

Mass Spectrometry as a Tool to Enhance “-omics” Education †

Michael J. Wolyniak^{1*}, Nathan S. Reyna², Ruth Plymale², Welkin H. Pope³, and Daniel E. Westholm⁴
¹Department of Biology, Hampden-Sydney College, Hampden-Sydney, VA 23943, ²Department of Biology, Ouachita Baptist University, Arkadelphia, AR 71998, ³Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260, ⁴Department of Biology, The College of St. Scholastica, Duluth, MN 55811

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 (3), and 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 I 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 open reading frames (ORFs) (these sequences may be obtained through a database like GenBank or annotations of a specific model system database). 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

*Corresponding author. Mailing address: Department of Biology, Hampden-Sydney College, Box 183, Hampden-Sydney, VA 23943. Phone: 434-223-6175. E-mail: mwolyniak@hsc.edu.

Received: 27 August 2017, Accepted: 27 November 2017, Published: 16 February 2018

†Supplemental materials available at <http://asmscience.org/jmbe>

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.

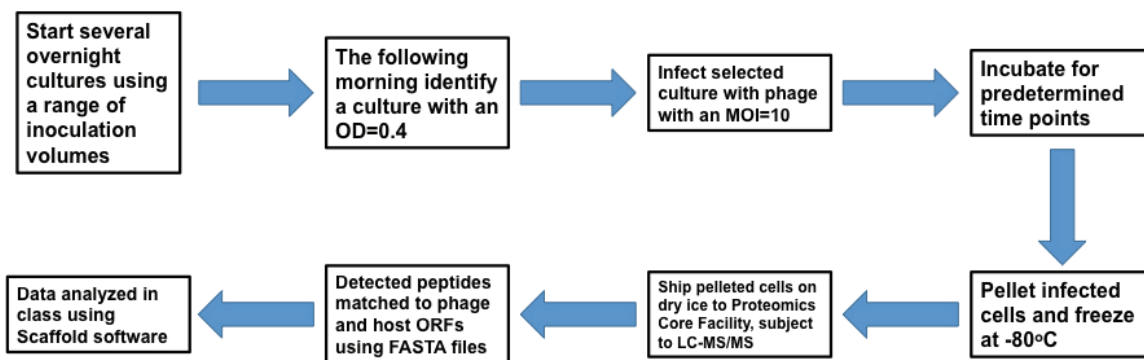


FIGURE 1. Mass spectrometry experimental protocol flowchart. OD = optical density; MOI = multiplicity of infection; LC-MS/MS = liquid chromatography-tandem mass spectrometry; ORF = open reading frame.

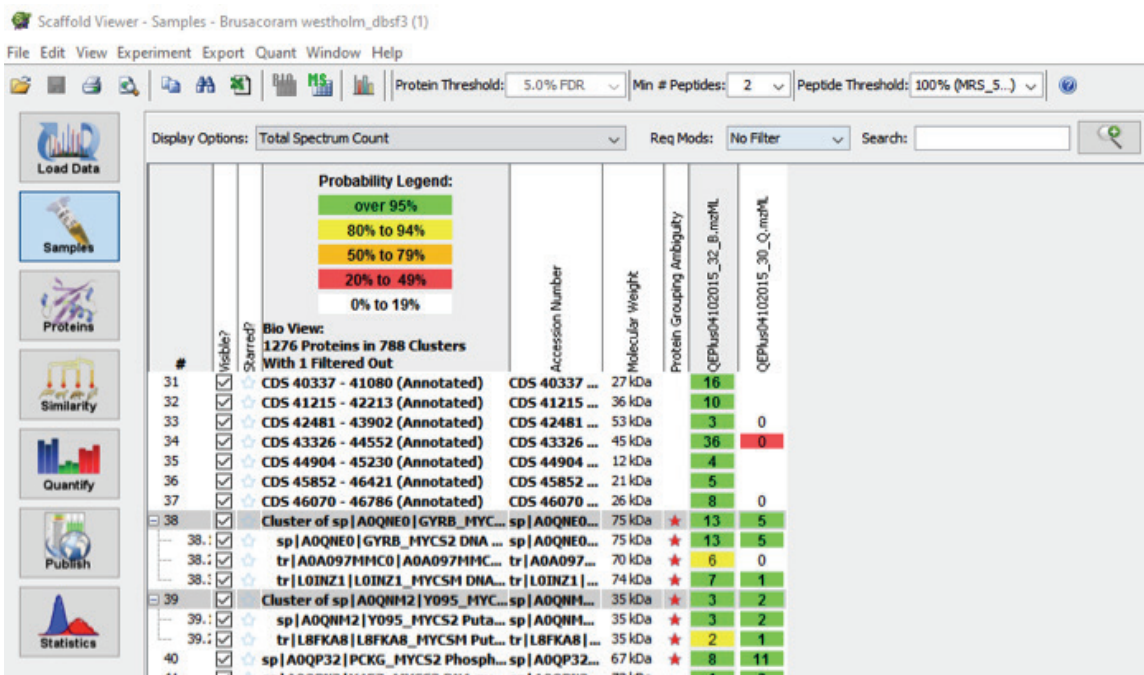


FIGURE 2. SCAFFOLD Viewer Sample display window. Gene product names beginning with CDS are linked to the mycobacteriophage Brusacoram. All others are of host *Mycobacterium smegmatis* or other origin.

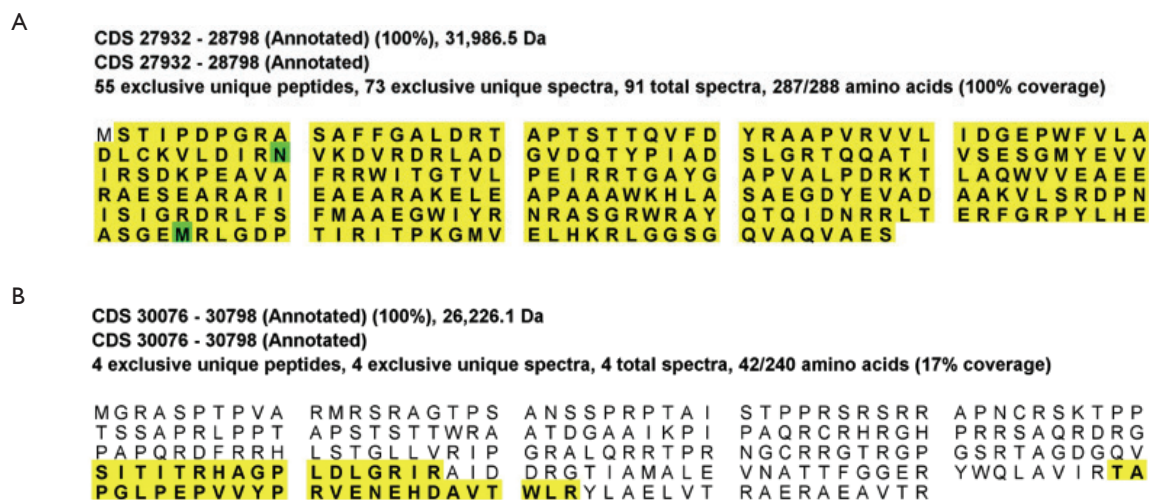


FIGURE 3. SCAFFOLD Viewer output for a representative bacteriophage infection experiment using the bacteriophage Brusacoram. (A) Representative recovered peptide from the mass spectrometry reading. Yellow highlights indicate that LC-MS/MS detected peptide overlap with the gene product. Green highlights 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.

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 I: 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

- Jordan TC, Burnett SH, Carson S, Caruso SM, Clase K, DeJong RJ, Dennehy JJ, Denver DR, Dunbar D, Elgin SCR, Findley AM,

- Gissendanner CR, Golebiewska UP, Guild N, Hartzog GA, Grillo WH, Hollowell GP, Hughes LE, Johnson A, King RA, Lewis LO, Li W, Rosenzweig F, Rubin MR, Saha MS, Sandoz J, Shaffer CD, Taylor B, Temple L, Vazquez E, Ware VC, Barker LP, Bradley KW, Jacobs-Sera D, Pope WH, Russell DA, Cresawn SG, Lopatto D, Bailey CP, Hatfull GF. 2014. A broadly implementable research course in phage discovery and genomics for first-year undergraduate students. *mBio* 5:e01051-13.
2. Hatfull GF. 2015. Innovations in undergraduate science education: going viral. *J Virol* 89:8111–8113.
 3. Elgin SC, Hauser C, Holzen TM, Jones C, Kleinschmit A, Leatherman J, The Genomics Education Partnership. 2017. The GEP: crowd-sourcing big data analysis with undergraduates. *Trends Genet* 33:81–85.
 4. Davis E, Sloan T, Aurelius K, Barbour A, Bodey E, Clark B, Dennis C, Drown R, Fleming M, Humbert A, Glasgo E, Kerns T, Lingro K, McMillin M, Meyer A, Pope B, Stalevicz A, Steffen B, Steindl A, Williams C, Wimberly C, Zenas R, Butela K, Wildschutte H. 2017. Antibiotic discovery throughout the small world initiative: a molecular strategy to identify biosynthetic gene clusters involved in antagonistic activity. *MicrobiologyOpen* 6:e00435.
 5. Zhang G, Annan RS, Carr SA, Neubert TA. 2010. Overview of peptide and protein analysis by mass spectrometry. *Curr Protoc Protein Sci Chapter16;Unit16.1:10–21*.
 6. Mageeney C, Pope WH, Harrison M, Moran D, Cross T, Jacobs-Sera D, Hendrix RW, Dunbar D, Hatfull GF. 2012. Mycobacteriophage Marvin: a new singleton phage with an unusual genome organization. *J Virol* 86:4762–4775.
 7. Grose JH, Belnap DM, Jensen JD, Mathis AD, Prince JT, Merrill BD, Burnett SH, Breakwell DP. 2014. The genomes, proteomes, and structures of three novel phages that infect the *Bacillus cereus* group and carry putative virulence factors. *J Virol* 88:11846–11860.
 8. Pope WH, Jacobs-Sera D, Russell DA, Rubin DH, Kajee A, Msibi ZNP, Larsen MH, Jacobs WR Jr, Lawrence JG, Hendrix RW, Hatfull GF. 2014. Genomics and proteomics of mycobacteriophage Patience, an accidental tourist in the *Mycobacterium* neighborhood. *mBio* 5:e02145-14.
 9. Searle BC. 2010. Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics* 10:1265–1269.
 10. Stock NL, March RE. 2014. Hands-on electrospray ionization-mass spectrometry for upper-level undergraduate and graduate students. *J Chem Educ* 91:1244–1247.
 11. Bedard L, Boyd A, Dyer N, Golay Z, Smith-Kinnaman W, Alakhras N, Mosley AL. 2017. Undergraduate student research in quantitative analysis of transcription elongation perturbation networks using mass spectrometry. *FASEB J* 31(1 Suppl):752–758.
 12. Kappler U, Rowland SL, Pedwell RK. 2017. A unique large-scale undergraduate research experience in molecular systems biology for non-mathematics majors. *Biochem Mol Biol Educ* 45:235–248.