ANNUAL REVIEWS

Annual Review of Virology

SEA-PHAGES and SEA-GENES: Advancing Virology and Science Education

Danielle M. Heller,[1] Viknesh Sivanathan,[1] David J. Asai,[1] and Graham F. Hatfull[2]

[1]Center for the Advancement of Science Leadership and Culture, Howard Hughes Medical Institute, Chevy Chase, Maryland, USA

[2]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, USA; email: gfh@pitt.edu

Annu. Rev. Virol. 2024. 11:6.1–6.20

The *Annual Review of Virology* is online at virology.annualreviews.org

https://doi.org/10.1146/annurev-virology-113023-110757

Keywords

bacteriophages, course-based research, SEA-PHAGES, SEA-GENES, equity in STEMM education

Abstract

Research opportunities for undergraduate students are strongly advantageous, but implementation at a large scale presents numerous challenges. The enormous diversity of the bacteriophage population and a supportive programmatic structure provide opportunities to engage early-career undergraduates in phage discovery, genomics, and genetics. The Science Education Alliance (SEA) is an inclusive Research-Education Community (iREC) providing centralized programmatic support for students and faculty without prior experience in virology at institutions from community colleges to research-active universities to participate in two course-based projects, SEA-PHAGES (SEA Phage Hunters Advancing Genomic and Evolutionary Science) and SEA-GENES (SEA Gene-function Exploration by a Network of Emerging Scientists). Since 2008, the SEA has supported more than 50,000 undergraduate researchers who have isolated more than 23,000 bacteriophages of which more than 4,500 are fully sequenced and annotated. Students have functionally characterized hundreds of phage genes, and the phage collection has fueled the therapeutic use of phages for treatment of *Mycobacterium* infections. Participation in the SEA promotes student persistence in science education, and its inclusivity promotes a more equitable scientific community.

INTRODUCTION

Scaling Access to Undergraduate Research

PEERs: persons excluded because of ethnicity or race; in US science, PEERs include persons identifying as Black, Indigenous, and Latine

Participating in research and experiencing the thrill of scientific discovery can be transformative experiences for an early-career scientist. Indeed, students who participate in research as undergraduates have been shown to earn higher course grades and have better mastery of scientific content and skills, and they are more likely to earn a degree in science, technology, engineering, mathematics, and medicine (STEMM) fields than their peers (1). The earlier the exposure to research, the better, with duration and depth of the research experience correlating with increased interest in STEMM careers (2). Undergraduate research experiences (UREs) help to integrate novice scientists into the culture and community of STEMM, exposing them to fundamental concepts such as collaboration and iteration, building their network, and fostering student confidence and belief that they too can be scientists (1). Importantly, the positive outcomes of undergraduate research participation are seen across many demographic groups (3–5), particularly for students who identify as persons excluded due to ethnicity or race (PEERs) (6). These demonstrated benefits have made the expansion of UREs a key aim of national efforts to improve student agency in STEMM education and build more equitable STEMM systems (1, 3). As the global community faces unprecedented challenges, growing a diverse STEMM workforce and fostering a healthy scientific ecosystem in which it can thrive are of the utmost importance.

An effective URE should align with the authentic, multifaceted experience of a professional scientist (1). Students working under the guidance of a mentor should build ownership over a project that seeks to address a scientific question of interest to the field and broader society. Through this experience, research students should gain mastery in specific techniques and exposure to fundamental processes such as experimental design, troubleshooting and replication, and communication of findings. UREs have been developed in many different forms to emphasize all or most of these aims, and they vary across many dimensions, including duration, mentorship format, and student expectations and outcomes (for a review, see 1) (**Table 1**).

Table 1 Features of undergraduate research experiences

	Apprentice-based research experience: research done in a faculty lab or as part of an internship	Course-based research experience: faculty-developed research embedded in course curricula at an institution	Inclusive Research-Education community: a network of institutions supported by a centralized infrastructure, implementing collaborative course-based research
Mentor capacity	One mentor for one or a few students	One mentor for many students	One mentor for many students
Cost	High cost per student	Lower cost per student	Lower cost per student
Student access	Available for only a selected subset of often later-career students	Open to all students in a course, with no prior research experience required	Open to all students in a course, with no prior research experience required
Institutional access	Limited to institutions with investigator-driven research and necessary infrastructure	Limited to institutions with research infrastructure and often requiring significant faculty development	Accessible to institutions without extensive research infrastructure or investigator-driven research

Student access:
: Barrier for student access.
: Condition for increased access.

In the traditional apprentice-based model, a student conducts independent research under the guidance of an experienced investigator, exploring a scientific question that feeds into the larger aims of an ongoing research project while becoming socialized in the customs and processes of science. Apprenticeships can be especially beneficial for persons belonging to groups historically excluded from science, granting them access to senior and peer mentors who can support their development and persistence (7). A 2018 longitudinal study found that participation in mentored research, and the social integration it provided, was strongly predictive of increased persistence in STEMM for students who identify as Black, Indigenous, and Latine (5). However, large numbers of students do not have access to these experiences for two reasons. First, many students attend colleges where the necessary research infrastructure is not available. This includes community colleges, which enroll more than 40% of all undergraduate students in the United States and serve disproportionately high numbers of women, PEERs, first-generation students, low-income students, and students with disabilities (8, 9). Second, even at those institutions with active research programs, limited mentor capacity, outsized demands on faculty time, and the scarcity-driven funding landscape make undergraduate research apprenticeships highly competitive, favoring later-career students with prior research experience and disadvantaging students with competing demands (e.g., employment or family obligations) on their time outside of class (10).

Faculty: persons teaching STEMM courses including tenure-track and non-tenure-track full-time and part-time faculty, research scientists, and adjunct instructors

Embedding research within STEMM courses can help to dismantle these barriers and scale access to research opportunities, enabling faculty to mentor more students during class hours. Note that throughout this text, we use the terms faculty and instructor broadly to include science educators of all ranks. With the right scientific project, course-based research experiences (CREs) can achieve all the critical components of an URE and can be situated early in the introductory lab experience, thereby serving as a gateway to additional research opportunities (10–12). A successful CRE is designed to advance equity so that it is accessible to novice researchers, requiring no previous lab experience or course prerequisites and employing a set of techniques that can be mastered easily by students over time (13). It should offer ample opportunities for discovery, teamwork, and problem-solving and go beyond inquiry-based learning to ask a question that directly contributes to a recognized scientific problem, with data deposited in public databases and findings communicated through presentations and publications (14). The science of CREs can also be scaffolded to provide multiple semesters of research experience. For example, in the Freshman Research Initiative (FRI) at the University of Texas, first-year students take a research methods course followed by two semesters of connected course-based research. Participation in the FRI was seen to increase and equalize STEMM retention rates across gender, race, and first-generation status (4). The FRI is one of several examples of course-based strategies that can be broadly impactful for students (4, 12, 15). Benefits for those teaching CREs have also been documented, with faculty reporting increases in their own sense of science identity and community and positive shifts in their relationships with students, noting more opportunities for mentorship and deeper scientific conversations than with traditional curricula (16, 17).

The Inclusive Research-Education Community Model

The CRE model has gained significant traction over the past decade or so, and there is ongoing discussion on how to best implement this model, practically and pedagogically, and how to fairly assess outcomes (1, 18). This has led to significant heterogeneity in format, content, and scale across different CREs (for examples, see **https://serc.carleton.edu/curenet**). Barriers persist within this model too, with access largely restricted to better-resourced institutions with higher research capacity and faculty with pertinent research skills. Furthermore, individual faculty-developed CREs are often difficult to sustain in the face of fluctuating budgets, institutional

practices, and time constraints, and their scalability is typically constrained to only a small number of laboratory sections (19).

An inclusive Research-Education Community (iREC) builds on the CRE model to scale student involvement and promote broad inclusion in undergraduate research, including a greater diversity of institutions, faculty that may not have extensive expertise in the research topic, and effectively all students without any form of preselection (12). The iREC model achieves this by centralizing much of the programmatic administration, resources, and scientific direction (**Figure 1**). Individual institutions joining an iREC can thus provide course-based research opportunities to undergraduate students, adopting a common set of protocols to address core scientific questions, and iREC participation is not restricted to high–research activity universities (12, 15, 19, 20). Community colleges (two-year and four-year) and tribal colleges can advance research questions alongside research-active institutions within an interactive community (**Table 1**).

There are several notable examples of programs that follow an iREC structure, including the Tiny Earth network supported by scientists at the University of Wisconsin-Madison (21), the Genomics Education Partnership (GEP) supported by a centralized programmatic team at the University of Alabama and Washington University (22), and the Science Education Alliance (SEA) led by a partnership between staff and scientists from the Howard Hughes Medical Institute (HHMI), the University of Pittsburgh, and James Madison University (12). In each of these, a centralized infrastructure supports key programmatic functions including curriculum and research tool development, onboarding and training of new faculty instructors, management and publication of scientific data, and sponsoring opportunities for community-building and scientific exchange. The centralized programmatic elements benefit from financial support, which can be considerable, but because relatively large numbers of students can participate, the cost per student can be minimal (12). Research opportunities are thus scalable on more campuses, and faculty have a scientific framework and technical support to sustain research with large numbers of their students, year after year. A survey of faculty participating in the GEP found that while institutional barriers such as high teaching loads and space restrictions were unavoidable, the centralized GEP program support and community were crucial for helping faculty stay on top of the evolving science and curriculum, troubleshoot the research, and ultimately sustain the CRE (19).

Not all research questions are equally suitable for the iREC structure. Ideally, the central research question of an iREC is one that does not require substantial prior knowledge or technical expertise and employs approaches that can be implemented by many students in parallel, thus enabling a systematic approach to larger scientific questions (13). Students at the more than 200 colleges and universities that participate in the GEP use bioinformatics to collectively annotate *Drosophila* genomes and learn more about the evolution and functions of eukaryotic genes (22). In Tiny Earth, undergraduate researchers worldwide participate in the hunt for new antibiotics against bacterial pathogens, with students collecting and analyzing thousands of environmental isolates to identify promising samples and compounds for further development (21). In the SEA, thousands of undergraduate scientists each year conduct research to collectively explore the remarkable diversity of the bacteriophage population, one of the most discovery-rich systems on our planet, and one with wide-reaching implications for ecology, biotechnology, and human health (23). The discovery and genomic characterization of the phage population is particularly well suited to the iREC structure.

THE SEA IREC COMMUNITY

The HHMI SEA, established in 2008, is an iREC composed of educators and scientists from colleges and universities across the United States who implement two accessible, collaborative

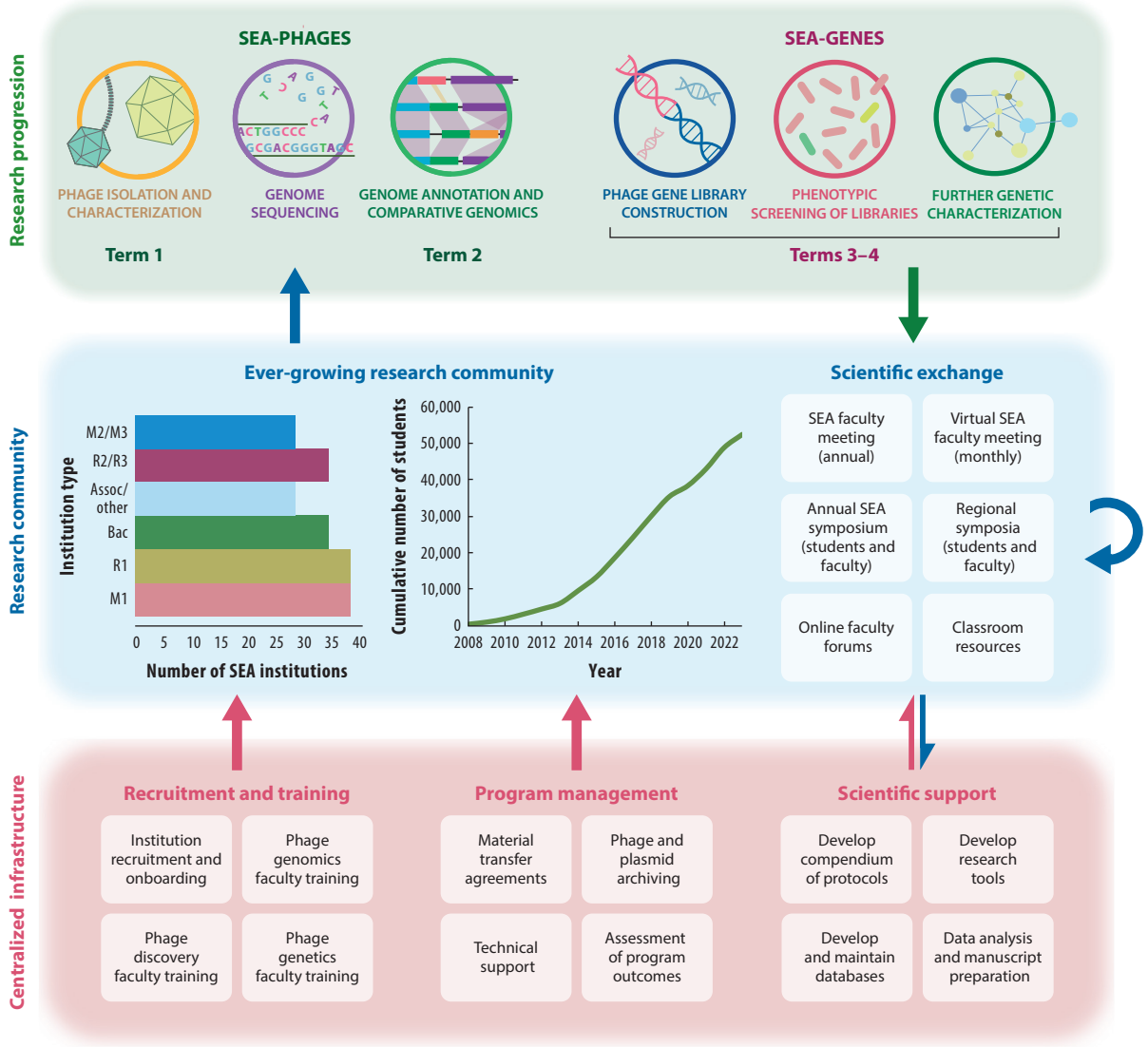

Figure 1

The SEA iREC. A centralized support structure (*pink box* and *arrows*) supports a growing community of faculty and students to implement a cohesive research program that progresses over multiple academic terms (*green box*). In the blue box, the bar graph represents the total number of each type of US institution (by Carnegie Classification) that have participated in the SEA since 2008, including associate's colleges and others (Assoc./Other); baccalaureate colleges (Bacc.); M1–M3, larger, medium, and smaller master's colleges and universities, respectively; and R1–R3, doctoral universities with highest, higher, and moderate research activity, respectively. The line graph shows the cumulative number of students that have participated in SEA research since 2008. Scientific exchange occurs at many levels of the iREC, with opportunities for faculty development and collaboration, scientific meetings, and continual refinement of the program science. Abbreviations: iREC, inclusive Research-Education Community; SEA, Science Education Alliance; SEA-GENES, SEA Gene-function Exploration by a Network of Emerging Scientists; SEA-PHAGES, SEA Phage Hunters Advancing Genomic and Evolutionary Science.

course-based research projects to build collections of valuable research outputs to drive outsized advances in phage biology and genomics. The SEA Phage Hunters Advancing Genomic and Evolutionary Science, or SEA-PHAGES, project was established in 2008 as an expansion of the earlier Phage Hunters Integrating Research and Education (PHIRE) program at the University of Pittsburgh (13, 14, 24). The principal aim of the PHIRE and SEA-PHAGES programs is to understand bacteriophage diversity, evolution, and origins through phage discovery and genomic analyses. In SEA-PHAGES, beginning undergraduates spend one year (two semesters) isolating novel bacteriophages from the local environment and conducting bioinformatic analyses to annotate a subset of the discovered genomes, ultimately depositing these annotations in GenBank (11, 12) (**Figure 1**). Since 2019, SEA students have had the opportunity to engage in a second year of research, participating in the SEA Gene-function Exploration by a Network of Emerging Scientists, or SEA-GENES, project to build phage gene libraries and systematically screen them for gene functions of interest in the bacterial host (25). The principal aim of SEA-GENES is to elucidate phage gene functions. In both projects, SEA researchers work in parallel, one phage at a time, one gene at a time, to collectively generate an archive with ∼24,000 phages and a genomic data set that presents the opportunity to explore phage evolution and gene function with stunning resolution.

SEA-PHAGES: Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science

SEA-GENES: Science Education Alliance Gene-function Exploration by a Network of Emerging Scientists

Living in a Phage World: A Discovery-Rich System for Course-Based Research

Phage biology offers the type of astronomical figures that can inspire awe in novice and seasoned scientists alike. It is estimated that there are roughly 10^{31} viral particles on our planet, with bacteriophages the dominant class, outnumbering their bacterial hosts by at least an order of magnitude and embroiling them in an estimated 10^{23} infections per second (26–28). These dynamic infections have enormous ecological impacts, churning through an estimated 20% of the microbial biomass in the ocean every day (29). Viral ecologists have found a diverse array of phages in every ecosystem, from geothermal vents at the bottom of the ocean to the mucosal surfaces of the human gut (30). These phages vary dramatically in their morphologies and genomic content. Single-stranded DNA marine microviruses encoding as few as four open reading frames have been described (31), while huge phages abundant in animal gut microbiomes have genomes of more than 200 kilobases and exhibit many unusual features such as use of alternative genetic codes and construction of complex nucleus-like structures to compartmentalize viral genome replication (32–34).

Phages have long been a rich source for uncovering the foundational principles of molecular biology and continue to yield discoveries that often upend our assumptions and inspire important technologies. Indeed, the past two decades have ushered in a new golden era for phage biology. The application of deep sequencing technologies has revealed a truly enormous reservoir of genetic diversity that defies traditional taxonomic classification (35), with pronounced sequence variation seen between phages that infect different and even the same bacterial hosts (30, 36). Metagenomic studies have generated incredible snapshots of viral genetic diversity within many different ecosystems, getting around the challenge of identifying and culturing the cognate host to describe the composition, abundance, and global distribution of phage populations (30, 37, 38).

Isolation and characterization of single phages remain powerful complements to these deep sequencing techniques, providing full genome assemblies, better resolution of genetic variation between related genomes, and the ability to experimentally interrogate the biology of the phage and its identified host (30, 39–41). Most phage genes that have been discovered encode hypothetical proteins of unknown function, and though recent advances in machine learning hold significant promise for bioinformatic prediction of protein structure and function, access to these phage-host systems is critical for the further characterization of complex phage-host interactions (36, 42). For instance, work in recent years has revealed a surprisingly diverse arsenal of

defense systems in bacterial genomes, forming an ancient immune system deployed to thwart infection by phages and a slew of phage-encoded counter-defense systems; a complement of genomics and experimental methods have been critical in identifying these systems and in elucidating the complex details of their mechanisms (43, 44). Access to more characterized phages is also crucial as the community turns increasingly toward developing phage-based therapeutics for treatment of antibiotic-resistant infections (45), and efforts to create banks of diverse phages infecting a spectrum of bacterial human pathogens across the globe are gaining speed (46).

In the classroom, phage research incorporates the learning of foundational tenets of biology coupled with real-world relevance. With selection of appropriate bacterial hosts, the isolation and study of novel phages from the environment is an endeavor that is well-matched for beginning undergraduate researchers in the classroom (13). It involves technically approachable procedures that can be performed in parallel by many students and offers a reasonable promise of novel discovery for each student that participates. The sheer size and remarkable diversity of the phage population make it likely that a SEA student will find a virus or annotate a gene that has never been seen before, thereby fueling student interest and project ownership (14).

The iREC Structure: Supporting Faculty in Scaling High-Impact Research

The SEA advances equity and inclusion by intentionally investing in a centralized infrastructure and practices that enable early undergraduate students from all backgrounds and at all institution types to engage in phage discovery and genomic exploration (12, 15) (**Figure 1**). Since 2008, a total of 216 institutions have joined the SEA programs, and 175 remain active, although not all schools teach the course every year. In the 2022–2023 academic year, faculty from more than 140 different institutions led more than 6,000 undergraduates in research through the SEA-PHAGES and SEA-GENES projects. The SEA iREC also supports sustainable growth of the program: In total, since 2008 more than 50,000 undergraduates, most in their first or second year of college, from more than 200 institutions have conducted research in the SEA (**Figure 1**). Schools are very likely to participate in the SEA for multiple years, and the SEA typically welcomes faculty from about a dozen new two-year and four-year colleges and universities each year.

The SEA iREC helps promote cost efficiencies, with administrative costs supported through centralized and sustained investment by HHMI rather than depending on funding by individual institutions; these administrative expenditures are incurred largely independently of student number, and thus program growth yields a lower per-student cost (12). The SEA also negotiates reagent discounts and donations from commercial partners, reducing per-student research costs for schools such that they are equivalent to costs for traditional lab activities.

Each school that participates in the SEA essentially operates as an independent research group, with instructors mentoring teams of undergraduate researchers in the classroom in investigating common aims and gathering data using common techniques (12–14, 25). This consortium of SEA schools is supported by a collaborative team comprising HHMI program staff and scientists and research teams at the University of Pittsburgh and James Madison University. The support provided by this partnership encompasses scientific leadership, scientific and pedagogical exchange, community-building, and other administrative support. SEA program scientists establish, coordinate, and adapt the overarching research aims for SEA projects and develop and distribute the research reagents, protocols, and bioinformatic tools used by faculty and students across the program. Program staff also review, curate, and support dissemination of student-generated data in open-access databases and publications, ensuring that high-quality science continues to drive the program research (12).

Each year, the SEA trains instructors from new SEA member institutions on the foundational knowledge and research skills needed to implement these course-based research projects, prior to their first implementation. In the case of SEA-PHAGES, this occurs over two, week-long in-person faculty training workshops, whereas for the SEA-GENES project, faculty are trained virtually, working through research protocols at their home institutions and meeting with program staff each week over the course of several months. Beyond this training, SEA staff continue to provide all faculty with on-demand technical support, helping schools adapt and optimize the research for their institutional context and troubleshoot protocols or tools if issues arise. This in-depth scientific training and support eliminate the need for instructors to have prior expertise in the specific field of study, enabling faculty from a range of research backgrounds and current research activity statuses to participate.

Scientific Exchange in the SEA iREC

To promote scientific exchange within this community, the SEA offers programming for faculty and students to advance their scientific skills and forge cross-institutional relationships and collaborations. This involves the hosting of community meetings—for example the national SEA Symposium, an annual science meeting held virtually since 2021—to which all SEA faculty and students are invited to attend and have opportunities to present talks or posters on their research. It also includes the SEA faculty meeting series, with virtual meetings held monthly and one in-person gathering every summer, where faculty discuss topics related to their scientific and educational research and engage in cross-institutional collaborations. For instance, at the in-person SEA faculty meeting in June 2023, ~175 instructors gathered over four days at the HHMI campus in Chevy Chase, Maryland, for a mix of scientific research presentations, a faculty poster session, and skill-building workshops on AlphaFold (47), ChimeraX (48), and genomic data visualizations, as well as dedicated working time for writing collaborative scientific manuscripts. At past meetings, faculty have collaborated on comparative genomics studies of phages infecting specific host genera, producing reports coauthored by many faculty describing the diversity and relationships of these student-discovered phages (39, 41). Asynchronous interactions are also supported by the SEA, including multiple online forums for community discussions, resource sharing, and troubleshooting. The SEA coordinates assessment and communication of student and faculty outcomes in the program, deploying the Persistence in the Sciences (PITS) survey to evaluate variables associated with student persistence (49), and facilitating community discussion and scholarship on best practices for teaching and learning in course-based research (16, 50). Altogether, the centralized structure of the SEA iREC sustains the coordinated collaboration of faculty and students from many diverse institutions, ensuring that the research they collectively pursue remains relevant and rigorous as they explore the bacteriophage population (12).

SEA RESEARCH: MOBILIZING UNDERGRADUATE SCIENTISTS TO EXPLORE PHAGE DIVERSITY

SEA-PHAGES Part I: Phage Discovery

The SEA-PHAGES and SEA-GENES projects engage undergraduate researchers in up to two years of mentored research that scaffolds foundational but accessible techniques of microbiology, molecular biology, genetics, and genomics to explore the diversity of phage populations through investigation of an evolving set of scientific aims. The progression of the SEA research is outlined in **Figure 2**. In the first semester of the SEA-PHAGES project, students sample the environmental diversity of phages, isolating novel phages from local environments that infect specific actinobacterial hosts. Actinobacteria, and therefore their phages, are ubiquitous in terrestrial and

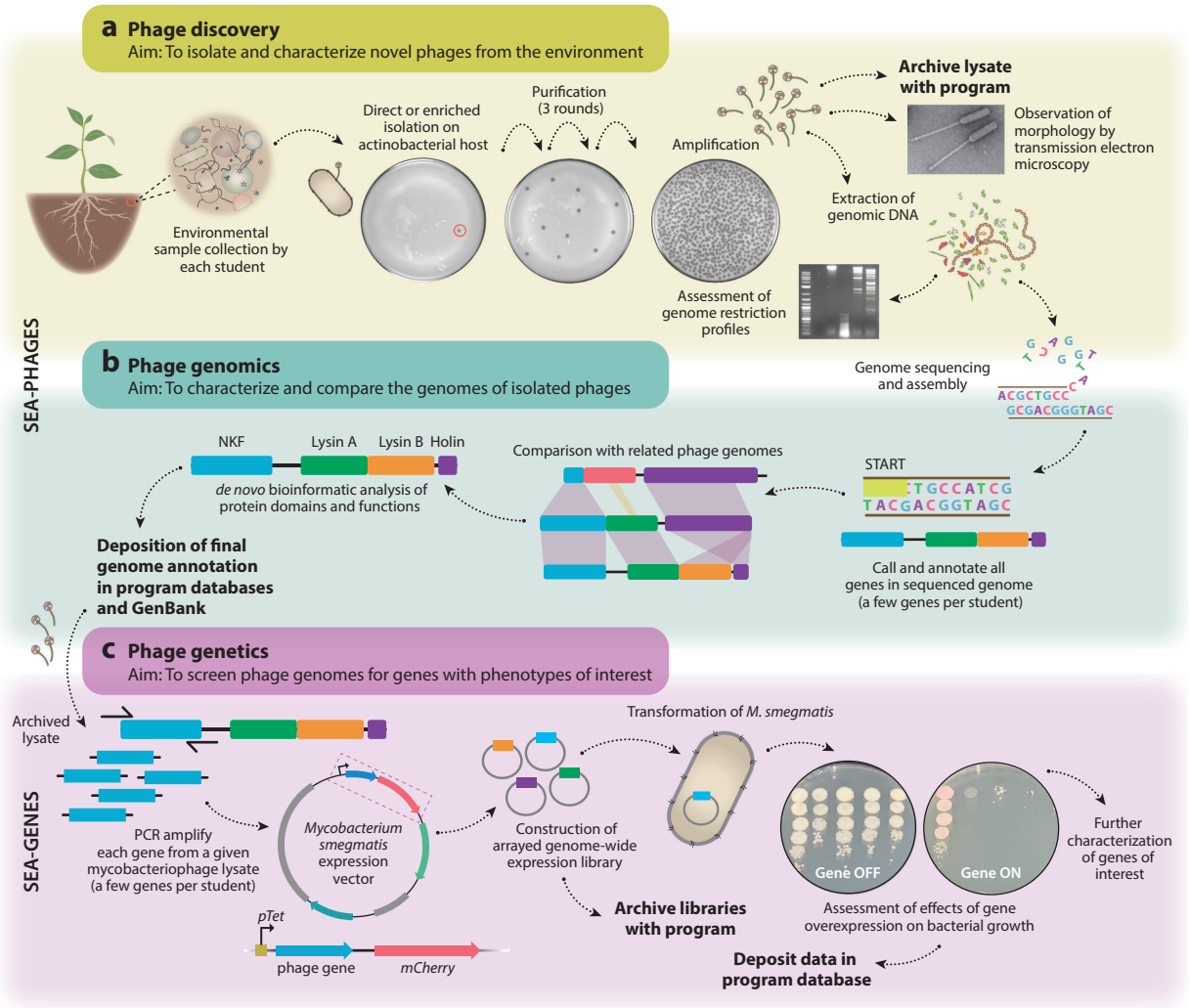

Figure 2

The SEA discovery sequence. SEA student researchers participate in a rich exploration of bacteriophages and their genomes. This research sequence, mapped here, is multifaceted and yields important student-generated research outputs (highlighted in *bold text*). Abbreviations: SEA, Science Education Alliance; SEA-GENES, SEA Gene-function Exploration by a Network of Emerging Scientists; SEA-PHAGES, SEA Phage Hunters Advancing Genomic and Evolutionary Science.

aquatic ecosystems (51). This phylum of high-GC gram-positive bacteria offers many strains of *Mycobacteria*, *Gordonia*, *Arthrobacter*, *Streptomyces*, *Rhodococcus*, and *Microbacterium* species that are safe and simple to culture in the lab but related to bacteria with important medical or technological implications. Specific actinobacterial strains are vetted by the program as having appropriate phage isolation rates, yielding phages in at least ~10% of tested samples from multiple locations. Using a set of core methods such as serial dilution and plaque assays, students identify phages in diverse environmental samples and iteratively take them through multiple rounds of purification and amplification to generate a high-titer lysate for long-term storage in a centrally maintained archive at the University of Pittsburgh. Students name their phages (e.g., Et2Brutus, Corndog, and

PineapplePizza) and do further characterization, including genomic DNA extraction, restriction enzyme digestion, and transmission electron microscopy.

Actinobacteriophage: a bacteriophage that infects a bacterial host in the phylum *Actinobacteria*

SEA-PHAGES Part II: Phage Genomics

Between the first and second term, two genomes (typically) per course section are sequenced and assembled by SEA program scientists. In the second term, teams of students apply a range of bioinformatic methods to collaboratively annotate these complete genomes, typically identifying 50–200 putative genes, exploring their sequence relationships, and predicting the functions of the gene products. Due to the small size of many phage genes and the abundance of genes of unknown function, phage genomes present substantial challenges for genome annotation. Manually curated annotations are typically more accurate than automated annotations, and the annotation experience provides in-depth exposure of students to the core tenets of molecular biology. Trained SEA faculty experts and SEA program scientists team up to review student annotations before they are ultimately submitted to GenBank and deposited in two program databases, PhagesDB [**https://phagesdb.org/** (52)] and Phamerator [**https://phamerator.org/** (53)]. The SEA supports opportunities for faculty and students to coauthor and publish their phage genomic information, including as genome announcements in the peer-reviewed journal *Microbiology Resource Announcements*.

SEA-GENES Part III: Phage Genetics

Since its advent in the fall of 2019, a subset of SEA schools (as of the autumn of 2023, 30 schools, ∼20% of active SEA schools) implement the SEA-GENES project over one or two terms, extending research opportunities for mid-level undergraduate students as part of their curricula for genetics, microbiology, or molecular biology (25). The science of SEA-GENES is a natural expansion of the SEA-PHAGES project, going beyond in silico gene function predictions to begin experimentally characterizing the functions of the genes encoded by the actinobacteriophage population. Each school investigates a unique phage from the SEA archives capable of infecting the model host strain *Mycobacterium smegmatis* mc^2155. Collectively, students employ molecular cloning techniques to insert each of the open-reading frames from that phage's genome into a specially engineered expression vector; each student or student pair tackles ∼5 different genes, collaboratively building an arrayed, genome-wide plasmid library that is archived with the program. The students then collaboratively deploy these libraries in systematic overexpression screens that employ accessible plate-based assays to identify phage-encoded products that affect growth and behavior of the mycobacterial host. SEA-GENES provides students and faculty with a growing set of extra genetic tools and support for further exploration of the sequence determinants and cellular targets of these gene products of interest. As in SEA-PHAGES, the SEA promotes dissemination of these student-led findings, with data deposited in a recently launched open-access database, GenesDB (**https://genesdb.org/**), and support for preprinting and publication of screening data sets as mutant screen reports in the Genetics Society of America journal *G3* (25, 54–56).

Assessing the Effects of the SEA iREC on Student Learning

Each module of the SEA research sequence satisfies all dimensions of an authentic URE (1) (**Table 2**). Students apply a common set of protocols iteratively, gaining mastery of key techniques such as DNA extraction, PCR, and BLAST (Basic Local Alignment Search Tool) grounded in the context of real research questions. Faculty instructors and peers provide mentorship through the roadblocks and ambiguities of research and help cultivate student science identity and self-efficacy (12, 16). Student discovery and collaboration are integral to SEA research at multiple levels, with

Table 2 Research attributes of the SEA iREC

	Attributes	
Defining feature of URE[a]	***SEA-PHAGES***	**SEA-GENES**
Engage students in research practices	Cohorts of students are mentored by instructors on data generation and analysis related to collective scientific aims.	
Aim to generate novel information with an emphasis on discovery and innovation	Students discover, name, and characterize novel phages from local environments. Genomes are annotated and submitted to GenBank.	Students construct gene libraries and systematically screen them to discover genes that influence bacterial growth.
Focus on significant, relevant problems of interest to STEMM researchers	It creates an enormous archive of phages for further application mining and a rich data set for high-resolution genomics.	It generates large arrayed and pooled phage gene libraries and screens large numbers of phage genes to identify a subset that elicit specific host phenotypes for further characterization.
Emphasize and expect collaboration and teamwork	Students collaborate to annotate a complete phage genome over the course of a semester. Students often adopt phages from each other if roadblocks are encountered.	Consecutive cohorts of students at each school collaborate in genomic library construction and genetic screens over multiple course implementations.
Involve iterative refinement of experimental design, experimental questions, or data obtained	Faculty mentor students in troubleshooting and optimizing experiments (e.g., generating a lysate with high enough titer for genomic DNA extraction and sequencing).	Faculty mentor students in optimizing PCR, cloning, and bacterial growth. Students perform multiple independent repeats of phenotypic assays to confirm screen candidates.
Allow students to master specific research techniques	A set of technically simple techniques (e.g., serial dilution, plaque assay, BLAST) are applied in a variety of contexts to enable mastery.	Common cloning techniques are applied to construct phage gene plasmids for ∼5 genes per student. Genes are screened in simple plate-based assays iteratively, with high likelihood for diverse outcomes.
Help students engage in reflection about the problems being investigated and process	Students keep notebooks to record and reflect on their lab work. SEA faculty have collaboratively developed and employed innovative pedagogical resources to engage students in reflection and research skills such as reading primary literature. Student-generated data are reviewed by peers, faculty, and SEA expert scientists with opportunities for revision.	
Require communication of results	Students present posters at the national SEA Symposium where they can interact with peers, faculty, and SEA scientists. Many schools offering additional institutional or regional opportunities to present. All student data are submitted to public databases, and students coauthor brief publications on their findings.	

[a]Adapted from the definition of an URE presented in the National Academies of Science, Engineering, and Medicine's *Undergraduate Research Experiences for STEM Students: Successes, Challenges, and Opportunities* (1).

Abbreviations: BLAST, Basic Local Alignment Search Tool; iREC, inclusive Research-Education Community; SEA, Science Education Alliance; SEA-GENES, SEA Gene-function Exploration by a Network of Emerging Scientists; SEA-PHAGES, SEA Phage Hunters Advancing Genomic and Evolutionary Science; STEMM, science, technology, engineering, mathematics, and medicine; URE, undergraduate research experience.

individual students contributing reagents and data to collaboratively build a class-wide database and, together, a population-wide data set. When students that participated in the SEA-PHAGES project as part of their early biology curricula were administered pre- and postcourse surveys testing a range of biological concepts, SEA-PHAGES students showed performance gains comparable to a matched cohort of students taking traditional lab courses (11); this suggests that there is no learning loss when we abandon prescripted lab activities that survey a wide range of biological topics and instead focus on a single complex research topic. Results from assessments administered by the program each year show promising gains for student persistence in STEMM

(11, 12). The PITS assessment survey measures five psychosocial variables—project ownership, self-efficacy, science identity, scientific community values, and networking—that are highly correlated with student intention to stay in science (49, 57). Students that participate in SEA-PHAGES show significant increases in these variables as compared to their matched peers in traditional labs. Although not all five variables increase evenly across all student groups surveyed, overall, positive gains are seen broadly, including for women, PEERs, and first-generation students at both two-year and four-year colleges (12, 15). A similar survey is currently being deployed for SEA-GENES students, and given the common network and similarities in research design between SEA-PHAGES and SEA-GENES, we predict similar positive student outcomes. Thus, the SEA offers large numbers of students a high-impact and inclusive research experience, integrating them into the research ecosystem early and positioning them to make important scientific contributions.

THE SEA AT THE LEADING EDGE OF PHAGE BIOLOGY

A Detailed View of Diverse Actinobacteriophages

Each year, the SEA produces a rich set of research outputs that, collectively and over time, meld the scale of metagenomics data sets with the granularity of complete characterization of isolated phages. In the autumn of 2023, the number of phages contained in the Actinobacteriophage Database, almost all of them contributed by SEA students, was more than 24,000, and more than 4,200 fully sequenced and annotated genomes have been submitted to GenBank (**https://phagesdb.org/**). SEA students and faculty have reported these findings to the community in almost 200 genome announcement publications, and SEA-generated data have resulted in at least 70 additional scientific manuscripts (**https://seaphages.org/publications/**). The SEA data set, very likely the largest and most detailed of its kind, has provided remarkable insight into the complexities of phage evolution (36).

The host range of the SEA phage collection spans at least 16 different genera of actinobacteria with very large numbers of phages capable of infecting common host strains such as *M. smegmatis* (>2,200 sequenced phages), *Gordonia terrae* (>500 sequenced phages), and *Microbacterium foliorum* (>500 sequenced phages). These phages vary in their morphologies, with siphoviruses, podoviruses, myoviruses, and tectiviruses, and phages with isometric or prolate capsids, all represented in the collection although in different distributions across hosts. For example, although phages infecting mycobacteria are the most deeply sampled in the data set, the vast majority are siphoviruses with long, flexible tails, with myoviruses constrained to a single clade and no podoviruses identified to date, likely reflecting evolutionary constraints imposed by the complex mycobacterial cell envelope (58).

The collection of student-isolated viruses continues to yield new advances in our understanding of phages. For example, high-resolution cryo–electron microscopy structures of diverse mycobacteriophages show that their capsids display novel variations of canonical protein folds and are decorated by a suite of accessory proteins (59, 60). Extensive glycosylation of capsid and tail structures, mediated by viral-encoded glycosyltransferases, has also been observed within this population, with evidence suggesting that this and possibly other surface modifications may play an important role in the ability of phages to infect bacteria and influence their interactions with the immune system of a mammalian host (61). Analyzing the profiles of host resistance to various phages has also enabled further probing of the cellular determinants of phage infection. Isolation of *M. smegmatis* strains resistant to infection by phage Fionnbharth revealed disruption of the gene coding for Lsr2 (62), a conserved nucleoid-associated protein involved in host genome organization (63); host Lsr2 was found to be broadly critical for genomic replication and productive

infection by many mycobacteriophages. This study applied live-cell fluorescence microscopy to generate a detailed visualization of the spatiotemporal dynamics of mycobacteriophage infection, also revealing that phages preferentially target regions of new cell-wall synthesis on the mycobacterial surface (62). Additional progress has been made in narrowing down which components of the complex mycobacterial surface are critical for phage recognition and infection (64). A recent study found that isolated strains of *Mycobacterium abscessus* that are resistant to infection by phages Muddy and BPs have alterations in the pathway producing trehalose polyphleates, suggesting an important role for this class of surface-exposed lipids in phage adsorption and host susceptibility (65).

The Remarkable Genomic Diversity of the Actinobacteriophage Population

Genomic analyses reveal that the actinobacteriophage population encodes thousands of genes that can be sorted into nearly 30,000 groups of related products, or phamilies (phams), based on protein sequence similarity (53, 66). These phams are arranged in a highly mosaic fashion within actinobacteriophage genomes, with units of one or more genes found in different genomic contexts within distantly related genomes (36, 67–69). Using pairwise gene content comparisons (i.e., the proportion of phams shared between two phages), actinobacteriophages can be grouped in different clusters based on their genomic relationships (36, 70). As the number of sequenced phages has grown, the computational strategies used to sort these highly mosaic genomes into meaningful groupings have evolved to accommodate the ever-increasing complexity (70). Mycobacteriophages alone form 31 different clusters (A–Z, AA–AE), 12 of which can be further divided into subclusters (40, 70), whereas phages infecting *Microbacterium* or *Gordonia* spp. form 12 and 14 clusters, respectively (39, 41) (**Figure 3**). Cluster A, the largest cluster in the database, is defined by a mixture of more than 750 phages infecting *Gordonia* or *Mycobacterium* spp. At the other extreme are many phages that remain as singletons without significant gene content similarity to other phages.

Phamilies (phams): a phamily, or pham, is a group of actinobacteriophage genes related to each other, according to amino acid sequence relatedness

Cluster: a group of bacteriophages related to each other and sharing at least 35% of genes with at least one other cluster member

Subcluster: a subset of phages more closely related to each other than to other cluster members

Singleton: a phage with no other known close relatives

Comparative analyses underscore the many different mechanisms that phages have evolved to carry out core viral functions. For instance, actinobacteriophage lysis cassettes are highly diverse, encoding various arrangements of holin and endolysin proteins with distinct enzymatic domains and cell-envelope targets (71, 72). Many actinobacteriophages are temperate, in some cases forming visibly turbid plaques and/or encoding genes like immunity repressor or integrases associated with lysogeny. However, several variations on the classical lambda model have been observed in this population. In some phages, the lysis-lysogeny decision is determined by site-specific integration of the phage genome, with integration leading to truncation of an ssrA-like proteolysis tag and subsequent production of a stabilized repressor protein (73), whereas some temperate Cluster A phages lack the integrase machinery altogether, instead encoding ParAB-like segregation systems and propagating as plasmids (74). Characterization of several mycobacteriophage-encoded systems for DNA partitioning and recombination has spawned useful genetic technologies, including an expanded arsenal of multi-copy plasmids, integrating vectors, and recombineering systems for further engineering of mycobacteria and their phages (74–77).

The vast majority of the phams encoded by this diverse population have no recognizable domains or sequence features and cannot be assigned function through bioinformatic means alone. Transcriptomic studies have revealed when and to what extent these genes are expressed over the course of the phage life cycle, identifying interesting patterns of gene expression (78–80). Notably, one analysis of *M. smegmatis* lysogens formed by Cluster N phages revealed expression of diverse sets of genes in addition to the immunity repressor genes, including some genes related to toxin-antitoxin or restriction modification systems (80); several of these expressed modules were shown to confer defense against lytic or temperate phages of varying relatedness to the prophage. Indeed,

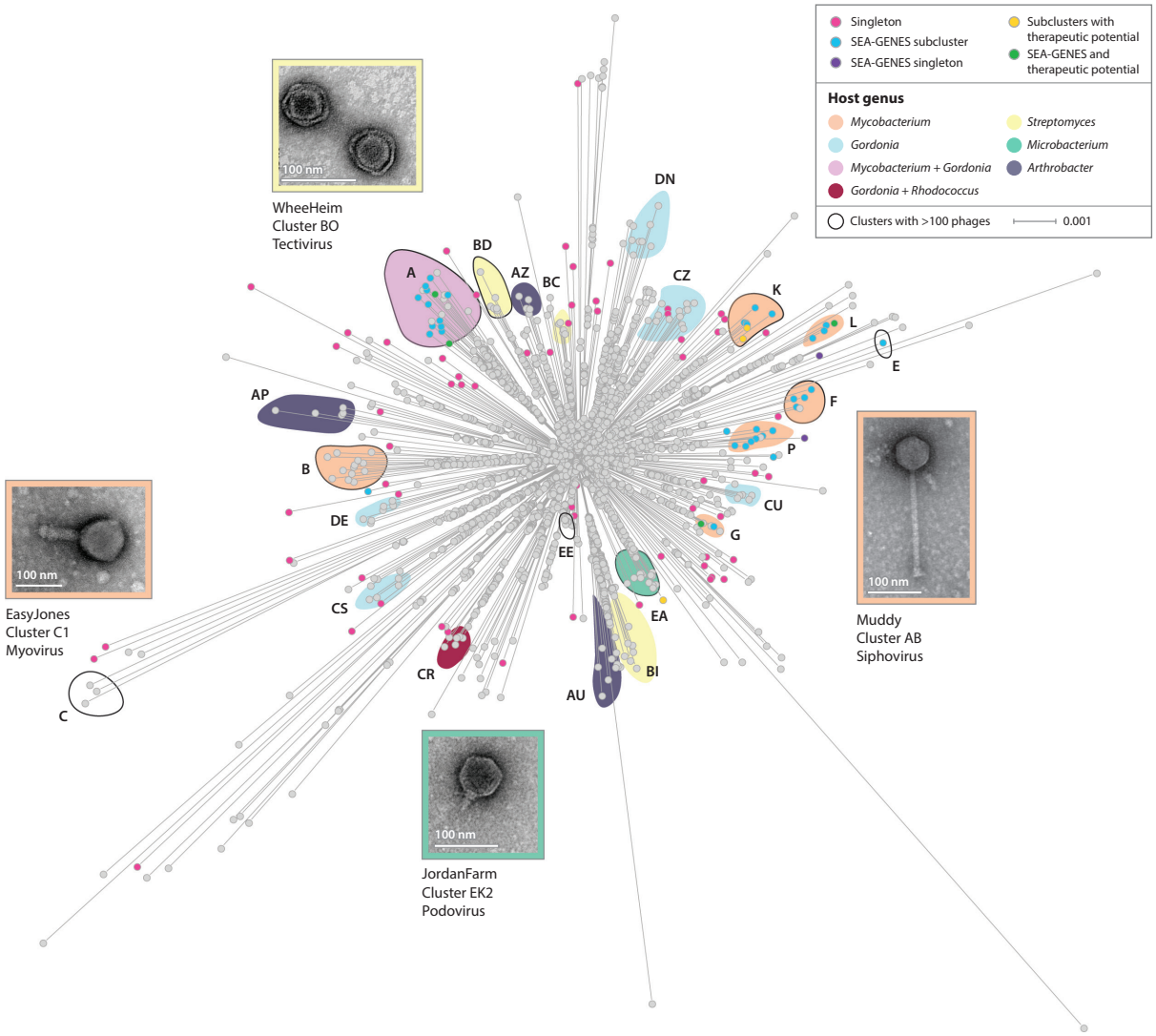

Figure 3

Actinobacteriophage diversity. A network phylogeny comparing the actinobacteriophages discovered by the SEA was generated using SplitsTree (95). A representative phage from each cluster and subcluster and all singletons are shown as a circular node with relationships drawn based on shared gene content. Nodes are colored to illustrate whether a subcluster or singleton (*pink*) is currently being studied in the SEA-GENES project (*blue* or *purple*), has phages that have been applied in recent phage therapy cases (*yellow*) (92), or both (*green*). Those clusters that can be subdivided into four or more subclusters are highlighted with shapes colored to represent the host genera infected by those phages (i.e., Clusters A, AP, AU, AZ, B, BC, BD, BI, CR, CS, CU, CZ, DE, DN, EA, F, G, K, L, and P), and those clusters that have more than 100 phages assigned to them are outlined in black (i.e., Clusters A, B, BD, C, E, EA, EE, F, and K). Representative transmission electron microscopy images from several phages across this phylogeny also highlight the morphological diversity of this population. Abbreviations: SEA, Science Education Alliance; SEA-GENES, SEA Gene-function Exploration by a Network of Emerging Scientists; SEA-PHAGES, SEA Phage Hunters Advancing Genomic and Evolutionary Science.

this prophage-mediated defense appears to be a common phenomenon, with diverse phage defense and counter-defense systems found throughout actinobacteriophage and prophage genomes (80–83).

SEA-GENES and related screening efforts have made headway in identifying genes whose products affect host growth (54–56, 84, 85). In the autumn of 2023, 50 different mycobacteriophage genomes, representing 12 different clusters and encoding >5,000 total genes, are being screened in the SEA-GENES project, and first reports from these efforts have identified an abundance of phage-encoded growth inhibitors that vary in their size, sequence, and predicted function—intriguing candidates for uncovering novel phage-host interactions. For example, a genome-wide screen of Cluster K phage Waterfoul revealed that 32 out of 94 encoded genes were capable of inhibiting *M. smegmatis* growth to some extent upon overexpression, including 6 genes that fully abolished host growth (56). Screening of a second Cluster K phage, Hammy, revealed an additional 24 growth inhibitors, 12 of which were conserved in phage Waterfoul (55). Thus, parallel screening in related genomes enables comparative analyses to determine the conservation, genomic arrangement, and sequence determinants of growth-inhibitory effects (54, 55).

From the Classroom to the Clinic

Although the main aim of the SEA is to understand phage diversity, evolution, and origins, the rich data sets, phage collections, and scientific foundation assembled over the past 15 years hold great promise for many genetic and clinical applications. Among these is the potential therapeutic use of phages to treat life-threatening *Mycobacterium* infections. The most obvious target diseases are tuberculosis caused by *Mycobacterium tuberculosis* infections and nontuberculous mycobacterium (NTM) infections caused by *Mycobacterium abscessus*, *Mycobacterium avium*, and a variety of other strains. In the first report of phage treatment of a human mycobacterial infection, a cocktail of three engineered phages was administered to a 15-year-old patient with cystic fibrosis and a disseminated, multidrug-resistant *M. abscessus* infection, with significant improvement in clinical outcomes observed in the patient after several weeks (86). Because few phages (if any) had at that time been isolated using *M. abscessus* as a host, the SEA collection of mycobacteriophages provided a critical resource. Although nearly all were isolated on related host *M. smegmatis*, extensive screening identified three with therapeutic potential, all of which had been identified by students in integrated research-education programs (86). Two of these phages are temperate and therefore had to be engineered to be obligatorily lytic to ensure potent antimicrobial activity; this was facilitated by the extensive understanding of mycobacteriophage life cycles (73, 87–89) and prior development of tools for efficient phage genome engineering made possible by student discoveries (75, 76).

Additional examples of the use of phages for NTM therapies provide further encouragement (45, 86, 90–92). Out of 20 recent compassionate use cases for patients with drug-resistant nontuberculous mycobacterial infections, administration of single phages or phage cocktails yielded favorable microbiological and clinical outcomes for 11 patients with no resistance to the phages observed in any of the cases (92). Challenges remain, including our incomplete understanding of the complex interactions between phages, bacteria, and the mammalian immune system, and the fundamental tool set and knowledge generated in the SEA will be pivotal in addressing them. Already, the annotation principles honed by the SEA-PHAGES project are being applied to discover, annotate, and analyze the prophage-ome of *M. abscessus*, unlocking new information about the infective landscape of clinically important strains (64, 83, 93, 94).

Although only a small proportion of phages in the SEA collection have been deployed therapeutically to date, all of the phages have contributed important biological context and insights. As

phages are further adapted for clinical use, naturally occurring viruses may serve mostly as intermediates in the development of totally synthetic phage-like nanomachines engineered for optimal antimicrobial activity and lacking genes that may interfere with effective therapeutic use. The collective knowledge of phage diversity and gene function generated by the SEA will continue to be important for achieving this goal.

CONCLUDING REMARKS: TRANSFORMING THE INTERFACE OF SCIENCE AND SCIENCE EDUCATION

Within the scientific ecosystem, research and education often exist in separated silos, thereby limiting opportunities to participate in research to a small number of selected students at a small number of selective institutions. The SEA iREC models the symbiosis that is possible when research and science education are integrated in an equity-centered manner. Equipping faculty at diverse institutions with the framework, support, and training to implement an in-depth investigation of phage diversity with their students is an effective model for scaling access to impactful research experiences and promoting greater inclusion within STEMM. The SEA also illustrates an especially productive strategy for doing science: employing course-based research to mobilize large numbers of faculty and students as active participants in the research community produces a large and detailed data set that is greatly enriched by the connectivity of a diverse human network. In the SEA, contributions by early undergraduate scientists have led to important scientific insights, transforming our fundamental understanding of the genetic diversity of the phage population, giving rise to life-saving therapeutics, and connecting them to the broader scientific enterprise. This scientific arc is the lifeblood of the SEA, with these evolving questions driving student excitement and engagement and enhancing their educational experience. The clinical use of phages in recent years to treat patients adds yet another dimension to student motivation. There are undoubtedly other scientific problems out there that would benefit from a similar iREC strategy, scaling data collection while simultaneously investing in a more diverse and better-prepared STEMM workforce for tomorrow.

DISCLOSURE STATEMENT

The Howard Hughes Medical Institute supports the SEA-PHAGES and SEA-GENES programs and provides funding to G.F.H. to support the SEA-PHAGES program. D.M.H., V.S., and D.J.A. are or were employees of HHMI.

ACKNOWLEDGMENTS

We thank Steve Cresawn and David Hanauer for their important contributions to the SEA program along with key SEA staff Debbie Jacobs-Sera, Dan Russell, Billy Biederman, and Pushpa Ramakrishna. We also thank the faculty and students in the SEA-PHAGES and SEA-GENES communities for their efforts in phage discovery, genomics, and characterization. This work was supported by grants from the National Institutes of Health (grant GM131729 to G.F.H.) and HHMI (grant GT12053 to G.F.H.).

LITERATURE CITED

1. Natl. Acad. Sci. Technol. Eng. Med. 2017. *Undergraduate Research Experiences for STEM Students: Successes, Challenges, and Opportunities*. Washington, DC: Natl. Acad. Press
2. Russell SH, Hancock MP, McCullough J. 2007. Benefits of undergraduate research experiences. *Science* 316(5824):548–49

3. Natl. Acad. Sci. Technol. Eng. Med. 2023. *Advancing Antiracism, Diversity, Equity, and Inclusion in STEMM Organizations: Beyond Broadening Participation*. Washington, DC: Natl. Acad. Press
4. Rodenbusch SE, Hernandez PR, Simmons SL, Dolan EL. 2016. Early engagement in course-based research increases graduation rates and completion of science, engineering, and mathematics degrees. *CBE—Life Sci. Educ.* 15(2):ar20
5. Estrada M, Hernandez PR, Schultz PW. 2018. A longitudinal study of how quality mentorship and research experience integrate underrepresented minorities into STEM careers. *CBE—Life Sci. Educ.* 17(1):ar9
6. **Asai DJ. 2020. Race matters. *Cell* 181(4):754–57**
7. Eagan MK, Sharkness J, Hurtado S, Mosqueda CM, Chang MJ. 2011. Engaging undergraduates in science research: not just about faculty willingness. *Res. High Educ.* 52(2):151–77
8. Community Coll. Res. Cent. 2024. Community college FAQs. *Community College Research Center*. **https://ccrc.tc.columbia.edu/community-college-faqs.html**
9. Am. Assoc. Community Coll. 2023. Fast facts 2023. *American Association of Community Colleges*. **https://www.aacc.nche.edu/wp-content/uploads/2023/03/AACC2023_FastFacts.pdf**
10. Bangera G, Brownell SE. 2014. Course-based undergraduate research experiences can make scientific research more inclusive. *CBE—Life Sci. Educ.* 13(4):602–6
11. Jordan TC, Burnett SH, Carson S, Caruso SM, Clase K, et al. 2014. A broadly implementable research course in phage discovery and genomics for first-year undergraduate students. *mBio* 5(1):e01051-13
12. **Hanauer DI, Graham MJ, SEA-PHAGES, Betancur L, Bobrownicki A, et al. 2017. An inclusive Research Education Community (iREC): impact of the SEA-PHAGES program on research outcomes and student learning. *PNAS* 114(51):13531–36**
13. **Hatfull GF, Pedulla ML, Jacobs-Sera D, Cichon PM, Foley A, et al. 2006. Exploring the mycobacteriophage metaproteome: phage genomics as an educational platform. *PLOS Genet.* 2(6):e92**
14. Hanauer DI, Jacobs-Sera D, Pedulla ML, Cresawn SG, Hendrix RW, Hatfull GF. 2006. Teaching scientific inquiry. *Science* 314(5807):1880–81
15. Hanauer DI, Graham MJ, Jacobs-Sera D, Garlena RA, Russell DA, et al. 2022. Broadening access to STEM through the community college: investigating the role of course-based research experiences (CREs). *CBE—Life Sci. Educ.* 21(2):ar38
16. Hanauer DI, Graham MJ, Arnold RJ, Ayuk MA, Balish MF, et al. 2022. Instructional models for course-based research experience (CRE) teaching. *CBE—Life Sci. Educ.* 21(1):ar8
17. DeChenne-Peters SE, Scheuermann NL. 2022. Faculty experiences during the implementation of an introductory biology course-based undergraduate research experience (CURE). *CBE—Life Sci. Educ.* 21(4):ar70
18. Leonetti CT, Lindberg H, Schwake DO, Cotter RL. 2023. A call to assess the impacts of course-based undergraduate research experiences for career and technical education, allied health, and underrepresented students at community colleges. *CBE—Life Sci. Educ.* 22(1):ar4
19. Lopatto D, Hauser C, Jones CJ, Paetkau D, Chandrasekaran V, et al. 2014. A central support system can facilitate implementation and sustainability of a classroom-based undergraduate research experience (CURE) in genomics. *CBE—Life Sci. Educ.* 13(4):711–23
20. Croonquist P, Falkenberg V, Minkovsky N, Sawa A, Skerritt M, et al. 2023. The genomics education partnership: first findings on genomics research in community colleges. *Scholarsh. Pr. Undergrad. Res.* 6(3):17–28
21. Hurley A, Chevrette MG, Acharya DD, Lozano GL, Garavito M, et al. 2021. Tiny Earth: a big idea for STEM education and antibiotic discovery. *mBio* 12(1):e03432-20
22. Shaffer CD, Alvarez C, Bailey C, Barnard D, Bhalla S, et al. 2010. The Genomics Education Partnership: successful integration of research into laboratory classes at a diverse group of undergraduate institutions. *CBE—Life Sci. Educ.* 9(1):55–69
23. Hatfull GF. 2015. Dark matter of the biosphere: the amazing world of bacteriophage diversity. *J. Virol.* 89(16):8107–10
24. Hatfull GF, Racaniello V. 2014. PHIRE and TWiV: experiences in bringing virology to new audiences. *Annu. Rev. Virol.* 1:37–53

6. A commentary emphasizing changing the culture of science by putting inclusive diversity at the center.

12. Describes the SEA-PHAGES program, with evidence for scientific and educational effect.

13. A description of phage genomics as an educational platform and seven key attributes of undergraduate research experiences.

25. Heller D, Sivanathan V. 2022. Publishing student-led discoveries in genetics. *G3* 12(8):jkac141
26. Hendrix RW, Smith MCM, Burns RN, Ford ME, Hatfull GF. 1999. Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *PNAS* 96(5):2192–97
27. Bar-On YM, Phillips R, Milo R. 2018. The biomass distribution on Earth. *PNAS* 115(25):6506–11
28. Mushegian AR. 2020. Are there 10^{31} virus particles on Earth, or more, or fewer? *J. Bacteriol.* 202(9):10-1128
29. Suttle CA. 2007. Marine viruses—major players in the global ecosystem. *Nat. Rev. Microbiol.* 5(10):801–12
30. Dion MB, Oechslin F, Moineau S. 2020. Phage diversity, genomics and phylogeny. *Nat. Rev. Microbiol.* 18(3):125–38
31. Ndela EO, Roux S, Henke C, Sczyrba A, Ngando TS, et al. 2022. Reekeekee- and roodoodooviruses, two different *Microviridae* clades constituted by the smallest DNA phages. *Virus Evol.* 9(1):veac123
32. Devoto AE, Santini JM, Olm MR, Anantharaman K, Munk P, et al. 2019. Megaphages infect *Prevotella* and variants are widespread in gut microbiomes. *Nat. Microbiol.* 4(4):693–700
33. Al-Shayeb B, Sachdeva R, Chen L-X, Ward F, Munk P, et al. 2020. Clades of huge phages from across Earth's ecosystems. *Nature* 578(7795):425–31
34. Harding KR, Kyte N, Fineran PC. 2023. Jumbo phages. *Curr. Biol.* 33(14):R750–51
35. Turner D, Shkoporov AN, Lood C, Millard AD, Dutilh BE, et al. 2023. Abolishment of morphology-based taxa and change to binomial species names: 2022 taxonomy update of the ICTV Bacterial Viruses Subcommittee. *Arch. Virol.* 168(2):74
36. Hatfull GF. 2020. Actinobacteriophages: genomics, dynamics, and applications. *Annu. Rev. Virol.* 7:37–61
37. Brum JR, Sullivan MB. 2015. Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nat. Rev. Microbiol.* 13(3):147–59
38. Mirzaei MK, Maurice CF. 2017. Ménage à trois in the human gut: interactions between host, bacteria and phages. *Nat. Rev. Microbiol.* 15(7):397–408
39. Pope WH, Mavrich TN, Garlena RA, Guerrero-Bustamante CA, Jacobs-Sera D, et al. 2017. Bacteriophages of *Gordonia* spp. display a spectrum of diversity and genetic relationships. *mBio* 8(4):e01069-17
40. Pope WH, Bowman CA, Russell DA, Jacobs-Sera D, Asai DJ, et al. 2015. Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. *eLife* 4:e06416
41. Jacobs-Sera D, Abad LA, Alvey RM, Anders KR, Aull HG, et al. 2020. Genomic diversity of bacteriophages infecting *Microbacterium* spp. *PLOS ONE* 15(6):e0234636
42. Clokie MRJ, Millard AD, Letarov AV, Heaphy S. 2011. Phages in nature. *Bacteriophage* 1(1):31–45
43. Hampton HG, Watson BNJ, Fineran PC. 2020. The arms race between bacteria and their phage foes. *Nature* 577(7790):327–36
44. Georjon H, Bernheim A. 2023. The highly diverse antiphage defence systems of bacteria. *Nat. Rev. Microbiol.* 21(10):686–700
45. Hatfull GF, Dedrick RM, Schooley RT. 2021. Phage therapy for antibiotic-resistant bacterial infections. *Annu. Rev. Med.* 73:197–211
46. Nagel T, Musila L, Muthoni M, Nikolich M, Nakavuma JL, Clokie MR. 2022. Phage banks as potential tools to rapidly and cost-effectively manage antimicrobial resistance in the developing world. *Curr. Opin. Virol.* 53:101208
47. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2022. ColabFold: making protein folding accessible to all. *Nat. Methods* 19(6):679–82
48. Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, et al. 2023. UCSF ChimeraX: tools for structure building and analysis. *Protein Sci.* 32(11):e4792
49. Hanauer DI, Graham MJ, Hatfull GF. 2016. A measure of college student persistence in the sciences (PITS). *CBE—Life Sci. Educ.* 15(4):ar54
50. Hanauer DI, Zhang T, Graham MJ, Adams SD, Ahumada-Santos YP, et al. 2023. Models of classroom assessment for course-based research experiences. *Front. Educ.* 8:1279921
51. Barka EA, Vatsa P, Sanchez L, Gaveau-Vaillant N, Jacquard C, et al. 2016. Taxonomy, physiology, and natural products of *Actinobacteria*. *Microbiol. Mol. Biol. Rev.* 80(1):1–43
52. Russell DA, Hatfull GF. 2016. PhagesDB: the actinobacteriophage database. *Bioinformatics* 33(5):784–86
53. Cresawn SG, Bogel M, Day N, Jacobs-Sera D, Hendrix RW, Hatfull GF. 2011. Phamerator: a bioinformatic tool for comparative bacteriophage genomics. *BMC Bioinform.* 12(1):395

54. Pollenz RS, Barnhill K, Biggs A, Bland J, Carter V, et al. 2023. A genome-wide cytotoxicity screen of Cluster F1 mycobacteriophage Girr reveals novel inhibitors of *Mycobacterium smegmatis* growth. bioRxiv 2023.08.04.552056. **https://doi.org/10.1101/2023.08.04.552056**
55. Amaya I, Edwards K, Wise BM, Bhattacharyya A, Pablo CHD, et al. 2023. A genome-wide overexpression screen reveals *Mycobacterium smegmatis* growth inhibitors encoded by mycobacteriophage Hammy. *G3* 12:jkad240
56. Heller D, Amaya I, Mohamed A, Ali I, Mavrodi D, et al. 2022. Systematic overexpression of genes encoded by mycobacteriophage Waterfoul reveals novel inhibitors of mycobacterial growth. *G3* 12(8):jkac140
57. Hanauer DI, Hatfull G. 2015. Measuring networking as an outcome variable in undergraduate research experiences. *CBE—Life Sci. Educ.* 14(4):ar38
58. Dulberger CL, Rubin EJ, Boutte CC. 2020. The mycobacterial cell envelope—a moving target. *Nat. Rev. Microbiol.* 18(1):47–59
59. Podgorski J, Calabrese J, Alexandrescu L, Jacobs-Sera D, Pope W, et al. 2020. Structures of three actinobacteriophage capsids: roles of symmetry and accessory proteins. *Viruses* 12(3):294
60. Podgorski JM, Freeman K, Gosselin S, Huet A, Conway JF, et al. 2023. A structural dendrogram of the actinobacteriophage major capsid proteins provides important structural insights into the evolution of capsid stability. *Structure* 31(3):282–94.e5
61. Freeman KG, Robotham AC, Parks OB, Abad L, Jacobs-Sera D, et al. 2023. Virion glycosylation influences mycobacteriophage immune recognition. *Cell Host Microbe* 31(7):1216–31.e6
62. Dulberger CL, Guerrero-Bustamante CA, Owen SV, Wilson S, Wuo MG, et al. 2023. Mycobacterial nucleoid-associated protein Lsr2 is required for productive mycobacteriophage infection. *Nat. Microbiol.* 8(4):695–710
63. Gordon BRG, Li Y, Wang L, Sintsova A, van Bakel H, et al. 2010. Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in *Mycobacterium tuberculosis*. *PNAS* 107(11):5154–59
64. Dedrick RM, Smith BE, Garlena RA, Russell DA, Aull HG, et al. 2021. *Mycobacterium abscessus* strain morphotype determines phage susceptibility, the repertoire of therapeutically useful phages, and phage resistance. *mBio* 12(2):e03431-20
65. Wetzel KS, Illouz M, Abad L, Aull HG, Russell DA, et al. 2023. Therapeutically useful mycobacteriophages BPs and Muddy require trehalose polyphleates. *Nat. Microbiol.* 8(9):1717–31
66. Gauthier CH, Cresawn SG, Hatfull GF. 2022. PhaMMseqs: a new pipeline for constructing phage gene phamilies using MMseqs2. *G3* 12(11):jkac233
67. Pedulla ML, Ford ME, Houtz JM, Karthikeyan T, Wadsworth C, et al. 2003. Origins of highly mosaic mycobacteriophage genomes. *Cell* 113(2):171–82
68. Hatfull GF, Hendrix RW. 2011. Bacteriophages and their genomes. *Curr. Opin. Virol.* 1(4):298–303
69. Mavrich TN, Hatfull GF. 2017. Bacteriophage evolution differs by host, lifestyle and genome. *Nat. Microbiol.* 2(9):17112
70. Gauthier CH, Hatfull GF. 2023. PhamClust: a phage genome clustering tool using proteomic equivalence. *mSystems* 8(5):e00443-23
71. Payne KM, Hatfull GF. 2012. Mycobacteriophage endolysins: diverse and modular enzymes with multiple catalytic activities. *PLOS ONE* 7(3):e34052
72. Pollenz RS, Bland J, Pope WH. 2022. Bioinformatic characterization of endolysins and holin-like membrane proteins in the lysis cassette of phages that infect *Gordonia rubripertincta*. *PLOS ONE* 17(11):e0276603
73. 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(2):237–48
74. Wetzel KS, Aull HG, Zack KM, Garlena RA, Hatfull GF. 2020. Protein-mediated and RNA-based origins of replication of extrachromosomal mycobacterial prophages. *mBio* 11(2):e00385-20
75. Wetzel KS, Guerrero-Bustamante CA, Dedrick RM, Ko C-C, Freeman KG, et al. 2021. CRISPY-BRED and CRISPY-BRIP: efficient bacteriophage engineering. *Sci. Rep.* 11(1):6796
76. Marinelli LJ, Piuri M, Swigoňová Z, Balachandran A, Oldfield LM, et al. 2008. BRED: a simple and powerful tool for constructing mutant and recombinant bacteriophage genomes. *PLOS ONE* 3(12):e3957

56. Describes the characterization of phage genes in the SEA-GENES program.

77. Dedrick RM, Marinelli LJ, Newton GL, Pogliano K, Pogliano J, Hatfull GF. 2013. Functional requirements for bacteriophage growth: gene essentiality and expression in mycobacteriophage Giles. *Mol. Microbiol.* 88(3):577–89
78. Dedrick RM, Mavrich TN, Ng WL, Hatfull GF. 2017. Expression and evolutionary patterns of mycobacteriophage D29 and its temperate close relatives. *BMC Microbiol.* 17(1):225
79. Dedrick RM, Bustamante CAG, Garlena RA, Pinches RS, Cornely K, Hatfull GF. 2019. Mycobacteriophage ZoeJ: a broad host-range close relative of mycobacteriophage TM4. *Tuberculosis* 115:14–23
80. Dedrick RM, Jacobs-Sera D, Bustamante CAG, Garlena RA, Mavrich TN, et al. 2017. Prophage-mediated defence against viral attack and viral counter-defence. *Nat. Microbiol.* 2(3):16251
81. Gentile GM, Wetzel KS, Dedrick RM, Montgomery MT, Garlena RA, et al. 2019. More evidence of collusion: a new prophage-mediated viral defense system encoded by mycobacteriophage Sbash. *mBio* 10(2):e00196-19
82. Montgomery MT, Bustamante CAG, Dedrick RM, Jacobs-Sera D, Hatfull GF. 2019. Yet more evidence of collusion: a new viral defense system encoded by *Gordonia* phage CarolAnn. *mBio* 10(2):e02417-18
83. Dedrick RM, Aull HG, Jacobs-Sera D, Garlena RA, Russell DA, et al. 2021. The prophage and plasmid mobilome as a likely driver of *Mycobacterium abscessus* diversity. *mBio* 12(2):e03441-20
84. Ko C-C, Hatfull GF. 2020. Identification of mycobacteriophage toxic genes reveals new features of mycobacterial physiology and morphology. *Sci. Rep.* 10(1):14670
85. Binsabaan SA, Freeman KG, Hatfull GF, VanDemark AP. 2023. The cytotoxic mycobacteriophage protein Phaedrus gp82 interacts with and modulates the activity of the host ATPase, MoxR. *J. Mol. Biol.* 435(20):168261

86. Reports the first therapeutic use of bacteriophages for a mycobacterial infection.

86. Dedrick RM, Guerrero-Bustamante CA, Garlena RA, Russell DA, Ford K, et al. 2019. Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant *Mycobacterium abscessus*. *Nat. Med.* 25(5):730–33

87. Petrova ZO, Broussard GW, Hatfull GF. 2015. Mycobacteriophage-repressor-mediated immunity as a selectable genetic marker: Adephagia and BPs repressor selection. *Microbiology* 161(8):1539–51
88. Donnelly-Wu MK, Jacobs WR, Hatfull GF. 1993. Superinfection immunity of mycobacteriophage L5: applications for genetic transformation of mycobacteria. *Mol. Microbiol.* 7(3):407–17
89. McGinnis RJ, Brambley CA, Stamey B, Green WC, Gragg KN, et al. 2022. A monomeric mycobacteriophage immunity repressor utilizes two domains to recognize an asymmetric DNA sequence. *Nat. Commun.* 13(1):4105
90. Little JS, Dedrick RM, Freeman KG, Cristinziano M, Smith BE, et al. 2022. Bacteriophage treatment of disseminated cutaneous *Mycobacterium chelonae* infection. *Nat. Commun.* 13(1):2313
91. Nick JA, Dedrick RM, Gray AL, Vladar EK, Smith BE, et al. 2022. Host and pathogen response to bacteriophage engineered against *Mycobacterium abscessus* lung infection. *Cell* 185:1860–74
92. Dedrick RM, Smith BE, Cristinziano M, Freeman KG, Jacobs-Sera D, et al. 2022. Phage therapy of *Mycobacterium* infections: compassionate-use of phages in twenty patients with drug-resistant mycobacterial disease. *Clin. Infect. Dis.* 76(1):103–12
93. Gauthier CH, Abad L, Venbakkam AK, Malnak J, Russell DA, Hatfull GF. 2022. DEPhT: a novel approach for efficient prophage discovery and precise extraction. *Nucleic Acids Res.* 50(13):e75
94. Abad L, Gauthier CH, Florian I, Jacobs-Sera D, Hatfull GF. 2023. The heterogenous and diverse population of prophages in *Mycobacterium* genomes. *mSystems* 8:e0044623
95. Huson DH. 1998. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* 14(1):68–73