INTRODUCTION

As technological advances improve the ability to study biological problems from a systemic perspective, undergraduate training in “-omics” fields, bioinformatics, and the use of “big data” is becoming unavoidable. This training is a particular challenge for instructors at non-research-intensive institutions, including community colleges and liberal arts colleges, who usually lack the infrastructure or resources necessary to produce engaging and accessible “-omics” laboratory experiences on their own. However, these challenges can often be offset by incorporating projects into a course-based research experience (CRE). Through CREs, instructors can design and implement large-scale projects within a classroom that, in a traditional apprentice model, would be limited to one or two students. Thus, CREs gain dual benefits over individual research experiences: increased opportunity for multiple students to engage in authentic research, and reduction of the cost per student.

Collaboration between schools and/or involvement with existing research adds feasibility and credibility to a CRE. A number of collaborative undergraduate research initiatives, such as the Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) (1, 2), the Genomics Education Partnership (SEA-PHAGES) (1, 2) the Genomics Education Partnership (SEA-PHAGES) (1, 2), the Genomics Education Partnership (SEA-PHAGES) (1, 2), the Genomics Education Partnership (SEA-PHAGES) (1, 2), the Genomics Education Partnership (SEA-PHAGES) (1, 2), the Genomics Education Partnership (SEA-PHAGES) (1, 2), the Genomics Education Partnership (SEA-PHAGES) (1, 2), the Genomics Education Partnership (SEA-PHAGES) (1, 2), the Genomics Education Partnership (SEA-PHAGES) (1, 2), the Genomics Education Partnership (SEA-PHAGES) (1, 2), the Genomics Education Partnership (1, 2) the Small World Initiative (4), have allowed many institutions to take advantage of a crowdsourcing approach to bring authentic “-omics” research into their classrooms. As these technologies become cheaper and more abundant, resource-limited institutions may draw inspiration by developing laboratory projects that allow students to explore connections between bioinformatics data on a computer screen and results from laboratory benchwork.

Recent advances in mass spectrometry have facilitated efficient and inexpensive identification of protein components within a cellular sample (5). This opens up an array of possibilities for instructors to develop laboratory activities in which students can compare their own proteomic data with genomic and transcriptomic data found in publicly available databases. Our classroom implementation of this model involves proteomic analysis of Mycobacterium smegmatis infected with bacteriophages. However, it is important to note that this model is applicable to any project where the goal is to link database analysis with proteomic data generated through benchwork.

PROCEDURE

Information and protocols on how to obtain and culture M. smegmatis and isolate bacteriophages can be found at http://phagesdb.org/workflow. Our mass spectrometry protocol is derived from those previously published (6–8), and a detailed procedure is provided in Appendix 1 and summarized in Figure 1. Briefly, student teams design comparative experimental conditions for infection of an M. smegmatis culture (time, temperature, etc.). Students then generate a time-course of infected cell pellets collected from liquid host cultures infected with phage at a high multiplicity of infection. Frozen cell pellets are sent to a proteomics core facility for processing and data analysis, including trypsin digestion, followed by peptide detection using high pressure liquid chromatography-tandem mass spectrometry (HPLC-MS/MS). Peptide mass/charge spectra are matched to a user-submitted custom database of protein sequences that includes predicted phage and host proteins. The results of the analysis may be compiled by the core facility into an .sf3 format summary file that can then be viewed by students using freeware such as SCAFFOLD Viewer (http://www.proteomesoftware.com/products/scaffold/download/) (9). SCAFFOLD Viewer provides user-friendly visualization of data spectra and interactive statistical thresholds for protein and peptide identification that facilitates comparisons between biologically related samples. Using SCAFFOLD Viewer, students can explore proteomics data in a user-friendly interface and gain insights into the biological processes that underlie the observed changes in the proteome.

©2018 Author(s). Published by the American Society for Microbiology. This is an Open Access article distributed under the terms of the Creative Commons Attribution-Noncommercial-NoDerivatives 4.0 International license (https://creativecommons.org/licenses/by-nc-nd/4.0/ and https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode), which grants the public the nonexclusive right to copy, distribute, or display the published work.
Viewer, our students have analyzed the proteins present in each of their samples with respect to their experimental parameters and used the information to consider how particular genes or gene families may contribute to bacteriophage infection of *M. smegmatis*. (See example screenshots in Figs. 2 and 3.) Paired data sets (0 time point/mock infected) were not used since we were not interested in a quantitative analysis of gene expression for this specific experiment. The entire workflow is relatively low cost; beyond the initial costs for cell growth etc., core facility costs for sample processing and data analysis run approximately $300 per sample.

**Safety issues**

All biological samples used in this example are biosafety level 1 (BSL1) and should therefore be utilized in conjunction with the American Society for Microbiology’s BSL1 guidelines for teaching laboratories (https://www.asm.org/images/asm_biosafety_guidelines-FINAL.pdf). There are no additional safety concerns to address with respect to this exercise, though it is important to provide students with some basic training in sterile technique prior to beginning the work.

---

**FIGURE 1.** Mass spectrometry experimental protocol flowchart. OD = optical density; MOI = multiplicity of infection; LC-MS/MS = liquid chromatography-tandem mass spectrometry; ORF = open reading frame.

**FIGURE 2.** SCAFFOLD Viewer Sample display window. Gene product names beginning with CDS are linked to the mycobacteriophage Brusacoram. All others are of host *Mycobacterium smegmatis* or other origin.
CONCLUSION

We describe one mechanism to allow students to develop and answer their own research questions within the context of “-omics” techniques and bioinformatics. By having students perform a “wet lab” mass spectrometry experiment in conjunction with a bioinformatic investigation, we anchor abstract data with real-world observations. Although our proteomic data sets are not as large as metagenomics or transcriptomic “big data” sets, many of the fundamental components of big data analysis, including database selection, signal, noise, statistical thresholds, and validation, are all present, making this an excellent introduction to the field. This CRE approach allows students to develop a hypothesis based on in silico analysis and test its validity using a “wet lab” experiment.

This example is based on work done in conjunction with the SEA-PHAGES initiative. Although SEA-PHAGES represents an outstanding way to introduce authentic collaborative research into the biology classroom, it is not the only way to construct an engaging “-omics”-based laboratory project. Mass spectrometry has been utilized by several groups in the development of engaging “-omics”-based CUREs (course-based undergraduate research experiences) (10–12); however, this particular SEA-PHAGES-based model shows great promise in its accessibility to institutions limited by budget and infrastructure. Although the generation of mass spectrometry data requires access to an instrument and trained technician to perform the necessary proteomics procedures, advances in technology have brought the costs of this work at many core facilities down to levels that are accessible to most classroom laboratory budgets. Adoption of this type of project in place of other laboratory activities and requisite supplies makes this CRE less financially daunting.

The only limitation to the adoption of this model, then, becomes the lack of expertise of the faculty in working with mass spectrometry data, which may lead to misinterpretations. Thus, we recommend discussing experimental plans with a prospective Core Facility to discover the best approach to generating data that will be of use to your students.

SUPPLEMENTAL MATERIALS

Appendix 1: Experimental protocol for mass spectrometry of *M. smegmatis* cell pellet infected by mycobacteriophage

ACKNOWLEDGMENTS

We are grateful for the support of Graham Hatfull, Debbie Jacobs-Sera, Dan Russell, and the rest of the HHMI SEA-PHAGES leadership team. Mass spectrometry and undergraduate research at Ouachita Baptist University were supported by grants from the National Center for Research Resources (P20RR016460) through the National Institute of General Medical Sciences and the National Institutes of Health (Grant # P20GM103429) and the National Science Foundation (grant# IIA-1457888). The authors declare that there are no conflicts of interest.

REFERENCES


FIGURE 3. SCAFFOLD Viewer output for a representative bacteriophage infection experiment using the bacteriophage Brusacoram. (A) Representative recovered peptide from the mass spectrometry reading. Yellow highlights indicate that LC-MS/MS detected peptide overlap with the gene product. Green highlights indicate modified amino acids. (B) In this case, a much smaller percentage of the predicted ORF was detected. Here, four peptides were detected that overlap with this ORF. A minimum of two detected peptides are required to confirm protein expression. ORF = open reading frame; LC-MS/MS = liquid chromatography-tandem mass spectrometry.


