National Genomics Research Initiative

FIRST ANNUAL SYMPOSIUM

HOWARD HUGHES MEDICAL INSTITUTE
Science Education Alliance
Janelia Farm Research Campus
Ashburn, Virginia
June 21, 2009

Abstracts

Office of Grants and Special Programs
Peter J. Bruns, Ph.D.
Vice President

Science Education Alliance
Tuajuanda C. Jordan, Ph.D.
Director

Lucia P. Barker, Ph.D.
Program Officer

Kevin W. Bradley
Research Technician

Razi Khaja
Bioinformatics Specialist

Melvina F. Lewis
Program Assistant

Credits
Electron Micrograph on the cover:
Richard Fetter, Ph.D.
Shared Resource Manager, Electron Microscopy
HHMI's Janelia Farm Research Campus

Aaron Best, Ph.D.
Hope College

Student Talks

Faculty Posters

A SURVEY OF MYCOBACTERIOPHAGES AT OREGON STATE UNIVERSITY

Anjali R. Menon[1], *Andrew Woodall*[1], *Christine E. Schnitzler*[2], *Dr. Barbara J. Taylor*[2], *Dr. Dee R. Denver*[2], *Robert B. Bloom*[1], *Alyssa B. Carey*[1], *Lauren W. Forbes*[1], *Roberto A. Garcia*[1], *Shannon E. Goff*[1], *Daryl M. Khaw*[1], *Junghee Kim*[1], *Nick C. Lowery*[1], *Rachel A. Miller*[1], *Elvis T. Nguyen*[1], *Tamsen M. Polley*[1], *Erica J. Puopolo*[1], *Jonathan W. Shepardson*[1].

[1]*BI211-213 Principles of Biology Course*, [2]*Zoology Department*, *Oregon State University*, *Corvallis*, *OR 97331*

Bacteriophages are estimated to be among the most numerous and diverse entities in the biosphere. There has been increasing interest in Mycobacteriophage species, as the *Mycobacterium* genus of Actinobacteria includes pathogens known to cause mammalian tuberculosis and leprosy. To approach the questions of morphological and genetic diversity, fourteen students in the Phage Genomics laboratory in the BI211-213 Principles of Biology course conducted a survey of mycobacteriophage species from sites within two compost recycling centers at Oregon State University. After enrichment, we isolated and plaque purified different phages from 10 isolates. Eleven of the phages were lytic, producing clear plaques, and three were lysogenic, producing turbid or bulls-eye plaque morphology. The DNA from high-titer lysates was restricted with six different restriction enzymes and subsequently run on an agarose gel, revealing fourteen genetically distinct phage genomes. The morphology of the phage was determined by staining phage samples with phosphotungstic acid and viewing with an electron microscope. All phages were *Siphoviridae*, with an average capsid diameter of 75 nm, and an average tail length of 171 nm. Both DNA and high-titer lysate were archived for future reference. Genomic DNA from three different phage samples were sent to the Joint Genome Institute for consideration and one, *Colbert*, was selected for sequencing and assembly. Using Glimmer and GeneMark programs to identify putative genes, and Apollo to call genes, we found that the 67 kbp Colbert genome has 99 called genes. Once the genome had been annotated, the genes and genomes were compared using Phamerator analysis. Based on the similarity of the Colbert genome to other sequenced phage genomes, we have tentatively concluded that Colbert belongs to the "B" cluster of phages. In summary, we found that all phages were *Siphoviridae*, regardless of whether the individual phages were lytic or lysogenic, had distinct restriction digest patterns and that Colbert was likely a new member of the B cluster of mycobacteriophages.

CHARACTERIZATION OF ISLAND-3 AND ITS RELATIVES IN MYCOBACTERIOPHAGE CLUSTER I

Turi A. Alcoser, Lisa M. Alexander, Ian M. Bayles, Keshav Budwal, Ian W. Campbell, ***Lianne B. Cohen****, Belle E.V. English, Laura Z. Filliger, Tyler M. Fox, Stephanie L. Guerra, Siping He, Kaitlin E. Healy, Paul G. Jasinto, Andrew J. Medenbach, Rachel E. Pferdehirt, Michael J. Reiss, Judith Savitskaya, Madhav K. Shroff, Jasper Thompson, Hannah S. Wirtshafter, David A. Zaidins, Kathryn E. Sheldon, Jonathan W. Jarvik, A. Javier Lopez.*

Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA 15213
lbcohen@andrew.cmu.edu

Mycobacteriophage Island-3 was isolated from a soil sample taken on the campus of Carnegie Mellon University in September 2008. It is a syphovirus with a prolate ellipsoid head morphology shared with two other mycobacteriophages, Brujita and Che9c, which previously defined phage Cluster I. Sequencing confirmed the relationship among the three phages. Island-3 and Brujita are almost identical except for two blocks of approximately 300 bp on the right arm. All three phages contain integrase and excisase genes but they differ at the putative attachment site. Island 3 and Brujita contain identical palindromes downstream of their integrase genes that are likely *att*P sites. This sequence is present in the genomes of several Lactobacilli but not in any sequenced Mycobacterium, calling into question whether these phages integrate into the Mycobacterium genome or even lysogenize at all. To investigate this issue, we isolated candidate lysogens of Island-3 in *M. smegmatis* by screening for surviving colonies after infection at high m.o.i. We confirmed that these are lysogens by three tests: immunity to infection by the parent phage, behavior as infectious centers on lawns of *M. smegmatis*, and detection of phage DNA in the cells by PCR. However, inverse PCR assays designed to test integration via the putative *att*P site failed to detect insertion in the bacterial genome. In contrast to Island-3 and Brujita, Che9c contains a complete tyrosine tRNA sequence immediately downstream of its integrase gene. A homologous tRNA gene is present in the *M. smegmatis* genome, raising the possibility that Che9c uses this sequence for integration. This hypothesis was tested by isolating Che9c lysogens and verifying that the prophage had inserted within the bacterial tRNA gene. The three phages of Cluster I exhibit complex immunity relationships: whereas Island-3 lysogens are resistant only to Island-3 and the closely related Brujita, Che9c lysogens are resistant to infection by all three members of Cluster I. The basis for this difference is not understood but may indicate that Che9c has a phage exclusion system that is not based on classical (repressor-based) immunity.

GENOMIC AND MORPHOLOGIC CHARACTERIZATION OF THE *MYCOBACTERIUM SMEGMATIS* BACTERIOPHAGE EAGLE

Katherine L. Belfield, *Jillian C. Trent, Zein Al-Atrache, Lynn O. Lewis and Kathryn E. Loesser-Casey.*
Department of Biology, University of Mary Washington, Fredericksburg, VA 22401

Background: The University of Mary Washington participated in a new program funded and sponsored by the HHMI Science Education Alliance (SEA) where first year students isolated and characterized novel bacteriophages from the environment using *Mycobacterium smegmatis* as a host.

Methods: Bacteriophages were isolated by 40 students from soil samples collected on the campus of the University of Mary Washington using an enrichment protocol. Each student picked and purified plaques until a single phage was isolated. Each student then isolated DNA digested the DNA with restriction endonucleases and compared digestion patterns. Several phage DNA samples were submitted to JGI for approval, and Mycobacteriophage Eagle was chosen, by JGI, for sequencing. Sequencing of Eagle was difficult because of the GC rich regions and "hard stops", which led to difficulty in generating a single contig. Therefore annotation began late in the second semester. While waiting for the JGI finished sequence, students designed primers for an ambiguous region of the Eagle genome, and sequencing of the ambiguous region was attempted on the UMW campus.

Results: Mycobacteriophage Eagle has a hexagonal capsid, approximately 44 nm in diameter, and a long, flexible tail approximately 350 nm long. The genome is approximately 52 kbp long and apparently linear (though the ends are being re-sequenced). Using Apollo, 42 genes were called and annotated (using an E value cut off of 10^{-25}). Annotated gene products included putative major and minor tail proteins, a putative tape measure protein, and putative enzymes DNA polymerase and integrase, among others. After annotation, the order of gene products appears to be the reverse of most other published Mycobacteriophages. The first half of the genome appears to be oriented in one direction, while the second half is oriented in the opposite direction.

Conclusion: The matches found with other mycobacteriophages using BLASTx were almost exclusively those in the A2 cluster, leading us to believe that Mycobacteriophage Eagle should be classified in the A2 cluster.

GENOMICS OF MYCOBACTERIOPHAGES FOUND IN THE ENVIRONS OF UC SANTA CRUZ

Trevor P. Sughrue, *Agustin Borjon, Kimberly R. Davis, Garrett V. Hagopian, Nathan A. Holz, Tomas Kasza, Victor Mac, Kaylee M. Nicholson, Lauren R. Rodriguez, Jeffrey D. Rubin, Jessica A. Ruby, Erica M. Shepard, Clark L. Straub, Brandon C. Yee, Janine Ilagan, John Paul Donohue, Melissa Jurica, Lourdes Valenzuela, Manuel Ares, Jr., Grant A. Hartzog.*

Department of Molecular, Cell & Developmental Biology, University of California, Santa Cruz, CA 95064

Sampling forest soil, gardens, fountains and compost piles on or near the UCSC campus, we isolated 14 bacteriophages that infect the benign soil bacterium *Mycobacterium smegmatis.* Both lytic and lysogenic bacteriophages were purified, and they exhibited a range of plaque morphologies and growth characteristics. DNA restriction analyses of their genomes indicates that they represent at least 12 genetically nonidentical viruses. Electron microscopy of the viral particles shows that they vary in their size and structure. In collaboration with the HHMI and the Joint Genome Institute, we sequenced the genome of one bacteriophage, LRRHood, which exhibits the contractile tail and large capsid morphology of the myovirus class. LRRHood has a circular double-stranded DNA genome of 154,349 basepairs. LRRHood is a member of the C1 subgroup of mycobacteriophages and is most closely related to mycobacteriophage Cali. We annotated 221 protein encoding genes, 32 tRNAs and one tmRNA gene in the LRRHood genome. To facilitate study of LRRHood, we created a Phage Genome Browser. Using this browser and Phamerator, we identified insertions and deletions in the LRRHood genome relative to other mycobacteriophages. The largest is 1327 bp and contains 4 genes, including one encoding an apparent transcriptional repressor protein. The inserted sequences are absent from other group C1 mycobacteriophages, but are present in group A1 and A2 mycophages, at 96-98% identity. The insert in LRRHood is similar to 985 bp of Fruitloop and Ms6 and a total of 820 bp from two nearby regions of KBG, with an overlapping segment of 585 bp found in both that encodes the repressor . The entire inserted region in LRRHood is bounded by a 29 bp direct repeat present in a single copy in the other C1 phages, suggesting that insertion proceeded through repair of a staggered break in the DNA. LRRHood is unusual in possessing such a large segment of group A phage genes. We speculate that to the inserted segment may confer immunity to infection by group A1 and A2 mycobacteriophages.

THE ISOLATION AND CHARACTERIZATION OF MYCOBACTERIOPHAGES J-GLADIATOR, UNCLE HOWIE, AND ANGELICA

Suchita Rastogi, *Martin Y. Fan, Ho Yee Joyce Fung, Michael B. Schultz, Joseph E. Marcus, Samantha T. Alford, Alexander G. Anderson, William B. Barshop, Lisa Deng, Vincent J. Huang, Peter M. Hynes, Patrick C. Ng, Hannah S. Rabinowitz, Corwin N. Rhyan, Erika F. Sims, Emilie G. Weisser, Bo Zhang, and Thomas Shull.*

Department of Biology, Washington University, St. Louis, MO 63130

Three novel mycobacteriophages, J-Gladiator, Uncle Howie, and Angelica, were isolated from soil samples from St. Louis, Missouri, purified on plates of *Mycobacterium smegmatis*, and initially characterized based on plaque morphology and electron micrographs. Subsequent genome sequencing revealed that J-Gladiator, Uncle Howie, and Angelica respectively consisted of 49,681, 68,016, and 51,533 base pairs; 93, 98, and 73 genes; 61.44%, 66.5%, and 66.38% GC contents; and 3, 0, and 1 putative tRNA genes. Although all three phages had many novel genes with no homology to previously sequenced organisms, the homologies that did exist and the overall genome organization suggest that J-Gladiator belongs to mycobacteriophage cluster A2, Uncle Howie to cluster B1, and Angelica perhaps to an entirely new cluster of its own. Where homology was found, putative function could be assigned to the phages' gene products. Additional gene characterization of J-Gladiator in particular suggests multiple mechanisms to increase the rate of viral replication. It also intimates a complex evolutionary history for mycobacteriophages marked by both horizontal and vertical gene transfer.

ISOLATION AND GENOME SEQUENCING OF A NOVEL MYCOBACTERIOPHAGE, PUMPKIN

Amanda Barber, Rebekah Chew, Jeffrey Corajod, ***Anne Georges****, Catherine Harmon, Erin Hildebrandt, Erica Jansen, Justin Knutter, Bruce Kraay, Jennifer LaRoche, Courtney Long, Christine Murray, Caitlin Peirce, Luke Peterson, Andrew Rose, Dale Schipper, Jessica Simmons, Robert Sjoholm, Ingrid Slette, Angelica Willis,* ***Aaron A. Best*** *and Joseph Stukey.*

Department of Biology, Hope College, Holland, MI 49423

Twenty first-year college students, participating in a new HHMI-sponsored educational initiative to involve beginning undergraduate students in authentic scientific research, successfully isolated and partially characterized 20 mycobacteriophages from soil samples collected near Holland, Michigan. The newly isolated bacteriophages infected the host *Mycobacterium smegmatis* and yielded plaque morphologies ranging in size, shape, and clarity. Both lytic and temperate phages are represented in this collection. High-titer (> 10^9 pfu/ml) lysates of plaque-purified phage were used to prepare genomic DNA samples for restriction digest analyses with 6 enzymes. Seven of the phage did not produce a restriction pattern with these enzymes, 10 yielded unique restriction patterns, and 2 yielded apparently identical patterns. However, plaque morphologies of the latter two were distinct, suggesting they are distinct phages. One isolate was chosen for complete genome sequencing and comparative genomic analysis. Mycobacteriophage Pumpkin yielded a complex plaque morphology exemplified by a clear center circle of 1-2 mm diameter surrounded by a similar-sized turbid area, circular to comet in shape, after 24 hours incubation at 37°C. Comparison of the restriction digest pattern for Pumpkin with 60 existing mycobacteriophage genomes suggested that Pumpkin is most closely related to cluster E comprised of mycobacteriophages 244, Cjw1, Porky and Kostya. Draft genome sequence data confirm this relationship, but reveal features unique to Pumpkin. The genome of Pumpkin is approximately 74 Kb, 63% GC, and contains ~140 genes in agreement with the genome characteristics of cluster E members. A detailed anlaysis of the complete genome sequence and comparison with sequenced members of cluster E is the subject of the second semester of this year-long course and is presented.

ISOLATION AND GENOMIC ANALYSIS OF SIX NOVEL SOIL BACTERIOPHAGES

Kathryn M. Sinclair, *James Madison University, Harrisonburg, VA 22807*
Louise Temple, Department of Integrated Science and Technology, James Madison University, Harrisonburg, VA 22807
Steven G. Cresawn, Department of Biology, James Madison University, Harrisonburg, VA 22807

Background: Bacteriophages are ubiquitous organisms and are thought to outnumber all other forms of life on the planet combined. As such, they are both ecologically significant and highly available for study. As pathogens of bacteria, phages have great clinical potential as therapeutics, diagnostics, and as a source of proteins with the capacity to manipulate the host cell. Genomes of dsDNA-containing bacteriophages are often mosaics. A distinguishing feature of the mosaic architecture is the presence of genome segments that are highly conserved between two or more phages but are punctuated by short stretches of sequence that are unrelated between the genomes. A genomics approach to studying bacteriophages can shed light on their clinical utility, genome evolution, and environmental significance.

Methods: Phages were extracted from soil and propagated on lawns of *Mycobacterium smegmatis* or *Bacillus pumilis*. Phage particles were observed using a TEM, and all were of the siphoviridial morphotype. Phage DNA was sequenced using either the Sanger or pyrosequencing method. Assembled genome sequences were viewed in Sequencher and Consed, and annotated with Glimmer, GeneMark, Apollo, and BLAST. Once annotated, the sequenced genomes were compared with other phages using Phamerator.

Results: Twenty-five mycobacteriophages were isolated in this study and one, Duke, was sequenced by the Sanger method. Another mycobacteriophage, Maury, and two *B. pumilis* phages, BC and Polaris, were sequenced via pyrosequencing and analyzed. Lastly, a novel prophage of *B. pumilis*, Pleiades, was discovered, subjected to pyrosequencing, and shown to be functional.

Conclusions: While the genetic diversity of mycobacteriophages is high, it appears that a limited number of patterns exist from which a mycobacteriophage genome can be built. Phage Maury fits the pattern of a cluster E phage. Two large contigs emerged from the Duke sample. One resembles a cluster B3 phage, the other a novel cluster. The Bacillus phages have similar relationships. The Polaris genome is essentially a subset of the BC genome, while Pleiades resembles nothing in GenBank.

ISOLATION AND GENOMIC CHARACTERIZATION OF MYCOBACTERIOPHAGE SP. HOPE

Marie Anderson[†], *Tara Avery*†, *Tiffany Bryan*, *Ashiya Buckels*[†], *Keithara Davis*[†], *Breyon Dixon*[†], *Arianna Forbes*[†], *Zindzi George*[†], *Audrianna Gordon*[†], ***Shalaina Griffin***[†], *Tiffany Grigley*[€], *Raven Hardy*[†], *Lateria Haynes*[†], *KaTerri Kelly*[†], *Kinnari Matheson*, *Jazmyn McCloud*[†], *Morgan McKenzie*[†], *Sterling Mitchell*[†], *Orianne Morrison*[†], *Lindsay Parnell*[†], *Brianna Peppers*[†], *Christina Roberts*[†], *Yetta Robinson*[†], *Kerona Sharpe*[†], *Blaire Spaulding*[†], *Starrissa Winters*[†], *Melina Zuniga*[†], *Charles Hardnett*[‡],*and Cynthia Bauerle*[†].

Departments of Biology[†], *Economics*[€], *and Computer Science*[‡], *Spelman College*, *Atlanta, GA 30314*

We report the characterization of a novel bacterial virus isolated from soil collected at Arabia Mountain Heritage Park, GA in September 2008. Initial isolates were obtained by enrichment culture with host bacterium *Mycobacterium smegmatis*. Multiple dilution passages confirmed the purity of each phage preparation. Mycobacteriophage Hope exhibits medium-sized clear plaques on bacterial-seeded lawns. Electron microscopy using Phosphotungstic acid negative staining revealed a thin-walled hexagonal capsid with a long flexible tail characteristic of the siphoviridae family of mycobacteriophages. RFLP analysis of isolated Hope genomic DNA confirmed a unique phage compared to all other isolates in the lab as well as control phage D29.

Hope DNA was submitted to the Los Alamos National Laboratory (LANL) for DNA sequencing in December 2008, and finished sequence was returned in February 2009. Hope genome is a double-stranded linear molecule of 41,901 base pairs, containing 58-59 putative genes. BLAST analysis of individual ORFs reveals strong similarity to previously characterized BPs and Halo mycophages (100% and 99% identity at the nucleotide level respectively), placing Hope in the G subcluster of siphoviridae. Comparison of Hope, BPS, and Halo genomes reveals a specific rearrangement near the 3' end involving a 439 bp fragment which contains Hope hypothetical protein gene gp53. This fragment is highly conserved in G subcluster isolates and is also observed in the genomes of mycobacteriophages from F and I subclusters. Similarly, a conserved 1283 bp fragment containing part of Hope gp29 appears to be distributed in members of the B1 subcluster. While Hope genome contributes no new nucleotide sequence information to the expanding mycobacteriophage database, it represents a novel arrangement of G-type mycobacteriophage genomes.

ISOLATION AND GENOMIC SEQUENCE CHARACTERIZATION OF MYCOBACTERIOPHAGE SP. PEACHES FROM A SOIL SAMPLE IN NORTHEAST LOUISIANA

John Robert Warner, *Jeremy S. Harmson, Chris R. Gissendanner, and Ann M. Findley.*
Department of Biology, University of Louisiana at Monroe, Monroe, LA 71209

During the fall 2008 semester, ULM freshman Biology students successfully isolated fifteen distinct bacteriophages via direct plating or enrichment regimes. Isolates were subjected to spot test analysis, repetitive phage purification plating, and an empirical testing protocol that led to a ten plate infection and harvesting of high titer lysates (10^9 – 10^{10} pfu/ml). Lysates were processed for transmission electron microscopy with negative staining and 13 of the 15 isolates were found to display the siphoviral morphotype. DNA was isolated from each phage and characterized by restriction digestion analysis using the *Bam*HI, *Cla*I, *Eco*RI, *Hae*III, and *Hind*III endonucleases. Mycobacteriophage sp. Peaches was submitted for library construction and genome sequencing at the DOE Joint Genome Institute. The initial phase of sequencing generated 3 contigs and 1 scaffold with a total genome size of over 50kb. During the spring 2009 semester, the SEA-NGRI freshmen annotated the Peaches genome. This annotation process included a finishing analysis of the draft sequence, gene calling, and assignment of predicted gene functions using the Sequencher, Glimmer/GeneMark and Phamerator programs. The finished Peaches genome places it within the A2 cluster of known Mycobacteriophages and is most closely related to Bxz2. Peaches has a final completed genome size of 51376 bp including a 10bp 3' overhang, 85 open reading frames (ORFs), and a GC content of 63.8722%. The left arm of the genome primarily encodes structural genes (i.e., major/minor tail, portal, scaffold, and tapemeasure proteins) and utilizes the + strand. The right arm is primarily read from the – strand and contains ORFs with similarity to other phage genes of known functions and ORFs encoding gene products of unknown function. The Peaches genome contains 10 orphams and examples of morons. Putative non-structural ORFs (via BLASTX analyses) include DNA polymerase I, DNAB-like helicase, DNA primase, ser/thr phosphatase, ribonucleotide reductase, and S-integrase. Phamily diagrams were constructed to explicate Peaches functional relatedness to known Mycobacteriophages (e.g., Pham 23 – S-integrase).

ISOLATION, CHARACTERIZATION AND GENOME SEQUENCE OF A NOVEL BACTERIOPHAGE INFECTING *MYCOBACTERIUM SMEGMATIS*

J. Walton, *O. Acoreza, S. Applin, S. Askew, P. Barkanic, C. Bergin, J. Carter, M. Del Toral, M. Farnham, K. Gordon, A. Jackson, K. Hartman, S. Harvey, J. Hudson, K. Javier, R. Lopez, X. Lu, C. Mayfield, M. McDonough, E. McConnell, A. Nunes, A. Perdedo, A. Perz, C. Ruiz, S. Schall, A. Sims, T. Tran, K. Walker, H. Whelan, V. Villanueva, R. Blue, K. Williamson, M. Saha,and M. Forsyth.*

College of William & Mary, Department of Biology, Williamsburg, VA

Bacteriophages have been and continue to be useful in exploring the molecular biology and physiology of bacteria, and are now becoming useful genetic systems for comparative genomics and viral evolution studies. In this study, part of the Howard Hughes Medical Institute Science Education Alliance's (HHMI SEA) National Genome Research Initiative, we describe the isolation and complete genome sequence of one unique bacteriophage infecting *Mycobacterium smegmatis* strain mc^2155 from an environmental source, designated CrimD. CrimD appears to be a member of the Siphoviridae based on transmission electron microscopy. The isolate selected for total genome sequence determination, CrimD, was isolated from a submerged soil within a freshwater stream on the campus of The College of William and Mary in Williamsburg, VA (N 37^o 16.243 W 76^o 42.876). The capsid diameter is ~54 nm and the tail length is ~158 nm. The genome sequence was determined by shotgun sequencing. The genome is 59,798 base pairs in length encoding approximately 90 putative protein products and one putative tRNA product. 27 genes comprising greater than 11% of the genome show no known homology to any sequences in the NCBI database. As such, CrimD exemplifies the genetic diversity in viral soil ecology.

PHINDING PHAGE: ISOLATION AND CHARACTERIZATION OF NOVEL MYCOBACTERIOPHAGES IN SAN DIEGO

Hannah Wang, *Tina Lu, Marcella Erb, Kit Pogliano and Joe Pogliano.*
Division of Biological Sciences, University of California, San Diego, La Jolla, CA 92093-0377

In the fall of 2008, the students of BIMM171A isolated, purified, and characterized phage infecting host *Mycobacterium smegmatis* in San Diego, CA and surrounding areas. Mycobacteriophages were abundant in compost and could be isolated from other environments after enrichment. Nineteen bacteriophages were isolated and characterized using plaque assays, DNA restriction analysis and electron microscopy. Plaque assays suggested that both lytic and lysogenic phages were isolated. Electron Microscopy demonstrated that all the phages were double stranded DNA phages, with 18 *Siphoviridae* and 1 *Myoviridae.* One bacteriophage isolated from direct plating on compost (ET08) was selected for genome sequencing and annotation due to its predicted large genome size and because it was the only *Myoviridae* identified. ET08 was shown to belong to the Cluster C1 contractile tailed mycobacteriophage family. The ET08 genome is predicted to encode 33 tRNAs and 1 tmRNA, which releases nascent chains from stalled ribosomes. The genome also encodes 4 inteins, which are protein-splicing domains that are excised from a host protein after translation. Two of these inteins were inserted into conserved genes that lack inteins in other sequenced cluster C1 phage, while two were inserted into the same target gene in several related phage. Phamerator analysis indicated that one ET08 gene product was a novel Phamily, but tBlastn analysis demonstrated that a similar peptide was encoded by other Cluster C1 phage but that this ORF was not annotated as a gene. Thus, ET08 shares extensive genetic similarity with other Cluster C1 mycobacteriophages.

PUHLTONIOPHAGE ISOLATION, ANNOTATION & PHAMILY IDENTIFICATION

Rebecca Goldstein, *James Sandoz, Steven Caruso.*
University of Maryland, Baltimore County, Baltimore, MD 21250

The goal of this research project was to isolate and characterize a novel bacteriophage, Puhltoniophage. This phage was isolated from a soil sample at UMBC on its host bacteria *Mycobacterium smegmatis*. This lytic phage belongs to the family *siphoviridae* because of its long tail. Data mining software programs (such as Genemark and Glimmer) and specific criteria were used to identify genes in the 68.4 kbp genome. Ninety-four genes were called, eight of which were found to have known functions. Phamerator determined that the most common homologies to Puhltoniophage were Mycobacterium phages Chah, Orion, and PG1. It was additionally found to belong to phamily B1. This was accomplished by comparing it against a database of already-characterized genomes. This project is important because phages can be used therapeutically: medical researchers seek one that can kill pathogens such as *M. tuberculosis*. Also, scientists can use our annotation as a reference to classify other phage genomes.

Faculty Posters

ANALYSIS OF MYCOBACTERIOPHAGE GENOME SEQUENCES AS PART OF THE NGRI GENOME RESEARCH INITIATIVE AT WASHINGTON UNIVERSITY IN ST. LOUIS

Kathleen A. Weston Hafer, *Christopher D. Shaffer, and Sarah C.R. Elgin.*
Department of Biology, Washington University, St. Louis, MO 63130

Seventeen students continued in the second semester of the NGRI course at Washington University (WUSTL). The 3 credit course met for five hours (on two days) each week. In the first semester, each of nine pairs of students isolated and characterized a mycobacteriophage. One phage was sent to JGI, and two other phages were sequenced at WUSTL. In the second semester, students were organized into groups of five or six (assigned by the instructor). When surveyed at the end of the semester, students generally were happy with their groups, thought the size was right, and thought that the groups had worked well together. Throughout the semester, each group worked on improving the draft sequence, calling genes, annotating genes and researching gene function for their phage. At the end of the semester two of the three phage genomes were nearly publication ready, while the third still needs work to finish the sequence. Nonetheless, the majority of the genes could be identified and analyzed in all cases.

At WUSTL, potential biology majors take their first biology course second semester of their freshman year. Thus we had to teach concepts early in the course that students would learn later in Principles of Biology I. Techniques we used to teach material in the NGRI course included short lectures/presentations, problem sets, quizzes, guest lectures, and analysis of primary literature. At the end of the semester groups presented their work both as a class oral presentation and as a poster at a WUSTL sponsored student research symposium; each individual wrote a final paper. Student surveys at the end of the year suggest high satisfaction with the course. Generally there were lots of successes, and some things that need to be done differently, including better spacing of guest lectures and assignments to allow more independent work time at the end of the semester. Students were not concerned about overlap in content between the SEA course and Principles of Biology I, stating this led to better understanding, but were frustrated that the introductory lab taught several skills they had already mastered. We are considering ways to achieve better integration between the two courses.

BIOLOGY 125-126: PHAGE HUNTERS AT THE UNIVERSITY OF MARY WASHINGTON (OR HOW YOU TOO CAN TEACH THIS COURSE WITH 40+ STUDENTS)

Lynn O. Lewis *and* ***Kathryn E. Loesser-Casey****.*
Department of Biology, University of Mary Washington, Fredericksburg, VA 22401

Biology 125-126 was taught as a course for first year students, and is a replacement for our introductory Biology 121-122 sequence. Students were not selected or screened for enrollment in this course, so we had prospective majors as well as students simply fulfilling their science requirement. This meant a wide variety of interests and abilities. The course was designed as three sections of 15 students each but we actually had a total of 40 students complete the first semester. Two sections were taught by Lewis in the Fall, and Loesser-Casey taught one. Because a few students did not continue, each professor taught one larger section during the Spring. Both instructors worked together to ensure that all students received the same information and instruction as much as possible, including giving identical exams. Further, both instructors were frequently present during lab periods to make labs run more smoothly. This also allowed us to utilize our different strengths for all sections. What did we learn? There were a lot of students to keep track of in the first semester, since students were not working at the same pace. Having adequate assistance with lab preparation is essential, and having student lab aides actually in the lab sections would have been helpful (we have recruited two of our current students to act as lab aides for next year). The second semester, even with larger sections, was much easier since all students were working on the same thing at the same time. Because our phage was difficult to sequence, we had time in the spring semester to add several traditional introductory biology labs, which allowed us to ensure that our students received some of the more essential laboratory experiences they would have received in our traditional course. What were our successes? All 40 of our Fall students isolated bacteriophages, although some were probably duplicates. Thirty-eight out of 40 students were able to isolate DNA from their phage. All 35 Spring students had electron micrographs of their phages. And finally, our impression is that a number of students have decided to major in Biology after completing this course.

DISSECTING THE TOOLS OF BIOINFORMATICS: BUILDING SPELFINDER, A PRIMER IDENTIFICATION TOOL IN PYTHON

Charles R. Hardnett and ***Cynthia Bauerle****.*
Departments of Computer Science and Biology, Spelman College, Atlanta, GA, 30314

As participant in the National Genomics Research Initiative supported by HHMI, the Biology department at Spelman College offered a course-based research opportunity for first year science students to isolate and characterize novel mycobacteriophages. In fall 2008, twenty first year students isolated novel mycobacteriophages and conducted standard microscopic and molecular characterization. In spring 2009, fourteen of these students, joined by six advanced biology majors, conducted genomic annotation of one phage isolate using standard sequence analysis approaches and bioinformatic tools available via the Internet. While prior experience *applying* bioinformatics tools varied among the cohort, no students described any prior experience in computer programming. To help students better understand the construction and utility of standard molecular analysis tools, we developed a course activity in which students used the Python programming language to build a primer identification tool that they named SpelFinder. Students were introduced to basic elements of the Python language in a hands-on group exercise, and then used these elements to build simple algorithms. Ultimately, students assembled algorithms into a script that analyzes an input nucleotide sequence and reports potential primer target sites as output. The program was based on an algorithm that reflected the following set of relevant biological criteria:

- Primer length between 18-30 nucleotides,
- GC content is 40% ≤ [GC] ≤60%,
- Sequence may not contain stretches of four or more repeating nucleotides, and
- T_m for the primer is 55°C≤ T_m ≤ 75°C, as calculated by the equation:

$$Tm = [(GC)_{primer} \times 4°C] + [(A+T)_{primer} \times 2°C]$$

Writing the program *de novo* in Python introduced students to basic programming strategies for developing functional algorithms useful in standard bioinformatics tools. Specifically, students were challenged to understand how biological criteria may be translated into a set of rules that drives the algorithm of a sequence analysis software program.

eBIOLOGY I & II - HUNTING PHAGES AT UMBC

Caruso, S., Sandoz, J., and Kelsey, J.
University of Maryland, Baltimore County, Baltimore, MD 21250

The UMBC version of the National Genomics Research Initiative course was offered over two fourteen-week semesters to a total of 26 students. All but five of the participating students were non-STEM majors. Each student was able to isolate a phage using either the standard isolation or enrichment procedure and obtain an electron micrograph of their phage. Nearly every student was able to produce a DNA sample of their bacteriophages, and several were able to produce quantities that made their phage candidates for sequencing by the Joint Genome Institute (JGI). The selected bacteriophage was successfully sequenced by JGI, and annotated by students in the second semester. Students were successfully introduced to the scientific method, biological concepts, and experienced hands-on use of *in vitro* and *in silico* research tools.

HOPE COLLEGE PHAGE GENOMICS RESEARCH COURSE: FIRST-YEAR IMPLEMENTATION AND RESULTS

Aaron A. Best and ***Joseph Stukey****.*
Department of Biology, Hope College, Holland, MI 49423

This past academic year (2008-09) Hope College of Holland, Michigan, offered a phage genomics research laboratory course to 20 first-year college students as part of a new science education experiment, designed and supported by the HHMI and SEA, to test the hypothesis that educational instruction, in the context of an authentic scientific research project, will lead to improved student learning and interest in science. Students were selected from a restricted pool of applicants earning high scores on the ACT and/or SAT college entrance exams. Lab sessions ran twice a week for a total of 5 hours in the fall (phage/DNA isolation semester) and 3 hours in the spring (bioinformatics work). The essential content of the course was delivered, largely in sequence and as described in the manual, and all major goals of the course were achieved. We used the timely shift in course content from phage/DNA isolation in the fall to bioinformatics-based genomic analysis in the spring to introduce several changes involving course mechanics and classroom dynamics to broaden the student learning experience. In the fall students worked independently and presented their findings in research article style written reports whereas in the spring students were organized in teams of five, analyzed only a section of the genome, and highlighted their findings in student-driven journal article discussions and group oral presentations. Multiple efforts were made to bring interdisciplinary coverage to select topics of discussion and to include additional laboratory and analysis skills consistent with the Hope College Biology core course sequence subject matter. About 80% of the students have enrolled in the final course of the Biology core sequence, historically an indicator that students will declare Biology as their major field of study.

IMPLEMENTATION OF THE NGRI RESEARCH COURSE AT UNIVERSITY OF CALIFORNIA, SAN DIEGO

Joe Pogliano and *Kit Pogliano*.
Division of Biological Sciences, University of California, San Diego, La Jolla, CA 92093-0377

Implementation of the NGRI course started in Spring, 2008. We solicited applications from two pools of students, those entering UCSD as Regents Scholars or honors students and those from six science preparatory schools in low income neighborhoods with high numbers of underrepresented minority students. After Fall quarter, two members of the latter group were supplied with tutoring to provide them with the foundation in basic biology necessary to succeed in the course. In fall quarter, the class met in a laboratory twice a week (5 hours on Tuesdays and 4 hours on Thursdays). Students had 24 hr access to the laboratory and many came into the lab daily to continue plaque purification. Fall quarter was intense, since we needed to submit our genomic DNA less than 8 weeks after the course began. We screened 491 samples by direct plating. Mycobacteriophages were reliably obtained only from compost heaps (20/21 compost heaps gave plaques). We performed 33 enrichment cultures, with 16 giving plaques. One phage, ET08, was chosen for sequencing because it was a large *Myoviridae* and had the highest yield and quality of DNA. During the winter quarter, instruction focused on central concepts of molecular biology, phage biology, bioinformatic theory and exercises, to prepare students for annotation. Original research articles were read and discussed to help students with the transition from reading a textbook to reading scientific literature and to familiarize students with commonly used scientific methods and modern concepts of phage biology. Annotation of the 155kb ET08 genome was performed in spring quarter. The genome was divided into 4 uniform pieces that were annotated by 4 teams of five students each. Each of the ~240 genes encoded by ET08 was discussed by the students in a series of lab meeting style presentations. Difficult annotations (short ORFs or those with several possible start sites) were resolved as a group after summarizing the alternatives in a table format and discussing possible alternatives. At the end of the quarter, students prepared a poster and presented the results of their research at an undergraduate research symposium at UCSD.

IMPLEMENTATION OF SEA BACTERIOPHAGE GENOMICS LABORATORY COURSE AT CARNEGIE MELLON UNIVERSITY

Kathryn E. Sheldon, A. Javier Lopez, ***Jonathan W. Jarvik****.*
Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA 15213
jarvik@cmu.edu

We designed this course to stimulate interest in biological research among freshmen in the Mellon College of Science. We selected 10 female and 11 male students using an application essay, without regard to academic record or major. We provided opportunities for individual and team work, to develop self-confidence and collaborative skills. The students represented a broad range of backgrounds, interests and personalities, but all developed high levels of engagement and proficiency. In the Fall, each student isolated one mycobacteriophage and characterized its morphology and restriction pattern. The group chose one phage ("Island-3") for genome sequencing due to its distinctive morphology. We supplemented these activities with experiments to develop a less abstract understanding of infection, replication and viral assembly. Experiments included scanning and transmission EM of plaque sections, in which phage adsorption and intracellular intermediates in viral assembly could be observed. The Spring, which was to be devoted to genome annotation, presented challenges due to delays in receiving the completed sequence and the fact that Island-3 was nearly identical to a known phage. We maintained student engagement in two ways. First, students carried out their own finishing experiments to fill large gaps and resolve poor quality regions in the preliminary sequence. Second, we compared the genomes of Island-3 and its two relatives to generate and test hypotheses about their life cycles and regulation, including ability to lysogenize, integration into the bacterial genome, and patterns of immunity. This approach was successful and had multiple benefits: 1) the value of the bioinformatic analysis became much more obvious; 2) students experienced the entire research cycle, including initial observations, generation of hypotheses, design of tests and experimental validation; 3) the value of combining computational and experimental approaches was highlighted by experiments that revealed novel unexpected behavior; 4) students made novel discoveries about the biology of their phage and its relatives, which rescued the course from an initial disappointment at having isolated a known phage.

IMPLEMENTING A FRESHMAN HONORS LAB IN PHAGE GENOMICS AT UC SANTA CRUZ

Lourdes Valenzuela, Manuel Ares, Jr., ***Grant A. Hartzog****.*
Department of Molecular, Cell & Developmental Biology, University of California, Santa Cruz, CA 95064

UC Santa Cruz is a research intensive university with ~14,500 undergraduate students. UCSC has an admirable record of recruiting, supporting and retaining students from educationally disadvantaged backgrounds, but does less well at meeting the needs of and retaining its best students. To address the needs of the better-prepared and most motivated students entering UCSC, we implemented a year-long honors lab in phage genomics. We targeted entering students by contacting those with scores ≥4 on the AP Biology exam and by making presentations at freshman orientation events. A diverse group of 15 students was selected using an application and interview process designed to distinguish strictly premedical students from those with broader educational interests and motivations. The goal of this course was to provide students with an authentic research experience in which they would take primary responsibility for their own progress and education. We sought to provide an experience that was distinct from and orthogonal to the passive lecture format typical of large universities. We avoided lectures and kept discussions focused on experimental goals, data interpretation and hypothesis testing. We emphasized process over content and resisted the temptation to provide comprehensive coverage of topics peripherally related to the research questions we were addressing. Since this class was taken in addition to a normal academic load of courses, we were careful to provide these experiences in a way that did not harm the student's progress in their other courses. Wherever possible, we explicitly contrasted the experiential learning of the phage class with the more traditional teaching and learning methods used in standard large lecture classes. Although one student transferred out of UCSC following the fall quarter for nonacademic reasons, the remaining students isolated and characterized their own phages. We took advantage of UCSC's strength in genomics to develop the UCSC Phage Genome Browser, providing an additional opportunity to discuss genomics and engage students interested in bioinformatics.

INITIAL IMPLEMENTATION OF THE NGRI GENOME RESEARCH INITIATIVE AT WASHINGTON UNIVERSITY IN ST. LOUIS

Kathleen A. Weston Hafer, *Christopher D. Shaffer, and Sarah C.R. Elgin.*
Department of Biology, Washington University, St. Louis, MO 63130

Eighteen students participated in the inaugural NGRI research course at Washington University in St. Louis (WUSTL). The course was designated a freshman "Focus" course, for which incoming students indicated interest in response to a mailing describing various freshman opportunities. Students were ultimately chosen for participation by the WUSTL Freshman Dean, who matched students with opportunities. The three credit course met for five hours each week (a three hour block on one day and two hours on the other). At the first class meeting, students submitted course schedules. Based on those schedules, the instructors paired students so that each pair would have a member available to work briefly on "non-scheduled" days. This allowed timely purification of isolated phage, and was only needed for about four weeks of the semester. By early October, each pair of students had isolated and purified a phage starting from a soil sample, and produced a high titer lysate for that phage. By mid-November, each group had produced electron micrographs of their phage, and purified and analyzed phage DNA. In addition to the lab work, class time included short presentations on basic biology and lab topics, guest speakers, and two quizzes. Several literature readings were assigned, both Scientific American articles and original literature; for each reading assignment students wrote a 1-2 page reading response, and participated in a classroom discussion of the reading. After student presentations to highlight the attributes of each phage, one phage was chosen for sequencing at JGI, and two other phages were chosen for sequencing at the Washington University Genome Sequencing Center. At the end of the course, student survey results showed that students found the course to be demanding, but in general were very satisfied with the experience. Students were truly invested in the project and anxious to get the results of the genome sequencing. Subsequent analysis of the phage genomes has confirmed that all three phages are novel; they are different from each other and from all previously sequenced mycobacteriophages.

LEARNING BY DOING: A YEAR IN THE PHAGE GENOMICS LABORATORY AT OREGON STATE UNIVERSITY

Dee R. Denver, ***Barbara J. Taylor****, Christy E. Schnitzler, Andrew Woodall.*
Departments of Zoology and Biology, Oregon State University, Corvallis, OR 97331

The guiding principle motivating our implementation of the Phage Genomics laboratory (PGL) was to have students engage in an early primary research experience. The PGL was one of two Honors laboratories associated with a large, year-long Introductory Biology course, Principles of Biology BI 211-213. The fourteen students, who were accepted into the laboratory, were immersed into the collection, isolation, purification and detailed characterization of their mycobacteriophage from the first day. Initially, we started each session with short demonstrations and lectures to introduce techniques, complemented by student reports on their weekly experimental progress. We supplied additional support until students became adept at the techniques and trouble-shooting their own problems. In the second term, we introduced research papers for discussion and provided instruction on molecular biology and computational analysis because these topics were not part of their course sequence. Once our phage genome sequence, Colbert, was finished, each pair of students annotated segments of the genome and the entire class held an annotation 'jamboree', modeled after those held for other genomes. Following Phamerator-based gene family analysis, the students conducted a module on molecular evolutionary analysis for gene phamilies in *Colbert* using Molecular Evolutionary Genetic Analysis (MEGA) and DNA Sequence Polymorphism (DnaSP) programs in order to assess whether natural selection might be acting on particular genes. To close out the year, each student chose one of four different task force teams. Two teams developed communication tools: a web site (Team Communication), which when completed will be housed on the OSU Biology Web Page and a poster (Team Poster), which will be posted on campus. Team Wet Lab tested other sites for the presence of mycobacteriophage and Team Genome prepared and ran another mycobacteriophage genome for Solexa sequencing. All in all, our students were excited by their discoveries, developed skills in critical thinking and research, met challenges, worked very hard and came away with a terrific research experience.

THE PHAGE GENOME RESEARCH INITIATIVE AT THE COLLEGE OF WILLIAM AND MARY: ASSESSMENT OF YEAR 1

Mark H. Forsyth, Margaret S. Saha, and *Kurt Williamson*.
College of William & Mary, Department of Biology, Williamsburg, VA 23185

Background: Two significant challenges in undergraduate science education are (1) incorporating substantive research experiences into freshmen introductory laboratory classes to engender and maintain excitement and interest in science, and (2) retaining students with diverse backgrounds during and following large introductory science courses. The HHMI Phage Genome Research Initiative at the College of William and Mary has focused on specifically addressing both of these issues.

Methods: The WM PGRI program targeted enrollments from admitted William and Mary students potentially interested in science who participated in summer bridge programs. In addition, in order to ensure a diverse group of participants, students who actively sought out the opportunity to join the program and those who successfully competed via an application process were accepted into the class. We employed a strategy for intense faculty involvement in the progress of the students in the PGRI as well as in all other courses, including review sessions, meetings for study strategies, and post-exam reviews. Additional strategies for creating a genuine research experience included: a non-syllabus driven course; grading based upon performance and notebooks; conducting classes as lab meetings; inviting notable speakers (Dr. Hatfull) for a seminar; and taking students to a national meeting where they presented the results of their own research efforts.

Results: Current data suggest preliminary success in both fronts. Retention of "at risk" students increased over previous years; with some exceptions, these students also performed better in their introductory science classes. Overall, student evaluations of the laboratory were extremely enthusiastic, particularly compared to similar evaluations of the standard introductory laboratory. Notably, a significantly higher percentage of students from the PGRI lab applied to and were accepted for our HHMI Freshmen Research Program.

PHAGE GENOME SEQUENCE PHAILURE RESULTS IN CREATIVE USE OF LOCAL COLLABORATIONS

Louise Temple, Department of Integrated Science and Technology, James Madison University, Harrisonburg, VA 22801
Steven G. Cresawn, Department of Biology, James Madison University, Harrisonburg, VA 22801

Background and hypothesis: Bacteriophage discovery, biology, and genomic analysis have been used successfully in the past by undergraduate educators, including the authors of this submission. Supported by the HHMI-SEA program, we began an effort to increase enrollment and retention in the STEM disciplines, with this course being one of several approaches. The hypothesis being tested is that original, interesting research experiences for entering college students will increase numbers of scientists being trained.

Methods: Our students were self-selected from any major, and the course did not substitute for any other in their curriculum. We used "just-in-time" teaching of phage biology and laboratory procedures. Students were required to manage their own phage and keep a notebook, and they were graded on "participation". Quizzes were given on basic information, and occasionally readings were required. In the bioinformatics semester, students worked in pairs and submitted assignments, for which they were given credit. This grading scheme resulted in a grade of "A" for almost every student, except those with poor attendance or failing to submit assignments.

A unique feature of our course was a collaboration with software engineering students. In groups of 4, they were assigned to interact with a real-world problem in software usage. Each of the groups took on a problem related to our work and produced a report about how the problem could be addressed.

Results and Conclusions: Our retention rate between semesters one and two was 82%, compared to 53% in a control group of students advancing between two semesters of the traditional biology curriculum. Anecdotal evidence indicates that there was a high level of satisfaction with the course, and the number of students interested in continuing to work during the summer (8/26) is impressive, considering there is no compensation offered. We know of at least 4 more who are continuing in research in the fall.

USE OF SUPPLEMENTAL INSTRUCTIONAL MATERIAL FOR THE ULM-HHMI NGRI LABORATORY CLASS

Chris R. Gissendanner *and* ***Ann M. Findley.***
Department of Biology, University of Louisiana at Monroe, Monroe, LA 71209

Throughout the two semester NGRI laboratory course sequence, a variety of supplemental instructional approaches were implemented to provide students with content tutorials, to summarize data collection efforts and provide guidance for next-step decisions, to explain the mechanics of computer algorithms, and to explore ethical issues associated with the biotechnology used during the course of the project. Brief introductions of phage life cycle biology, genome organization and experimental protocol flowcharts were followed by content quizzes that probed student understanding and encouraged follow-up questions. The DNA sequencing video was presented multiple times to generate a step-by-step understanding of pertinent procedural information and engender an appreciation for the biotechnology employed in modern sequencing protocols. MathBench modules (www.mathbench.umd.edu) exploring probability calculations through BLAST ("BLAST and (Im)probability") and restriction enzymes and gel electrophoresis through plasmid structure ("Chopping Up Plasmids") were employed as interactive tutorial exercises to enrich student understanding of important program tools. Since Mycobacteriophage sp. Peaches was only cut with *Hae*III during restriction digestion, NEBcutter V2.0 (New England BioLabs) was used to determine restriction enzymes that would cut the Peaches genome and generate gel outcomes from such digestions. The topic of comparative genomics was explored with students by introducing the Human Genome Project and the importance of model organisms ("Scanning Life's Matrix: Genes, Proteins, and Small Molecules" – HHMI-Holiday Lectures on Science series). Finally, the ethical, legal, and social implications of the Human Genome Project were probed by student viewing and critical assessment of *GATTACA*. The success of our combined approaches as reflected by student assessment instruments and performance measures in the introductory biology lecture and laboratory course sequence will be presented.

HOWARD HUGHES MEDICAL INSTITUTE
Office of Grants and Special Programs

Science Education Alliance
4000 Jones Bridge Road
Chevy Chase, MD 20815-6789
www.hhmi.org/sea