Testing Re-emergence of Specialized Φ6 Bacteriophage

Abbey J. Isaac[a], Taylor P. Andrews[b], Siobain Duffy[c]
[a] Department of Biology, Brookdale Community College
[b] Microbial Biology Graduate Program, Rutgers School of Environmental and Biological Sciences
[c] Department of Evolution and Ecology, Rutgers School of Environmental and Biological Sciences

Abstract

RNA viruses have high mutation rates, which contribute to their ability to rapidly host shift. Using specialized Φ6 strains as a model for pathogenic eukaryotic RNA viruses, we aim to determine how easy it is for a virus to reverse the narrowing of its host range in favor of expanding it. Through the experimental evolution of these strains, we found that host range specialization is easily reversed with Φ6. We also identified regions of high genomic variance using Shannon entropy: vital precursor data to eventually help identify which mutations were directly involved in this host-range shift, thus allowing reemergence.

Background

Host shifting and RNA viruses: RNA viruses mutate at a faster rate than their hosts, introducing the possibility for natural selection to act upon a viral population. (Duffy S. 2018). In the case of pathogenic viruses, such as SARS CoV-2 and Influenza, the mutations they have acquired have allowed for host shifts, resulting in infections that have caused economic and health disasters both locally and globally. (Vilcek S. 2020, Woolhouse ME, Haydon DT, Antia R. 2005). Host shifting is a pathway viruses can take to optimize their fitness on novel hosts and outcompete related viruses. (Duffy, Siobain et al. 2007). Viral fitness is a measurement of the likelihood of a virus to produce infectious progeny in a given environment. (Wargo, A. R., & Kurath, G. 2012).

Observing and studying mutations in pathogenic viruses can pose health and ethical risks, particularly in terms of reemergence and gain-of-function research. This has led ~~to~~ Φ6, a

bacteriophage that infects various pathovars of plant pathogens *Pseudomonas syringae*, to be a model organism for viral evolution studies. Due to the fast generation time of their hosts and lack of pathogenesis in humans, it has proved widely useful for numerous experiments, especially our focus to analyze how mutations in a virus's genome affect host range (Dennehy J. J. (2009), Ford, Brian Elwood 2015).

Φ6: Φ6 is a bacteriophage in the Cystoviridae family, which is the *only* known family of phages to possess a lipid bilayer and a segmented double-stranded RNA genome. This is unique, as only some eukaryotic viruses typically have a lipid bilayer present in their structure. (Mindich, L. 1988). Φ6 shares, overall, more similarity to the structure of eukaryotic viruses than bacteriophages, especially in how it reproduces within a cell and lyses out of it. (Gottlieb, P., & Alimova, A. 2022).

Specialism vs. Generalism: A virus, based on host range, can be classified as a specialist or a generalist. A specialist infects only one or few hosts, while generalists infect multiple hosts within a related range and more easily emerge on new, unrelated hosts (Duffy et. a; 2007). It has been shown that Φ6 can narrow host range when exposed to only one host over a period of time. The mutations that narrow host range also increase the virus' fitness on the host on which it was grown – the narrowed host range was an unselected side effect of a virus increasing its reproductive ability on the single host it was evolving on. (Duffy et. al 2007) What has not been widely studied is the fate of viruses with those narrowed host ranges – how is the evolutionary future of a virus impacted by the host range mutations it already carries in its genome?

The impact that pre-existing mutations have on the effects of other mutations is called epistasis. Viral genomes with host range mutations may not be able to further change their host ranges in the same way as a 'wildtype' virus because of epistasis. (Zhao et al 2019).

This study examines three descendants of the ancestral E8G strain in Duffy et al 2007 that had evolved on a novel host (*Pseudomonas pseudoalcaligenes*) for 30 days. Three of these evolved populations lost host range compared to the E8G ancestor: two populations lost the ability to infect *P. syringae* pv *tomato*, due to the same mutation in the Φ6 spike protein, and one population lost the ability to infect *P. syringae* pv *tomato* and *P. syringae* pv *phaseolicola* due to a different mutation in the spike protein.

By reintroducing specialist strains of Φ6 to challenge hosts outside of their narrowed host range, we can trigger a host range expansion to occur. If they broaden their host range – will it be by reversion or other compensatory mutations?

Methods and Materials

Media: All LC (lysogeny broth, adjusted to pH 7.5) media used to grow bacteria and evolve phage was made in the lab. The standard recipe for the LC broth is composed of 1L of Millipore purified water, 10g NaCl, 10g Tryptone, and 5g of yeast pH balanced to 7.5 with 1M solution of NaOH. Agar was added from media broth at varying concentrations to procure two kinds—plate agar, which used 15g of agar per 1L of LC broth (1.5%), and the softer top agar, which was 0.53g of Bactoagar per 75mL of LC broth (0.7%). All media was autoclaved to ensure sterility.

Legend * = Bacterial host – = Specialist strain ✔ = permissible host ✖ = challenge host	3eN-	1eNd20-	1eNd30-	E8G- (ancestor)
ERA*	✔	✔	✔	✔
PP*	✔	✖	✖	✔
TOM*	✖	✖	✖	✔
ATRO*	✖	✖	✖	✖

Table 1: Initial host range of specialist strains tested. Ancestral strain E8G was added as a reference, even though not a part of the experiment. Checkmarks represent a permissive host, which means the phage is well adapted to replicating within the host. X's represent the challenge hosts, or hosts the strain cannot infect.

Microorganism Strains: Φ6 strains and various *Pseudomonas syringae* pathovars were present in the Duffy Lab (original sources in Duffy et al 2006). The three E8G Φ6 isolates were selected, named 3eN, 1eNd20, and 1eNd30, where the N stands for narrow host range (Duffy et. al 2007). All of these mutants previously evolved from an ancestral strain with an expanded host range mutation in its spike protein (E8G), to specialize to *Pseudomonas pseudoalcaligenes* ERA (East River isolate A) due to acquiring single mutations (A31T in 3eN, evolved for 30 days; G247A in 1eN, already present at 20 days of evolution and still there after 30 days total). Our three challenge bacterial hosts were selected from multiple pathovars of *Pseudomonas syringae*: *P. syringae* pv *tomato* (TOM), *P. syringae* pv *phaseolicola* (PP), and *P. syringae* pv *atrofaciens* (ATRO). The first two challenge hosts are part of the host range of the E8G ancestor of the three narrowed strains, and 3eN can still infect *P. syringae* pv *phaseolicola*. *P. syringe* pv *atrofaciens* is a novel host for all of the viruses in this study (Table 1).

Experimental evolution: In order to create an environment that would promote a host range expansion, we introduced the specialized strains to the challenge hosts. To ensure genetic diversity that could contain host range broadening mutations, high concentrations of phage were introduced to a fixed concentration of bacteria**.** To ensure statistical rigor, ~~each~~ 6 strain-host trial was done twice in triplicate, for a total of 6 trials. Starting lysates were generated by inoculating 3mL of molten top agar with 5µL of permissible host *P. pseudoalcaligenes* ERA liquid culture and 3 µL of mutant strain freezer stock stored in 40% glycerol. Once vortexed to ensure proper suspension of microorganisms, top agar was poured onto an agar plate using pour plate technique and incubated at 25ºC for 18hrs.

After incubation, the top agar is scraped off, combined with 3mL LC broth, vortexed to liberate phage, and centrifuged for 10 mins at 3000rpm. The resulting supernatant was passed through a 0.22µL filter to remove bacteria from the sample. The resulting lysate is then serially diluted in LC broth from 10^{-1} to 10^{-5} and inoculated with each challenge host. After 18 hours of incubation at 25ºC, visible plaques were counted. Plaques (zones of clearing) represent the effects of one phage able to infect a bacteria and propagate into a circular clearing before the bacteria reach stationary phase. Plaque counts under 300 were optimal to reduce difficulties of discerning overlapping plaques —anything above 300 was considered Too Many to Count (TMTC). Plating the phage at various dilutions allowed for accurate quantification (Figure 2).

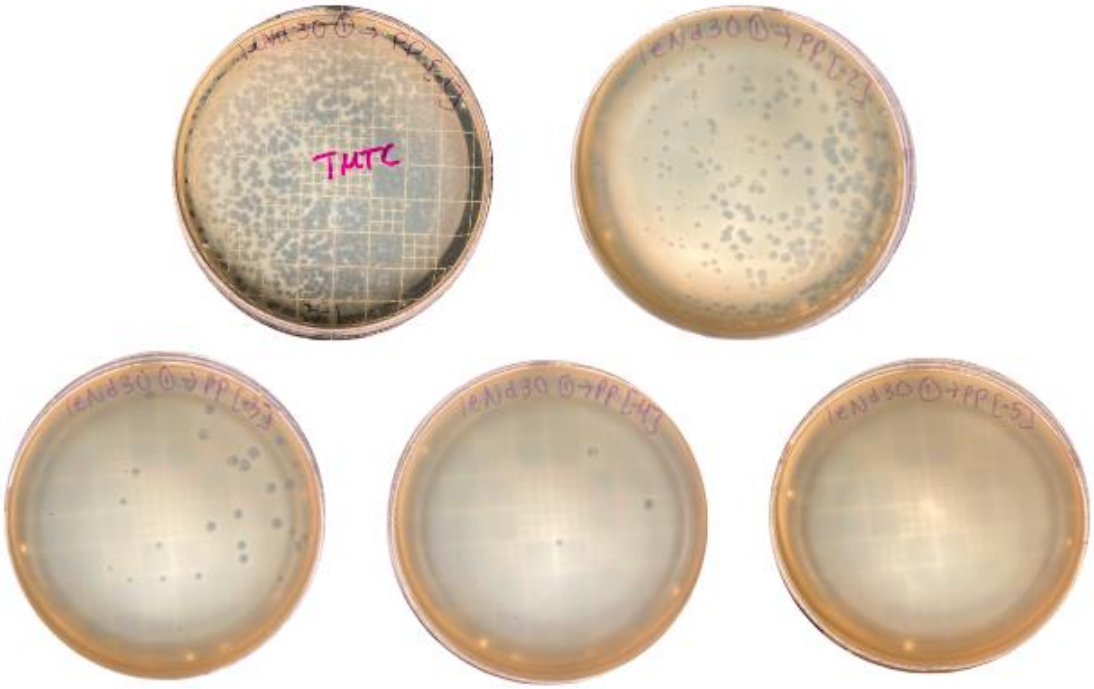

Figure 2: A sample of the 1st 1eNd30 replicate grown on *P. syringae* pv *phaseolicola* from 10^{-1} to 10^{-5} dilution. The first plate had too many plaques to count (TMTC).

Statistical analysis: All data sets were organized in Excel and analyzed in R-Studio. From the plaque count number, viral concentration (in PFU/mL) was calculated by multiplying the number of plaques counted by the reciprocal of the dilution of the phage plated with the host.

$$Viral\ concentration\ = \frac{\#\ of\ plaques\ formed\ on\ pour\ plate}{Dilution\ of\ phage\ plated\ on\ pour\ plate}$$

From the PFU/mL of the 6 trials, the mutational frequency was calculated by converting PFU/mL values to log(10), then dividing the PFU/mL of the strain on a challenge host by the PFU/mL on the permissible host. This produced 6 mutational frequencies for 8 trials (excluding *P. pseudoalcaligenes*). Values for each trial were also averaged and recorded.

An ANOVA test followed by a Tukey posthoc was done to compare the mutational frequencies of 3en, 1eNd20, and 1eNd30 grown on *P. syringae* pv *tomato* and *P. syringae* pv *atrofaciens* ($\boldsymbol{\alpha}$ = 0.05). When examining viral mutational frequency on *P. syringae* pv *phaseliocola*, a two-tailed t-

test was done instead to compare 1eNd20 and 1eNd30, as there were not enough datasets to constitute an ANOVA. ($\boldsymbol{\alpha}$ = 0.05)

RNA extraction and sequencing: A Qiagen QIAamp Viral RNA Mini Kit was used to isolate RNA from high titer lysates from each specialized strain tested against each host (11 samples total). The resulting RNA samples were purified using gel electrophoresis to remove background contamination from the *Pseudomonas* host genome. 24µL of RNA with 4µL of 6X loading dye were run through 0.8% low-temperature melt agarose gel (with 5uL EtBR/75mL 1xTAE gel) to isolate Φ6 genomic RNA. The gel was run at 150V for 45 minutes, the gel was placed under UV light.

Bands from the sample were cut out using a clean razor blade, then placed into a 1.5mL tube. Tubes were incubated at 70°C for 10 minutes (or until agarose is completely molten), then were transferred to a 42°C water bath to equilibrate for 5 minutes. 1 U of Agarase per 100mg of molten agarose was added, and samples were incubated again at 42°C for 30 mins. Sodium acetate was added to the sample to produce a 3 M concentration at a pH of 5.35, placed on ice for 5 minutes, then centrifuged at 10,000rpm for 10 mins to help pellet undigested carbohydrates from the agarose. The supernatant was transferred to a clean 1.5 mL tube, then 2.5 volumes of ethanol were added.

In order for RNA to be viable for sequencing, a concentration greater than 15ng/mL was required, which was not met from the samples after the gel extraction protocol. A way to counteract this was by following an RNA precipitation protocol (Walker, S. E., & Lorsch, J. 2013), which would increase the concentration of the sample. This was done by adding 0.1 volumes of 3 M sodium acetate, then mixing. 2.5 volumes of 100% ethanol per sample was added, then left to incubate in a bath of dry ice and ethanol for 25 minutes. Samples were centrifuged at 12,000g for 15

minutes–supernatant was carefully removed, and the resultant RNA pellet was washed with 2.5 volumes of 70%, then left to soak for 2 mins. Samples were centrifuged again at 12,000g for 2 minutes, then ethanol was removed. The resultant pellet was left to dry for 1 hour at room temperature, then the pellet was hydrated with RNAse-free water to dissolve RNA.

The concentration of the resulting purified RNA was measured with a nanodrop (measured in ng/µL) and subsequently sent out for cDNA conversion and 150bp paired-end Illumina sequencing (SeqCenter, Pittsburgh, PA). Received sequencing data samples were processed on Galaxy (usegalaxy.org) using the following steps: Both copies of the received sequence sample were uploaded along with a concatenated wild-type Φ6 genome as a reference, and the received sequence had *Pseudomonas* contamination. It was run through the BWA-MEMw tool to map long base-pair reads. The bam and bai outputs were downloaded and renamed avoiding spaces. The file was then uploaded to IGVtools to run a count. The file was then reformatted to a .wig file to open up the processed sequencing data in Excel.

Shannon Entropy: From the processed files of the sequenced genomes, Shannon Entropy was calculated using R-studio, using this equation. (Zhao et al. 2019).

$$H(X) = -\sum_{i=1}^{n} P(xi)\ log\ P(x)$$

The change in the Shannon entropy was calculated by subtracting the Shannon entropy on the viral strain grown on the permissible host from the Shannon entropy of the strain grown on a challenge host. Not only does this help eliminate inconsistent data points from Illumina sequencing, but it also highlights where the variance occurs along the gene segment. The Varscan tool in Galaxy was used for variant detection to identify specific mutations found in the analyzed Φ6 genome.

Results

Φ6 specialism is easily reversed: When the specialized strains of Φ6 were exposed to challenge hosts, plaques were able to form at high plating concentrations (10[1]-10[4]), which indicates that reemergence is possible and host range narrowing is reversible in Φ6. This could be due to the fast mutation rate of RNA viruses like Φ6 (Duffy S. 2018). The faster a virus is able to replicate its genome, the faster mutations can occur and accumulate, increasing evolution speed compared to other viruses.

Another significant contributing factor to this high mutation rate is a lack of nucleotide repair. For instance, RNA viruses generally lack 3' exonuclease proofreading activity (Sanjuán R. & Domingo-Calap 2015); exonucleases are present in DNA polymerases and act as proofreaders, fixing mistakes made during replication. A lack of this activity, as seen in most RNA viruses, can lead to more mutations accumulating, thus genetic variation.

Finally, the use of naive bacteria for every strain-host test prevented possible bacterial resistance from forming in the host. (Oechslin F. 2018) Bacteria that develop this resistance have a harder time being infected by viruses, and since we want the virus to infect the host, this would be counterintuitive.

Specialist strains with wider initial host range might have higher mutational frequency: When comparing the average mutational frequency between 3eN, 1eNd20, and 1eNd30, it was observed across all challenge hosts that 3eN had the highest mutational frequency on both *P. syringae* pv *tomato* and *P. syringae* pv *atrofaciens*. However, when comparing 1eNd20 and 1eNd30, no statistically significant difference was found between mutational frequencies on *P. syringae* pv *phaseolicola*, *P. syringae* pv *tomato*, or *P. syringae* pv *atrofaciens*.

A notable discovery was a virus with a broader host range might have an easier time reemerging than a more specialized virus, with a narrower host range (Duffy et. al 2006), and the statistically significant difference in 3eN's mutational frequency compared to 1eNd20 and 1eNd30 leans toward this observation. 3eN only had one extra permissible host (PP) than 1eNd20 and 1eNd30, so while it was only slightly more generalist than 1eN, it did have a broader host range. As generalists are considered more likely to emerge on new hosts, one might assume the extra host in 3eN's permissible host range could have contributed to the highest mutational frequency of the three strains. (Zhao, L., & Duffy, S. 2019). (Figure 3)

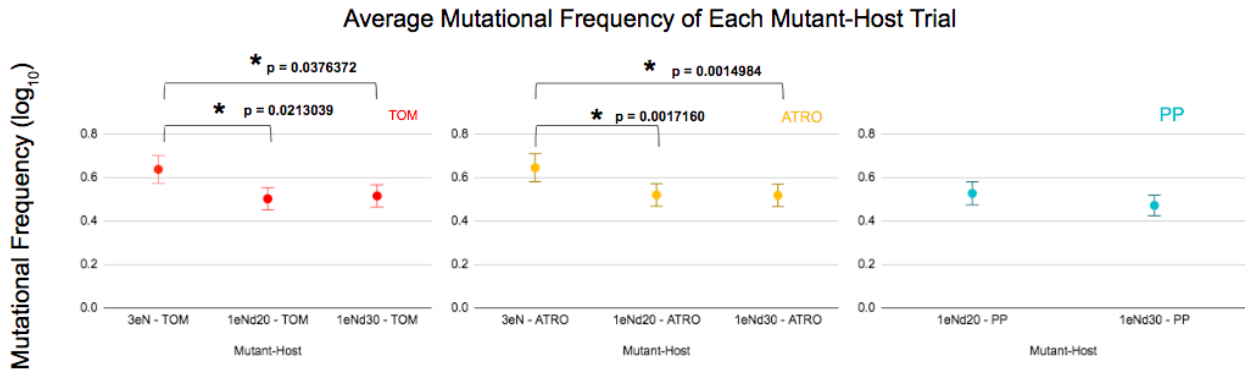

Figure 3: Average mutational frequency graphs generated in Excel and separated by challenge host. The asterisk (*) and brackets ([ ) represent significant differences (p values listed above bracket) between 3eN's mutational frequency compared to 1eNd20 and 1eNd30 on *P. syringae* pv *tomato* and *P. syringae* pv *atrofaciens.*

Shannon Entropy: The figures below depict the change in Shannon entropy comparing 1eNd20 and 1eNd30 grown on *P. pseudoalcaligenes* to the challenge hosts. High positive values indicate sites where there was more nucleotide diversity after exposure to a challenge host than before exposure to a challenge host – sites likely to be involved in broadening host range and improved fitness on the challenge host. (Zhao et al. 2019, Mullick et al. 2021).

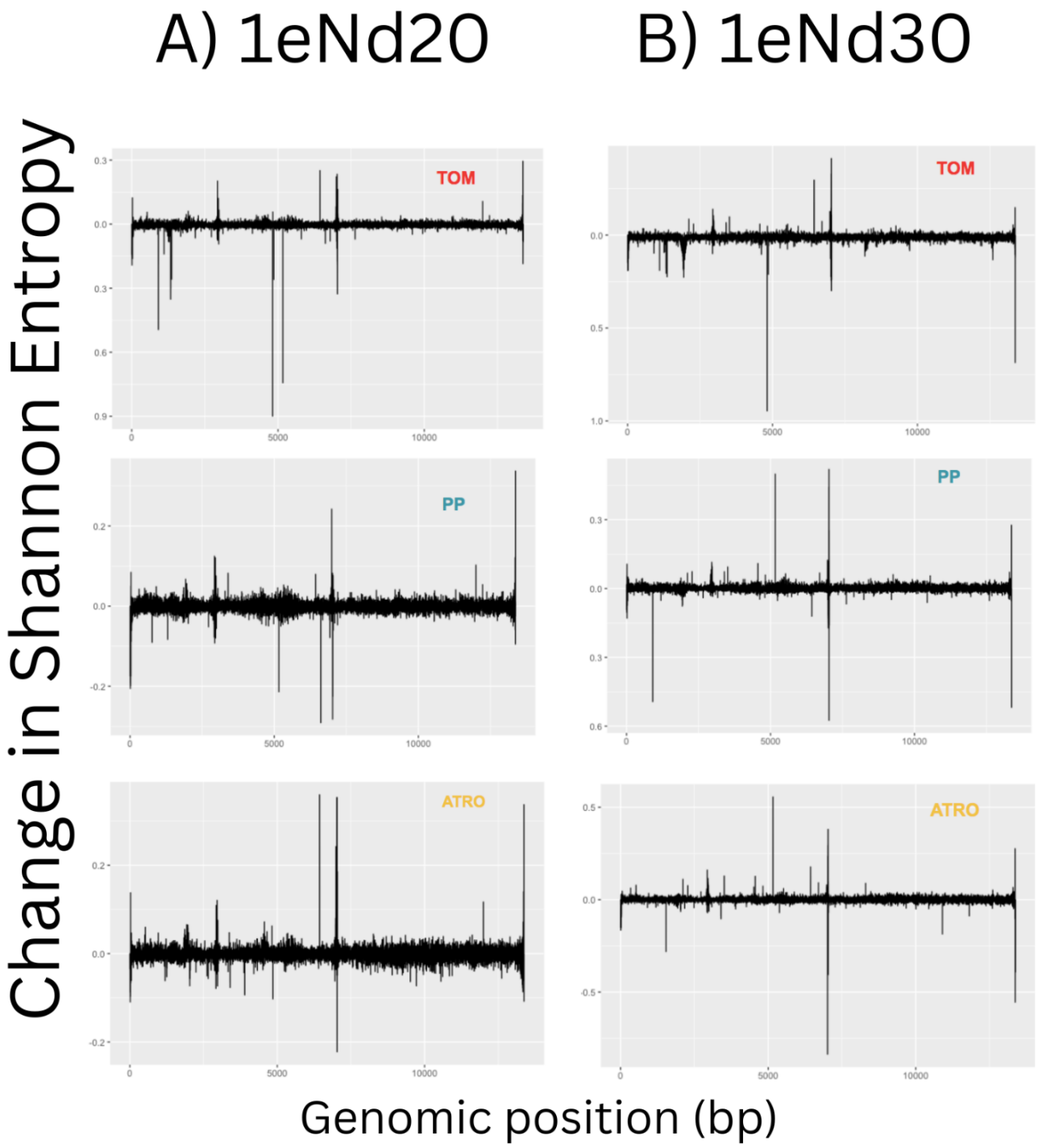

Figure 4: Change in Shannon Entropy when comparing the Shannon entropy of a challenge host to the permissive host ERA. The three segments of the Φ6 genome are artificially concatenated to show all of the data in a single figure: S, then M, then L. A) 1eNd20's change in Shannon entropy on TOM (*P. syringae* pv tomato), PP (*P. syringae* pv phaseolicola), and ATRO (*P. syringae* pv atrofaciens). B) 1eNd30's change in Shannon entropy on TOM, PP, ATRO. See table 2 for mutation location on this graph.

Referring to Figure (4), the Shannon entropy data sheet proved useful in identifying mutations

across the genome, when cross-referenced with the VarScan output. Certain high spikes on the graph did not relate to mutations, as their location was too close to the end segments of Φ6.

A) 1end20

TOM

Base pair position	Change in H	Mutation
5164	0.743324329	A247G
1328	0.351897715	K192R

PP

Base pair position	Change in H	Mutation
765	0.013902416	G4G
5164	0.054000161	A247G

ATRO

Base pair position	Change in H	Mutation
4855	0.102652003	K144R
1196	0.061412492	E148A

B) 1end30

TOM

Base pair position	Change in H	Mutation
4813	0.094668866	Q130R
1363	0.219460843	Q8R

PP

Base pair position	Change in H	Mutation
4563	-0.109270248	T47S
4822	-0.059012919	V133A

ATRO

Base pair position	Change in H	Mutation
3400	0.064543569	V28V

Table 2: Single nucleotide variants that were found at sites with mostly highly positive change in Shannon entropies in the genome of 1eNd20 and 1eNd30 after interaction with each challenge host (also seen in Figure 4). The base pair location correlates to what position along the concatenated genome it is located. Most of these changes were on the P3 spike protein (T47S, Q130R, V133A, K144R, A247G), which aids in host attachment.

Limitations and Future Work

Time: In order to fully analyze the cause and effect of mutations accumulated in the three strains tested, there was not sufficient time to do so in the span of the internship. The 4th descendant of the ancestral E8G (E4 isolate evolved on P. pseudoalcalgenes for 30 days) was a strain that was supposed to be a part of the experimental strain-host trials but was omitted due to time constraints. Current work on testing 4eN on challenge hosts is continuing as of now. Additionally, due to poor sequencing coverage of processed 3eN data, it had to be omitted as well. Solutions to this could be redoing the challenge host trials on 3eN and applying the same data analysis

pipeline.

Gel extraction: The gel extraction protocol had to be repeated multiple times and optimized to produce RNA in sufficient concentration and quality. The modified protocol extended the melting time of cut-out agarose bands to ensure all agarose was molten, as the 10 minutes stated in the initial protocol was not sufficient enough to melt the agarose in our 11 samples, but despite this resultant samples still had low concentrations. This was improved via the RNA precipitation protocol.

Further Sequencing Analysis: The process of analyzing sequencing data and interpreting the information into quantifiable values took much longer than initially anticipated. We were able to get a first glance at what the data entails, but as the data was received and processed too late, more significant information the mutations were omitted from the project. While the Shannon Entropy graph was generated, it still needs cleaning, as sites with very few mapping reads (less than 100bp) can skew results and appear as significant sites of variation until further analysis of the segment of the genome proves the opposite. Cleaning the data set would optimize the presentation of results by eliminating the frequently high change in Shannon Entropy values that were artefactual – due to low coverage at the ends of the three genomic segments.

Future work also aims to simplify figures and deliver genomic data better by focusing on small, medium, and large segments, plotting out where mutations occurred on the segment, and deeper analysis of possible epistatic constraints resulting from the viral mutations. Also, further extension discussion of the mutations that affect what protein.

Conclusion

Re-emergence of specialized Φ6 is possible when exposed to a new host thanks to host-shift mutations that are more likely to occur due to the nature of the RNA genome. Specialized RNA viruses that have more than 1 permissible host might have a higher mutational frequency, which correlates to the higher chance for host-range expansion of generalists. Ultimately, the future investigation of this experiment can shed insight into further epistatic constraints that affect host-range shifts of not only pho6, but pathogenic RNA viruses that harm eukaryotes—which include humans, like us.

Acknowledgments

This work was supported by the New Jersey Space Grant Consortium, Brookdale Community College, and Rutgers: School of Environmental and Biological Sciences. I would like to thank my mentors–Taylor Andrews, Siobain Duffy, and the Duffy Lab for allowing this research project to be possible, and for their support, tutelage, and passion for education. I would like to thank my advisors Professor Ochs and Professor Teodorescu for helping me along the way and always being the most supportive influence.

References

Dennehy J. J. (2009). Bacteriophages as model organisms for virus emergence research. Trends in microbiology, 17(10), 450–457. https://doi.org/10.1016/j.tim.2009.07.006

Duffy S. (2018). Why are RNA virus mutation rates so damn high?. PLoS biology, 16(8), e3000003. https://doi.org/10.1371/journal.pbio.3000003

Duffy, S., Turner, P. E., & Burch, C. L. (2006). Pleiotropic costs of niche expansion in the RNA bacteriophage phi 6. Genetics, 172(2), 751–757. https://doi.org/10.1534/genetics.105.051136

Duffy, Siobain et al. “Evolution of host specificity drives reproductive isolation among RNA viruses.” Evolution; international journal of organic evolution vol. 61,11 (2007): 2614-22. doi:10.1111/j.1558-5646.2007.00226.x

Ford, B. E., Sun, B., Carpino, J., Chapler, E. S., Ching, J., Choi, Y., Jhun, K., Kim, J. D., Lallos, G. G., Morgenstern, R., Singh, S., Theja, S., & Dennehy, J. J. (2013). Frequency and Fitness Consequences of Bacteriophage Φ6 Host Range Mutations. PLoS ONE, 9(11). https://doi.org/10.1371/journal.pone.0113078

Ford, Brian Elwood, "Pseudomonas Bacteriophage Phi6 as a Model for Virus Emergence" (2015). CUNY Academic Works.

Gottlieb, P., & Alimova, A. (2022). RNA Packaging in the Cystovirus Bacteriophages: Dynamic Interactions during Capsid Maturation. International Journal of Molecular Sciences, 23(5). https://doi.org/10.3390/ijms23052677
https://academicworks.cuny.edu/gc_etds/924

McLeish, M. J., Fraile, A., & García-Arenal, F. (2018). Ecological Complexity in Plant Virus Host Range Evolution. Advances in Virus Research, 101, 293-339. https://doi.org/10.1016/bs.aivir.2018.02.009

Mindich, L. (1988). Bacteriophage ϕ6: A Unique Virus Having a Lipid-Containing Membrane and a Genome Composed of Three dsRNA Segments. Advances in Virus Research, 35, 137-176. https://doi.org/10.1016/S0065-3527(08)60710-1

Mindich, L., Nemhauser, I., Gottlieb, P., Romantschuk, M., Carton, J., Frucht, S., Strassman, J., Bamford, D. H., & Kalkkinen, N. (1988). Nucleotide sequence of the large double-stranded RNA segment of bacteriophage phi 6: Genes specifying the viral replicase and transcriptase. Journal of Virology, 62(4), 1180-1185. https://doi.org/10.1128/jvi.62.4.1180-1185.1988

Mullick, B., Magar, R., Jhunjhunwala, A., & Farimani, A. B. (2021). Understanding mutation hotspots for the SARS-CoV-2 spike protein using Shannon Entropy and K-means clustering. Computers in Biology and Medicine, 138, 104915.

Oechslin F. (2018). Resistance Development to Bacteriophages Occurring during Bacteriophage Therapy. Viruses, 10(7), 351. https://doi.org/10.3390/v10070351

Sanjuán, R., & Domingo-Calap, P. (2015). Mechanisms of viral mutation. Cellular and Molecular Life Sciences, 73(23), 4433-4448. https://doi.org/10.1007/s00018-016-2299-6

Vilcek S. (2020). SARS-CoV-2: Zoonotic origin of pandemic coronavirus. Acta virologica, 64(3), 281–287. https://doi.org/10.4149/av_2020_302

Walker, S. E., & Lorsch, J. (2013). RNA Purification – Precipitation Methods. Methods in Enzymology, 530, 337-343. https://doi.org/10.1016/B978-0-12-420037-1.00019-1

Wargo, A. R., & Kurath, G. (2012). Viral fitness: definitions, measurement, and current insights. Current opinion in virology, 2(5), 538–545. https://doi.org/10.1016/j.coviro.2012.07.007

Woolhouse ME, Haydon DT, Antia R. 2005. Emerging pathogens: the epidemiology and evolution of species jumps. Trends Ecol Evol 20:238–244

Zhao, L., & Duffy, S. (2019). Gauging genetic diversity of generalists: A test of genetic and ecological generalism with RNA virus experimental evolution. Virus Evolution, 5(1). https://doi.org/10.1093/ve/vez019

Zhao, L., Seth-Pasricha, M., Stemate, D., Crespo-Bellido, A., Gagnon, J., Draghi, J., & Duffy, S. (2019). Existing Host Range Mutations Constrain Further Emergence of RNA Viruses. Journal of virology, 93(4), e01385-18. https://doi.org/10.1128/JVI.01385-18