool that allows visualization of sequence features ¹⁵ .	1. Students were asked to download the <i>Caulobacter segnis</i> genome and identify the potential sequencing errors. 2. Project report on the HeLa. Strategies on identification of the difference between healthy and non-healthy cells. Ways of identifying HPV 18 contamination in HeLa cells
Comparative Genomics	(i) Strategies to identify prokaryotic and eukaryotic genes (ii) Strategies for genome comparison: genome size, genomic signature, gene order analyses through sequence alignment	MAUVE: Multiple genome aligner to compare genomes for evolutionary events and rearrangements ¹⁶ .	Comparative analyses of 'Odorant binding proteins' among strains of <i>Drosophila melanogaster</i> and <i>Apis mellifera</i> . Students performed homology comparisons and constructed phylogenetic trees to observe OBP diversification across the genomes.
Metagenomics	1. Importance of metagenomics across research domains. 2. Exploring types of research questions answered by metagenomic based studies 3. How to set up metagenomic studies, data extraction, submission and analyses through MG-RAST pipeline	MG-RAST pipeline: It provides an automated quality control, annotation, comparative analysis and archiving service of metagenomic and amplicon sequences using a combination of several bioinformatics tools ¹⁷ . STAMP: software package for analyzing taxonomic and metabolic profiles by choosing appropriate statistical techniques ¹⁸ .	Comparison and analyses of the Global Ocean Sampling Expedition data available at the MG-RAST data repository. Students were also introduced to statistical hypothesis testing within data sets and between data sets.
Introduction to statistics	(i) Descriptive and Inferential statistics. (ii) Univariate and Bivariate analyses (iii) ANOVA and PCA	R Statistical package: Students were introduced to the R package and were given cheat sheets on how to load, access, and manipulate biological data.	Students were introduced to these concepts and then allowed to work on their comparative metagenomics data analyses projects.
Transcriptomics	Students were introduced to the RNA sequencing technologies and analyzed data from an RNAi knock-down experiment of the <i>pasilla</i> splicing factor gene in <i>Drosophila</i> ¹⁹ .	R Statistical package ²⁰	Students detected differentially expressed genes using R packages and learned how to take confounding factors into account in differential expression analysis. They were also introduced to different visualization packages in R.
Proteomics	Students were introduced to protein diversity characterization using proteomics. The dataset used for this module was from Bioconductor Conference held at Stanford in July 2016.	R Statistical packages	Student used R/Bioconductor packages to explore, process, visualize, and understand mass spectrometry-based proteomics data.
Cancer data analyses	This module was offered by Dr. Phillip Buckhaults (Director of the Cancer Genetics laboratory at the University of South Carolina)	UCSC Cancer genomics browser ²¹ , TCGA ²² , Gene set enrichment analysis ²³	Students were reintroduced to RNASeq analysis and its role in generation of cervical cancer data for Dr. Buckhaults' recent paper ²⁴ . They were also shown the features of UCSCS Cancer genome browser. Students analyzed TCGA database for gene expression association analyses for Glioblastoma. Further data mining was carried out using Gene set enrichment analyses were carried out for previously identified genes to check for statistical importance.

* All the presentations associated with each module, course assignments and problem assignments are available for access in the supplementary section of the paper. The final projects that were presented as posters and talks are not available for access at this time.

Table 4. Student pre class and expected user groups.

Student	Pre-class User group	Expected user group
1.	Bioinformatics Tool User	Bioinformatics Tool User
2	Bioinformatics Tool User	Bioinformatics Data scientist
3	Bioinformatics Tool User	Bioinformatics tool user, Bioinformatics Data scientist
4	Bioinformatics Tool User	Bioinformatics Data scientist
5	Bioinformatics Tool User	Bioinformatics Data scientist
6	Bioinformatics Tool User	Bioinformatics Tool User

In our class, every student had a different learning curve. We determined the competency of a student per module by their successful completion of the problem set and or the project. The first objective of the course was to expose the students to not just one living system but many including Bacterial, Human, *Drosophila*. The other objective was to introduce the students to the R computational platform²⁰. Our initial challenge was to address the problems faced by the students in using the platform for the first time. We wanted the students to understand the intricacies of using R as a programming language but if we repeat this class, we will have the codes for the students as R- markdown documents. We would also have additional R assignments at the beginning of the course and out of class help sessions to help students get comfortable using R.

A major challenge was to identify ways to map the competencies required to the expectations of the course at both the undergraduate and graduate levels. Since we had a small number of students, we designed and delivered a structured curriculum that integrated both the continuously changing and stable technological platforms using model systems that were used by at least one student for every module.

As the important goal of the course was to address the needs of the students, we designed the current model of ‘multi-project’ modules of biological data analyses. Due to the small class size, we were able to give personalized attention to every student. In the future, a big change that we would incorporate would be to separate the projects and problems assigned to graduate and undergraduate students. Generally, the undergraduate students do not have their own data while the graduate students usually have or are in the process of obtaining data that they want to analyze. Therefore, we would either have separate sections for the graduate and undergraduate students or we would have a combined lecture but separate recitation section where the students would apply what they have learned in the lecture portion of the class. The graduate students would be encouraged to develop projects that are relevant to their research while the undergraduates would work in groups on projects designed by the instructor.

Keypoints

- This course was designed to address the students need to analyze ‘omic’ data sets at University of South Carolina
- It was divided into seven modules with practical tasks at the end of each module.
- Students designed their projects and presented it as papers, posters and talks at The ISCB- RSG-SEUSA symposium.

Data availability

Dataset 1: Pre-class surveys [10.5256/f1000research.16310.d218863](https://doi.org/10.5256/f1000research.16310.d218863)²⁵

Dataset 2: Post-class surveys [10.5256/f1000research.16310.d218864](https://doi.org/10.5256/f1000research.16310.d218864)²⁶

Ethical considerations

The authors have posted the pre-class survey answers of students who have consented to have their responses published anonymously. All identifying information has been edited from the responses. The post-class survey responses are given as a feedback to the instructors, also anonymously, through an online survey carried out by the university.

Grant information

The author(s) declared that no grants were involved in supporting this work.

Acknowledgements

The authors would like to thank Dr. Phillip Buckhaults for the design, conception and delivery of the lectures on “Cancer Genomics”. The authors would also like to thank all the attendees, participants and professors of the Departments of Biological Sciences and Computer Science of University of South Carolina for participating in the first ‘ISCB-RSG-SEUSA’ symposium held this past December of 2017 at Columbia, SC.

Supplementary material

Supplementary File 1: Course syllabus and teaching materials

[Click here to access the data](#)

References

1. Searls DB: **The roots of bioinformatics.** *PLoS Comput Biol.* 2010; **6**(6): e1000809.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
2. Barone L, Williams J, Micklos D: **Unmet needs for analyzing biological big data: A survey of 704 NSF principal investigators.** *bioRxiv.* 2017; 108555.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
3. Madlung A: **Assessing an effective undergraduate module teaching applied bioinformatics to biology students.** *PLoS Comput Biol.* 2018; **14**(1): e1005872.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. Dinsdale E, Elgin SC, Grandgenett N, et al.: **NIBLSE: A Network for Integrating Bioinformatics into Life Sciences Education.** *CBE Life Sci Educ.* 2015; **14**(4): 1e3.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
5. Via A, Blicher T, Bongcam-Rudloff E, et al.: **Best practices in bioinformatics training for life scientists.** *CBE Life Sci Educ.* 2013; **14**(5): 528–37.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
6. Cresiski RH: **Undergraduate bioinformatics workshops provide perceived skills.** *J Microbiol Biol Educ.* 2014; **15**(2): 292–4.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
7. Banta LM, Crespi EJ, Nehm RH, et al.: **Integrating genomics research throughout the undergraduate curriculum: a collection of inquiry-based genomics lab modules.** *CBE Life Sci Educ.* 2012; **11**(3): 203–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
8. Attwood TK, Blackford S, Brazas MD, et al.: **A global perspective on evolving bioinformatics and data science training needs.** *Brief Bioinform.* 2017; bbx100.
[PubMed Abstract](#) | [Publisher Full Text](#)
9. Emery LR, Morgan SL: **The application of project-based learning in bioinformatics training.** *PLoS Comput Biol.* 2017; **13**(8): e1005620.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. Luo J: **Teaching the ABCs of bioinformatics: a brief introduction to the Applied Bioinformatics Course.** *Brief Bioinform.* 2014; **15**(6): 1004–13.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
11. Altmäe S, Esteban FJ, Stavreus-Evers A, et al.: **Guidelines for the design, analysis and interpretation of 'omics' data: focus on human endometrium.** *Hum Reprod Update.* 2014; **20**(1): 12–28.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
12. Boekel J, Chilton JM, Cooke IR, et al.: **Multi-omic data analysis using Galaxy.** *Nat Biotechnol.* 2015; **33**(2): 137–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
13. Mulder N, Schwartz R, Brazas MD, et al.: **The development and application of bioinformatics core competencies to improve bioinformatics training and education.** *PLoS Comput Biol.* 2018; **14**(2): e1005772.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
14. Welch L, Lewitter F, Schwartz R, et al.: **Bioinformatics curriculum guidelines: toward a definition of core competencies.** *PLoS Comput Biol.* 2014; **10**(3): e1003496.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
15. Carver T, Harris SR, Berriman M, et al.: **Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data.** *Bioinformatics.* 2012; **28**(4): 464–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
16. Darling AE, Tritt A, Eisen JA, et al.: **Mauve assembly metrics.** *Bioinformatics.* 2011; **27**(19): 2756–7.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
17. Meyer F, Paarmann D, D'Souza M, et al.: **The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes.** *BMC Bioinformatics.* 2008; **9**(1): 386.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
18. Parks DH, Beiko RG: **Identifying biologically relevant differences between metagenomic communities.** *Bioinformatics.* 2010; **26**(6): 715–721.
[PubMed Abstract](#) | [Publisher Full Text](#)
19. Brooks AN, Yang L, Duff MO, et al.: **Conservation of an RNA regulatory map between *Drosophila* and mammals.** *Genome Res.* 2011; **21**(2): 193–202.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
20. R Core Team: **R: A language and environment for statistical computing.** R Foundation for Statistical Computing. Vienna, Austria. 2014.
[Reference Source](#)
21. Goldman M, Craft B, Swatloski T, et al.: **The UCSC Cancer Genomics Browser: update 2015.** *Nucleic Acids Res.* 2015; **43**(Database issue): D812–817.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
22. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, et al.: **The Cancer Genome Atlas Pan-Cancer analysis project.** *Nat Genet.* 2013; **45**(10): 1113–20.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
23. Subramanian A, Tamayo P, Mootha VK, et al.: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proc Natl Acad Sci U S A.* 2005; **102**(43): 15545–50.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
24. Banister CE, Liu C, Pirisi L, et al.: **Identification and characterization of HPV-independent cervical cancers.** *Oncotarget.* 2017; **8**(8): 13375–86.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
25. Saarunya G, Ely B: **Dataset 1 in: Design and implementation of semester long project and problem based bioinformatics course.** *F1000Research.* 2018.
<http://www.doi.org/10.5256/f1000research.16310.d218863>
26. Saarunya G, Ely B: **Dataset 2 in: Design and implementation of semester long project and problem based bioinformatics course.** *F1000Research.* 2018.
<http://www.doi.org/10.5256/f1000research.16310.d218864>

Open Peer Review

Current Peer Review Status: ? ? ?

Version 1

Reviewer Report 17 December 2018

<https://doi.org/10.5256/f1000research.17818.r40462>

© 2018 Via A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Allegra Via

National Research Council of Italy (CNR), Institute of Molecular Biology and Pathology (IBPM), c/o Department of Biochemical Sciences "A. Rossi Fanelli", Sapienza University of Rome, Rome, Italy

The paper describes a semester long bioinformatics course targeting graduate seniors and graduate students who were bench scientists in need for learning how to analyse data generated across different 'omic technologies'.

I find it weird that "The authors haven't come across a course that incorporates multi-omics data analyses in a single semester." If not in a single course, some curricula offer multi-omics data tools and analyses spread in more than one course. A comparison of the presented course with such curricula would be of great interest, as well as a discussion on the convenience of integrating such large amount of bioinformatics materials in a Biological Sciences curriculum.

There is much discussion in the field on what is the best strategy to incorporate Bioinformatics in Life Sciences curricula and I wonder whether an overload of different topics, techniques, approaches, methods would be successful in contexts where instructor could not work individually with students.

Table 3 displays a number of features of the course's modules. However, a well structured program of each module is missing. As for reproducibility, a lesson plan describing how much time was allocated to each classroom activity (lectures, work in group, hands-on, work on individual projects, types and frequency of formative assessments, etc.) would help. Teaching materials provided in the Supplementary materials are not structured at all. Teaching materials are organised in modules, but navigating modules it is very difficult to understand how to use the various files. There is no homogeneity in file names and a "readme" file describing the content of each folder (and how to use it in reproducing the course) is missing. Slides are not annotated. In summary, materials are not reusable in the current form and the course would not be reproducible based on them and on the information provided in the article. The teaching techniques/strategies used in the classroom were not described/discussed, apart from mentioning the importance of the individual work with students. I think the article would benefit from more details on the course design and from the description of the pedagogical

approaches the instructors adopted to teach programming and computational skills to bench scientists.

I understand that a key point was the small number of students. Nevertheless, most courses with a small number of students and motivated instructors usually produce successful results. One big challenge is when the number is high. It would be interesting if the authors could reason on how their course could be translated into one for a bigger group of students. What should be definitely changed? Which other strategies could be adopted (peer instruction? Helpers?)?

Finally, the authors use a lot of the term “competency/competencies”. There is currently quite a lot of debate around the convenience of using competencies to describe the outcomes of courses. Indeed, competencies can hardly be assessed and mapped on a learning trajectory. By completing a single course, students may develop knowledge, skills and abilities (KSAs), which are measurable and accessible objects and the development of which can be followed along a learning trajectory, rather than competencies. Could the authors comment on this?

Here are more specific points:

1. p.3 – Re the following sentence: “Practitioners of bioinformatics now add techniques from statistics, information science and engineering to develop algorithms and build predictive models to understand the dynamics within a biological system.” In my experience, practitioners of bioinformatics have always added techniques from statistics, information theory and engineering to develop algorithms to predict the functioning of biological systems. The paradigm shift caused by the rapid advances in sequencing technologies is of different kind in my opinion: in the first place, bioinformatics has become the only approach to make sense of the deluge of biological data the authors refer to. Moreover, the storage, management, sharing, annotation, “fairfication” of the enormous amount of data produced, poses important technological challenges and emphasizes the need for new professions.
2. p. 3 – In the sentence: “Practitioners of bioinformatics...”, “Practitioners” should be changed to “Practitioner”. Please, check the whole manuscript for typos/misspellings.
3. p. 3 – The authors put the sentence: “However, one of the biggest challenges is the heterogeneity of the backgrounds of the course participants” in opposition to the previous one on ISCB competencies (“However,...”). In contrast, I believe that Bioinformatics core competencies listed in Mulder et al. indirectly express the high degree of heterogeneity of backgrounds in bioinformatics.
4. p.3 – Re the sentence: “In fact, there are three different types of user groups that employ bioinformatics in their research”, I would not define Bioinformatics Engineers as bioinformatics users, but rather developers and managers/maintainers of computational tools.
5. p.4, Table 1 – There is another relevant group of bioinformatics practitioners: those who take care of and manage data, bioinformatics resources and their interoperability and develop standards, data quality metrics, ontologies, annotation, etc. The “big data issue” is especially relevant in the “omics” field and, in my opinion, it would be good if the authors could mention this fourth group, even though none of their students did belong to it.
6. p.3, In the sentence: “We sent a three-question survey (Table 2) to all the participants to understand their reasoning for registering in the course.” I suggest that the authors replace “reasoning” with “motivations” or “reasons”.
7. p.3, in the sentence “We also identified the following three course outcomes for the

- students." The authors say "course outcomes". What is a course outcome? I suspect they mean "learning outcomes". There is quite a lot of confusion in the field around the definition and usage of "learning objectives", "learning outcomes" and "teaching objectives". I suggest that the authors replace "course outcomes" with "learning outcomes".
8. p.3, Re Learning outcomes. The literature provides quite precise rules to write learning outcomes. You can use the sentence "by the end of the course, students will (*NOT should*) be able to" followed by an "actionable verb", namely a verb expressing an action or a behaviour that can be (at least in principle) assessed. The verbs used in learning outcomes I ("identify" and "implement") are of this type, whereas some verbs used in II and III are not ("be comfortable", "elicit"). Moreover, it is a good practice to write learning outcomes that are as much specific as possible in terms of both the cognitive complexity level they express and their content. For example, in learning outcome I, "identify" and "implement" express two different levels of cognitive complexity and learning outcome II includes a large variety of contents.
  9. p.3, Learning outcome II. What do the authors mean by "command line programming"? Do they mean "Linux shell scripting" or "navigating files and directories using the command line shell"? To be able to use R statistical packages implies to be able to do (at least some) R programming. I suggest that the authors specify this.
  10. p.4, the footnote of Table 2 is misleading. What does it mean that the authors did not have the information about programming experience in the pre-class survey answers? Did they asked question 1 in the pre-class survey (as stated in the manuscript) or in class (as stated in the footnote)? Were the 7 responses about programming experience? If so, this means that the authors got 2 answers in class and 7 answers in the pre-class survey. Is this correct? Or the pre-labsurvey is another thing? Very confusing.
  11. Table 2. Survey questions sent out to the students - As question 1 is about "programming experience", please notice that "using bioinformatics software" is not "programming".
  12. For consistency with answers to questions 1 and 2, please specify the distribution of answers to question 3.
  13. p.4, Re the sentence: "Based on the responses of the students, we assigned potential user groups as explained in Table 1 at the start of the class with their expected competency levels at the end of the class.", I have three main concerns: 1) I don't see where competency levels at the end of the class are listed (unless the authors are now calling "competency levels" what they called "characteristics" in Table 1. Should this be the case, in no way can students acquire the characteristics listed in Table 1 by completing the course described in this paper; 2) Competencies are yes/no objects, which means either an individual has a competency or they don't have it. Therefore, it may be problematic to talk about "competency levels"; it may be perhaps more appropriate to talk about knowledge, skills or abilities (KSAs) levels; 3) If by "class" you mean a series of lectures on a subject, could you specify at the end of which class (a module? The entire course?) you defined "expected competency levels"? As a side note, a single class can possibly increase the level of a KSA, surely not allow students to acquire a competency.
  14. p. 4: in this sentence: "Successful completion of the project assigned to every student by the end of a course module determined their competency of the course." It is not clear what do the authors mean by "competency of the course". Do they mean that the competency acquired in a module determined students' competency in the whole course?
  15. p. 6: In the sentence: "We determined the competency of a student per module by their successful completion of the problem set and or the project." what do the authors mean by "successful completion of the problem set and or the project"? There were students who

did not successfully complete the project? How did instructors grade them?

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

No

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Protein structural bioinformatics, protein structure and function prediction and analysis, and protein interactions. Programming and software development. Science of learning, educational psychology, cognitive sciences, and (bioinformatics) curriculum development.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 10 December 2018

<https://doi.org/10.5256/f1000research.17818.r40513>

© 2018 Pauley M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Mark A. Pauley

National Science Foundation, Alexandria, VA, USA

“Design and Implementation of Semester Long Project and Problem based Bioinformatics Course” describes a “multi-omics” bioinformatics course at the University of South Carolina intended for advanced undergraduates and graduate students. The course was implemented in Fall 2017; nine students took it. Per the authors, the primary learning objective of the class was to introduce

students “to the breadth and depth of the field of Bioinformatics for ‘omics’ data analyses.” The course was divided into seven modules (e.g., “Genome Assembly and Annotation,” “Comparative Genomics”). Each module had an associated graded problem set, and students completed a research project at the end of the course. A three-question, pre-course survey was used to place students into user groups—bioinformatics tool users, bioinformatics data scientists, and bioinformatics engineers.

The article has many strengths. The authors make a compelling case for the need for courses like it to prepare students for graduate school and to address the need for specialists in the field, and they do a good job of putting their course in the context of other bioinformatics education efforts. The contents of the course are clearly laid out (Table 3), and the authors provide a large amount of material (syllabus, slide decks, problem sets) developed for the class as a supplementary file—both will be invaluable for others wishing to implement the entire course or parts of it. As how a course could be improved is often more instructive than what went well, their discussion of potential changes in subsequent iterations of the class is very helpful. Finally, the article is clearly written and easy to read.

That said, the manuscript has several issues that should be addressed. First, a number of references are potentially mis-cited. For example, References 6 and 7 cite a Global Organization for Bioinformatics Learning (GOBLET) study that showed that basic data stewardship skills are only taught in 25% of education programs. However, neither of these papers mention the GOBLET survey or the 25% statistic. In addition, References 11 and 12 do not deal with bioinformatics courses and Reference 15 does not discuss the competencies of different bioinformatics users as their use would imply. Similarly, I am concerned about the bioinformatics user groups given in Table 1. Specifically, the descriptions of the three groups are very similar to the three personas described in Reference 14, and the name of one (bioinformatics engineer) is the same (the names of the other two are almost the same). In short, it’s not clear if the authors are restating the results of Reference 14 or are proposing a slightly different grouping. Although the posted resources are clearly an important contribution, I found them to be incomplete in one important aspect. In particular, the authors state that every module had a problem set/project associated with it, but this was missing from three of the seven modules. Furthermore, a brief description of the final research projects the students worked on would be helpful as it would indicate what the students were able to do at the end of the semester.

In addition to the above, very little is provided in terms of results. One of the results seems to be the placement of students into the three user groups. However, how the results of the pre-course survey were used to place the students into these groups and if and how they impacted the way in which the course was taught is not clear. Similarly, Table 4 and the corresponding description of it in the narrative, particularly the use of the word “expected,” is confusing. Does Column 3 of the table refer to the group a given student was in at the end of the semester or where they were expected to be at some other point in the semester? In any event, how was this determined? Although the course evaluation is helpful in understanding how students felt the course went, I would have liked to have seen more assessment results, particularly if the learning objectives of the course had been met. In general, the paper would be strengthened by the results of another iteration of the course, one in which the proposed changes had been made and the learning gains of the students were assessed.

As previously mentioned, the article is well-written. However, I did notice two small errors. The

first sentence of "Course design" should probably be "We had nine students register for the course." Also, "bioinformatics" is incorrectly capitalized in "This course was designed to provide a structured Bioinformatics course. . .".

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bioinformatics education, bioinformatics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 02 November 2018

<https://doi.org/10.5256/f1000research.17818.r39760>

© 2018 Schwartz R. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Russell Schwartz

Department of Biological Sciences and Computational Biology Department, Carnegie Mellon University , Pittsburgh, PA, USA

Saarunya and Ely describe a problem-based bioinformatics course designed to meet a need for "next generation data scientists" in the life sciences, a need identified by many current efforts in life sciences education. Case studies of course development efforts like this can be valuable to those seeking to develop similar courses or incorporate those courses into their curricula and

looking for ideas or for pitfalls to avoid. The authors do a good service for the field in putting out their efforts and lessons learned in a form from which other educators can benefit. The specific effort here is a nice example of a small project-focused course serving a cohort with some diversity of backgrounds and immediate training needs. While it presents just one small example, that description might reasonably apply to courses many training programs are developing or would like to develop. In addition to the article itself, the supplementary material includes a full syllabus, lecture slides, assignments, and some supplementary materials, increasing its value to others looking to develop course materials in this space.

The authors make a good case for the need for new courses along these lines. They back that need up well with appropriate citations to the relevant literature on life sciences and bioinformatics education. The manuscript provides a good background on prior efforts to characterize the need for bioinformatics training, identify the specific skills required by future life scientists, and how those skills are or are not being provided in practice. The authors further give reasonable consideration to challenges to the design of bioinformatics curricula that they expected to confront in this effort. On the latter point, they might also refer to Williams *et al.* (2017¹), which identified a number of other recurring challenges to bioinformatics education in the life sciences. Others in the field might appreciate the perspective of these authors on whether any of the challenges Williams *et al.* identified were encountered in their effort and, if so, how they were overcome.

The course itself covers a nice range of topics in applied bioinformatics, which might be expected to meet the needs of a diverse set of likely users. The course materials provided in the supplement might therefore find a good audience. One general concern, though, is that the supplementary materials contain some third-party resources, for which it might be more appropriate to include a reference or link rather than the material itself. The teaching approach is fairly applied, with a lot of focus on specific data resources and software, although with some attention to principles behind these resources. While some user communities might favor an approach more grounded in the principles and theory, the focus here seems typical of many bioinformatics courses aimed primarily at biology students. The authors might do a bit more to justify the balance of focus on practice versus theory, with reference to efforts at identifying specific bioinformatics competencies needed by their likely user community, several of which the paper cites.

The Results present some interesting material in the form of a pre-class survey and post-class course evaluation material. While the cohort here is a single small sample, some useful lessons can be drawn about the diversity of backgrounds and needs of even a small group like this. The paper would be considerably stronger with some more serious assessment of whether the learning objectives of the course were met. That is a non-trivial undertaking and cannot be done retroactively, but might be worth considering for a future iteration of the class if it is being continued. The materials do include results of a university-run course evaluation, which provide some indication of how students felt about the course, although that is different from showing how successfully they learned the material. This post-class evaluation makes for some interesting reading, although if it is being included with the paper, it might bear some comment in the Results and Discussion.

It would be useful also to see some comparison to other similar course material available in publicly accessible forms. While that is a difficult moving target, comparing to a few alternatives from prominent course repositories or MOOCs, particularly to highlight the unusual or especially

innovative features of this course, would be valuable.

The paper does a nice job of presenting some lessons learned in the Discussion. It is commendable that the authors spend some time on what did not work so well in this class and consider how it might be done differently in the future. One would ideally like to see this taken further via a more comprehensive formative assessment process – with problems identified via a formal assessment, solutions proposed, and those solutions demonstrated to be effective in a re-assessment. It is understandable that that may be beyond the scope of a one-off paper like this, though, and it is nonetheless easy to see how others developing a class in this domain might benefit from the advice given here to avoid some of the same pitfalls.

Beyond these more specific technical points, the document is clear and generally well-written. I noted just a couple of minor errors:

- p. 4: “International society of Computational biology” should be “International Society for Computational Biology”.
- p. 4: “Regional student group – Southeast USA” should be “Regional Student Group – Southeast USA”.

References

1. Williams J, Drew J, Galindo-Gonzalez S, Robic S, et al.: Barriers to Integration of Bioinformatics into Undergraduate Life Sciences Education. *bioRxiv*. 2017. [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 26 Nov 2018

Geetha Saarunya

The authors would like to thank Dr. Schwartz for his in-depth and insightful feedback on the paper. Following are the comments from the authors, which will be incorporated into the final version of the paper after the second and third referees' feedback:

The authors recognize the contributions made by 'Williams *et al.*.*' in identifying the challenges of introducing bioinformatics to life-science students. These issues are already addressed in the paper in the following ways:

(i) Faculty issues (training): The authors' training and background gave them an opportunity to design a multi-project/problem based course. But the post-module projects/problem sets were based on the background of the students. And this was possible because of the small class size.

(ii) Faculty issue (time): This course was designed with inputs from the students based on their needs and training. Hence a lot of time was spent on the course design followed by making changes/adjustments to the course during the implementation.

(iii) Student issue (Background skills): The authors addressed the gaps in student's computational and statistical training by offering additional learning modules. The authors have also addressed the problems faced by the students and ways to tackle them in the future under 'Discussion' section.

(iv) Student issue (Interest): As an applied Bioinformatics course, the students had an opportunity to apply their learning to solve problems and projects in their area of interest/background. Active engagement and participation of the students was encouraged throughout the course by timely submission of projects and problem sets.

2. The authors recognize the need to have a better competency assessment of the students' pre- and post-course. In future, this can be accomplished in the form of pre-course problem solving and post-course problem solving to ensure that the students meet the set learning objectives. The course in the current format had the student's research, design, address and present their learning (with emphasis on critical evaluation and problem solving) in the form of a project presented as a talk/poster in the research symposium held at the end of the semester. To protect the student's data/projects, the final posters and presentations are not included in this paper.

3. As most of the participants were classified as 'Bioinformatics tool users' the authors chose to focus on applied bioinformatics as opposed to Bioinformatics theory. In order to have a bioinformatics focused theory class designed to address every 'omic' problem, the authors believe that it would be prudent to have just one or two modules together and introduce theory and problem/projects pertaining to the same.

4. The authors have cited the third-party resources in the main paper with reference numbers in the supplementary materials. The authors will add the supplementary

references in supplementary section and main references in the main paper.

5. The course design and challenges addressed in this paper are pertaining to the small class size and may not accurately reflect the challenges faced at the level of MOOC learning. But the authors can add references to MOOC courses that offer similar style of training in the background section.

*Reference:

* Williams J, Drew J, Galindo-Gonzalez S, Robic S, Dinsdale E, Morgan W, Triplett E, Burnette J, Donovan S, Elgin S, Fowlks E, Goodman A, Grandgenett N, Goller C, Hauser C, Jungck J, Newman J, Pearson W, Ryder E, Wilson Sayres M, Sierk M, Smith T, Tosado-Acevedo R, Tapprich W, Tobin T, Toro-Martínez A, Welch L, Wright R, Ebenbach D, McWilliams M, Rosenwald A, Pauley M: Barriers to Integration of Bioinformatics into Undergraduate Life Sciences Education. *bioRxiv*. 2017

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research