Persistence in the SEA of Bioinformatics with Phage Genome Annotations

Michèle Barmoy, Jeff Norman, and John Drummond

June 2023

INTRODUCTION

SEA-PHAGES (Science Education Alliance – Phage Hunters Advancing Genomics and Evolutionary Science) is a two-semester Course-Based Research Experience (CRE) administered by the University of Pittsburgh and the Howard Hughes Medical Institute (HHMI). In the first semester, undergraduate students collect soil samples and learn simple microbiology techniques to isolate bacteriophage (hereafter "phage", which are viruses that infect bacteria). Students extract DNA from their phage and one or two samples from the class are sent to the University of Pittsburgh for sequencing. The sequence is returned for the second semester when students annotate the phage genome using bioinformatic tools (programs and databases) to identify the genes and their functions. The final annotated genome is submitted to public databases allowing student researchers to contribute to the field of genomics. Student participation in the SEA-PHAGES CRE has been shown to increase retention in the sciences when compared to other biology laboratory courses.

GENOME ANNOTATION METHOD

The phage genome is auto-annotated using DNA Master (a database manager and DNA sequence editor) which integrates results of its own tools with external programs like Glimmer, GeneMark, BLAST and Aragorn. This draft annotation of the genome is then reviewed by the students to improve the annotation. Working in groups of two or three, on assigned genes, students refine the annotation each backchecking the other. PECAAN (a newer annotation program) is often used to facilitate this process. Later, students work in larger groups, again checking each other.

Using 15 Guiding Principles of Annotation from SEA-PHAGES and additional programs like Phamerator, Starterator, and HHpred evidence for each gene's start call and function is checked (Pope et. Al. 2017). While annotating, students answer three basic questions; is this a gene, where does the gene start, and what is its function. Students curate reports identify reading frames, coding potential and possible functions. Comparisons are made to similar genes in other genomes using resources such as the Actinobacteriophage database, Genbank, and Conserved Domain Database to collect evidence for the gene calls.

Students may encounter frameshifts, gaps in the genome (phage genomes are tightly packed) and possible tRNAs that need to be investigated. The SEA PHAGES website has forums where researchers (faculty and students) can ask questions and collaborate as any research team would. This is often the source of ideas for student posters (final project). The green boxes in figure 1 show the steps of the SEA PHAGES program that are done by students in the Discovery semester while the blue rectangle highlights the steps done during the Bioinformatics semester. The completed annotation is submitted to the University of Pittsburgh for quality control (pink box). The annotation is then reviewed by GenBank. After approval, the genome is made available in the GenBank database (www.ncbi.nlm.nih.gov/genbank), along with PhagesDB (https://phagesdb.org), both public data bases are used by researchers around the world.

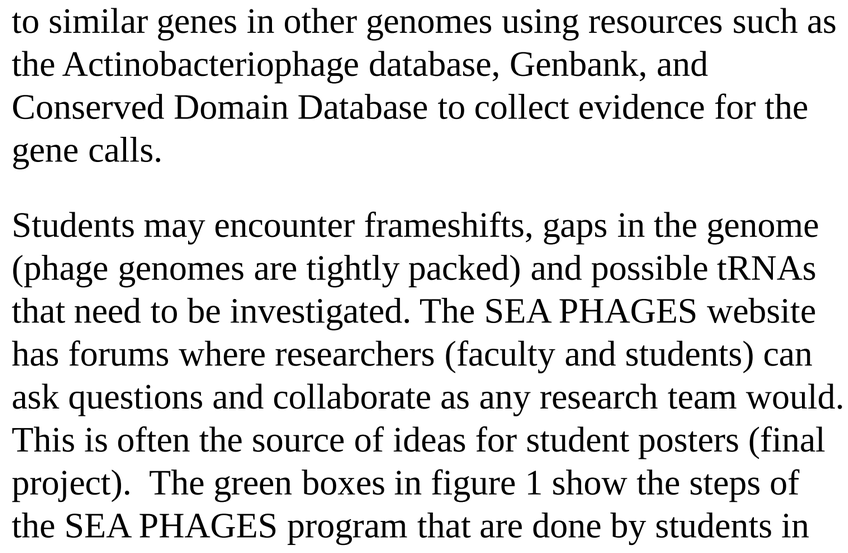


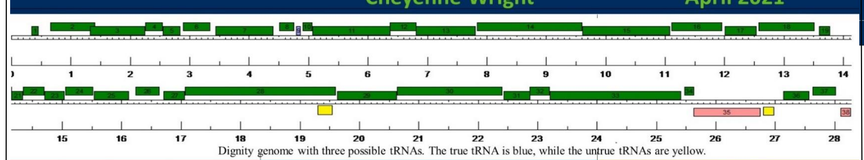
Figure 1: SEA PHAGES program flow chart highlighting work done by students during the bioinformatics semester (Pope, 2017).

POSTER PROJECT

The final project for the course is a poster presentation and SEA-PHAGES students have the opportunity to present their poster at the annual SEA Symposium. Two sample posters from Allegany College of Maryland are shown here.

tRNA and its Role in Bacteriophages

Cheyenne Wright — April 2021



Dignity genome with three possible tRNAs. The true tRNA is blue, while the untrue tRNAs are yellow.

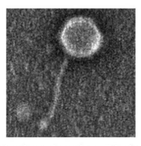
INTRODUCTION

tRNAs are molecules that are essential for protein synthesis and other functions within cells, and are structurally conserved across all domains of life, including some bacteriophages. tRNAs, and in some cases tRNA clusters, are found in some bacteriophages. These tRNAs are identified within phages during the annotation process using three main programs: Aragorn (v 1.1), Aragorn v1.2.38, and tRNA ScanSe 2.0.

tRNA presence was detected in the Phage Dignity of cluster A. The three programs were used to determine which of the possible tRNAs were actual tRNA molecules. Additionally, research was conducted into why tRNAs might be found in bacteriophage viruses that cannot replicate on their own.

RESEARCH QUESTIONS

1. Of the three possible tRNAs identified in Dignity, are any identified as actual tRNA?
2. What disqualified the others?
3. Why is tRNA found in phages?



Mycobacterium phage Dignity

METHODS

Three possible tRNAs were detected in phage Dignity. The three tRNA identifying programs, Aragorn version 1, version 2, and tRNA ScanSE 2.0, and criteria for including and excluding possible tRNA were used to analyze the possible tRNAs identified in Dignity. Sources and theories on why tRNA are in phages were researched to determine possible reasons for tRNAs in bacteriophages and specific clusters of phages.

RESULTS

- Three possible tRNAs in Dignity: tRNAs 96, 97, and 98.
- tRNA 96: identified in all three programs, from 4796-4870, 77 bp long, Cove Score of 43.98. Criteria: Cove Score above 35, identified in all three programs, correct length. Confirmed as a tRNA and included in the annotation. (SEAPHAGES)

- Other two possible tRNAs: It was thought that there were only two possible tRNAs in phage Dignity, both identified in PECAAN's version of Aragorn and tRNA Scan SE: tRNA 1/tRNA 96 (4796-4870), and tRNA 2/tRNA98 (26,864-26,938) because there were two tRNAs in PECAAN and two in DNAMasters' Aragorn. There was a different tRNA found in DNAMaster called tRNA 97 (19368-20146). Criteria: tRNA 98 only identified in tRNAScanSE, Cove Score of 10.57, too low to be considered.
- tRNA 97: listed as being 779 base pairs long. This is too long, tRNAs cannot be over 90 bp long, so it was not included.

- Research done into why tRNAs may be found in bacteriophages.
- Study conducted in 2016 by Delesalle et al. sought to test hypotheses on why tRNA genes might be found in phages. They state that phages can encode their own tRNA genes and examined specific clusters of bacteriophages to determine presence or absence of tRNAs.
- Cluster A was found to have equal numbers of phages with tRNAs and phages without them; most clusters either all have tRNAs, or all lack them. (Delesalle et al,. 2016)
- This study did not find that the presence or absence of tRNA affected the length of the genome, but found that the locations of tRNA genes were highly conserved within specific clusters. This suggests tRNA uses are dependent on where they are located in phages.
- Study states cluster A bacteriophages have tRNA earlier in the genome, typically before the lysin A gene. Since tRNA 96 appears to be between Genes 8 and 9, and Gene 10 was identified as lysin A, this appears consistent. (Delesalle et al, 2016)
- They also suggest that tRNA in phages may be from different bacteria and recombination events from bacterial host cells, and may exist in certain locations because of retention constraints. (Delesalle et al, 2016)

- Research project at the UMBC found a correlation between GC content in phages and tRNA presence.
- The *Streptomyces* phage they were studying had a lower GC content than other phages in its group, possibly because it had recently infected a non-Actinobacteriophage host with a lower GC content. However, the GC content difference from any host the phage could infect could be made up for by encoding its own tRNA molecules, which encode codons specific to the phage.
- This would allow the phage to infect a broader range of hosts, since a difference of codon usage in the bacterial host and the phage is compensated for by the phage encoding its own tRNA. (Alqahtani et al., 2021)

- Further supported by Bischoff et al. who propose tRNAs may be encoded in phages to bypass codon usage differences
- It may be beneficial to the phage to encode tRNAs that make the amino acids it needs rather than use the host's tRNAs.
- Further suggest that bacteriophages with more tRNAs in their genome are able to infect a larger number of hosts, those with both low GC content and high GC content, while phages with a low number of tRNAs do not have such a wide host range. (Bischoff et al, 2019).
- Supports the idea that encoding tRNA allows the phage a wider host range to compensate differences in codon usage.

CONCLUSIONS

Only one true tRNA was found in Phage Dignity, and it corresponded to all tRNA criteria listed on the SEAPHAGE database. The tRNA fit findings on tRNAs found in phages in Cluster A.

Many hypotheses have been proposed for the existence of tRNA in phages, but what seems to be the case in Cluster A is they share a specific location within the same cluster, are found in similar clusters, and are found before lysin A gene. Encoding its own tRNA appears to serve a beneficial role in infecting a wider host range, as differences in codon usage between bacterial host and bacteriophage can be overcome by phage-encoded tRNAs.

FUTURE RESEARCH QUESTIONS

1. Why are tRNAs found in some subclusters but not others? Does this limit their host range?
2. Why is cluster A the odd variety? Why is it the only cluster that contains some phages with tRNAs and some without?
3. Do phages incorporate themselves into a host as a tRNA?

REFERENCES

Delesalle , V, Tanke, N, Vill, A, Krukonis, G (2016). Testing hypotheses for the presence of tRNA genes in mycobacteriophage genomes; *Bacteriophage, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5056768/*

Bischoff, V, Zucker, F, Moraru, C (2019). Marine Bacteriophages: Transfer RNAs (tRNAs) in Phages and Their Role in Phage-Host Interactions, *Reference Module in Life Sciences*

Alqahtani, R, Csanadi-Schwartz, K, Dalal, A, Najmabadi, R, Qalawee, I, Sweeney, P (2021). There's no such thing as overpacking: do phages compensate for GC content mismatch by bringing their own tRNAs? https://whova.com/portal/webapp/seasy_20 2104/Artifact/3799

ACKNOWLEDGEMENTS

I would like to thank Dr. Steve Heninger and Dr. Michele Barmoy for their guidance through this project, and thank

SEA PHAGES

ALLEGANY COLLEGE of MARYLAND

The Role of Phage Rona in a Cocktail

By Mandy Chucci and Rebekah George

April 20, 2022

INTRODUCTION

Bacteriophages, which are viruses that infect bacteria, have interested scientists for a long time as a potential way of fighting bacterial disease. A **phage cocktail** is a combination of different bacteriophages that infect a bacterium simultaneously. Phage cocktails are often used in lieu of a single type of phages in the elimination of disease because bacteria are much less likely to develop a resistance to a combination of phages than to a single phage type. (Molina, F. et al., 2021)

Throughout past months, the genes of bacteriophage Rona have been annotated and reviewed to determine the structure and function of each. Using the information from the called genes, an investigation was performed to discover how Rona might be selected for a phage cocktail, and what bacteria would be affected by such a cocktail.

RESEARCH QUESTIONS

1. How is it decided what phages should be used in a phage cocktail for a specific disease?
2. What kinds of diseases/bacteria would a phage cocktail containing Rona be used to fight?
3. Where would the host come from, and are there similar hosts that phage Rona could also affect?

METHODS

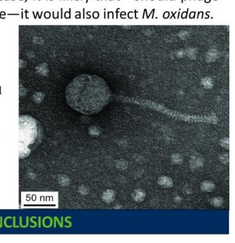
- Used online resources and the ACM library to research phage cocktails and Rona's host.
- Used phagesDB to find the host affected by Rona and other phages.

RESULTS

1. An ideal phage cocktail is usually between two and ten phages: if too few phages are used the bacteria can develop immunity, use restriction-modification systems (essentially destroying phage DNA), or spontaneously mutate. If too many phages are used there is an increased risk of toxic horizontal gene transfer and the transfer of virulence and antibiotic resistance genes. (Molina, F. et al., 2021)
2. Generalist phages (phages that can infect more than one type of bacteria) are preferred in cocktails because they have been found to infect hosts that specialized phages (phages that only infect one type of bacteria) were not able to infect. When selecting phages for a cocktail, the percent of successful infections and nestedness (meaning that phages in the cocktail should have specialist and generalist phages with a wide range in the number of hosts they infect (Beckett S. J., 2013)) should be considered. (Molina, F. et al., 2021)

RESULTS (cont.)

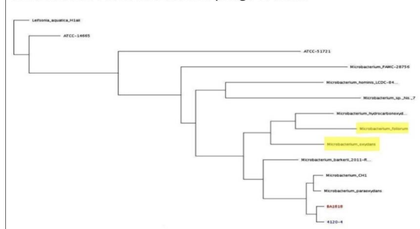
3. *The Actinobacteriophage Database* states that Rona infects the host strain *Microbacterium foliorum;* thus, a cocktail containing Rona would likely infect *M. foliorum* or its relatives *(n.d.).*
4. According to Chorost, *Microbacterium foliorum* has been mostly found in wound swabs and blood cultures (2018). In the experiment conducted by Chorost on *M. foliorum* and its relatives (shown in the phylogenetic tree below), a correlation between Central Venous Catheter (CVC) infections and these microbacteria were noted (2018). *M. oxydans*, one of *M. foliorum's* closest relatives, showed exceptionally high correlations with this type of infection and responded to antibiotics very similarly (Chorost, M. S. et al., 2018). This being the case, it is likely that—should phage Rona be a generalist phage—it would also infect *M. oxidans.*



Microbacteriophage Fizzles, Isolated from Microbacterium foliorum (Adams et al., 2022) [A photo of Rona was not available, but Fizzles is a similar phage.]

CONCLUSIONS

More research would be needed to determine Rona's efficiency in fighting infections or whether it would be a good match for a phage cocktail used against *M. foliorum.* Because Rona was just recently annotated, there are many unknown factors, and with current information, not much can be concluded about Rona's use in a phage cocktail.



Phylogenic tree highlighting the relationship of M. foliorum and other microbacteria, specifically M. oxydans, as they shared similar outcomes in an experiment on CVC infections. (Chorost M. S. et al., 2018).

REFERENCES

Adams, Skyler & Spotz, Gabrielle & Babcock, Riley & Butler, Chloe & Conger, Samantha & Crew, Madison & Garcia, Stephanie & Gonzalez, Joanna & Hodges, Jocelyn & Martinez, Alondra & Munoz, Samuel & O'Grady, Chloe & Quirl, Abigail & Sefcik, Kristin & Taylor, Tanner & Vazquez, Gustavo & Cox, Faith & Edwards, Dustin. (2022). Complete Genome Sequence of Bacteriophage Fizzles, Isolated from Microbacterium foliorum. https://www.researchgate.net/figure/Transmission-electron-microscopy-TEM-of-microbacteriophage-Fizzles-High-titer-lysate_fig1_357645511

Beckett, S. J., & Williams, H. T. (2013). Coevolutionary diversification creates nested-modular structure in phage-bacteria interaction networks. Interface focus, 3(6), 20130033. https://doi.org/10.1098/rsfs.2013.0033

Chorost, M. S., Smith, N. C., Hutter, J. N., Ong, A. C., Stam, J. A., McGann, P. T., Hinkle, M. K., Schaecher, K. E., & Kamau, E. (2018). Bacteraemia due to Microbacterium paraoxydans in a patient with chronic kidney disease, refractory hypertension and sarcoidosis. JMM case reports, 5(11), e005169. https://doi.org/10.1099/jmmcr.0.005169

Molina F., Simancas A., Ramirez M., Tabla R., Roa I., Rebollo J.E., 2/16/2021, A New Pipeline for Designing Phage Cocktails Based on Phage-Bacteria Infection Networks, https://www.frontiersin.org/articles/10.3389/fmicb.2021.564532/full#refer1

The Actinobacteriophage Database. *Microbacterium foliorum NRRL B-24224.* (n.d.). The Actinobacteriaphage Database. https://phagesdb.org/hosts/strains/65/

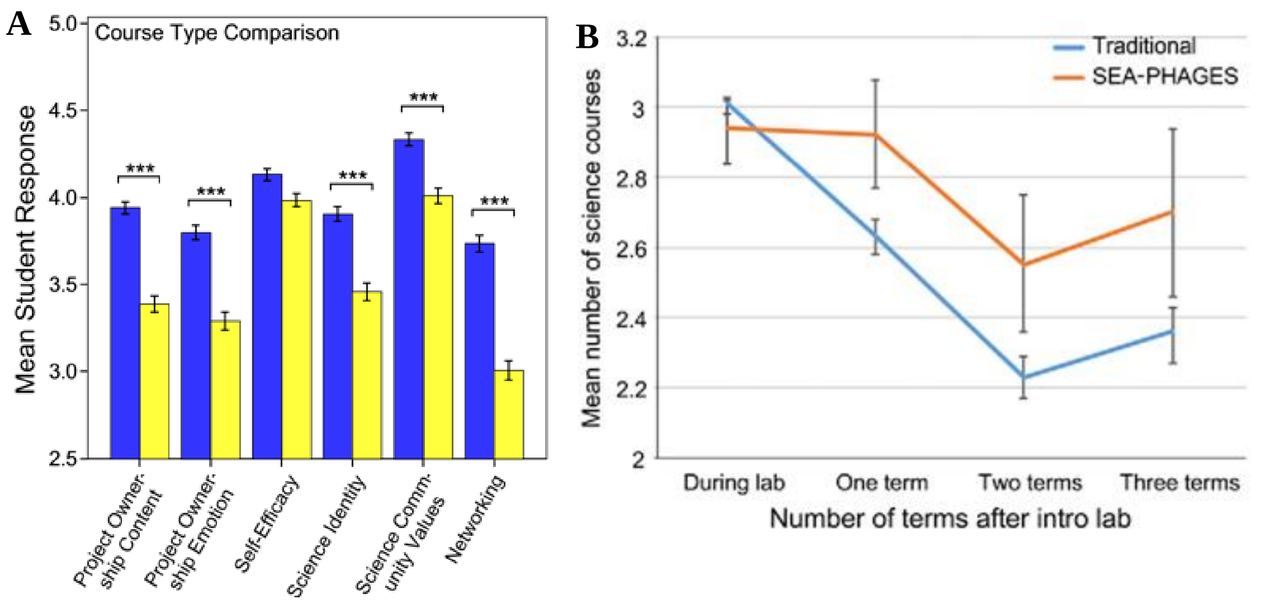
ACKNOWLEDGEMENTS

- Professor Michèle Barmoy
- Professor Steven Heninger
- Melissa Haut, ACM Science Study Lab
- The ACM Reading and Writing Center

ALLEGANY COLLEGE of MARYLAND

PERSISTENCE IN THE SCIENCES

Figure 2: (A) SEA PHAGES classes likely to continue in the sciences PITS survey measures compared to traditional laboratory classes. (B) Average number of science courses taken by students experiencing SEA-PHAGES (red) or a non-research laboratory course (blue) in three subsequent terms; 95% confidence intervals are shown (Hanauer, 2017).



Many CREs have been implemented instead of traditional labs based on reports that the experience of doing actual research provides insight into science careers and improves retention in STEM fields (Mader et al. 2017) . The SEA PHAGES program is one of only a few centralized inclusive Research Education Communities (iREC). This iREC provides a collaborative community where students and faculty regularly interact, train, and present their work with one another (Hanauer, 2017). SEA PHAGES surveys students each year on six key variables on student retention (summarized in table 1). According to survey administrators, SEA PHAGES students statistically out-perform traditional non-research labs on the survey (Hanauer, 2017). Some comparisons are shown in figure 2.

Table 1: Summary of variables of retention identified by SEA PHAGES.

Variable	Relevance for Course
Project Ownership Content	personal connection with their research.
Project Ownership Emotion	Excitement with their research
Self-Efficacy	belief in ability to perform scientific procedures
Science Identity	self-identification as a scientist.
Scientific Community Values	Acceptance that scientific knowledge is important.
Networking	sharing their research with other people.

REFERENCES

- Hanauer, David I. et al. 2017. An inclusive Research Education Community (iREC): Impact of the SEA-PHAGES program on research outcomes and student learning. PNAS. 114(51) 13531-13536.
- Mader, Catherine M. et al. 2017. Multi-Institutional, Multidisciplinary Study of the Impact of Course-Based Research Experiences. Journal of Microbiology & Biology Education. 18:ar2 1317.
- Pope, Welkin H., Russell, Daniel A., Creswan, Steven G., Hatfull, Graham F. 2017. SEA-PHAGES Bioinformatics Guide, University of Pittsburgh, The Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science, © 2017, Howard Hughes Medical Institute. All rights reserved. 4000 Jones Bridge Road, Chevy Chase, Maryland 20815.https://seaphagesbioinformatics.helpdocsonline.com/guiding-principles

ACKNOWLEDGEMENTS

ALLEGANY COLLEGE of MARYLAND

LAFAYETTE COLLEGE

SEA PHAGES

hhmi