

University of Montana

## ScholarWorks at University of Montana

---

Biological Sciences Faculty Publications

Biological Sciences

---

2-4-2014

### A Broadly Implementable Research Course in Phage Discovery and Genomics for First-Year Undergraduate Students

Tuajuanda C. Jordan  
*Lewis and Clark College*

Sandra H. Burnett  
*Brigham Young University - Utah*

Susan Carson  
*North Carolina State University at Raleigh*

Steven M. Caruso  
*University of Maryland - Baltimore County*

Karl Clase  
*Purdue University*

See this page for additional works at [https://scholarworks.umt.edu/biosci\\_pubs](https://scholarworks.umt.edu/biosci_pubs)



Part of the [Biology Commons](#), and the [Education Commons](#)

## Let us know how access to this document benefits you.

---

#### Recommended Citation

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(1):e01051-13. doi:10.1128/mBio.01051-13.

This Article is brought to you for free and open access by the Biological Sciences at ScholarWorks at University of Montana. It has been accepted for inclusion in Biological Sciences Faculty Publications by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact [scholarworks@mso.umt.edu](mailto:scholarworks@mso.umt.edu).

---

## Authors

Tuajuanda C. Jordan, Sandra H. Burnett, Susan Carson, Steven M. Caruso, Karl Clase, Randall J. DeJong, John J. Dennehy, Dee R. Denver, David Dunbar, Sarah C. R. Elgin, Ann M. Findley, Chris R. Gissendanner, Urszula P. Golebiewska, Nancy Guild, Grant A. Hartzog, Wendy H. Grillo, Gail P. Hollowell, Lee E. Hughes, Allison Johnson, Rodney A. King, Lynn O. Lewis t, Wei Li, Frank Rosenzweig, Michael R. Rubin, Margaret S. Saha, James Sandoz, Christopher D. Shaffer, Barbara Taylor, Louise Temple, Edwin Vazquez, Vassie C. Ware, Lucia P. Barker, Kevin W. Bradley, Deborah Jacobs-Sera, Welkin H. Pope, Daniel A. Russell, Steven G. Cresawn, David Lopatto, Cheryl P. Bailey, and Graham F. Hatfull

## RESEARCH ARTICLE

# A Broadly Implementable Research Course in Phage Discovery and Genomics for First-Year Undergraduate Students

Tuajuanda C. Jordan,<sup>a</sup> Sandra H. Burnett,<sup>b</sup> Susan Carson,<sup>c</sup> Steven M. Caruso,<sup>d</sup> Kari Clase,<sup>e</sup> Randall J. DeJong,<sup>f</sup> John J. Dennehy,<sup>g</sup> Dee R. Denver,<sup>h</sup> David Dunbar,<sup>i</sup> Sarah C. R. Elgin,<sup>j</sup> Ann M. Findley,<sup>k</sup> Chris R. Gissendanner,<sup>l</sup> Urszula P. Golebiewska,<sup>m</sup> Nancy Guild,<sup>n</sup> Grant A. Hartzog,<sup>o</sup> Wendy H. Grillo,<sup>p</sup> Gail P. Hollowell,<sup>p</sup> Lee E. Hughes,<sup>q</sup> Allison Johnson,<sup>r</sup> Rodney A. King,<sup>s</sup> Lynn O. Lewis,<sup>t</sup> Wei Li,<sup>u</sup> Frank Rosenzweig,<sup>v</sup> Michael R. Rubin,<sup>w</sup> Margaret S. Saha,<sup>x</sup> James Sandoz,<sup>d</sup> Christopher D. Shaffer,<sup>j</sup> Barbara Taylor,<sup>h</sup> Louise Temple,<sup>y</sup> Edwin Vazquez,<sup>w</sup> Vassie C. Ware,<sup>z</sup> Lucia P. Barker,<sup>aa</sup> Kevin W. Bradley,<sup>aa</sup> Deborah Jacobs-Sera,<sup>bb</sup> Welkin H. Pope,<sup>bb</sup> Daniel A. Russell,<sup>bb</sup> Steven G. Cresawn,<sup>cc</sup> David Lopatto,<sup>dd</sup> Cheryl P. Bailey,<sup>aa</sup> Graham F. Hatfull<sup>bb</sup>

College of Arts and Sciences, Lewis and Clark College, Portland, Oregon, USA<sup>a</sup>; Department of Microbiology and Molecular Biology, Brigham Young University, Provo, Utah, USA<sup>b</sup>; Biotechnology Program and Department of Plant and Microbial Biology, North Carolina State University, Raleigh, North Carolina, USA<sup>c</sup>; Department of Biological Sciences, University of Maryland, Baltimore County, Baltimore, Maryland, USA<sup>d</sup>; Technology Leadership and Innovation, Agricultural and Biological Engineering, Purdue University, West Lafayette, Indiana, USA<sup>e</sup>; Department of Biology, Calvin College, Grand Rapids, Michigan, USA<sup>f</sup>; Department of Biology, Queens College and the Graduate Center of the City University of New York, Flushing, New York, USA<sup>g</sup>; Department of Zoology, Oregon State University, Corvallis, Oregon, USA<sup>h</sup>; Department of Biology, Cabrini College, Radnor, Pennsylvania, USA<sup>i</sup>; Department of Biology, Washington University in St. Louis, St. Louis, Missouri, USA<sup>j</sup>; Department of Biology, University of Louisiana at Monroe, Monroe, Louisiana, USA<sup>k</sup>; Department of Basic Pharmaceutical Sciences, University of Louisiana at Monroe, Monroe, Louisiana, USA<sup>l</sup>; Department of Biology Sciences and Geology, Queensborough Community College, City University of New York, New York, USA<sup>m</sup>; Department of Molecular, Cellular and Developmental Biology, University of Colorado at Boulder, Boulder, Colorado, USA<sup>n</sup>; Department of Molecular, Cell and Developmental Biology, University of California, Santa Cruz, California, USA<sup>o</sup>; Department of Biology, North Carolina Central University, Durham, North Carolina, USA<sup>p</sup>; Department of Biological Sciences, University of North Texas, Denton, Texas, USA<sup>q</sup>; Center for the Study of Biological Complexity, Virginia Commonwealth University, Richmond, Virginia, USA<sup>r</sup>; Department of Biology, Western Kentucky University, Bowling Green, Kentucky, USA<sup>s</sup>; Department of Biological Sciences, University of Mary Washington, Fredericksburg, Virginia, USA<sup>t</sup>; Master of Physician Assistant Studies Program, Indiana University, Indianapolis, Indiana, USA<sup>u</sup>; Department of Biological Sciences, University of Montana, Missoula, Montana, USA<sup>v</sup>; Department of Biology, University of Puerto Rico at Cayey, Cayey, Puerto Rico<sup>w</sup>; Department of Biology, College of William and Mary, Williamsburg, Virginia, USA<sup>x</sup>; Department of Integrated Science and Technology, James Madison University, Harrisonburg, Virginia, USA<sup>y</sup>; Department of Biological Sciences, Lehigh University, Bethlehem, Pennsylvania, USA<sup>z</sup>; Undergraduate and Graduate Science Education, Howard Hughes Medical Institute, Chevy Chase, Maryland, USA<sup>aa</sup>; Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, USA<sup>bb</sup>; Department of Biology, James Madison University, Harrisonburg, Virginia, USA<sup>cc</sup>; Department of Psychology, Grinnell College, Grinnell, Iowa, USA<sup>dd</sup>

**ABSTRACT** Engaging large numbers of undergraduates in authentic scientific discovery is desirable but difficult to achieve. We have developed a general model in which faculty and teaching assistants from diverse academic institutions are trained to teach a research course for first-year undergraduate students focused on bacteriophage discovery and genomics. The course is situated within a broader scientific context aimed at understanding viral diversity, such that faculty and students are collaborators with established researchers in the field. The Howard Hughes Medical Institute (HHMI) Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) course has been widely implemented and has been taken by over 4,800 students at 73 institutions. We show here that this alliance-sourced model not only substantially advances the field of phage genomics but also stimulates students' interest in science, positively influences academic achievement, and enhances persistence in science, technology, engineering, and mathematics (STEM) disciplines. Broad application of this model by integrating other research areas with large numbers of early-career undergraduate students has the potential to be transformative in science education and research training.

**IMPORTANCE** Engagement of undergraduate students in scientific research at early stages in their careers presents an opportunity to excite students about science, technology, engineering, and mathematics (STEM) disciplines and promote continued interests in these areas. Many excellent course-based undergraduate research experiences have been developed, but scaling these to a broader impact with larger numbers of students is challenging. The Howard Hughes Medical Institute (HHMI) Science Education Alliance Phage Hunting Advancing Genomics and Evolutionary Science (SEA-PHAGES) program takes advantage of the huge size and diversity of the bacteriophage population to engage students in discovery of new viruses, genome annotation, and comparative genomics, with strong impacts on bacteriophage research, increased persistence in STEM fields, and student self-identification with learning gains, motivation, attitude, and career aspirations.

Received 6 December 2013 Accepted 12 December 2013 Published 4 February 2014

**Citation** 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(1):e01051-13. doi:10.1128/mBio.01051-13.

**Editor** Richard Losick, Harvard University

**Copyright** © 2014 Jordan et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution-Noncommercial-ShareAlike 3.0 Unported license](http://creativecommons.org/licenses/by-nc-sa/3.0/), which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

Address correspondence to Cheryl P. Bailey, [baileyc@hhmi.org](mailto:baileyc@hhmi.org), or Graham F. Hatfull, [gfh@pitt.edu](mailto:gfh@pitt.edu).

In 2012, the President's Council of Advisors on Science and Technology (PCAST) reported that there is a need for an additional one million science, technology, engineering, and mathematics (STEM) graduates in the United States over the next decade to meet U.S. economic demands (1). It was noted that even a modest increase in the persistence of STEM students in the first 2 years of their undergraduate education would alleviate much of this shortfall (1). Replacing conventional introductory laboratory courses with discovery-based research courses is a key recommendation that is expected to lead to enhanced retention. Providing authentic research experiences to undergraduate students and directing them toward careers in STEM is a priority of science education in the 21st century (1–4).

An abundance of evidence shows that involvement of undergraduate students in authentic research experiences has strong benefits for their engagement and interest in science (5–7) and that this often increases student interest in STEM careers (8). It is common for undergraduate students at research colleges and universities to participate in faculty-led research programs—especially during their last 2 years—with graduate students and postdoctoral researchers participating in their mentorship (9). Research experiences promote college retention (10), but the capacity for high-quality mentored undergraduate research within faculty research programs is limited, and this route is unlikely alone to satisfy the economic demands of the coming decade. There have been many successful efforts to develop classroom undergraduate research experiences (11–14; see also <http://www.sciencemag.org/site/special/ibi/> and <http://www.curen.net.org/>), but identifying authentic research experiences that scale to larger numbers of undergraduate students often proves elusive (4). Bioinformatic approaches engaging substantial numbers of students at diverse institutions have been described (15, 16) and are successful in providing research experiences (14) but do not include a wet-bench laboratory component.

Taking advantage of research infrastructures at research-intensive institutions to advance missions in undergraduate education is desirable, and community-oriented approaches have been developed (17, 18), although the potential is largely untapped. Some research projects are likely to be more suitable for undergraduate involvement than others, and identifying those both rich in discovery and accessible to early-career students is challenging (19). The Phage Hunters Integrating Research and Education (PHIRE) program, in which undergraduate and high school students isolate novel bacteriophages, sequence their genomes, annotate them, and analyze them from a comparative genomics perspective, is one response to this challenge (19–21). The approach takes advantage of the large, dynamic, old, and highly genetically diverse nature of the bacteriophage population (22, 23). Moreover, although phages play key roles in bacterial pathogenesis (24) and the global climate and ecology (25), we know remarkably little about them outside a few well-studied prototypes.

Phages can be easily isolated from the environment, and their relatively small genomes (40 to 150 kbp) are readily sequenced and annotated (26). Phage isolation requires little prior expert knowledge or technical skill, providing an accessible entry point for students from all backgrounds to engage in inquiry-based science (21). Each isolated phage is new, students can name their own phage, and a sense of ownership in their discovery helps to motivate them to explore the secrets of their phage by isolating

genomic DNA, determining its sequence, annotating gene predictions, and comparing the sequence to that of other known viruses (21). This programmatic transition from a broadly accessible and concrete introduction to sophisticated genomic analysis provides a rich and structured education platform (27), applicable to STEM and non-STEM students, including first-year undergraduates (28–30).

To investigate whether the PHIRE approach can be extended to environments beyond the expert phage-focused research laboratory, the Howard Hughes Medical Institute (HHMI), the University of Pittsburgh, and James Madison University investigated a framework enabling broad usage at diverse institutions, involving large numbers of undergraduate students and nonexpert instructors, and assessed its impact. The approach proved to be scalable (4,800 students at 73 schools over 5 years), it was implementable at research-intensive and research-poor institutions, generated gains in phage biology research, and enhanced student retention, and the student-reported gains were equivalent to those from an intense summer research experience.

## RESULTS

The attributes of the PHIRE program at the University of Pittsburgh demonstrate that phage discovery and genomics are a platform that supports engagement of students in authentic research without requiring prior mastery of anything other than very basic concepts and content material (21). We therefore examined whether this could be broadly implemented at institutions with a wide spectrum of missions and demographics, without a requirement for resident expertise in bacteriophage biology. Our core hypothesis was that student participation in this research would generate new insights into phage diversity and evolution while simultaneously elevating student engagement in science, stimulating overall academic performance, and encouraging persistence in STEM fields. Below, we report the structure of the HHMI Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) course and its impacts on both research advances and student learning.

**The SEA-PHAGES course.** The SEA-PHAGES course (formerly called the National Genomics Research Initiative) is a year-long research experience targeted at beginning college students. Classes typically enroll 18 to 24 students and are taught by one or two faculty members together with a student teaching assistant. In the first term, students isolate phages from locally collected soil samples using *Mycobacterium smegmatis* as the primary bacterial host, a nonpathogenic strain relevant to understanding *Mycobacterium tuberculosis*. Students purify and characterize their phages, visualize them with electron microscopy, and extract and purify the DNA. The genome of one phage isolate is sequenced between terms, and in the second term, students annotate the genome using bioinformatics tools to define putative genes, understand genomic arrangements, and predict protein functions. Sequence and annotation quality is expertly reviewed and collated on the PhagesDB database (<http://www.phagesdb.org>) and submitted to GenBank. The Phamerator program (31) is used to explore genome relationships, and all phage samples are archived for use by the research community.

The SEA-PHAGES course curriculum aims to introduce students to research methods and approaches, experimental design, and data interpretation but does not seek to instruct students in content matter outside the immediate biological context. But, as

**TABLE 1** Diversity of institutions participating in SEA-PHAGES

Carnegie classification <sup>a</sup>	No. of schools
Research universities; very high/high research activity	30
Master's degree-granting colleges and universities	18
Baccalaureate colleges	22
Associate's degree-granting colleges	3

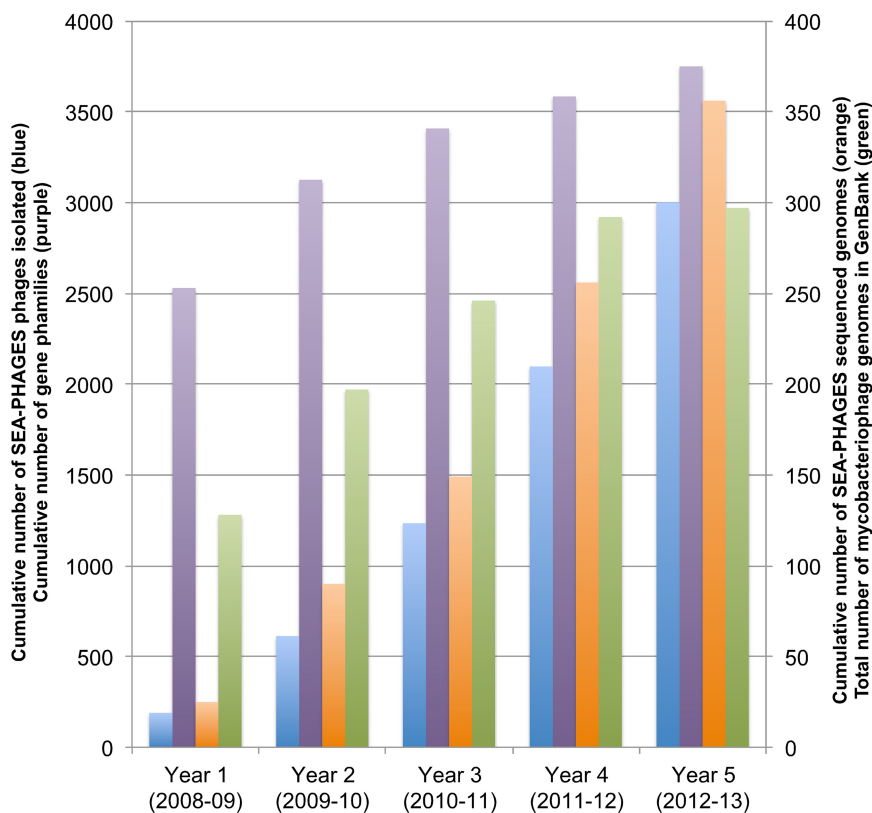
<sup>a</sup> Schools offering the SEA-PHAGES course are organized according to their classification by the Carnegie Foundation for the Advancement of Teaching (2010).

students are direct participants in scientific discovery, the goal is to engage, excite, increase the confidence of, and draw students into a cycle of self-motivation. If successful, we predicted that this would translate into enhanced performance in other STEM classes, greater retention within STEM training, and an increase in the numbers of students seeking continued research experiences beyond their freshman year.

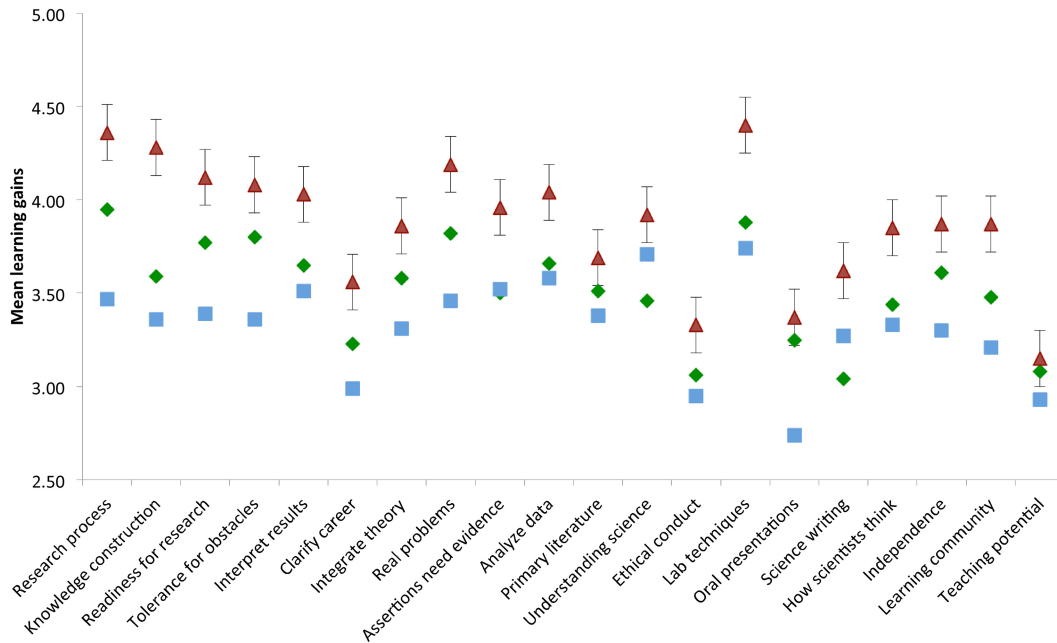
Program faculty and teaching assistants are trained at two weeklong workshops, one for each term of the course. Detailed manuals are provided, and community discussions are facilitated by a wiki site. Students and faculty present their findings at an annual SEA-PHAGES Research Symposium, at regional and national meetings, and through peer-reviewed publications. In the 5 years of the program, more than 4,800 students have participated (1,800 in 2012–2013), including STEM majors, non-STEM majors, honors students, and “typical” students. The number of participating schools has grown to more than 70 institutions (see Table S1 in the supplemental material), ranging from community

colleges to research universities (Table 1). As can be seen from these program design features, the educational model of the SEA-PHAGES program integrates course-based learning within a framework of scientific activity, including a real-world scientific research agenda, professional networking, and scientific dissemination of results. In this way, the cost-effectiveness of course-based learning is combined with professional science with mutual benefits.

**Gains in understanding viral diversity.** The contributions of the SEA-PHAGES students have been essential to our current understanding of the diversity of mycobacteriophages, demonstrating the substantial impact of the distributed approach compared to what would be accomplished by a single laboratory, and have resulted in several publications with student authors (29, 31–39). Since the start of the program in 2008, SEA-PHAGES students have isolated 3,000 new phages (with global positioning system [GPS] coordinates recorded) and characterized their phages by DNA restriction analysis and electron microscopy. More than 450 mycobacteriophage genomes have been sequenced and annotated, and more than 350 sequences have been deposited in GenBank (Fig. 1). These genomes include many distinctly different types and numerous complex variants (40), and the entire genome collection codes for over 48,000 genes representing 3,780 sequence phamilies (a group of proteins sharing similarity to at least one other above threshold BlastP and Clustal values [31]). Correlations between genome and geography or time of isolation have been explored (35, 41), as well as the evolutionary mechanisms



**FIG 1** SEA-PHAGES students contribute to scientific knowledge. Results are from the first 5 years of the SEA-PHAGES program isolating new phages, showing the cumulative numbers of phages isolated (blue), cumulative numbers of genomes sequenced (orange), cumulative numbers of gene phamilies (purple), and total numbers of mycobacteriophages in GenBank (green). Not all genomes sequenced and annotated in year 5 are yet available in GenBank.



**FIG 2** Student evaluation of learning gains. Mean learning gains for common survey items on the SURE (green diamonds), CURE (blue squares), and the SEA-PHAGES (red triangles) assessment instruments are shown. The SURE survey data represent 2,358 students who completed summer research in 2009; the CURE survey data represent 476 students who were enrolled in science courses that were described by their instructors as without a research element (data collected for fall 2007 through spring 2009); the SEA-PHAGES data represent 121 students who evaluated their course following the academic year 2008–2009. Error bars represent 2 standard errors around the mean.

contributing to the pervasive genome mosaicism (33). The genomes contain numerous examples of biological intrigue, including novel inteins, introns, mobile elements, immunity systems, and regulatory schemes (33–35, 42–45), as well as potential for developing new tools for understanding tuberculosis (46–49).

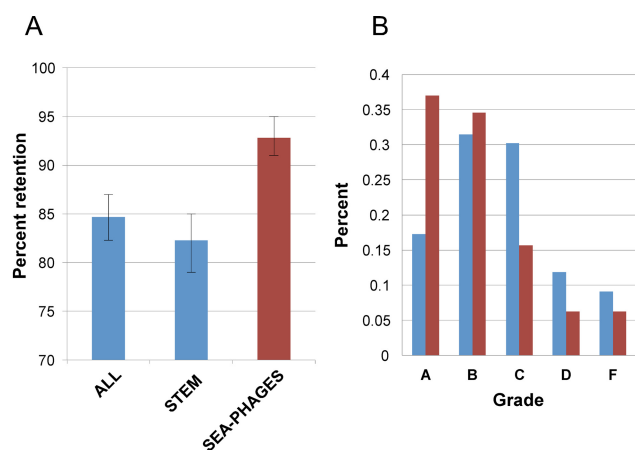
The diversity of phages known to infect a single common host is remarkable; there are many thousands of potential bacterial hosts for phage isolation, and host range studies suggest that simply using a different strain of the same bacterial species will result in distinct profiles of diversity (38). With an estimated  $10^{31}$  phage particles in the biosphere and a population that turns over every few days (23), there is an inexhaustible reservoir for discovery.

**Impacts on student education and retention.** The Survey of Undergraduate Research Experience (SURE) and the Classroom Undergraduate Research Experience (CURE) measure the students' assessment of their understanding of science and scientists, confidence in their ability to perform research, and their perceived gains in skills (50). The self-perceptions of learning gains, motivation and attitude, and career aspirations of the SEA-PHAGES course participants were assessed with pre- and postcourse SURE-like surveys (see Fig. S1 in the supplemental material). Twenty of the SEA-PHAGES survey items are shared with the regular SURE and CURE surveys, allowing the comparison of the SEA-PHAGES students' learning gains with those of students who engaged in a dedicated summer research experience (SURE) and students who completed traditional science courses with no research element (CURE) (Fig. 2). The SEA-PHAGES students scored as well as or better on all 20 learning gains compared to the SURE students, reflecting benefits at least equivalent to those accrued through a summer-long apprentice-based undergraduate research experience. The increase in scientific self-efficacy reported by the SEA-

PHAGES students is likely to be directly related to their retention in science (51).

To analyze the effect of the SEA-PHAGES course on student persistence, we compared retention of students enrolled in the SEA-PHAGES course (77% first-year students and 95% STEM majors) with two benchmark statistics: the retention of all students and the retention of STEM majors with the same number of years of college experience and enrolled at the same school (Fig. 3A), important parameters given the typical rates for student attrition between first- and second-year STEM undergraduates (52). Data were from 27 comparisons from 20 institutions and show clearly that SEA-PHAGES students matriculated into the second year at significantly higher rates than did either benchmark group. Thus, early engagement in a research experience improves student retention into the second year. The positive impacts of this course-based research experience are similar to what has been reported for apprentice-based research experiences (5, 53), represent an effective response to the call to action in the National Science Foundation (NSF) *Vision and Change* and PCAST reports (1, 4), and provide validation for this educational model on a larger scale.

Anticipating that research-stimulated motivation will influence student performance in other courses, we selected six schools that substituted the SEA-PHAGES course for a regular biology laboratory and compared the grades of participating students in the accompanying biology lecture course (Fig. 3B). We limited this analysis to schools that enrolled "typical" students into the PHAGES lab sections rather than those aimed at honors students or students at academic risk. The biology lecture course grades of SEA-PHAGES students were compared directly to those of peers enrolled in the same lecture course but in the regular biology



**FIG 3** (A) Retention of SEA-PHAGES participants (red) compared to other students at the same institution (blue), year 1 to year 2 of their college experience. Retention data were gathered from 20 institutions, with some institutions contributing data from multiple years, resulting in 27 sets of comparison data. Retention data were analyzed with a between-group analysis of variance with 3 levels of the independent variable (all majors, STEM majors, and SEA-PHAGES students) for 171 reports. The result was interpreted as significant at the 0.05 level. (B) SEA-PHAGES students (red) perform better than peers (blue) in traditional laboratory sections in the introductory lecture course. Results are for 127 SEA-PHAGES students and 1,120 students in the traditional laboratory course from six institutions. In the lecture course, SEA-PHAGES students averaged 2.95 on a 4.0 scale, compared to the 2.58 average of students in traditional lab sections. This difference was significant ( $t = 2.64$ ;  $P < 0.05$ ).

laboratory. As is the case with most applied research, students were not randomly assigned to conditions, and even among these “typical” students, there may have been some self-selection for registration in the SEA-PHAGES course. We observed substantial differences in both the average grades and the grade distribution of SEA-PHAGES students relative to those of students in traditional lab sections (Fig. 3B), and although these data are preliminary and warrant further study, they suggest that there could be broad educational benefits to the SEA-PHAGES experience. Because of the concern that SEA-PHAGES students might suffer from lack of exposure to a broader coverage of subject matter in the regular laboratory course, we developed a 25-item pre- and postcourse survey of biological concepts (see Fig. S2 in the supplemental material) which was administered to students before and after the laboratory courses. There was no significant difference in performances on the test between SEA-PHAGES students and the comparison group of students (see Fig. S3). Both groups improved from pretest to posttest, and there was no significant difference between the groups in terms of the extent of their improvement. The lack of exposure to additional topics in the SEA-PHAGES course thus had no obvious detrimental effect.

## DISCUSSION

The HHMI SEA-PHAGES program provides a general model for accomplishing improvements in the persistence of students in science by transforming a small-scale scientific inquiry into a cross-institution education platform that engages first-year students. The outcomes are consistent and robust, benefitting diverse groups of students across a variety of institutions. The materials costs are similar to those of other inquiry-based courses, and many institutions have implemented the course without external

support, other than assistance with sequencing costs and programmatic and scientific support from HHMI and the University of Pittsburgh (some schools received direct external support for materials during their first 3 years in the program). The size and diversity of the phage population provide an inexhaustible wealth of biological novelty that imposes no obvious limit on the number of students who can participate. Future opportunities include further broadening the implementation of the SEA-PHAGES course as well as extending the model to development of similar projects in which scientific discovery, project ownership, and simple entry points can be implemented at the first-year college level. Meeting these opportunities will lead to a broad and sustainable enhancement of undergraduate science education, an advancement of scientific knowledge, and an increase of student persistence in science.

## MATERIALS AND METHODS

**Participants.** The study was conducted with SEA-PHAGES faculty and students in the United States and the Commonwealth of Puerto Rico. David Lopatto and participant institutions obtained appropriate institutional review board (IRB) approval. SEA-PHAGES faculty are trained in a weeklong workshop focusing on *in situ* procedures and pedagogy in preparation for the fall semester and a weeklong workshop focusing on *in silico* bioinformatics tools in preparation for the spring semester. Faculty and students are invited to a SEA-PHAGES National Symposium to present their scientific findings. The SEA office conducts annual site visits and provides continuous technical support for institutions year-round. The SEA Wiki maintains an up-to-date depository for announcements, communication forums for faculty and students, curriculum resources, instructional materials, and research archives. SEA-PHAGES faculty members recruited comparison group students on a volunteer basis to enhance the validity of statistical analysis. The comparison group students were recruited among students taking introductory laboratory courses. Except for the student grade analysis, comparison group students cannot be matched to SEA-PHAGES students on each campus, so statistical analysis was limited to quasiexperimental analysis based on a nonequivalent comparison group. Systemic Research sent out invitations to all consenting students’ e-mail addresses individually.

**Analysis.** During academic year 2009–2010, different aspects of the SEA-PHAGES and comparison group were measured. White/Caucasian students made up the majority of each group, 66% of SEA-PHAGES students and 76% of comparison group students. The majority of both groups lived in suburban communities (66% SEA-PHAGES and 64% comparison group students), attended public high schools (83% SEA-PHAGES and 83% comparison group students), and were in their first year in college (SEA-PHAGES, 77% first-year students, 18% sophomores; comparison group, 70% first-year students, 20% sophomores). There were a higher percentage of male students in the SEA-PHAGES course (38%) than in the comparison group (29%), but in both groups, female students were the clear majority.

**Retention rates.** The Institutional Annual Survey measures student retention rates by tracking full-time, first-time entering students who are seeking bachelor’s degrees. The Institutional Annual Survey was conducted among institutions during November to December. Retention rates were calculated for students returning in fall 2008 and fall 2009. An analysis of variance was performed over 3 groups (all majors, STEM majors, and SEA-PHAGES students). The data were reported by institution and category, including 63 reports for all majors, 43 reports for STEM majors, and 65 reports for SEA-PHAGES students.

**The SEA CURE survey.** The Classroom Undergraduate Research Experience (CURE) survey was specially adapted to the SEA-PHAGES program by David Lopatto (Grinnell College, Grinnell, IA). The CURE survey consists of multiple sections, including institution, class, demographics, science-related activities, major and minor concentration, post-

graduate academic goals, experiences in laboratory course elements, experience in research, engagement in activities or endeavors, course benefit, learning experience in laboratory experiments and tools, overall course evaluation, and opinions about science. Systemic Research added a few questions to the postcourse CURE survey to collect data regarding students' SEA-PHAGES course satisfaction, SEA Wiki access and utilization, SEA-PHAGES research paper and presentation experience, and general comments. The survey was administered twice a year: the presurvey at the beginning of the fall semester and the postsurvey at the end of the spring semester. As with the Biological Concepts Survey (BCS), Systemic Research developed the online survey forms using the Vovici EFM Community Professional website. The pre- and postcourse survey invitations were e-mailed to individual students according to their academic calendars. Using Vovici's survey follow-up feature, three reminder e-mails were sent after the initial invitations. The collected survey responses were securely saved in a dedicated Vovici HHMI website and Systemic Research's NGRI student database. The SURE survey data represent 2,358 students who completed summer research in 2009; the CURE survey data represent 476 students who evaluated science courses that were described by their instructors as without a research element (data collected fall 2007 through spring 2009); the SEA-PHAGES data represent 121 students who evaluated their course following the academic year 2008–2009. Mean learning gains were calculated for each category of the 20 items common to both the CURE and SURE surveys.

**Grades.** Eleven institutions submitted their SEA-PHAGES students' laboratory and introductory biology course performance data for fall 2008 and spring 2009 in the academic year 2008–2009 and fall 2009 and spring 2010 in the academic year 2009–2010. Letter grade distributions for both SEA-PHAGES and comparison students were collected. Six institutions had matched data that were utilized in the analysis, with 127 SEA-PHAGES and 1,120 comparison student grades. For statistical analysis, the letter grades were assigned numerical values from 4 (grade A) to 0 (grade F). *t* tests were performed comparing the mean grades received by SEA-PHAGES students and comparison group students in the biology lecture course.

**Biological methods.** Mycobacteriophage isolation was performed using *Mycobacterium smegmatis* mc<sup>2</sup>155 as a host, and phages were identified as PFU either by direct plating on bacterial lawns or after enrichment in the presence of *M. smegmatis*. Following purification and amplification, DNA was isolated and sequenced using Sanger, 454, or Ion Torrent technologies, using a shotgun approach followed by targeted sequencing to validate ambiguities and determine genome ends. Genome annotations were performed using various software platforms, including GBrowse (54), Apollo (55), DNAMaster (<http://cobamide2.bio.pitt.edu/>), Glimmer (56), GeneMark (57), and analysis programs available at the National Center for Biotechnology Information (NCBI). Comparative genomics used Phamerator (31) and Gepard (58). Assembled genome sequences and genome annotations were subjected to expert review prior to submission to GenBank. Detailed methods for phage isolation, sequencing, and analysis are available on PhagesDB (<http://phagesdb.org>).

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.01051-13/-/DCSupplemental>.

- Figure S1, PDF file, 0.2 MB.
- Figure S2, PDF file, 0.3 MB.
- Figure S3, PDF file, 0.1 MB.
- Table S1, PDF file, 0.1 MB.

## ACKNOWLEDGMENTS

We thank Peter Bruns, Tom Cech, and Malcolm Campbell for their vision, advice, and support in the development of the Science Education Alliance; Matte Conte and Ed Lee for the SEA workflow; Jeffrey Lawrence for development of DNAMaster for genome annotation; Melvina Lewis for professional editing and design of the SEA manuals and programs; David Hanauer and David Asai for critical comments on the manuscript;

and William Sandy and Nelly Gregorian of Fluentresearch for their helpful discussions on how to assess the pedagogical impact of this experiment. The first 3 years (2008 to 2011) of data were collected and analyzed by Systemic Research, Inc., Mansfield, MA.

## REFERENCES

1. President's Council on Science and Technology. 2012. Report to the President. Engage to excel: producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. Executive Office of the President, Washington, DC.
2. National Academy of Sciences, National Academy of Engineering, and Institute of Medicine of the National Academies. 2007. Rising above the gathering storm: energizing and employing America for a brighter economic future. National Academies Press, Washington, DC.
3. National Research Council of the National Academies. 2003. BIO2010: transforming undergraduate education for future research biologists. National Academies Press, Washington, DC.
4. American Association for the Advancement of Science. 2009. Vision and change in undergraduate biology education: a call to action. American Association for the Advancement of Science, Washington, DC.
5. Seymour E, Hunter A-B, Laursen SL, DeAntoni T. 2004. Establishing the benefits of undergraduate research for undergraduates in the sciences: first findings from a three-year study. *Sci. Educ.* 88:493–594. <http://dx.doi.org/10.1002/sce.10131>.
6. Eagan MK, Hurtado S, Chang MJ, Garcia GA, Herrera FA, Garibay JC. 2013. Making a difference in science education: the impact of undergraduate research programs. *Am. Educ. Res. J.* 50:683–713. <http://dx.doi.org/10.3102/0002831213482038>.
7. Lopatto D. 2007. Undergraduate research experiences support science career decisions and active learning. *CBE Life Sci. Educ.* 6:297–306. <http://dx.doi.org/10.1187/cbe.07-06-0039>.
8. Russell SH, Hancock MP, McCullough J. 2007. The pipeline. Benefits of undergraduate research experiences. *Science* 316:548–549. <http://dx.doi.org/10.1126/science.1140384>.
9. Prunuske AJ, Wilson J, Walls M, Clarke B. 2013. Experiences of mentors training underrepresented undergraduates in the research laboratory. *CBE Life Sci. Educ.* 12:403–409. <http://dx.doi.org/10.1187/cbe.13-02-0043>.
10. Nagda BA, Gregerman SR, Lerner JS, von Hippel W, Jonides J. 1998. Undergraduate student-faculty research partnerships affect student retention. *Rev. Higher Ed.* 22:55.
11. Singer SR, Schwarz JA, Manduca CA, Fox SP, Iverson ER, Taylor BJ, Cannon SB, May GD, Maki SL, Farmer AD, Doyle JJ. 2013. IBI series winner. Keeping an eye on biology. *Science* 339:408–409. <http://dx.doi.org/10.1126/science.1229848>.
12. Chen J, Call GB, Beyer E, Bui C, Cespedes A, Chan A, Chan J, Chan S, Chhabra A, Dang P, Deravanesian A, Hermogeno B, Jen J, Kim E, Lee E, Lewis G, Marshall J, Regalia K, Shadpour F, Shemmassian A, Spivey K, Wells M, Wu J, Yamauchi Y, Yavari A, Abrams A, Abramson A, Amado L, Anderson J, Bashour K, Bibikova E, Bookatz A, Brewer S, Bui N, Calvillo S, Cao J, Chang A, Chang D, Chang Y, Chen Y, Choi J, Chou J, Datta S, Davarifar A, Desai P, Fabrikant J, Farnad S, Fu K, et al. 2005. Discovery-based science education: functional genomic dissection in *Drosophila* by undergraduate researchers. *PLoS Biol.* 3:e59. <http://dx.doi.org/10.1371/journal.pbio.0030059>.
13. Bascom-Slack CA, Arnold AE, Strobel SA. 2012. IBI series winner. Student-directed discovery of the plant microbiome and its products. *Science* 338:485–486. <http://dx.doi.org/10.1126/science.1215227>.
14. Lopatto D, Alvarez C, Barnard D, Chandrasekaran C, Chung HM, Du C, Eckdahl T, Goodman AL, Hauser C, Jones CJ, Kopp OR, Kuleck GA, McNeil G, Morris R, Myka JL, Nagengast A, Overvoorde PJ, Poet JL, Reed K, Regisford G, Revie D, Rosenwald A, Saville K, Shaw M, Skuse GR, Smith C, Smith M, Spratt M, Stamm J, Thompson JS, Wilson BA, Witkowski C, Youngblom J, Leung W, Shaffer CD, Buhler J, Mardis E, Elgin SC. 2008. Undergraduate research. Genomics education partnership. *Science* 322:684–685. <http://dx.doi.org/10.1126/science.1165351>.
15. Goodner B, Wheeler C, Hall P, Slater S. 2003. Massively parallel undergraduates for bacterial genome finishing. *ASM News* 69:584–585.
16. Shaffer CD, Alvarez C, Bailey C, Barnard D, Bhalla S, Chandrasekaran C, Chandrasekaran V, Chung HM, Dorer DR, Du C, Eckdahl TT, Poet JL, Frohlich D, Goodman AL, Gosser Y, Hauser C, Hoopes LL, Johnson D, Jones CJ, Kaehler M, Kokan N, Kopp OR, Kuleck GA, McNeil G,



- Moss R, Myka JL, Nagengast A, Morris R, Overvoorde PJ, Shoop E, Parrish S, Reed K, Regisford EG, Revie D, Rosenwald AG, Saville K, Schroeder S, Shaw M, Skuse G, Smith C, Smith M, Spana EP, Spratt M, Stamm J, Thompson JS, Wawersik M, Wilson BA, Youngblom J, Leung W, Buhler J, Mardis ER, Lopatto D, Elgin SC. 2010. The genomics education partnership: successful integration of research into laboratory classes at a diverse group of undergraduate institutions. *CBE Life Sci. Educ.* 9:55–69. <http://dx.doi.org/10.1187/09-11-0087>.
17. Desai KV, Gatson SN, Stiles TW, Stewart RH, Laine GA, Quick CM. 2008. Integrating research and education at research-intensive universities with research-intensive communities. *Adv. Physiol. Educ.* 32: 136–141. <http://dx.doi.org/10.1152/advan.90112.2008>.
  18. Balster N, Pfund C, Rediske R, Branchaw J. 2010. Entering research: a course that creates community and structure for beginning undergraduate researchers in the STEM disciplines. *CBE Life Sci. Educ.* 9:108–118. <http://dx.doi.org/10.1187/cbe.09-10-0073>.
  19. Hanauer DI, Jacobs-Sera D, Pedulla ML, Cresawn SG, Hendrix RW, Hatfull GF. 2006. Inquiry learning. Teaching scientific inquiry. *Science* 314:1880–1881. <http://dx.doi.org/10.1126/science.1136796>.
  20. Hanauer D, Hatfull GF, Jacobs-Sera D. 2009. Active assessment: assessing scientific inquiry. Springer Verlag, New York, NY.
  21. Hatfull GF, Pedulla ML, Jacobs-Sera D, Cichon PM, Foley A, Ford ME, Gonda RM, Houtz JM, Hryckowian AJ, Kelchner VA, Namburi S, Pajcini KV, Popovich MG, Schleicher DT, Simanek BZ, Smith AL, Zdanowicz GM, Kumar V, Peebles CL, Jacobs WR, Jr, Lawrence JG, Hendrix RW. 2006. Exploring the mycobacteriophage metaproteome: phage genomics as an educational platform. *PLoS Genet.* 2:e92. <http://dx.doi.org/10.1371/journal.pgen.0020092>.
  22. Hambly E, Suttle CA. 2005. The virosphere, diversity, and genetic exchange within phage communities. *Curr. Opin. Microbiol.* 8:444–450. <http://dx.doi.org/10.1016/j.mib.2005.06.005>.
  23. Hatfull GF, Hendrix RW. 2011. Bacteriophages and their genomes. *Curr. Opin. Virol.* 1:298–303. <http://dx.doi.org/10.1016/j.coviro.2011.06.009>.
  24. Casas V, Maloy S. 2011. Role of bacteriophage-encoded exotoxins in the evolution of bacterial pathogens. *Future Microbiol.* 6:1461–1473. <http://dx.doi.org/10.2217/fmb.11.124>.
  25. Danovaro R, Corinaldesi C, Dell'anno A, Fuhrman JA, Middelburg JJ, Noble RT, Suttle CA. 2011. Marine viruses and global climate change. *FEMS Microbiol. Rev.* 35:993–1034. <http://dx.doi.org/10.1111/j.1574-6976.2010.00258.x>.
  26. Mokili JL, Rohwer F, Dutilh BE. 2012. Metagenomics and future perspectives in virus discovery. *Curr. Opin. Virol.* 2:63–77. <http://dx.doi.org/10.1016/j.coviro.2011.12.004>.
  27. Hatfull GF. 2010. Bacteriophage research: gateway to learning science. *Microbe* 5:243–250.
  28. Harrison M, Dunbar D, Ratmanský L, Boyd K, Lopatto D. 2011. Classroom-based science research at the introductory level: changes in career choices and attitude. *CBE Life Sci. Educ.* 10:279–286. <http://dx.doi.org/10.1187/cbe.10-12-0151>.
  29. Caruso SM, Sandoz J, Kelsey J. 2009. Non-STEM undergraduates become enthusiastic phage-hunters. *CBE Life Sci. Educ.* 8:278–282. <http://dx.doi.org/10.1187/cbe.09-07-0052>.
  30. Temple L, Cresawn SG, Monroe JD. 2010. Genomics and bioinformatics in undergraduate curricula: contexts for hybrid laboratory/lecture courses for entering and advanced science students. *Biochem. Mol. Biol. Educ.* 38:23–28. <http://dx.doi.org/10.1002/bmb.20359>.
  31. Cresawn SG, Bogel M, Day N, Jacobs-Sera D, Hendrix RW, Hatfull GF. 2011. Phamerator: a bioinformatic tool for comparative bacteriophage genomics. *BMC Bioinformatics* 12:395. <http://dx.doi.org/10.1186/1471-2105-12-395>.
  32. Mageaney 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. <http://dx.doi.org/10.1128/JVI.00075-12>.
  33. Pope WH, Ferreira CM, Jacobs-Sera D, Benjamin RC, Davis AJ, DeJong RJ, Elgin SC, Guilfoile FR, Forsyth MH, Harris AD, Harvey SE, Hughes LE, Hynes PM, Jackson AS, Jatal MD, MacMurray EA, Manley CM, McDonough MJ, Mosier JL, Osterbann LJ, Rabinowitz HS, Rhyan CN, Russell DA, Saha MS, Shaffer CD, Simon SE, Sims EF, Tovar IG, Weisser EG, Wertz JT, Weston-Hafer KA, Williamson KE, Zhang B, Cresawn SG, Jain P, Piuri M, Jacobs WR, Jr, Hendrix RW, Hatfull GF. 2011. Cluster K mycobacteriophages: insights into the evolutionary origins of mycobacteriophage TM4. *PLoS One* 6:e26750. <http://dx.doi.org/10.1371/journal.pone.0026750>.
  34. Pope WH, Jacobs-Sera D, Best AA, Broussard GW, Connerly PL, Dedrick RM, Kremer TA, Offner S, Ogiefo AH, Pizzorno MC, Rockenbach K, Russell DA, Stowe EL, Stukey J, Thibault SA, Conway JF, Hendrix RW, Hatfull GF. 2013. Cluster J mycobacteriophages: intron splicing in capsid and tail genes. *PLoS One* 8:e69273. <http://dx.doi.org/10.1371/journal.pone.0069273>.
  35. Pope WH, Jacobs-Sera D, Russell DA, Peebles CL, Al-Atrache Z, Alcoser TA, Alexander LM, Alfano MB, Alford ST, Amy NE, Anderson MD, Anderson AG, Ang AA, Ares M, Jr, Barber AJ, Barker LP, Barrett JM, Barshop WD, Bauerle CM, Bayles IM, Belfield KL, Best AA, Borjon A, Jr, Bowman CA, Boyer CA, Bradley KW, Bradley VA, Broadway LN, Budwal K, Busby KN, Campbell IW, Campbell AM, Carey A, Caruso SM, Chew RD, Cockburn CL, Cohen LB, Corajod JM, Cresawn SG, Davis KR, Deng L, Denver DR, Dixon BR, Ekram S, Elgin SCR, Engelsen AE, English BEV, Erb ML, Estrada C, et al. 2011. Expanding the diversity of mycobacteriophages: insights into genome architecture and Evolution. *PLoS One* 6:e16329. <http://dx.doi.org/10.1371/journal.pone.0016329>.
  36. Hatfull GF, Jacobs-Sera D, Lawrence JG, Pope WH, Russell DA, Ko CC, Weber RJ, Patel MC, Germane KL, Edgar RH, Hoyte NN, Bowman CA, Tantoco AT, Paladin EC, Myers MS, Smith AL, Grace MS, Pham TT, O'Brien MB, Vogelsberger AM, Hryckowian AJ, Wynalek JL, Doniskeller H, Bogel MW, Peebles CL, Cresawn SG, Hendrix RW. 2010. Comparative genomic analysis of 60 mycobacteriophage genomes: genome clustering, gene acquisition, and gene size. *J. Mol. Biol.* 397: 119–143. <http://dx.doi.org/10.1016/j.jmb.2010.01.011>.
  37. Smith KC, Castro-Nallar E, Fisher JN, Breakwell DP, Grose JH, Burnett SH. 2013. Phage cluster relationships identified through single gene analysis. *BMC Genomics* 14:410. <http://dx.doi.org/10.1186/1471-2164-14-410>.
  38. Jacobs-Sera D, Marinelli LJ, Bowman C, Broussard GW, Guerrero Bustamante C, Boyle MM, Petrova ZO, Dedrick RM, Pope WH, Science Education Alliance Phage Hunters Advancing Genomics And Evolutionary Science Sea-Phages Program, Modlin RL, Hendrix RW, Hatfull GF. 2012. On the nature of mycobacteriophage diversity and host preference. *Virology* 434:187–201. <http://dx.doi.org/10.1016/j.virol.2012.09.026>.
  39. Lorenz L, Lins B, Barrett J, Montgomery A, Trapani S, Schindler A, Christie GE, Cresawn SG, Temple L. 2013. Genomic characterization of six novel *Bacillus pumilus* bacteriophages. *Virology* 444:374–383. <http://dx.doi.org/10.1016/j.virol.2013.07.004>.
  40. Hatfull GF, Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science Program, KwaZulu-Natal Research Institute for Tuberculosis and HIV Mycobacterial Genetics Course Students, Phage Hunters Integrating Research and Education Program. 2012. Complete genome sequences of 138 mycobacteriophages. *J. Virol.* 86:2382–2384. <http://dx.doi.org/10.1128/JVI.06870-11>.
  41. Hatfull GF, Cresawn SG, Hendrix RW. 2008. Comparative genomics of the mycobacteriophages: insights into bacteriophage evolution. *Res. Microbiol.* 159:332–339. <http://dx.doi.org/10.1016/j.resmic.2008.04.008>.
  42. Broussard GW, Oldfield LM, Villanueva VM, Lunt BL, Shine EE, Hatfull GF. 2013. Integration-dependent bacteriophage immunity provides insights into the evolution of genetic switches. *Mol. Cell* 49:237–248. <http://dx.doi.org/10.1016/j.molcel.2012.11.012>.
  43. Sampson T, Broussard GW, Marinelli LJ, Jacobs-Sera D, Ray M, Ko CC, Russell D, Hendrix RW, Hatfull GF. 2009. Mycobacteriophages BPs, angel and halo: comparative genomics reveals a novel class of ultra-small mobile genetic elements. *Microbiology* 155:2962–2977. <http://dx.doi.org/10.1099/mic.0.030486-0>.
  44. Tori K, Dassa B, Johnson MA, Southworth MW, Brace LE, Ishino Y, Pietrokovski S, Perler FB. 2010. Splicing of the mycobacteriophage Bethlehem DnaB intein: identification of a new mechanistic class of inteins that contain an obligate block F nucleophile. *J. Biol. Chem.* 285:2515–2526. <http://dx.doi.org/10.1074/jbc.M109.069567>.
  45. Tori K, Perler FB. 2011. Expanding the definition of class 3 inteins and their proposed phage origin. *J. Bacteriol.* 193:2035–2041. <http://dx.doi.org/10.1128/JB.01407-10>.
  46. Banaiee N, Bobadilla-Del-Valle M, Bardarov S, Jr, Riska PF, Small PM, Ponce-De-Leon A, Jacobs WR, Jr, Hatfull GF, Sifuentes-Osornio J. 2001. Luciferase reporter mycobacteriophages for detection, identification, and antibiotic susceptibility testing of *Mycobacterium tuberculosis* in Mexico. *J. Clin. Microbiol.* 39:3883–3888. <http://dx.doi.org/10.1128/JCM.39.11.3883-3888.2001>.

47. Jacobs WR, Jr, Barletta RG, Udani R, Chan J, Kalkut G, Sosne G, Kieser T, Sarkis GJ, Hatfull GF, Bloom BR. 1993. Rapid assessment of drug susceptibilities of *Mycobacterium tuberculosis* by means of luciferase reporter phages. *Science* 260:819–822. <http://dx.doi.org/10.1126/science.8484123>.
48. van Kessel JC, Hatfull GF. 2007. Recombineering in *Mycobacterium tuberculosis*. *Nat. Methods* 4:147–152. <http://dx.doi.org/10.1038/nmeth996>.
49. Lee MH, Pascopella L, Jacobs WR, Jr, Hatfull GF. 1991. Site-specific integration of mycobacteriophage L5: integration-proficient vectors for *Mycobacterium smegmatis*, *Mycobacterium tuberculosis*, and bacille Calmette-Guerin. *Proc. Natl. Acad. Sci. U. S. A.* 88:3111–3115.
50. Lopatto D. 2004. Survey of undergraduate research experiences (SURE): first findings. *Cell Biol. Educ.* 3:270–277. <http://dx.doi.org/10.1187/cbe.04-07-0045>.
51. Estrada-Hollenbeck M, Woodcock A, Hernandez PR, Schultz PW. 2011. Toward a model of social influence that explains minority student integration into the scientific community. *J. Educ. Psychol.* 103:206–222. <http://dx.doi.org/10.1037/a0020743>.
52. Reason RD, Terenzini PT, Domingo RJ. 2006. First things first: developing academic competence in the first year of college. *Res. High. Educ.* 47:149–175. <http://dx.doi.org/10.1007/s11162-005-8884-4>.
53. Hunter A-B, Laursen SL, Seymour E. 2007. Becoming a scientist: the role of undergraduate research in students' cognitive, personal, and professional development. *Sci. Educ.* 91:36–74. <http://dx.doi.org/10.1002/sce.20173>.
54. Stein LD, Mungall C, Shu S, Caudy M, Mangone M, Day A, Nickerson E, Stajich JE, Harris TW, Arva A, Lewis S. 2002. The generic genome browser: a building block for a model organism system database. *Genome Res.* 12:1599–1610. <http://dx.doi.org/10.1101/gr.403602>.
55. Lewis SE, Searle SM, Harris N, Gibson M, Lyer V, Richter J, Wiel C, Bayraktaroglu L, Birney E, Crosby MA, Kaminker JS, Matthews BB, Prochnik SE, Smithy CD, Tupy JL, Rubin GM, Misra S, Mungall CJ, Clamp ME. 2002. Apollo: a sequence annotation editor. *Genome Biol.* 3:research0082. <http://dx.doi.org/10.1186/gb-2002-3-12-research0082>.
56. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* 27:4636–4641. <http://dx.doi.org/10.1093/nar/27.23.4636>.
57. Borodovsky M, McIninch J. 1993. GeneMark: parallel gene recognition for both DNA strands. *Comput. Chem.* 17:123–133. [http://dx.doi.org/10.1016/0098-1354\(93\)80010-K](http://dx.doi.org/10.1016/0098-1354(93)80010-K).
58. Krumsiek J, Arnold R, Rattei T. 2007. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* 23:1026–1028. <http://dx.doi.org/10.1093/bioinformatics/btm039>.