Exploring Bacteriophage Biology

Table of Contents

1 Introduction

Genome annotation is just the beginning of phage genome exploration. In our experience, a novice annotator can complete the process of gene identification and initial functional predictions within approximately 10-15 hours of work (more or less depending on genome size and experience), with the first genes taking more time, and later genes requiring less, as the annotator becomes familiar with the process, along with the core tenets of molecular biology.

Within the SEA-PHAGES program we hope that the annotation process can be completed with sufficient time remaining in term to engage in the exploration of the biology of the mycobacteriophages. There are fundamentally two directions. The first is the application of bioinformatic approaches, and the second is a return to wet-lab experiments. The computational approaches are likely to be more readily applicable, because of both time-constraints and space/facilities coordination. Some participants may be interested in exploring these approaches in separate courses for sophomore, junior, or senior students.

Below, we have briefly described a variety of questions and approaches that you may want to consider. This is of course only a sampling of the potential approaches and we strongly encourage you to use these as a stimulus to develop your own. In addition, there is a burgeoning literature on mycobacteriophage biology and we encourage you to access it for background information and for ideas for future projects. The descriptions below are necessarily described only very briefly, and we have tried to point you in the direction of some tools that you might use to apply them. Please let us know if we can provide more detailed help.

Finally, the Pittsburgh group has significant experience in exploring the biology of mycobacteriophages and we strongly encourage you to make use of this as you navigate this stage of your SEA-PHAGES program.

2 SEA-PHAGES coordination

The end-of-year Annual Symposium will focus on presentations organized by the outcome of phage discovery and genomics rather than a school-by-school focus. We therefore strongly encourage faculty and students to identify other SEA-PHAGES participants that have a common interest in a particular phage type (such as Cluster or Subcluster) or a particular set of genes, phams, or a biological process (integration, expression, genome organization, etc.). We anticipate that groups of students and faculty will present these findings as groups with a common interest.

3 Timeline

SEA-PHAGES schools have to accommodate a variety of timelines depending on your local academic calendar. We have provided some examples of timelines below that might be helpful for you as you plan your Spring classes.

3.1   Terms: Courses running Early Jan-End of April

Begin with Etude, the practice genome. Cover basic biology and phage biology concepts. Discuss Next-Gen sequencing technology (four wks).

Feb 1: receive class genome. Annotate through March 1(four weeks. Students who already practiced with Etude should be familiar enough with the software and concepts to finish quickly).

March 1- End of Class

Beyond Annotation explorations!

3.2 Quarters: Courses running Feb-May

Get genome on day1 of class. Can either start with Etude, or teach annotation directly on class genome. Cover basic biology concepts, and finish annotation in (4-6) wks;

Leaves six weeks for beyond annotation explorations.

SAMPLE SYLLABUS, ANNOTATION PORTION OF CLASS

W	Date	Activity / Assignments *
1	1/04,05	Lecture 1: **Phage biology, including dirt and phagehunting program** Com. lab: NCBI databases, phagesdb.org; Introduction to Lab notebook I (Reading: DNA Master annotation guide section 1, 3, 4)
2	1/09,10	Lecture 2: **Next generation sequencing** Comp. lab: auto-annotation, Glimmer/GeneMark, Intro to DNA Master (Reading: Pedulla et al 2003; DNA Master annotation guide sections 2, 6)
	1/11,12	Lecture 3: **Genome architecture**; **keeping electronic notes** Comp. lab: BLAST, Phamerator basic (Reading: DNA Master Annotation Guide section 5,7,8 through 8.4.2, 9 through 9.3)
3	1/16,17	*No class – Dr. Martin Luther King (University closed); 1/17 – add/drop period ends*
	1/18,19	Comp. lab: **Sequence alignments local vs global**/ continue annotation refinements (Reading for week 4: DNA Master annotation Guide section 10) Comp. lab: Continue annotation
4	1/23,24	Lecture 5: **Functional assignments**/ BLAST, Conserved Domains, HHPred Comp. lab: Continue annotation (Reading: DNA Master Annotation Guide, rest of section 8, 9)
	1/25,26	Lecture 6: **tRNAs, frameshifts**, **parasitic elements**; Com. Lab: continue gene assignments
5	1/30,31	Lecture 7: **Promoters, terminators, AttP sites**, Comp. lab: Wrap up gene assignments for merge (**homework 1 due: Genome Profile**)
	2/01,02	Lecture 8: Analyze discrepancies as a class; (Reading: Hatfull et al. 2010) (**homework 2 due: Merged genome discrepancies**)
6	2/06,07	Lecture 9: **Comparative genomics and Clusters**/DotPlots, Phamerator advanced functions, identifying genes of interest for research project Reading: Pace, 2009)
	2/08,09	Lecture 10: **Phylogeny** Com. Lab: Compare genomes in DNA Master, Gene vs genome phylogenies, ANIs, gene content, Splitstree Reading: Reading: Pope et al PLoS One 2011 Cluster K paper)
7	2/13,14	Comp. lab: Comparative Genomics computer time (Reading: PLoS One 2008)

	2/15,16	Lecture 12: Beginning of lectures for second portion of course, Recombineering for example **(Homework 3 due: Comparative Genome Map; Research Project Hypothesis with supporting bioinformatics)**
8	2/20,21	**MID TERM EXAM**

4 Bioinformatic Approaches

4.1 Annotate more than one!

Many schools sequence more than one of the phages isolated from their fall semester class. Annotation of the additional phages can happen subsequent to the first annotation, or at the same time (i.e. you can split the class in half, give each a genome, and then have them switch files and check each other's work).

4.2 Non-gene features

There are a number of features in a phage genome that are non-coding sequence—things like promoters, terminators, repeats, *attP* sites, and operator sites. While we typically don't include these in files for GenBank submission, it can be an interesting and informative class exercise to identify these other features.

4.2.1 Promoters

Mycobacteriophage promoters remain somewhat ill-defined and not especially easily to identify bioinformatically. However, we know that some are closely-related to a large class of promoters recognized by σ-70 (sigma-70), and can be searched for bioinformatically. DNA Master has a function to search for these. Highlight the sequence of interest in the sequence panel, and then click "DNA→ Promoter Prediction." Make sure the settings are for sigma-70 promoters, and then click analyze. Remember that you expect promoters to be located in intergenic regions, but also note that defining the start sites of genes immediately adjacent to intergenic regions is especially error-prone. You may want to look at refs (3, 9).

4.2.2 Terminators

Transcriptional terminators in bacteria typically come in a couple of varieties, those that require a factor such as Rho, and the Rho-independent stem-loop terminators. The first of these are not defined in mycobacteria or their phages and are not predictable bioinformatically. The second are predictable, and typically are composed of a base-paired stem (from 10-20 bp in length) with a 4-5 base loop followed by a run of 5-8 T residues (U in the RNA). Remember that the stem can contain U-G base pairs as well as G-C and A-T, and this can be a useful indicator as to which strand is being transcribed. The program "FindTerm" at www.softberry.com can predict these types of structures. These are described in several refs (5, 8).

4.2.3 Repeats:

There are two types of repeated sequences that you may want to search. Dispersed repeated sequences, and sets of repeats that are closely spaced, within in direct or indirect orientation. Typically, these are relatively short sequences, in the range of 10-30 bp. You may also find larger more extensive repeats, which are also of potential interest.

The best way to find repeats is through using BLASTN with the low-complexity filter turned off (you can do that in www.phagesdb.org). The top hit you get should be the phage genome against itself,

and when you look at the alignments you will see the end to end alignment, but below that you will see all of the shorter repeats. Finding a short segment (e.g. 10bp) that is repeated twice is unlikely to especially interesting (because this could be by chance rather than biologically meaningful) but if there are multiple copies of a repeated sequence, this can be tricky to see just by looking at these.

However, you can refine and advance your search using the "Scan" function in DNA Master (DNA→Scan). In the scan window type the sequence you are looking for and then click the Query button at the bottom of the window. The program will return the positions of exact or closely matching sequences. When/if you find repeats, analyze their location. Are they within a gene? Just upstream from certain genes? Overlapping start/stop codons or genes? Found only in phages of a certain cluster? This kind of information allows us to start to make a guess as to why the repeat is present or what kind of function it might have.

As an example, you might want to go to phagesdb.org, and go to the phage Adephagia, and do the BLASTN as described above. Then take a look at the ref (11).

Note that if interested in looking for short inverted repeats that you may find terminators!

4.2.4 Stoperators

A particular type of repeated sequences are the Stoperators described in the Cluster A genomes. These are described in more detail in the papers (1) and (12). If you have a Cluster A phage, do it have stoperators and are these the same or similar to the others? Are there other phages that have Stoperators that have yet to be identified.

4.2.5 *attP* sites:

Many phages are temperate and form stable lysogens. Most of these integrate their DNA into the bacterial chromosome when establishing lysogeny (though not all), and the phage-encoded integrases that catalyze integration are usually readily recognizable bioinformatically. Note, however, that there are two major types of integrases, the Tyrosine-Integrases (Int-Y) and the Serine-Integrases (Int-S). These are quite distinct from each other and use different types of DNA sites (i.e. attachment sites) for integration.

The *attP* (attachment phage) site is the sequence in the phage genome that is recognized by the integrase and at which strand exchange occurs in the site-specific recombination reaction that is catalyzed by Integrase. The *attP* site is typically located close to the integrase gene, but can be to either its 3' or 5' side. The *attP* site usually shares a short region of sequence homology to the bacterial attachment site (*attB*), known as the common core.

Phages that use tyrosine-integrases typically (but not always) use an *attB* site that overlaps a host tRNA gene (and if you can figure out why, please let us know!), and the *attP* and *attB* sites have 20-45bp in common. One way to think about this is that the phage genome carries not only the sequences needed for recombination, but also the complete 3' end of a tRNA gene, such that the tRNA gene is 'reconstructed' following integration. So to find these, you can simply BLASTN your entire phage sequence against *M. smegmatis*, using the BLAST function in phagesdb.org.

If you try this with, for example, mycobacteriophage L5, you will see the top hit reported is a 100% match of 43 bp to a chromosomal segment starting at 4,764,522. You can then see where this is in the *M. smegmatis* genome by going to the CMR database at

http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage.cgi

and entering 'smegmatis' in the organism name, and then follow the links through to rRNAs, and you will find that this common sequence overlaps with a host tRNA-gly gene. There's probably half-down

other routes to getting to this same answer! Also note that there is a usually a short imperfect inverted repeat associated with attP, although it is not entirely encompassed within the common core.

Check out papers such as (7, 10).

The *attP* sites for serine-integrases are a bit trickier, because the *attB* sites do not overlap tRNA genes and the common core (the sequence shared between *attB* and *attP*) can be as short as 2 bp! In this case, it may be simpler to try and find the short inverted repeats flanking the crossover site, focusing on the intergenic regions to the 3' and 5' sides of the integrase gene (all the ones found so far are on the 5' side).

You may want to look at papers such as (4, 6, 10, 14).

A word of warning! In some instances, the *attP* common core lies within an open reading frame. BPs is a good example. So it is not always within an intergenic region!

4.2.6   Insertion sequences and transposons

Mobile genetic elements such as transposons move from one genomic location to another at frequencies that are relatively high in evolutionary terms, although infrequently in terms of the number of generations of growth. As such, transposons can be identified as segments of DNA that are found in different locations but with very high levels of nucleotide similarity. These may also be identified by the BLASTN searches described above. Transposons are usually flanked by inverted repeats, although you may need to look for these by manual inspection of the sequences.

We have found relatively few mobile elements in the mycobacteriophages, but there's likely more out there! A good example of these is the MPME1 and MPME2 elements, and you should check out reference (13). There are some more distantly related transposons predicted in the Hatfull Advances in Virology review article.

4.3   Advanced Phamerator

Phamerator has a number of advanced functions, including the ability to sort its database by pham, gene, or domain. Sorting a database by pham allows you to tell quickly which phages have members in a certain pham.

From this menu it is also possible to draw Pham circles, which shows all the member phages of a database around the circumference of the circle, and connects the phages that contain a member of a particular pham with an internal arc.

In the map view in Phamerator, the display of conserved domains is especially useful (you probably used this in your annotation) in predicting gene functions, and the hover-over-Pham is also useful for seeing which other genomes in your display has members of that particular Pham.

4.4   Finding newly acquired genes

If you take a look at the paper (12) you will see an analysis where we examined the distribution of Phams and the size of the Phams within particular clusters. You can apply this to the phages you re interested in by using the Pham Table download option in Phamerator, where you will get a large excel sheet of which genomes contain which Pham. This is a large data set and you can extract a lot of data from it!

In addition, sometimes when you do the BLASTN searches, there are relatively large segments of nucleotide matches against other mycobacterial genomes, or even other genomes (using an NCBI web search). What are the genes within these segments, and can you define the extremities of the matches – do these correspond with gene starts/or and stops?

6

4.5 Identifying potential membrane proteins

You can use an output of all of your predicted protein sequences from DNA Master to run a batch search for proteins containing putative transmembrane domains using SOSUI at http://bp.nuap.nagoya-u.ac.jp/sosui/. You expect to see the holin appear in this list, and you will likely see the tapemeasure protein too. What else do you find and what does it mean?

4.6 DNA Master Genome Manger and Compare Genomes functions

DNA Master allows you to load phages genomes into your own local database, and then perform several advanced bioinformatic analyses through the "compare genomes" function. These include determining the average nucleotide identity between phages, examining gene content, analyzing nucleotide substitution rates to compare non-synonymous to synonymous values (Ka/Ks) and even building phylogeny trees.

To load files into your genome manager: If your .dnam5 file is open, click "Genome→ Add to Database". Otherwise, click "Tools→Genome Manager".

The genome manager has a number of tabs, the left-most one labeled "Browse". This tab allows you to view all the files in your local database. The fourth tab from the left is labeled "Retrieve". From this tab, you can search NCBI for GenBank files and import them directly into your local database or open them as .dnam5 files.

To perform genome comparisons between genomes within your genome manager, you first must enter a release code into a field in the program preferences. Click "File→Preferences", and then click the far right tab labeled "Miscellaneous". At the bottom of this window, there is a field labeled "Release Code". Enter "Watson" into this field, and then click "Apply". Then click "OK", and close the window.

Now click "Tools". The "Genome comparison" choice towards the bottom of the drop-down menu should be enabled (text should be in black and not light gray).

Click "Genome comparison→Manual". This will bring up a window that lists all of the genomes loaded into your local database on the left hand side, and an empty list on the right hand side. Click on genomes on the left to highlight them, then click "add" to put them into the right-hand list. Once all the genomes you want to compare are in the right-hand list, click "Analyze".

Depending on how many genomes you've added, and how long they are, this process can take quite a bit of time (an hour). For just a few genomes, it will be relatively quick—several minutes.

To read more about this, look in the DNA Master Help files for "Genome comparison".

4.7 Multiple sequence alignments:

You can align multiple sequences of whole related genomes or individual homologous genes by concatenating the sequences in .fasta format, and then opening the file in DNA Master through File→Open→Fasta Multiple Sequence File.

Once your file is open, click the "Multiple Alignment" tab. Select your alignment and phylogeny algorithms (do not leave these menus reading "Default" unless you have manually preset the default algorithm choice in the program preferences. If you don't specify the default algorithms in the preferences and you leave the menu reading "Default" nothing will happen when you click "Do Multiple Alignment".)

Once you are happy with your settings, click "Do Multiple Alignment".

For more details look up "Fasta File Viewer-Multiple Alignment Pane" in the DNA Master Help.

You can also do pair-wise alignments from the pair-wise alignment tab.

For more details look up "Fasta File Viewer" in the DNA Master Help.

4.8 Drawing Phylogeny Trees:

After you perform multiple sequence alignments, you can use the output to create Phylogeny trees. Click "Tools→Tree Viewer". From the "Make Distance Tree" tab (far right), you can load or paste .csv data. From the "Tree" tab, you can load saved Phylogeny data from a multiple sequence alignment in Newick tree format. Once you have loaded your data, you can draw your dendrogram (what we think of when we think of tree figures in papers) from the Dendrogram tab.

See the help files in DNA Master for "Tree Viewer" for more information.

You may want to consider some specific phylogenetic questions. Such as, for example, if your genome has a gene with sequence similarity to a host gene, then what is its nearest homologue. Can you predict which type of genome is might have been acquired from.

4.9 Start site preferences

Does your phage prefer to use ATG, GTG, or TTG as start codons? What percentage of each is used? Can you learn anything about potential preferred Ribosome Binding Sites by aligning and comparing the upstream sequences?

4.10 Is the GC% constant across the genome?

The GC% content sometime varies across a genome, and parts that depart significantly may be segments that have been acquired relatively recently from a source with a GC% content that is different to your average. You can examine this by using the function in DNA Master: DNA → Composition Window.

4.11 Design/test PCR primers specific to your phage.

Once upon a time in 2009, we only had 60 mycobacteriophages. One of our undergraduate phage hunters (Victoria Hohenstein, now a Graduate Student and TA for Johns Hopkins) designed PCR primers specific to each of the clusters and to each of the phages. This allowed us to test crude lysates quickly for contaminants, and to somewhat screen new phages for cluster assignments (not always). Now, at 243+ sequenced genomes, these primers could easily be outdated and incorrect. We've posted our primer list on phagesdb.org under "Phorum".

Will the previously designed 60-phage cluster primers still work on the 243+ collection? Can you design primers that will allow you to identify your phage?

The wet bench side of this is to actually order the primers and try the PCR. It is possible to amplify using crude phage lysate or even a picked plaque as a template. See our mycobacteriophage PCR protocol on phagesdb.org for more information.

5 Back to the bench

5.1 Phage Characterization

There are a multitude of relatively simple experiments that you can do to determine some basic characteristics of your phages. These include:

5.2 Permissive Temperature range for growth

Does your phage make plaques at 42°C? 20°C? See paper (11) for some examples.

5.3   Ion co-factors

Does your phage require Calcium to make plaques? Not all do. See paper (2).

5.4   Hydration

Some mycobacteriophages can be completely dried (place a drop of lysate on a petri dish, let it dry and leave it for a week, and then reconstitute with phage buffer,) and still make plaques.

5.5   Stability

Is your phage stable during storage conditions? Trying determining the titer as a function of storage time and see if it changes.

5.6   One-step growth curve:

How long does your phage take for one round of infection: adsorption to lysis?

This kind of experiment involves adding a known amount of phage to a known amount of cells in culture and then removing aliquots of infected cells and plating them with non-infected cells over the time course of the infection (these cells, also called infectious centers, will yield plaques). You can also pellet the infected cells from the samples, and titer the supernatant. At the end of one round of infection, you will see a sudden rise in the titer of the culture, as infected cells begin to lyse and release new particles. By keeping track of the number of cells and phage in your culture throughout the experiment you can determine things like the eclipse period (how long from adsorption to lysis?) and the burst size (how many phage come out of one infected cell?)

Many of the mycobacteriophage take approximately three hours to go from adsorption to lysis, and have a burst size somewhere between 100-500.

One of the original papers describing this type of experiment is from Ellis and Delbruck, 1939, in the Journal of General Physiology. You can find it online here: *jgp.rupress.org/content/22/3/365.full.pdf*

Ellis and Delbruck did their work with an *E. coli* phage, and the conditions for setting up a successful one-step growth curve with *M. smegmatis* are slightly different. It is very important to pre-concentrate the cells into a small volume during adsorption, and then to dilute the cells into a larger volume for the remainder of the infection. The best results we have seen published comes from:

Catalao et al, (2010) Molecular Microbiology 77, Issue 3, 672–686

5.7   Temperate vs lytic:

Does your phage make lysogens? How do you know you have a lysogen? The quickest way to make a lysogen is to spot 3ul of a serial dilution of your phage on plates with cells and top agar. With a temperate phage, one or some of the higher titer spots may grow a "mesa": a thick ring of cells surrounded by an area of clearing. The mesa is comprised primarily of lysogens. Streak out new colonies from the mesa, and test 4-6 for lysogeny.

5.8   Lysogeny Tests:

-Streak/patch your lysogen onto a plate and overlay with non-lysogenic cells and top agar. Lysogens release phage particles and so you will see an area of clearing around the lysogen patch.

-Grow your lysogen in liquid culture and titer the supernatant.

-Use your lysogen as plating cells and see if it is resistant to the phage that you created it with.

-Identify the AttP and AttB sites. Design PCR primers to amplify across the AttL (or AttR) sites of the lysogen (one primer in the phage, the other in the host). Use the lysogen as a template, and look for the correct size PCR product. Compare this reaction to the product from a set of primers that amplifies across the AttB site.

See phagesdb.org for more information.

5.9 Frequency of lysogeny:

How well does your phage make lysogens? Do 5 out of 10 infections yield lysogens? 8 out of 10? Plate 10^10 phage onto several plates, and let the liquid dry. Then plate a serial dilution of host cells onto both phage-seeded plates and regular plates. Consider all colonies that grow on the phage-seeded plates lysogens. Compare to the non-phage seeded plates. Did all the cells become lysogens? 50%?

5.10 Immunity assays: Can other phages infect your phage's lysogen?

Host range- We use *M. smegmatis* mc2 155, but there are several other BSL1 strains available from ATCC, as well as related species. See phagesdb.org for complete list.

5.11 Testing new phage specific PCR primers- see above

5.12 CsCl bands and SDS-PAGE protein gels

Prior to the use of the small plate prep DNA kits, we had to grow large quantities of phage and purify them on CsCl density gradients using ultracentrifugation to recover enough DNA to sequence a genome. If you have access to an ultracentrifuge, it is still worthwhile to CsCl band your phage and examine the particle proteins by SDS-PAGE, and/or mass-spectrometry or N-terminal sequencing. These types of experiment can provide experimental confirmation of some of your gene calls in your annotation, and allow you to actually determine which of your genes are in your phage particle. These protocols are all posted on phagesdb. These types of experiments could be performed in 2-4 weeks, depending on how often class meets and/or if students have access to the lab in off-hours.

5.13 BRED

Bacteriophage Recombineering of Electroporated DNA (BRED) is a technique by which it is possible to target and knock-out specific genes from your phage genome. The University of Pittsburgh Genomics Laboratory class spends six weeks annotating a genome and six weeks knocking out genes via BRED and examining the resulting phenotypes. Some genes are not possible to knock-out; these are then deemed "essential" phage genes. Performing BRED requires the design of several lengthy and complicated PCR primers per gene (these can run into the $100s of dollars) and an electroporator that can be manually set for a specific set of conditions (the standard conditions for *E. coli* will not work on Mycobacterium), so BRED may not be suitable for every classroom. Please see phagesdb.org for detailed protocols, and Marinelli et al, PLoS ONE. 2008; 3(12): e3957 for more information.

5.14 Cloning:

We have several expression vectors available that contain *E. coli* and *M. smegmatis* origins of replication, meaning that it is possible to construct your clone in *E. coli* using standard methods and then transfer it *M. smegmatis* and induce gene expression. After electroporation, *M. smegmatis* transformants will not be visible on plates incubated at 37°C until 3 or 4 days have passed. See phagesdb.org for protocols. Successful cloning and expression is likely to take 4-6 weeks in a classroom.

References

1.   **Brown, K. L., G. J. Sarkis, C. Wadsworth, and G. F. Hatfull.** 1997. Transcriptional silencing by the mycobacteriophage L5 repressor. Embo J **16:**5914-21.
2.   **Fullner, K. J., and G. F. Hatfull.** 1997. Mycobacteriophage L5 infection of Mycobacterium bovis BCG: implications for phage genetics in the slow-growing mycobacteria. Mol Microbiol **26:**755-66.
3.   **Garcia, M., M. Pimentel, and J. Moniz-Pereira.** 2002. Expression of Mycobacteriophage Ms6 lysis genes is driven by two sigma(70)-like promoters and is dependent on a transcription termination signal present in the leader RNA. J Bacteriol **184:**3034-43.
4.   **Ghosh, P., A. I. Kim, and G. F. Hatfull.** 2003. The orientation of mycobacteriophage Bxb1 integration is solely dependent on the central dinucleotide of attP and attB. Mol Cell **12:**1101-11.
5.   **Hatfull, G. F., and G. J. Sarkis.** 1993. DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics. Mol Microbiol **7:**395-405.
6.   **Kim, A. I., P. Ghosh, M. A. Aaron, L. A. Bibb, S. Jain, and G. F. Hatfull.** 2003. Mycobacteriophage Bxb1 integrates into the Mycobacterium smegmatis groEL1 gene. Mol Microbiol **50:**463-73.
7.   **Lee, M. H., and G. F. Hatfull.** 1993. Mycobacteriophage L5 integrase-mediated site-specific integration in vitro. J Bacteriol **175:**6836-41.
8.   **Morris, P., L. J. Marinelli, D. Jacobs-Sera, R. W. Hendrix, and G. F. Hatfull.** 2008. Genomic characterization of mycobacteriophage Giles: evidence for phage acquisition of host DNA by illegitimate recombination. J. Bacteriol. **190:**2172-82.
9.   **Nesbit, C. E., M. E. Levin, M. K. Donnelly-Wu, and G. F. Hatfull.** 1995. Transcriptional regulation of repressor synthesis in mycobacteriophage L5. Mol Microbiol **17:**1045-56.
10.   **Pham, T. T., D. Jacobs-Sera, M. L. Pedulla, R. W. Hendrix, and G. F. Hatfull.** 2007. Comparative genomic analysis of mycobacteriophage Tweety: evolutionary insights and construction of compatible site-specific integration vectors for mycobacteria. Microbiology **153:**2711-23.
11.   **Pope, W. H., C. M. Ferreira, D. Jacobs-Sera, R. C. Benjamin, A. J. Davis, R. J. DeJong, S. C. R. Elgin, F. R. Guilfoile, M. H. Forsyth, A. D. Harris, S. E. Harvey, L. E. Hughes, P. M. Hynes, A. S. Jackson, M. D. Jalal, E. A. MacMurray, C. M. Manley, M. J. McDonough, J. L. Mosier, L. J. Osterbann, H. S. Rabinowitz, C. N. Rhyan, D. A. Russell, M. S. Saha, C. D. Shaffer, S. E. Simon, E. F. Sims, I. G. Tovar, E. G. Weisser, J. T. Wertz, K. A. Weston-Hafer, K. E. Williamson, B. Zhang, S. G. Cresawn, P. Jain, M. Piuri, W. R. Jacobs, Jr., R. W. Hendrix, and G. F. Hatfull.** 2011. Cluster K Mycobacteriophages: Insights into the Evolutionary Origins of Mycobacteriophage TM4. PLoS ONE **6:**e26750.
12.   **Pope, W. H., D. Jacobs-Sera, D. A. Russell, C. L. Peebles, Z. Al-Atrache, T. A. Alcoser, L. M. Alexander, M. B. Alfano, S. T. Alford, N. E. Amy, M. D. Anderson, A. G. Anderson, A. A. S. Ang, M. Ares, Jr., A. J. Barber, L. P. Barker, J. M. Barrett, W. D. Barshop, C. M. Bauerle, I. M. Bayles, K. L. Belfield, A. A. Best, A. Borjon, Jr., C. A. Bowman, C. A. Boyer, K. W. Bradley, V. A. Bradley, L. N. Broadway, K. Budwal, K. N. Busby, I. W. Campbell, A. M. Campbell, A. Carey, S. M. Caruso, R. D. Chew, C. L. Cockburn, L. B. Cohen, J. M. Corajod, S. G. Cresawn, K. R. Davis, L. Deng, D. R. Denver, B. R. Dixon, S. Ekram, S. C. R. Elgin, A. E. Engelsen, B. E. V. English, M. L. Erb, C. Estrada, L. Z. Filliger, A. M. Findley, L. Forbes, M. H. Forsyth, T. M. Fox, M. J. Fritz, R. Garcia, Z. D. George, A. E. Georges, C. R. Gissendanner, S. Goff, R. Goldstein, K. C. Gordon, R. D. Green, S. L.**

Guerra, K. R. Guiney-Olsen, B. G. Guiza, L. Haghighat, G. V. Hagopian, C. J. Harmon, J. S. Harmson, G. A. Hartzog, S. E. Harvey, S. He, K. J. He, K. E. Healy, E. R. Higinbotham, E. N. Hildebrandt, J. H. Ho, G. M. Hogan, V. G. Hohenstein, N. A. Holz, V. J. Huang, E. L. Hufford, P. M. Hynes, A. S. Jackson, E. C. Jansen, J. Jarvik, P. G. Jasinto, T. C. Jordan, T. Kasza, M. A. Katelyn, J. S. Kelsey, L. A. Kerrigan, D. Khaw, J. Kim, J. Z. Knutter, C.-C. Ko, G. V. Larkin, J. R. Laroche, A. Latif, et al.** 2011. Expanding the Diversity of Mycobacteriophages: Insights into Genome Architecture and Evolution. PLoS ONE **6:**e16329.
13. **Sampson, T., G. W. Broussard, L. J. Marinelli, D. Jacobs-Sera, M. Ray, C. C. Ko, D. Russell, R. W. Hendrix, and G. F. Hatfull.** 2009. Mycobacteriophages BPs, Angel and Halo: comparative genomics reveals a novel class of ultra-small mobile genetic elements. Microbiology **155:**2962-77.
14. **Smith, M. C., and H. M. Thorpe.** 2002. Diversity in the serine recombinases. Mol Microbiol **44:**299-307.