

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

A trait-based approach to understanding the evolvability of viral host-range expansions

Permalink

<https://escholarship.org/uc/item/3wf539pm>

Author

Strobel, Hannah Megan

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

A trait-based approach to understanding the evolvability of viral host-range expansions

A Dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Biology

by

Hannah Megan Strobel

Committee in charge:

Professor Justin Meyer, Chair
Professor Randolph Hampton
Professor Sergey Kryazhimskiy
Professor Ulrich Muller
Professor Joseph Pogliano

2022

Copyright

Hannah Megan Strobel, 2022

All rights reserved

The Dissertation of Hannah Megan Strobel is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

TABLE OF CONTENTS

DISSERTATION APPROVAL PAGE	iii
TABLE OF CONTENTS.....	iv
LIST OF FIGURES	vi
LIST OF TABLES	viii
ACKNOWLEDGEMENTS	ix
VITA.....	xi
ABSTRACT OF THE DISSERTATION	xii
CHAPTER 1 - Literature Review - A trait-based approach to predicting viral host-range evolvability	1
1.1 Abstract.....	1
1.2 Introduction	1
1.3 Mutation rates	3
1.4 Recombination rates	8
1.5 Phenotypic robustness and heterogeneity.....	14
1.6 Conclusion.....	26
1.7 Acknowledgements	26
CHAPTER 2 - Conformational heterogeneity in a bacteriophage receptor binding protein opens an evolutionary pathway to expanded host range.....	38
2.1 Abstract.....	38
2.2 Introduction	38
2.3 Results	41
2.4 Discussion.....	49
2.5 Materials and Methods	53
2.6 Acknowledgements	59
2.7 References	59
CHAPTER 3 - Viral protein instability enhances host-range evolvability	64
3.1 Abstract.....	64

3.2 Introduction	64
3.3 Results	67
3.4 Discussion.....	86
3.5 Materials and methods.....	90
3.6 Acknowledgments	105
3.A Appendix	105
3.7 References	118
CHAPTER 4 - Evolvability is an important trait in the selection of bacteriophages for therapeutic use.....	122
4.2 Introduction	123
4.3 Materials and Methods	126
4.4 Results	131
4.5 Discussion.....	138
4.6 Acknowledgements	141
4.A Appendix	142
4.7 References	150
CONCLUSIONS AND FUTURE DIRECTIONS	154

LIST OF FIGURES

Figure 1.1 - An approach to predicting virus host-range expansion evolvability by examining intrinsic traits and constraints	22
Figure 1.2 - Mutations can alter the shape of virus-host protein interactomes	23
Figure 1.3 - Genome modularity may prevent the costs of recombination, facilitating host-range expansion	24
Figure 1.4 - Virus evolvability is in many ways shaped by the ability of viral proteins to evolve new functions. Stability is one trait that affects evolvability	25
Figure 2.1 Despite nearly identical AlphaFold predicted structures, the evolved protein is dramatically less soluble	46
Figure 2.2 In a thermal shift assay, the evolved J has a higher initial fluorescence and lower melting point than the ancestor	47
Figure 2.3 A mild heat treatment selectively reduces activity on the new receptor, while not affecting activity on the ancestral receptor	48
Figure 3.1 - Experimental system and selection for thermostability	79
Figure 3.2 - Evolution experiment with naturally occurring thermostable genotypes	80
Figure 3.3 - Evolution experiment with engineered library genotypes	81
Figure 3.4 - Mapping thermostabilizing and putative destabilizing mutations on a structural prediction of J's reactive region	82
Figure 3.5 -Trajectories of evolution of OmpF ⁺ in unmodified and stabilized backgrounds.....	83
Figure 3.6 - Growth rates of naturally occurring thermostable genotypes and engineered library genotypes	84
Figure 3.7 - Stability and evolvability of 5-mut and 5-mut T987A stable variant	85
Figure 3.A.1 - Phage titer during evolution experiment on naturally occurring thermostable genotypes	108
Figure 3.A.2 - Aligned AlphaFold predictions for ancestor and 6-mut J proteins	109
Figure 3.A.3 - Editing N1107K into all variant backgrounds	110
Figure 3.A.4 - Turbid clearings on plates correspond to weak growth in liquid.....	111
Figure 3.A.5 - Independently created T987C lysogen	112

Figure 3.A.6 - Productivity of stable genotypes with and without N1107K.....	113
Figure 4.1 - A trio of closely related phage genotypes demonstrate a three-way trade-off between stability, reproduction, and evolvability	134
Figure 4.2 - Bacterial suppression dynamics of the three genotypes	135
Figure 4.3 - Bacterial suppression dynamics of two genotypes differing only in their reproductive rates.....	136
Figure 4.4 - Environmental contingency of bacterial suppression dynamics.....	137
Figure 4.A.1 - Phage Titer from Figure 4.2.....	142
Figure 4.A.2 - Phage Titer from Figure 4.3.....	143
Figure 4.A.3 - Net reproductive rate of phages used in Figure 4.3	144
Figure 4.A.4 - Phage Titer from Figure 4.4.....	145

LIST OF TABLES

Table 3.A.1 - High throughput MAGE to engineer N1107K into genotypes that either did not evolve OmpF ⁺ in the evolution experiment or required two mutations to evolve OmpF ⁺	114
Table 3.A.2 - MAGE oligos	115
Table 3.A.3 - Unadjusted and adjusted decay rates for engineered library variants	116
Table 3.A.4 - λ growth rates on lamB ⁻ hosts measured in liquid culture on D6 of the replay experiment	117
Table 4.A.1 - Phage trait values used to create Figure 4.1A	146
Table 4.A.2 - Phage titers used to initiate suppression experiments	147
Table 4.A.3 Statistics from Figure 4.2, 4.3, and 4.4.....	148

ACKNOWLEDGEMENTS

I would first like to thank my doctoral advisor, Dr. Justin Meyer. In addition to being a great scientist, Dr. Meyer has also been a wonderful mentor. The last three years of my Ph.D. were a tumultuous time in the world, and yet my graduate experience was relatively smooth, and I credit a great deal of that to Dr. Meyer. I also thank my doctoral committee, Dr. Randolph Hampton, Dr. Sergey Kryazhimskiy, Dr. Ulrich Muller, and Dr. Joseph Pogliano for providing useful feedback on my work throughout the years. I am also grateful to Dr. Hampton for helping me learn protein techniques. For funding support, I would like to thank the Chris Wills Endowed Graduate Student Research Award and the Cell and Molecular Genetics Training Grant.

For their unwavering support and friendship, I would like to thank the other members of the Meyer lab: Animesh Gupta, Joshua Borin, Liz Stuart, and Elijah Horwitz, as well as the other graduate students in the Ecology, Behavior, and Evolution section. In particular, I would like to thank Elizabeth Bullard for reminding me that my value is not dependent on my work.

Finally, I give my sincere thanks to my family. My siblings, Molly Penzenik and Derek Strobel, have been a constant source of support and inspiration. I also thank Maria Tahir for her enduring friendship and for reminding me that the world is full of joy. I am incredibly grateful to Jason Good for his companionship throughout the past three years, and I am thrilled to embark on the next chapter together. And I thank my mother, Michelle Albertus, for bringing me up with unconditional love.

The Chapter 1, in full, is a reprint of the material as it appears in Strobel HM, Stuart EC, Meyer JR (2022). A trait-based approach to predicting viral host-range evolvability. Annual Review of Virology. In press. The dissertation author was the primary investigator and author of this paper.

Chapter 2, in full, is coauthored with Meyer, Justin R. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is a reprint of the material as it appears in Strobel HM, Horwitz EK, Meyer JR (2022). Viral protein instability enhances host-range evolvability. PLoS Genet 18: e1010030. The dissertation author was the primary researcher and author of this paper.

Chapter 4 in full, is a reprint of the material currently being prepared for publication. Horwitz EK, Strobel HM, Meyer JR (2022). Evolvability is a key trait in the selection of bacteriophages for therapeutic use. The dissertation author was one of the primary investigators and the second author of this paper.

VITA

- 2015 Bachelor of Science in Ecology and Evolutionary biology, Tulane University
- 2022 Doctor of Philosophy in Biology, University of California San Diego

PUBLICATIONS

- Horwitz EK, **Strobel HM**, Meyer JR (2022). Evolvability is a key trait in the selection of bacteriophages for therapeutic use. In prep.
- Strobel HM**, Stuart EC, Meyer JR. 2022. A trait-based approach to predicting viral host-range evolvability. *Annual Review of Virology*. In press.
- Strobel HM**, Horwitz EK, Meyer JR. 2022. Viral protein instability enhances host-range evolvability. *PLoS Genetics* 18: e1010030.
- Gupta A, Zaman L, **Strobel HM**, Gallie J, Burmeister AR, Kerr B, Tamar ES, Kishony R, Meyer JR. Host-parasite coevolution promotes innovation through deformations in fitness landscapes. *eLife*. In press.
- Strobel HM**, Hays SJ, Moody KN, Blum MJ, Heins DC. 2019. Estimating effective population size for a cestode parasite infecting three-spined sticklebacks. *Parasitology*, 146: 883-896.
- Strobel HM**, Alda F, Sprehn CG, Blum MJ, Heins DC. 2016. Geographic and host-mediated population genetic structure in a cestode parasite of the three-spined stickleback. *Biological Journal of the Linnean Society*, 119 (2): 381-396.

ABSTRACT OF THE DISSERTATION

A trait-based approach to understanding the evolvability of viral host-range expansions

by

Hannah Megan Strobel

Doctor of Philosophy in Biology

University of California San Diego, 2022

Professor Justin Meyer, Chair

For decades, scientists have been fascinated by the ease with which viruses, seemingly simple life forms, evolve new feats of innovation. One viral innovation relevant to humans is gaining infectivity on a new host type. Although numerous instances of viral host-range expansion have been documented, we still lack the ability to predict them reliably. In part, this is because many factors affect whether a host shift will occur, ranging from molecular interactions to host behavior. To increase the tractability of this problem, scientists are beginning to ask whether viruses might vary in their innate evolvability, or capacity for adaptive evolution. I focused on the role of stability, here defined as thermodynamic stability, an intrinsic trait of viral

proteins thought to enhance evolvability. Using the well-studied host-range expansion of bacteriophage λ , I first showed how mutations that confer expanded host range destabilize the receptor binding protein and allow it to assume alternative conformations with new binding activity. Then, I showed that among λ genotypes varying in stability, the most evolvable tended to be the most unstable, and the stable genotypes that did evolve gained destabilizing mutations. Instability promoted the evolution of new host range, in contrast to the widely cited consensus that stability enhances evolvability. I discovered one λ genotype that exhibited high stability and evolvability, but it grew poorly, suggesting a three-way tradeoff between stability, evolvability, and reproduction. This result led to another question: if traits that affect current fitness, like stability and reproductive rate, trade off with future evolutionary capacity, then which traits most influence the outcome of coevolutionary arms races between viruses and their hosts? I examined which traits influence λ 's ability to overcome host resistance and maintain infectivity on its coevolving host bacteria. The fast-reproducing, evolvable, but unstable genotype emerged most successful, suggesting that lineages that initially appear poorly adapted may give rise to progeny that persist due to their capacity to evolve. This work suggests that a positive, linear relationship between stability and evolvability does not hold in all scenarios, with important implications for predicting viral emergence and selecting genotypes for phage therapy.

CHAPTER 1

Literature Review - A trait-based approach to predicting viral host-range evolvability

1.1 Abstract

Predicting the evolution of virus host range has proven to be extremely difficult, in part because of the sheer diversity of viruses, each with unique biology and ecological interactions. We have not solved this problem, but to make the problem more tractable, we narrowed our focus to three traits intrinsic to all viruses that may play a role in host-range evolvability: mutation rate, recombination rate, and phenotypic heterogeneity. Although each trait should increase evolvability, they cannot do so unbounded because fitness trade-offs limit the ability of all three traits to maximize evolvability. By examining these constraints, we can begin to identify groups of viruses with suites of traits that make them especially concerning, as well as ecological and environmental conditions that might push evolution toward accelerating host-range expansion.

1.2 Introduction

Living through the ongoing severe acute respiratory syndrome 2 (SARS-CoV-2) pandemic, researchers and the public have become interested in understanding how viruses shift host species and the factors that contribute to disease emergence. The goal is to leverage this knowledge to design pre-pandemic intervention strategies. The specifics of how emergence works vary with each virus and novel host; however, there are universal hurdles all viruses must overcome to shift hosts. The virus must gain the ability to recognize and inject genetic material

into the new species' cells and inject genetic material into the cell. Once the genetic information is within the cell, it must be able to take control of cellular processes and replicate new copies of the viral genome, synthesize proteins and other molecules, and then assemble the parts into infectious particles. The virions must then escape the cell and find a new host. All of this must be completed while also avoiding host defenses. The full process involves hundreds of host-viral interactions working out in the virus's favor. For example, SARS-CoV-2 is known to interact with 332 human proteins to complete its infection cycle (Gordon et al. 2020; V'Kovski et al. 2021). Given how many opportunities there are for misalignments between the emerging virus and its new host, it would seem nearly impossible that host-shifts ever occur. Yet they do. This is in part because host species share common ancestry and maintain similar molecular pathways, which viruses adapted to other hosts can plug in to. Additionally, viruses have shown extraordinary ability to generate genetic variation that allows them to ameliorate host-incompatibilities and switch host species (Figure 1.1).

For this review, we focused on understanding the evolutionary aspects of the host-range expansion and viral emergence. Multiple excellent reviews have already been published on the evolution of host-range expansion that provide a thorough background on the subject (Hall et al. 2013; Peck et al. 2015; de Jonge et al. 2019; Rothenburg and Brennan 2020). Given this, we narrowed our focus to the topic of viral evolvability. Evolvability is the capacity of lifeforms to adapt and since viral adaptation plays an important role in host-range expansion, it is reasonable to expect that more evolvable viruses are more likely to emerge. Viral species are known to vary greatly in their evolvability (e.g., low mutation rates of DNA-based versus RNA-based viruses), but even viruses separated by just a single mutation can vary in their evolutionary potential (Strobel et al. 2022). Given the variation in evolvability and its likely importance in predicting

host-shifts, identifying drivers of evolvability could play an important role in identifying the factors that contribute to viral host-shifts.

To tackle the subject of drivers of host-range evolvability, we started by searching for viral traits that could enhance evolvability. We then focused on the subset that have a well-developed theory and/or direct experimental evidence to support their role and settled on mutation rate, recombination rate, and protein stability. For each trait, we explored the theory behind how the trait affects evolvability, the relationship between trait values and evolvability, and empirical evidence to support the trait's influence on evolvability. Next, we explored constraints on the traits' evolution by determining possible tradeoffs and the types of environments that could tip the scales towards evolving increased or decreased evolvability.

1.3 Mutation rates

Often the barrier for viruses to shift hosts is a small number of mutations that help the virus ameliorate incompatibilities with the new host (Figure 1.2). The following are a few examples to help visualize the role mutation plays in host-range expansion. Many bacterial cells are resistant to the bacteriophage λ because the cells do not express the outer membrane protein LamB, which λ uses as its receptor. Four mutations in the host-recognition protein can allow λ to gain access to these hosts by interacting with a completely new surface protein, OmpF (Meyer et al. 2012; Maddamsetti et al. 2018). Similar mutations in the tail fiber of other bacteriophage have been linked to host-range expansions to new strains (Tétart et al. 1996b; Yehl et al. 2019; Boon et al. 2020a) or entirely new species (Crill et al. 2000; Duffy et al. 2006; Duffy et al. 2007). This type of evolution also underlies mammalian viruses shifts to humans (Koel et al. 2013; Lu et al. 2013; Linster et al. 2014; Shi et al. 2014; Song et al. 2017a). Incompatibilities may also arise

internal to the cell too (Bradel-Tretheway et al. 2011; Haller et al. 2014), but these can also be repaired by relatively few mutations. For example, λ is unable to infect bacteria that lack DnaJ that is involved in λ -DNA replication and ManXYZ involved in transporting λ -DNA into the cytoplasm (Maynard et al. 2010). However, λ can overcome these missing elements and successfully infect the hosts by gaining mutations that allow the virus to no longer rely on these host proteins during replication (Gupta et al. 2020).

Given that so many host-virus cellular incompatibilities can be solved with relatively few mutations, it is reasonable to expect that viruses with higher mutation rates would be more likely to expand their host range. This effect should be amplified when a virus has multiple incompatibilities because the chance of gaining multiple mutations simultaneously is the product of the mutation rate (e.g., 10-fold increase in mutation rates will increase the probability of uncovering two mutations by 100-fold, 1,000-fold for three, and so on). Viruses also have enormous variation in mutation rates, with RNA viruses mutating on average 1 in 10^3 bases per replication, and DNA viruses at 1 in every 10^8 (Sanjuán 2012). Furthermore, viruses can evolve increased mutation rates through mutations that alter the activity of the enzymes that synthesize genetic material or that carry out proofreading (Elena and Sanjuán 2005). Given this variation and their capacity to evolve increased mutation rates, mutation rate is expected to be an important driver of host-range evolvability.

Increased mutation rates may increase host-range evolvability through less intuitive mechanisms as well. For example, a computational model of viral host-range evolution showed that high viral mutation rates may help viruses maintain broad host-ranges ((Fisher 2021). Under low host diversity conditions, in which a virus serially infects the same species, natural selection is expected to eliminate energetically costly traits that allow the virus to infect unavailable hosts.

Viruses with high mutation rates can maintain a broad host-range in the face of this selection pressure because host-range expansion mutations are generated faster than natural selection can purge them. This provides the viral population with genetic variation that can be readily employed if a new host is encountered.

If mutation is key to host-range evolvability, does that mean that viruses with the highest mutation rates are most likely to shift hosts? Not necessarily, since a large fraction of mutations are deleterious and can slow adaptive evolution. This cost to mutation is especially pronounced in viruses. It is estimated that 20-41% of mutations are lethal in five very different viruses (DNA and RNA viruses that infect bacterial, plant, and animal hosts) (Sanjuán 2010). These surprisingly high values were confirmed by a second study in which three DNA and three RNA bacteriophages were studied side-by-side. The average percentage of deleterious mutations in DNA viruses was 29% and 20% for RNA viruses (Domingo-Calap et al. 2009). Additional methods were used to measure the average non-lethal mutation fitness effect, which was deleterious for DNA (-0.027) and RNA (-0.047) viruses. These findings are also in line with a human virus, influenza A, where researchers found 31.6% of mutations are lethal, but most of the non-lethal mutations are deleterious (Visher et al. 2016). Taken together, viruses experience high numbers of deleterious mutations, and increasing mutation rates too high will impart a significant fitness cost on viruses and limit their evolvability.

Moreover, certain mutations can reshape viral fitness landscapes through epistasis and restrict the availability of host-range expansion mutations, This creates a second-order effect of the accumulation of mutations on reducing evolvability is that certain mutations can re-shape viral fitness landscapes through epistasis and restrict the availability host-range expansion mutations. This idea was explored through experiments conducted on three genotypes of

bacteriophage $\phi 6$ that typically infects *Pseudomonas. syringae pv. phaseolicola* but is known to evolve to infect multiple new *Pseudomonas* species (Zhao et al. 2019). Two of the $\phi 6$ genotypes studied had previously evolved the ability to infect the novel host *Pseudomonas syringae pv. Tomato* and the third was their ancestor. Through sequencing populations descended from these three genotypes, researchers found that the evolved genotypes had access to fewer mutations for expansion to a third host, *Pseudomonas syringae pv. atrofaciens*. This showed that the accumulation of mutations, even beneficial mutations, can alter viral genomes in ways that increase epistasis and can restrict their host-range evolvability.

Natural patterns of viral variation lead to the hypothesis that increases in mutation rates increase evolvability, but only to a certain point. In 2012 an analysis was published by Rafael Sanjuán that studied mutation rates measured for 84 different viruses that span the full spectrum of mutation rates (Sanjuán 2012). Sanjuán found a strong positive correlation between mutation rate and evolutionary rate, wherein an increase in mutation showed a proportional increase in evolutionary rate, up to a point after which it plateaued. Using a mathematical theory based on expected fitness effects of mutations, Sanjuán showed that evolutionary rate should begin to decline with increases in mutation rates. In line with this prediction are results from laboratory experiments that increase mutation rates beyond natural levels. These studies show that viral fitness rapidly declines beyond the predicted threshold and at high enough levels that viruses can even go extinct (Anderson et al. 2004; Bull et al. 2007; Domingo 52006). Indeed, lethal mutagenesis is the basis of viral therapies, including molnupiravir for SARS-CoV-2 treatment (Malone and Campbell 2021). Altogether this shows that viruses with higher natural mutation rates are more likely to evolve expanded host ranges, but this is expected to be true only to a certain point.

Most viruses exist below the threshold where increases in mutation rate will cause a decline in fitness. Thus, is presumably opportunity for many viruses to increase their mutation rates to possibly become more evolvable. How might the environment cause viruses to increase their mutation rates, either directly through exposure to mutagens, or indirectly by selecting for viruses with elevated mutation rates? This question is especially relevant in the context of anthropogenic global change because the changing environment could be changing in ways to cause higher mutation rates and ultimately increase the risk of emerging diseases. This topic was explored for the A/H1N1 strain influenza virus, where researchers assessed the effects of temperature, population density, precipitation, and social development on genomic substitution rate (Jiang et al. 2020). Researchers examined 11,721 cases of H1N1 from locations across the globe. They assessed the nucleotide substitution rate by comparing the genetic sequences of the focal viruses to the sequence of the earliest reported isolate from each location. Minimum annual temperature had a nonlinear association with mutation, with mutation peaking at 15°C. Population density was found to have a positive association with substitution rate, In contrast, no correlations were found between precipitation and social development.

Taken together, the environment has a role in shaping viral evolvability and global change can alter viral evolvability. This is particularly problematic because global change, such as deforestation, is predicted to also cause humans to encounter more zoonotic disease, which will further increase the chance of disease emergence. If those diseases have heightened mutation rates or are generally more evolvable, then this could help tip the scales towards increased frequency of host-shifts. More work along these lines is warranted, especially studies that more directly assess mutation rates and consider additional environmental variables and their interactions.

1.4 Recombination rates

Point mutations play an important role in driving host-range evolution. However, they require time and, if multiple mutations are required, relatively smooth paths of incremental gain in the fitness landscape to evolve (Burmeister et al. 2016; Gupta et al. 2021). Recombination provides an opportunity to transfer large amounts of genetic material between genomes, driving even more rapid genetic divergence. In the context of virus host-range evolution, recombination could facilitate increased evolvability of host-range expansions if genomes exchange genetic elements that confer infectivity on new hosts.

In viruses, there are many mechanisms for recombination that have been reviewed extensively elsewhere (Pérez-Losada et al. 2015), so we will briefly highlight just a few. In viruses with DNA genomes, recombination typically occurs via pathways related to DNA replication and repair (Young et al. 1984; Wilkinson and Weller 2004; Weller and Sawitzke 2014). Some viruses rely on host-encoded recombination machinery, while others encode their own recombination proteins (Bobay et al. 2013). Two of the most thoroughly understood virus-encoded recombination systems are the λ Red system of bacteriophage λ (Murphy 2016) and the T4 recombination system (Liu and Morrical 2010). Viruses with RNA genomes primarily use a copy choice mechanism, in which the RNA-dependent RNA polymerase or reverse transcriptase jumps from one piece of RNA to another (Worobey and Holmes 1999; Simon-Loriere and Holmes 2011). For both DNA and RNA viruses, the extent to which homology plays a role in determining the sites of recombination is highly variable; indeed there is evidence that recombination between viral genomes can occur even with substantial sequence divergence (Lai 1992; Morris et al. 2008; Bobay et al. 2013; De Paepe et al. 2014).

How might recombination rates themselves evolve? Elevated recombination rates may result from the use of virus-encoded recombination systems. In a study of lambdoid phages, those that had evolved their own recombination machinery tended to have more genomic mosaicism than related viruses that relied on host-encoded recombination systems (Bobay et al. 2013). For viruses that use host-recombination machinery, it is possible that modulating the use of host enzymes that ensure fidelity (Dudenhöffer et al. 1998) could also be used to increase or decrease recombination. In RNA viruses, recombination rate could be modulated indirectly via RNA secondary structure evolution (Simon-Loriere et al. 2010). For example, in human immunodeficiency virus, RNA secondary structure influences the rate of recombination (Galetto et al. 2006). If an increased recombination rate enables an ancestral virus to generate more diverse progeny, descendants with adaptive variation would also carry the trait of high recombination, causing it to be indirectly selected for (Tenailon et al. 2001).

Elevated recombination rates might enhance evolvability of host-range expansion (Patiño-Galindo et al. 2021). One group of animal viruses thought to be particularly prone to host-range evolution via recombination is the coronaviruses (Banner and Lai 1991; Fang et al. 2005; Jackwood et al. 2010; Decaro and Lorusso 2020). There have been three separate emergences of coronaviruses in humans this century, and there is evidence that in all three instances, the strains responsible arose via a combination of point mutations and recombination in the spike proteins. Recombination is most clearly implicated in the evolution of the strain that caused the 2003 severe acute respiratory syndrome coronavirus virus pandemic (Graham and Baric 2010; Hu et al. 2017). Recombination also clearly occurred in the recent evolutionary history of the strain of Middle East respiratory syndrome coronavirus that caused the 2012 outbreak (Dudas and Rambaut 2016), and subsequent recombination among strains circulating in

humans likely increased its transmissibility (Wang et al. 2015). Investigations into the origins of the strain that caused the SARS-CoV-2 pandemic are ongoing. One hypothesis is that recombination enabled emergence in humans by replacing the receptor binding motif (RBM) in a bat coronavirus with that of a pangolin coronavirus capable of binding the human angiotensin converting enzyme-2 (ACE2) receptors (Li et al. 2020). An alternative hypothesis is that SARS-CoV-2 is descended with little change from a bat coronavirus that already possessed the key ACE2-binding residues, and the immediate progenitors to SARS-CoV-2 have simply not been sampled (Boni et al. 2020).

There is evidence for recombination's role in host shifts in other animal viruses, and by mechanisms other than altered host-recognition. A well-known example occurred when an eastern equine encephalitis virus and a sindbis-like virus hybridized and produced a strain with new antigen specificity (Hahn et al. 1988). In baculoviruses, host-range expansion has been observed to occur via recombination in natural infections (Kondo and Maeda 1991), and host-range can be intentionally engineered via recombination of the helicase genes from two different baculovirus (Wu et al. 2004). Recombination has also been widespread in the evolution of plant geminiviruses and has likely contributed to host-switching in agriculturally important hosts (Padidam et al. 1999; Lefeuvre and Moriones 2015). Another intriguing possibility is that recombination might provide a mechanism for rapidly excising genomic elements that trigger host antiviral response (Aguado et al. 2018). Positive-stranded RNA viruses, which more readily undergo recombination compared to negative-stranded RNA viruses, were more evolvable in escaping host antiviral defenses (Aguado et al. 2018).

Extensive work on T-even bacteriophages has shown that genes encoding the tail fiber proteins that determine host-range readily recombine, even between relatively divergent

sequences (Tétart et al. 1996b; Tétart et al. 1998). Transferring entire tail fiber genes, or, more rarely, specific regions within genes, between phages with distinct host ranges conferred the host-range of the donor upon the recipient phage (Tétart et al. 1996b; Tétart et al. 1998; Mahichi et al. 2009). Recombinants were generated under laboratory conditions, but it is possible that similar re-shuffling of host-specificity regions could allow naturally evolving viruses to expand host-range.

Understanding the conditions that favor host-range evolvability in bacteriophages also has applications to phage therapy, in which it would be useful to intentionally broaden the host-range of phages. One approach to generating broad host-range phages uses conditions that favor recombination to accelerate host-range evolution (Burrowes et al. 2019). Iterative rounds of evolution with a cocktail of different phages yielded a phage with a host-range that is even broader than the sum of the ranges of the initial cocktail (Burrowes et al. 2019). Remarkably, the evolved phage with the broadest host-range underwent at least 48 recombination events between two of the initial cocktail strains (Burrowes et al. 2019). A different study exploring the use of “training” to pre-evolve λ phage for use in phage therapy applications identified a highly suppressive variant that contained both point mutations and a recombination in the host-recognition protein (Borin et al. 2021). Intriguingly, the recombination occurred not with a co-infecting phage but with a relict prophage encoded in the host genome (Borin et al. 2021), a phenomenon that has been observed elsewhere (Zhang et al. 2013).

Recombination may generate genetic variation favorable for evolution, but high rates of recombination likely come with costs, such as the production of defective progeny. One way this can occur is if recombination occurs within the coding sequence of a protein, resulting in a nonfunctional protein due to frameshifts or a chimera. The viability of chimeric proteins depends

on the similarity of the peptide sequences and the location of the breakpoint, but many chimeric proteins are non-viable due to disruptions in protein folding (Drummond et al. 2005; Lefevre et al. 2007). Recombination can also cause genome truncation, rendering some progeny incapable of completing a complete infection cycle (Poirier et al. 2015). In some viruses, the disadvantage is compounded because the non-viable particles can interfere with the production of viable particles (Giachetti and Holland 1989; Frensing et al. 2013). Another potential cost of recombination is the production of incompatible hybrids. Even in highly related virus genotypes recombination can create incompatibilities between genes (Sackman et al. 2015) or even within the same gene (Meyer et al. 2016). This was observed between two closely related λ genotypes that specialize on different *Escherichia coli* receptors (Meyer et al. 2016). When host-recognition protein mutations from different genotypes were engineered into a single hybrid protein, the resulting phage was inviable (Meyer et al. 2016).

Virus genomes may have evolved properties that minimize the costs of recombination. Computational and experimental studies have shown that intragenic recombination, occurring within protein coding sequences, appears to be less disruptive than might be expected (Drummond et al. 2005; Lefevre et al. 2007; Golden et al. 2014). It is not clear whether this is because non-viable chimeric proteins are purged by selection or because viruses have evolved recombination hotspots at domain boundaries where breakpoints will be less disruptive (Lefevre et al. 2007; Golden et al. 2014). It has also been observed that genome-wide recombination breakpoints are more likely to occur at gene boundaries than would be expected by chance (Lefevre et al. 2009). This could be partially explained by selection purging non-viable recombinants (Martin et al. 2011), but it is also possible that viral genomes may have also undergone evolution to favor genomic architecture that minimizes the disruption of co-evolved

genes with interacting functions. In maize streak virus, recombination was more favorable when the genome fragment being exchanged did not have extensive interactions with the rest of the genome (Martin et al. 2005). Many virus genomes are organized such that genes with related functions are positioned together (Hendrix et al. 2000). For example, in the genome of bacteriophage λ , the genes coding for the tail shaft proteins, tail tip proteins, and side tail fiber proteins are clustered together (Casjens and Hendrix 2015), and the same is true for T4-related phages (Comeau et al. 2007). It is thought that virus evolutionary history has been shaped by exchanging functional modules (Botstein 1980). Viruses with genomes characterized by spatial modularity might more easily transfer the elements necessary to exploit novel hosts, potentially making them more evolvable with respect to host-range expansion (Figure 1.3).

There are several host-related, ecological, and environmental factors that might shape the relationship between recombination and evolvability. High multiplicity of infection could enhance evolvability via recombination by increasing the potential for increasing sequence diversity (Bocharov et al. 2005). However, high MOI does not always result in increased viral evolvability. For example, in a study in which phage $\Phi 2$ was grown with its host bacteria, *Pseudomonas fluorescens*, under varying MOI, and infectivity did not evolve faster at high multiplicity of infection MOI (Hall et al. 2012). However, this study examined adaptation to the current host rather than host-range expansion. Because recombination can occur between related viruses when they co-infect the same host, industrial agriculture and the wildlife trade likely create conditions that facilitate the evolution of novel viruses through recombination (Priyadarsini et al. 2020).

1.5 Phenotypic robustness and heterogeneity

Mutation and recombination provide the genetic variation that can allow viruses to infect new hosts. However, virus host-shifts cannot be predicted solely based on the supply of genetic variation, in part because of higher-order mechanisms, such as genetic robustness and nongenetic phenotypic heterogeneity, which tune the extent to which genetic changes affect the virus phenotype. Genetic robustness suppresses the effects of mutation on the phenotype (Wagner 2005), which in the immediate term might be expected to hinder the production of novel phenotypes, potentially slowing adaptation. However, because a high fraction of mutations are deleterious (Eyre-Walker and Keightley 2007), robustness can favor the accumulation of cryptic genetic diversity that may lead to adaptation in the longer term (Hayden et al. 2011; Zheng et al. 2019). In essence, genetic robustness allows viruses to traverse otherwise insurmountable fitness valleys, enabling them to eventually ascend fitness peaks. The second mechanism, phenotypic heterogeneity, allows organisms to generate a range of phenotypes from a single genotype without underlying genetic variation, and therefore can be conceptualized as the opposite of robustness. A genotype with the ability to express multiple phenotypes can bypass the delay associated with acquiring adaptive mutations and therefore might experience greater adaptability (Holland et al. 2014; Bódi et al. 2017). Although robustness and heterogeneity tune phenotypic response in opposite directions, there is evidence that both can enhance evolvability. Most studies on this topic have been performed on enzymes, (Bloom et al. 2006; Bloom and Arnold 2009; Zheng et al. 2020), but because the structural components of viruses are mostly proteins, insights gained from these studies can inform drivers of viral evolvability.

A consensus has emerged that genetic robustness enhances enzyme evolvability (Payne and Wagner 2019). Because robustness can be difficult to characterize in living systems, many

studies approached the question by manipulating thermostability, a trait that is intrinsically linked to robustness. Thermostability is a measure of resistance to heat, and it is thought that thermostable proteins tend to also be robust to mutation (Besenmatter et al. 2007). Enzymes with high thermostability are buffered against the destabilizing effects of mutations, allowing the proteins to evolve more mutations, and increasing the likelihood that an adaptive mutation is uncovered (Bloom et al. 2006; Bloom and Arnold 2009; Zheng et al. 2020).

There is some evidence that robustness might also promote evolvability in viruses, but this evidence is not unequivocal. Robustness increased the evolvability of thermotolerance in bacteriophage $\phi 6$ (McBride et al. 2008), although robustness and thermotolerance are traits that tend to be correlated (Domingo-Calap et al. 2010), and it is possible that this pattern would not generalize to the evolvability of other traits. In one study in line with the stability-evolvability link, a vesicular stomatitis virus that had been selected for thermotolerance exhibited enhanced antigenic diversification and antibody escape (Presloid et al. 2016). However, in a different study comparing the host-range evolvability of two vesicular stomatitis viruses, the less robust strain host-range evolved faster (Cuevas et al. 2009).

In contrast to robustness, there is a less extensive literature on the role of phenotypic heterogeneity in protein evolvability. However, recent with advances in structural methods capable of detecting heterogeneity have renewed interest in the implications of heterogeneity for evolvability. Conceptually, it seems plausible that proteins with more structural heterogeneity may have immediate access to phenotypes that carry out new functions (Tokuriki and Tawfik 2009; Sikosek and Chan 2014). There is evidence to support this assertion in laboratory studies. For example, decreasing structural rigidity by removing amino acid contacts resulted in a more evolvable scaffold for designing a novel metallo- β -lactamase via directed evolution (Song et al.

2017b). Similarly, a mutation that stabilizes a non-native, alternative conformation in PSD95, a synaptic scaffolding protein, allows the protein to recognize multiple classes of ligands at once, thereby acting as an evolutionary bridge between an ancestral protein that recognizes only one class and a double mutant protein that recognizes only a second class (Raman et al. 2016).

Another study revealed that the ability of antibody proteins to achieve specific recognition of a diverse set of targets depends on precursors that are conformationally heterogeneous. Subsequent mutations introduce contacts that increase the structural rigidity of a single conformation, generating specificity to a single target in the mature antibody (Zimmermann et al. 2006). It is intriguing to speculate that this pattern may be mirrored in the evolution of other protein types, with “metamorphic” intermediates capable of folding into multiple conformations playing a key role in functional and structural transitions (Yadid et al. 2010).

One way in which phenotypic heterogeneity of proteins could impact virus evolvability is if receptor binding proteins, which are responsible for host-recognition, can evolve the capacity to produce multiple conformations with differing binding specificity, like the precursor antibodies (Zimmermann et al. 2006). There is some evidence that an analogous process occurs in bacteriophage λ during its evolution from a single-receptor specialist to a dual-receptor generalist followed by subsequent specialization on either the old or new receptor (Meyer et al. 2012; Meyer et al. 2016; Petrie et al. 2018). As predicted, the evolution from specialist to generalist was accompanied by a loss of stability, and the new specialist genotypes that evolved from the generalist regained stability (Petrie et al. 2018). The generalist also displayed properties consistent with the production of phenotypic heterogeneity because multiple phenotypic subpopulations of phage particles were detected in an isogenic culture (Petrie et al. 2018). In

addition, λ selected for enhanced thermostability were less evolvable and required additional destabilizing mutations to gain the use of the new receptor (Strobel et al. 2022).

If genetic robustness and phenotypic heterogeneity act in opposite directions, how can both promote evolvability? There have been several attempts to reconcile the effects of robustness and phenotypic heterogeneity through theoretical and computational models. One approach focused on the timescale on which new functions evolve, concluding that robustness is favorable in the long term but not necessarily in the short term (Elena and Sanjuán 2008). For this, robustness of an individual sequence is defined by the number of one-mutation-away sequences that encode the same phenotype. Use of the theoretical framework of neutral networks to demonstrated that robustness does in fact correlate negatively with the evolvability of that sequence. However, robust phenotypes, defined as those that can be encoded by many different sequences, facilitate the proliferation of diverse sequences. This, in turn, increases the likelihood that one of the many sequences encoding the phenotype will have one-mutation-away sequences that encode a novel phenotype (Wagner 2008). Under this framework, phenotypic robustness, but not genotype robustness, should promote evolvability. Phenotypic robustness could be encoded by disordered protein regions, in which structural changes may actually be tolerated more easily, allowing for more sequence diversity and rapid evolution. For example, Nodamura viral polymerase can tolerate high levels of sequence disruption in its structurally disordered C-terminal region without losing function (Gitlin et al. 2014). Finally, it may be that the effect of robustness depends on the extent of change in the distribution of fitness effects between an old and new environment (Stern et al. 2014). Through these efforts, it is understood that the relationship between stability and evolvability is likely far more complex than a simple linear correlation.

We consider whether fitness trade-offs provide another perspective on this problem. In our hypothetical model, the relationship between stability and evolvability is unimodal, with an optimal stability at which evolvability is highest (Figure 1.4A). From the optimum, evolvability decreases as a protein moves toward both extremes of stability, low and high (Figure 1.4A). This framework would reconcile conflicting empirical results from different studies, as the effect of changing stability (i.e. increasing or decreasing) would have a different effect on evolvability (i.e. either positive or negative) depending on where the protein is on the curve (Figure 1.4A). The exact shape of the curve might be different depending on the evolutionary history of the protein (Figure 1.4B) but is fundamentally driven by tradeoffs. Thus, at low stability, evolvability is constrained by unfolding and aggregation, whereas at high stability, evolvability is constrained by excessive rigidity preventing the exploration of novel folds. Applied to virus evolvability, these constraints on protein evolvability may manifest at various levels of the viral life cycle, particularly during particle production and transmission (Figure 1.1), which we will address next.

The main constraints restricting evolvability at low stability are the production of viable progeny particles and survival outside the host cell during transmission between hosts (Figure 1.1). During virus production inside the cell, particles that contain unstable proteins might not assemble properly, might be subject to degradation by host quality control machinery, or might be more reliant on host chaperones to fold properly (Aviner and Frydman 2020). Once the progeny viruses are produced and assembled, they are released from the intracellular environment into the external environment and must survive until encountering the next susceptible cell. Unstable viruses might be more likely to become deactivated by environmental forces outside the host, potentially hindering transmission between host individuals. There are

several documented cases of stabilizing mutations resulting in increased viral transmission rates (Sang et al. 2015; Yang et al. 2021b). In animal viruses, increased sensitivity to temperature might restrict a virus from infecting hosts with higher body temperatures (Yang et al. 2021a), inhibiting initial spillover to a novel host. The tradeoff between phenotypic heterogeneity and transmission raises the question of how enhanced evolvability could possibly outweigh the seemingly high cost of reduced transmissibility. One possibility is that chance destabilizing mutations introduce conformational heterogeneity into viral proteins, generating an incipient function, such as binding to a new host receptor. Additional mutations might then rapidly tune performance of the new function and re-stabilize the protein, minimizing the number of transmission events that would be required with the less stable particle. Evidence for this model comes from bacteriophage λ , in which destabilizing mutations that arose in a stable, single receptor specialist allowed activity on a novel receptor, and subsequent mutations restored stability and simultaneously increased specialization to the new receptor (Petrie et al. 2018).

A constraint on evolvability at high stability could be reduction in the capacity to produce the phenotypic flexibility necessary for novel activity. At the protein level, high stability has been associated with structural rigidity and reduced conformational flexibility (Vihinen 1987; Rathi et al. 2015). Consistent with this, stabilizing mutations sometimes reduce activity in enzymes (Wang et al. 2002; Studer et al. 2014), and mutations that confer new abilities are often destabilizing, although perhaps not more destabilizing than the average mutation (Tokuriki et al. 2008). The hypothesis that stability can constrain activity is also consistent with results from the non-enzymatic bacteriophage λ host-recognition protein. When stabilizing mutations were inserted into the receptor binding proteins of generalists able to use two different receptors, they

lost function on one receptor (Strobel et al. 2022). Together, these results suggest that high stability can constrain new activity.

Given these constraints, we can begin to speculate about the ecological and environmental conditions that might favor or impede virus evolvability via protein heterogeneity. Host density has been shown to be an important determinant of viral transmission (Marina et al. 2005). At high host density a viral particle may not need to persist for long periods of time outside of the cell before finding a new host. Thus, selection for stability would be relaxed and favor evolvability via phenotypic heterogeneity. Similarly, the diversity of hosts in the environment, specifically the ratio of susceptible to resistant hosts could shape the response. Host diversity has been shown to influence the evolution of host-range in bacteriophage T7 (Holtzman et al. 2020) and a novel phage ϕ JB01 (Sant et al. 2021), although potential links to robustness and phenotypic heterogeneity were not explored.

An abiotic condition that could influence evolvability by phenotypic heterogeneity is temperature. Under conditions of high temperature, it seems reasonable to predict that phenotypic heterogeneity would be more constrained due to selection for thermostability. This might manifest as a shift in the optimal stability for evolvability toward higher values (Figure 1.4B). Rapidly rising global temperatures will undoubtedly alter the selection pressures faced by viruses, and it is interesting to speculate about whether virus evolution might shift toward increasing thermotolerance (Barik 2020). However, because viruses are intracellular parasites, the extent to which they experience selection from the external environment might be modulated by host processes. Unstable viruses could potentially be shielded from misfolding by host protein-folding chaperones (Aviner and Frydman 2020). For example, an amino acid change in the influenza nucleoprotein known to destabilize the protein and enhance immune evasion

resulted in severe fitness costs at febrile temperatures, but only when a host heat shock factor was inhibited (Phillips et al. 2018). In viruses that use multiple host species to complete their life cycles, such as arboviruses, viral proteins face the additional challenge of folding and functioning at vastly different temperatures (Murrieta et al. 2021).

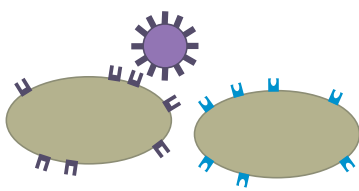
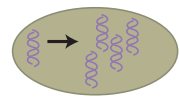
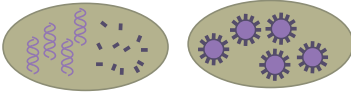
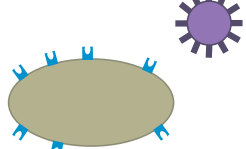
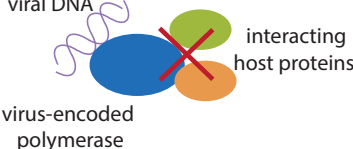
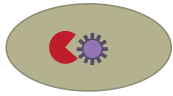
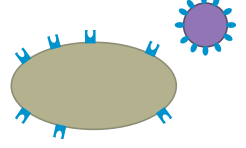

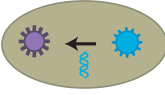
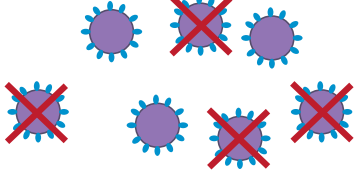


step in virus life cycle	Host recognition and attachment: 	Viral genome replication 	Production and assembly of viral progeny: 
potential virus-host misalignment	Virus receptor binding protein fails to bind receptors of new host 	Viral enzymes incompatible with host enzymes involved in replication viral DNA virus-encoded polymerase interacting host proteins 	Virus is inactivated by host anti-viral mechanisms 
a virus innovation counters misalignment	Protein heterogeneity in virus receptor binding protein (RBP) may confer activity on new receptor 	Mutation in viral protein ameliorates incompatibility between host and virus proteins 	Co-infection and recombination allow transfer of genes adapted to evade anti-viral measures 
constraints on virus's ability to evolve innovation	Altered viral RBP may be less stable, reducing its survival outside the host 	Excessive mutation in viral polymerase could inactivate it 	Recombination can disrupt co-evolved genes, leading to lower fitness hybrids 

Figure 1.1 An approach to predicting virus host-range expansion evolvability by examining intrinsic traits and constraints. In this review, we examine the many possible misalignments between a virus and a novel host species or cell type at various steps in the virus life cycle. We examine several intrinsic traits that can enhance the ability of viruses to evolve functional innovations that correct misalignments and permit infection of a new host. Traits that promote innovation can come with associated costs, causing evolvability to be constrained. These constraints enable speculation about the conditions that might promote host-range expansion. Several hypothetical examples are illustrated here to demonstrate how each trait could enhance evolvability by ameliorating misalignments, and how constraints could limit that ability. Abbreviation: RBP, receptor binding protein.

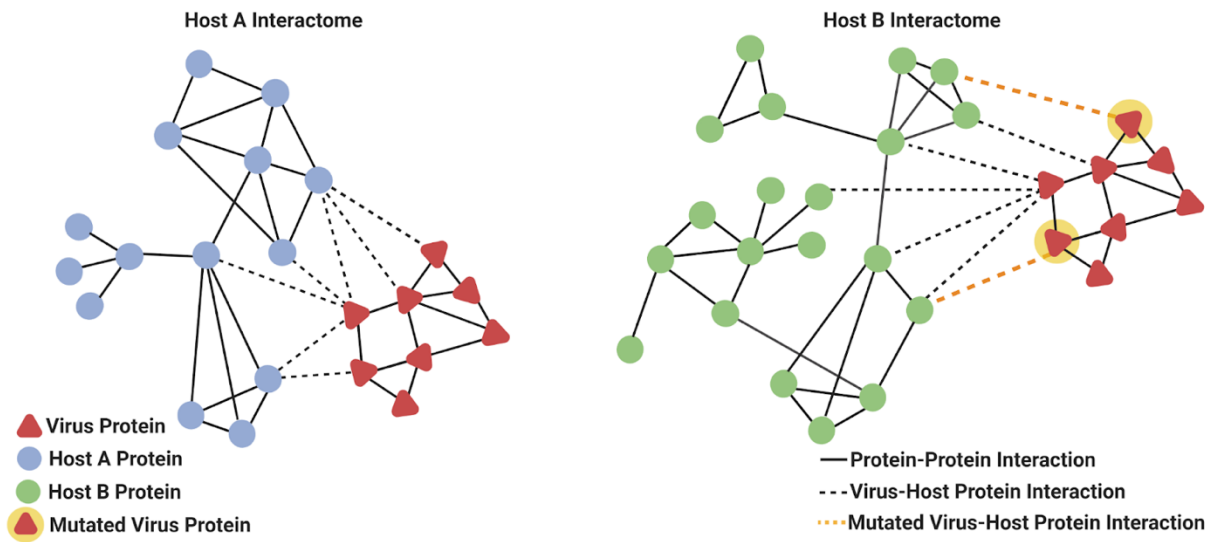


Figure 1.2 Mutations can alter the shape of virus-host protein interactomes. The interactions between the virus proteins from host A determine what the virus requires to mount an infection. If the interactome of another potential host is similar, then the virus may be able to gain infectivity on that host with few mutations. Here, we show a schematic representation of how a mutated virus protein might facilitate a host-range shift from its natural host, A, to a novel host, B. The mutated proteins gained the ability to interact with host B's proteins, allowing infection. Figure created with BioRender.com.

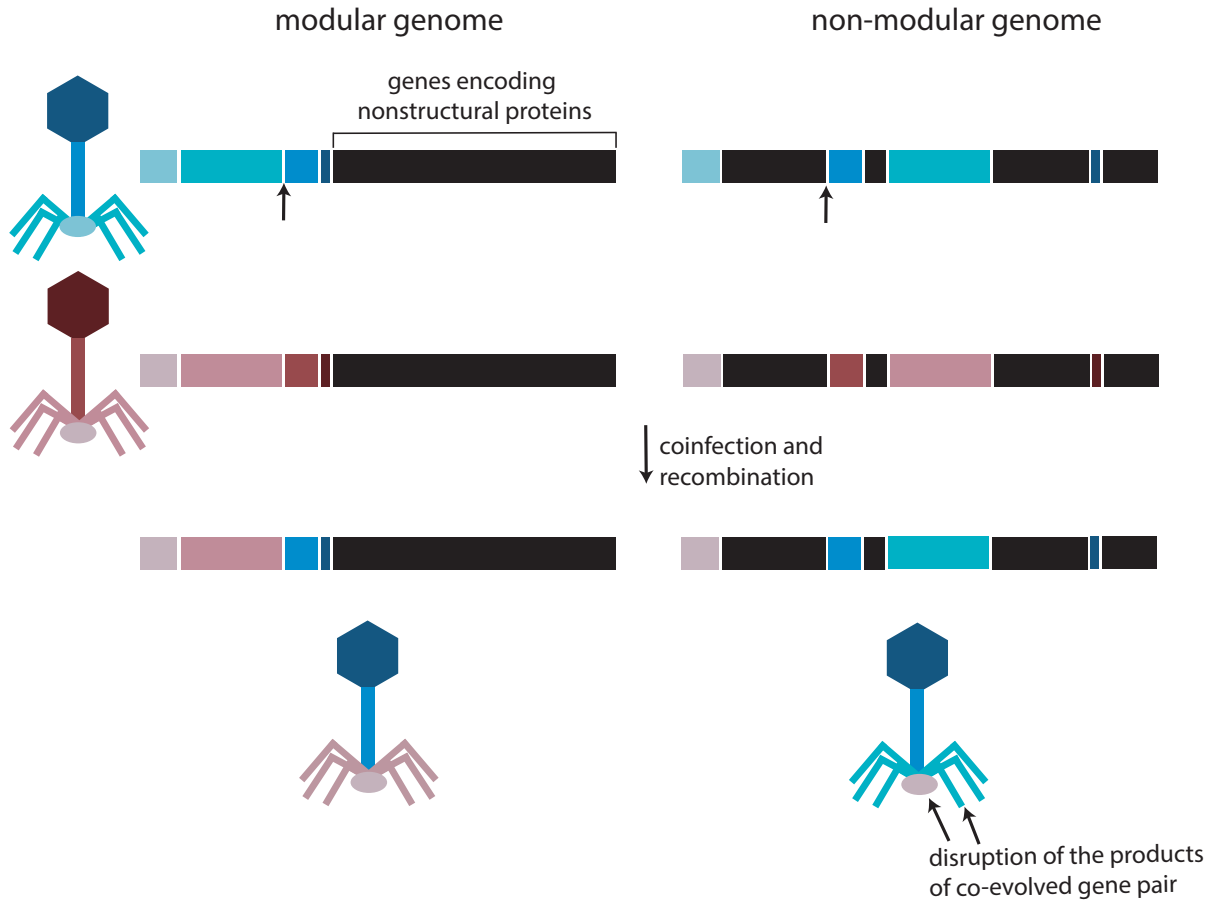


Figure 1.3 Genome modularity may prevent the costs of recombination, facilitating host-range expansion. Many viruses have genomes that are remarkably modular, with functionally similar proteins encoded by genes located in close proximity in the genome. In this hypothetical schematic, we show how this organization might facilitate the transfer of functional modules between virus genomes. Viruses with modular genomes might more readily exchange the genes that confer infectivity on novel host species or cell types, such as an ensemble of tail proteins in a bacteriophage. Modular organization would help avoid the cost of breaking up gene pairs that have coevolved to interact with each other, such as the tail tip protein and the tail fiber proteins of a bacteriophage.

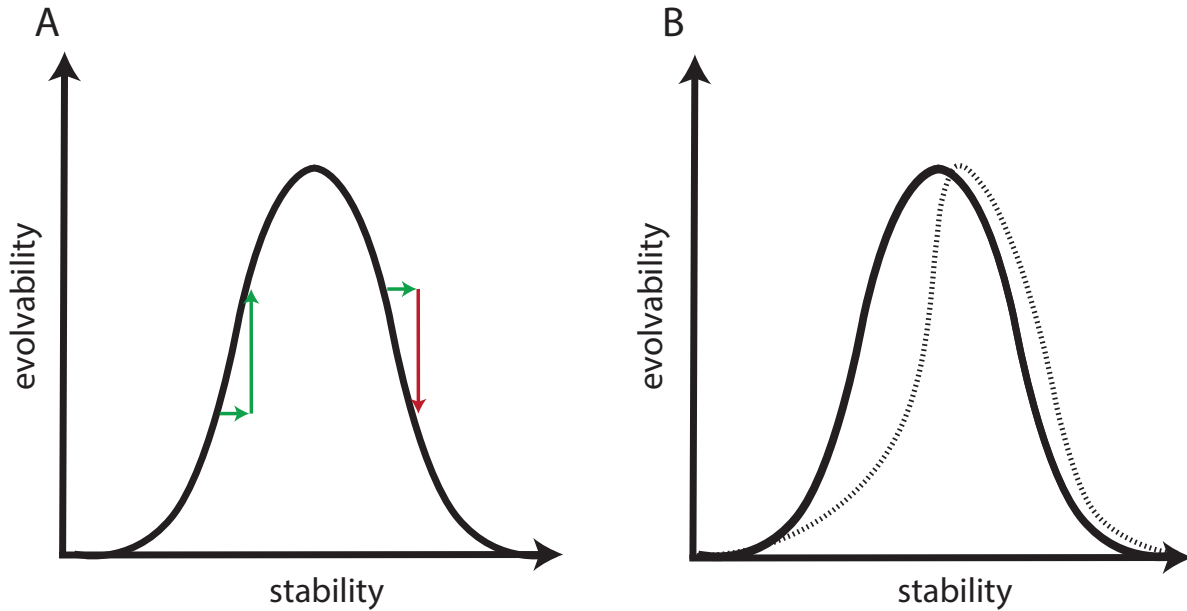


Figure 1.4 Virus evolvability is in many ways shaped by the ability of viral proteins to evolve new functions. Stability is a trait intrinsic to proteins that is thought to affect evolvability, but the relationship is complex. A) We propose that protein evolvability, and by extension virus evolvability, is constrained at both extremes. At low stability, evolvability is constrained by protein misfolding, whereas at high stability it is constrained by excessive rigidity that prevents structural dynamism. Depending on where a protein is on the continuum of stability, enhancing stability could increase or decrease evolvability. B) Proteins evolved under different environments might have differently shaped stability-evolvability curves. A protein evolved under thermal stress might have a right-shifted curve because low stability would be more costly.

1.6 Conclusion

A similar unimodal relationship between trait values and evolvability was uncovered for all three traits examined. Such relationships make it difficult to predict viral evolvability because it is often unclear whether the viral variants being scrutinized reside on the increasing or decreasing slopes of the relationship. Unimodal relationships suggest that there are limits to virus evolvability and that environmental pressures that push those limits could have a significant impact on reducing viral evolvability. These pressures could occur naturally and might indicate areas of less concern in order to focus surveillance efforts on locations where conditions favor evolvable viruses. Or the pressures could stem from human interventions designed to mitigate the risk of disease emergence. Certainly, more research is necessary to understand the drivers of viral host-range evolvability.

1.7 Acknowledgements

We thank Animesh Gupta, Katherine Petrie, and Joshua Borin for discussions. Funding via the National Science Foundation DEB1934515.

Chapter 1, in full, is a reprint of the material as it appears in Strobel HM, Stuart EC, Meyer JR (2022). A trait-based approach to predicting viral host-range evolvability. *Annual Review of Virology*. In press. The dissertation author was the primary investigator and author of this paper.

1.8 References

- Aguado, L. C., T. X. Jordan, E. Hsieh, D. Blanco-Melo, J. Heard, M. Panis, M. Vignuzzi, and B. R. tenOever. 2018. Homologous recombination is an intrinsic defense against antiviral RNA interference. *Proc Natl Acad Sci U S A* 115:E9211-e9219.
- Anderson, J. P., R. Daifuku, and L. A. Loeb. 2004. Viral error catastrophe by mutagenic nucleosides. *Annu Rev Microbiol* 58:183-205.
- Aviner, R. and J. Frydman. 2020. Proteostasis in Viral Infection: Unfolding the Complex Virus-Chaperone Interplay. *Cold Spring Harb Perspect Biol* 12.
- Banner, L. R. and M. M. Lai. 1991. Random nature of coronavirus RNA recombination in the absence of selection pressure. *Virology* 185:441-445.
- Barik, S. 2020. Evolution of Protein Structure and Stability in Global Warming. *Int J Mol Sci* 21.
- Besenmatter, W., P. Kast, and D. Hilvert. 2007. Relative tolerance of mesostable and thermostable protein homologs to extensive mutation. *Proteins* 66:500-506.
- Bloom, J. D. and F. H. Arnold. 2009. In the light of directed evolution: pathways of adaptive protein evolution. *Proc Natl Acad Sci U S A* 106 Suppl 1:9995-10000.
- Bloom, J. D., S. T. Labthavikul, C. R. Otey, and F. H. Arnold. 2006. Protein stability promotes evolvability. *Proc Natl Acad Sci U S A* 103:5869-5874.
- Bobay, L. M., M. Touchon, and E. P. Rocha. 2013. Manipulating or superseding host recombination functions: a dilemma that shapes phage evolvability. *PLoS Genet* 9:e1003825.
- Bocharov, G., N. J. Ford, J. Edwards, T. Breinig, S. Wain-Hobson, and A. Meyerhans. 2005. A genetic-algorithm approach to simulating human immunodeficiency virus evolution reveals the strong impact of multiply infected cells and recombination. *J Gen Virol* 86:3109-3118.
- Boni, M. F., P. Lemey, X. Jiang, T. T. Lam, B. W. Perry, T. A. Castoe, A. Rambaut, and D. L. Robertson. 2020. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nat Microbiol* 5:1408-1417.
- Boon, M., Holtappels, D., Lood, C., v. Noort, Vera, Lavigne, and Rob. 2020. Host Range Expansion of Pseudomonas Virus LUZ7 Is Driven by a Conserved Tail Fiber Mutation. *PHAGE* 1:87-90.
- Borin, J. M., S. Avrani, J. E. Barrick, K. L. Petrie, and J. R. Meyer. 2021. Coevolutionary phage training leads to greater bacterial suppression and delays the evolution of phage resistance. *Proc Natl Acad Sci U S A* 118.

Botstein, D. 1980. A theory of modular evolution for bacteriophages. *Ann N Y Acad Sci* 354:484-490.

Bradel-Tretheway, B. G., J. L. Mattiaccio, A. Krasnoselsky, C. Stevenson, D. Purdy, S. Dewhurst, and M. G. Katze. 2011. Comprehensive proteomic analysis of influenza virus polymerase complex reveals a novel association with mitochondrial proteins and RNA polymerase accessory factors. *J Virol* 85:8569-8581.

Bull, J. J., R. Sanjuán, and C. O. Wilke. 2007. Theory of lethal mutagenesis for viruses. *J Virol* 81:2930-2939.

Burmeister, A. R., R. E. Lenski, and J. R. Meyer. 2016. Host coevolution alters the adaptive landscape of a virus. *Proc Biol Sci* 283.

Burrowes, B. H., I. J. Molineux, and J. A. Fralick. 2019. Directed in Vitro Evolution of Therapeutic Bacteriophages: The Appelmans Protocol. *Viruses* 11.

Bódi, Z., Z. Farkas, D. Nevozhay, D. Kalapis, V. Lázár, B. Csörgő, Á. Nyerges, B. Szamecz, G. Fekete, B. Papp, H. Araújo, J. L. Oliveira, G. Moura, M. A. S. Santos, T. Székely, Jr., G. Balázsi, and C. Pál. 2017. Phenotypic heterogeneity promotes adaptive evolution. *PLoS Biol* 15:e2000644.

Casjens, S. R. and R. W. Hendrix. 2015. Bacteriophage lambda: Early pioneer and still relevant. *Virology* 479-480:310-330.

Comeau, A. M., C. Bertrand, A. Letarov, F. Tétart, and H. M. Krisch. 2007. Modular architecture of the T4 phage superfamily: a conserved core genome and a plastic periphery. *Virology* 362:384-396.

Crill, W. D., H. A. Wichman, and J. J. Bull. 2000. Evolutionary reversals during viral adaptation to alternating hosts. *Genetics* 154:27-37.

Cuevas, J. M., A. Moya, and R. Sanjuán. 2009. A genetic background with low mutational robustness is associated with increased adaptability to a novel host in an RNA virus. *J Evol Biol* 22:2041-2048.

de Jonge, P. A., F. L. Nobrega, S. J. J. Brouns, and B. E. Dutilh. 2019. Molecular and Evolutionary Determinants of Bacteriophage Host Range. *Trends Microbiol* 27:51-63.

De Paepe, M., G. Hutinet, O. Son, J. Amarir-Bouhram, S. Schbath, and M. A. Petit. 2014. Temperate phages acquire DNA from defective prophages by relaxed homologous recombination: the role of Rad52-like recombinases. *PLoS Genet* 10:e1004181.

Decaro, N. and A. Lorusso. 2020. Novel human coronavirus (SARS-CoV-2): A lesson from animal coronaviruses. *Vet Microbiol* 244:108693.

- Domingo, E. 2006. Quasispecies: concept and implications for virology. Springer, Berlin.
- Domingo-Calap, P., J. M. Cuevas, and R. Sanjuán. 2009. The fitness effects of random mutations in single-stranded DNA and RNA bacteriophages. *PLoS Genet* 5:e1000742.
- Domingo-Calap, P., M. Pereira-Gómez, and R. Sanjuán. 2010. Selection for thermostability can lead to the emergence of mutational robustness in an RNA virus. *J Evol Biol* 23:2453-2460.
- Drummond, D. A., J. J. Silberg, M. M. Meyer, C. O. Wilke, and F. H. Arnold. 2005. On the conservative nature of intragenic recombination. *Proc Natl Acad Sci U S A* 102:5380-5385.
- Dudas, G. and A. Rambaut. 2016. MERS-CoV recombination: implications about the reservoir and potential for adaptation. *Virus Evol* 2:vev023.
- Dudenhöffer, C., G. Rohaly, K. Will, W. Deppert, and L. Wiesmüller. 1998. Specific mismatch recognition in heteroduplex intermediates by p53 suggests a role in fidelity control of homologous recombination. *Mol Cell Biol* 18:5332-5342.
- Duffy, S., C. L. Burch, and P. E. Turner. 2007. Evolution of host specificity drives reproductive isolation among RNA viruses. *Evolution* 61:2614-2622.
- Duffy, S., P. E. Turner, and C. L. Burch. 2006. Pleiotropic costs of niche expansion in the RNA bacteriophage phi 6. *Genetics* 172:751-757.
- Elena, S. F. and R. Sanjuán. 2005. Adaptive value of high mutation rates of RNA viruses: separating causes from consequences. *J Virol* 79:11555-11558.
- Elena, S. F. and R. Sanjuán. 2008. The effect of genetic robustness on evolvability in digital organisms. *BMC Evol Biol* 8:284.
- Eyre-Walker, A. and P. D. Keightley. 2007. The distribution of fitness effects of new mutations. *Nat Rev Genet* 8:610-618.
- Fang, S. G., S. Shen, F. P. Tay, and D. X. Liu. 2005. Selection of and recombination between minor variants lead to the adaptation of an avian coronavirus to primate cells. *Biochem Biophys Res Commun* 336:417-423.
- Fisher, A. M. 2021. The evolutionary impact of population size, mutation rate and virulence on pathogen niche width. *J Evol Biol* 34:1256-1265.
- Frensing, T., F. S. Heldt, A. Pflugmacher, I. Behrendt, I. Jordan, D. Flockerzi, Y. Genzel, and U. Reichl. 2013. Continuous influenza virus production in cell culture shows a periodic accumulation of defective interfering particles. *PLoS One* 8:e72288.

Galetto, R., V. Giacomoni, M. Véron, and M. Negroni. 2006. Dissection of a circumscribed recombination hot spot in HIV-1 after a single infectious cycle. *J Biol Chem* 281:2711-2720.

Giachetti, C. and J. J. Holland. 1989. Vesicular stomatitis virus and its defective interfering particles exhibit in vitro transcriptional and replicative competition for purified L-NS polymerase molecules. *Virology* 170:264-267.

Gitlin, L., H. T, A. LaBarbera, M. Solovey, R. Andino, and 2014. Rapid evolution of virus sequences in intrinsically disordered protein regions. *PLoS pathogens* 10.

Golden, M., B. M. Muhire, Y. Semegni, and D. P. Martin. 2014. Patterns of recombination in HIV-1M are influenced by selection disfavouring the survival of recombinants with disrupted genomic RNA and protein structures. *PLoS One* 9:e100400.

Gordon, D. E., G. M. Jang, M. Bouhaddou, J. Xu, K. Obernier, K. M. White, M. J. O'Meara, V. V. Rezelj, J. Z. Guo, D. L. Swaney, T. A. Tummino, R. Hüttenhain, R. M. Kaake, A. L. Richards, B. Tutuncuoglu, H. Foussard, J. Batra, K. Haas, M. Modak, M. Kim, P. Haas, B. J. Polacco, H. Braberg, J. M. Fabius, M. Eckhardt, M. Soucheray, M. J. Bennett, M. Cakir, M. J. McGregor, Q. Li, B. Meyer, F. Roesch, T. Vallet, A. Mac Kain, L. Miorin, E. Moreno, Z. Z. C. Naing, Y. Zhou, S. Peng, Y. Shi, Z. Zhang, W. Shen, I. T. Kirby, J. E. Melnyk, J. S. Chorba, K. Lou, S. A. Dai, I. Barrio-Hernandez, D. Memon, C. Hernandez-Armenta, J. Lyu, C. J. P. Mathy, T. Perica, K. B. Pilla, S. J. Ganesan, D. J. Saltzberg, R. Rakesh, X. Liu, S. B. Rosenthal, L. Calviello, S. Venkataramanan, J. Liboy-Lugo, Y. Lin, X. P. Huang, Y. Liu, S. A. Wankowicz, M. Bohn, M. Safari, F. S. Ugur, C. Koh, N. S. Savar, Q. D. Tran, D. Shengjuler, S. J. Fletcher, M. C. O'Neal, Y. Cai, J. C. J. Chang, D. J. Broadhurst, S. Klippsten, P. P. Sharp, N. A. Wenzell, D. Kuzuoglu-Ozturk, H. Y. Wang, R. Trenker, J. M. Young, D. A. Caverro, J. Hiatt, T. L. Roth, U. Rathore, A. Subramanian, J. Noack, M. Hubert, R. M. Stroud, A. D. Frankel, O. S. Rosenberg, K. A. Verba, D. A. Agard, M. Ott, M. Emerman, N. Jura, M. von Zastrow, E. Verdin, A. Ashworth, O. Schwartz, C. d'Enfert, S. Mukherjee, M. Jacobson, H. S. Malik, D. G. Fujimori, T. Ideker, C. S. Craik, S. N. Floor, J. S. Fraser, J. D. Gross, A. Sali, B. L. Roth, D. Ruggero, J. Taunton, T. Kortemme, P. Beltrao, M. Vignuzzi, A. García-Sastre, K. M. Shokat, B. K. Shoichet and N. J. Krogan. 2020. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583:459-468.

Graham, R. L. and R. S. Baric. 2010. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. *J Virol* 84:3134-3146.

Gupta, A., A. N. Soto, S. J. Medina, K. L. Petrie, and J. R. Meyer. 2020. Bacteriophage lambda overcomes a perturbation in its host-viral genetic network through mutualism and evolution of life history traits. *Evolution* 74:764-774.

Gupta, A., L. Zaman, H. M. Strobel, J. Gallie, A. R. Burmeister, B. Kerr, E. S. Tamar, R. Kishony, and J. R. Meyer. 2021. Host-parasite coevolution promotes innovation through deformations in fitness landscapes. *bioRxiv:2021.2006.2025.449783*.

Hahn, C. S., S. Lustig, E. G. Strauss, and J. H. Strauss. 1988. Western equine encephalitis virus is a recombinant virus. *Proc Natl Acad Sci U S A* 85:5997-6001.

Hall, A. R., P. D. Scanlan, H. C. Leggett, and A. Buckling. 2012. Multiplicity of infection does not accelerate infectivity evolution of viral parasites in laboratory microcosms. *J Evol Biol* 25:409-415.

Hall, J. P., E. Harrison, and M. A. Brockhurst. 2013. Viral host-adaptation: insights from evolution experiments with phages. *Curr Opin Virol* 3:572-577.

Haller, S. L., C. Peng, G. McFadden, and S. Rothenburg. 2014. Poxviruses and the evolution of host range and virulence. *Infect Genet Evol* 21:15-40.

Hayden, E. J., E. Ferrada, and A. Wagner. 2011. Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme. *Nature* 474:92-95.

Hendrix, R. W., J. G. Lawrence, G. F. Hatfull, and S. Casjens. 2000. The origins and ongoing evolution of viruses. *Trends Microbiol* 8:504-508.

Holland, S. L., T. Reader, P. S. Dyer, and S. V. Avery. 2014. Phenotypic heterogeneity is a selected trait in natural yeast populations subject to environmental stress. *Environ Microbiol* 16:1729-1740.

Holtzman, T., R. Globus, S. Molshanski-Mor, A. Ben-Shem, I. Yosef, and U. Qimron. 2020. A continuous evolution system for contracting the host range of bacteriophage T7. *Sci Rep* 10:307.

Hu, B., L. P. Zeng, X. L. Yang, X. Y. Ge, W. Zhang, B. Li, J. Z. Xie, X. R. Shen, Y. Z. Zhang, N. Wang, D. S. Luo, X. S. Zheng, M. N. Wang, P. Daszak, L. F. Wang, J. Cui, and Z. L. Shi. 2017. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS Pathog* 13:e1006698.

Jackwood, M. W., T. O. Boynton, D. A. Hilt, E. T. McKinley, J. C. Kissinger, A. H. Paterson, J. Robertson, C. Lemke, A. W. McCall, S. M. Williams, J. W. Jackwood, and L. A. Byrd. 2010. Emergence of a group 3 coronavirus through recombination. *Virology* 398:98-108.

Jiang, D., Q. Wang, Z. Bai, H. Qi, J. Ma, W. Liu, F. Ding, and J. Li. 2020. Could Environment Affect the Mutation of H1N1 Influenza Virus? *Int J Environ Res Public Health*.

Koel, B. F., D. F. Burke, T. M. Bestebroer, S. van der Vliet, G. C. Zondag, G. Vervaet, E. Skepner, N. S. Lewis, M. I. Spronken, C. A. Russell, M. Y. Eropkin, A. C. Hurt, I. G. Barr, J. C. de Jong, G. F. Rimmelzwaan, A. D. Osterhaus, R. A. Fouchier, and D. J. Smith. 2013. Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution. *Science* 342:976-979.

- Kondo, A. and S. Maeda. 1991. Host range expansion by recombination of the baculoviruses *Bombyx mori* nuclear polyhedrosis virus and *Autographa californica* nuclear polyhedrosis virus. *J Virol* 65:3625-3632.
- Lai, M. M. 1992. RNA recombination in animal and plant viruses. *Microbiol Rev* 56:61-79.
- Lefevre, P., J. M. Lett, B. Reynaud, and D. P. Martin. 2007. Avoidance of protein fold disruption in natural virus recombinants. *PLoS Pathog* 3:e181.
- Lefevre, P., J. M. Lett, A. Varsani, and D. P. Martin. 2009. Widely conserved recombination patterns among single-stranded DNA viruses. *J Virol* 83:2697-2707.
- Lefevre, P. and E. Moriones. 2015. Recombination as a motor of host switches and virus emergence: geminiviruses as case studies. *Curr Opin Virol* 10:14-19.
- Li, X., E. E. Giorgi, M. H. Marichanegowda, B. Foley, C. Xiao, X. P. Kong, Y. Chen, S. Gnanakaran, B. Korber, and F. Gao. 2020. Emergence of SARS-CoV-2 through recombination and strong purifying selection. *Sci Adv* 6.
- Linster, M., S. van Boheemen, M. de Graaf, E. J. A. Schrauwen, P. Lexmond, B. Mänz, T. M. Bestebroer, J. Baumann, D. van Riel, G. F. Rimmelzwaan, A. Osterhaus, M. Matrosovich, R. A. M. Fouchier, and S. Herfst. 2014. Identification, characterization, and natural selection of mutations driving airborne transmission of A/H5N1 virus. *Cell* 157:329-339.
- Liu, J. and S. W. Morrical. 2010. Assembly and dynamics of the bacteriophage T4 homologous recombination machinery. *Virol J* 7:357.
- Lu, G., Y. Hu, Q. Wang, J. Qi, F. Gao, Y. Li, Y. Zhang, W. Zhang, Y. Yuan, J. Bao, B. Zhang, Y. Shi, J. Yan, and G. F. Gao. 2013. Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature* 500:227-231.
- Maddamsetti, R., D. T. Johnson, S. J. Spielman, K. L. Petrie, D. S. Marks, and J. R. Meyer. 2018. Gain-of-function experiments with bacteriophage lambda uncover residues under diversifying selection in nature. *Evolution* 72:2234-2243.
- Mahichi, F., A. J. Synnott, K. Yamamichi, T. Osada, and Y. Tanji. 2009. Site-specific recombination of T2 phage using IP008 long tail fiber genes provides a targeted method for expanding host range while retaining lytic activity. *FEMS Microbiol Lett* 295:211-217.
- Malone, B. and E. A. Campbell. 2021. Molnupiravir: coding for catastrophe. *Nat Struct Mol Biol* 28:706-708.
- Marina, C. F., I. Fernández-Salas, J. E. Ibarra, J. I. Arredondo-Jiménez, J. Valle, and T. Williams. 2005. Transmission dynamics of an iridescent virus in an experimental mosquito population: the role of host density. *Ecological Entomology* 30:376-382.

Martin, D. P., P. Lefeuvre, A. Varsani, M. Hoareau, J. Y. Semegni, B. Dijoux, C. Vincent, B. Reynaud, and J. M. Lett. 2011. Complex recombination patterns arising during geminivirus coinfections preserve and demarcate biologically important intra-genome interaction networks. *PLoS Pathog* 7:e1002203.

Martin, D. P., E. van der Walt, D. Posada, and E. P. Rybicki. 2005. The evolutionary value of recombination is constrained by genome modularity. *PLoS Genet* 1:e51.

Maynard, N. D., E. W. Birch, J. C. Sanghvi, L. Chen, M. V. Gutschow, and M. W. Covert. 2010. A forward-genetic screen and dynamic analysis of lambda phage host-dependencies reveals an extensive interaction network and a new anti-viral strategy. *PLoS Genet* 6:e1001017.

McBride, R. C., C. B. Ogbunugafor, and P. E. Turner. 2008. Robustness promotes evolvability of thermotolerance in an RNA virus. *BMC Evol Biol* 8:231.

Meyer, J. R., D. T. Dobias, S. J. Medina, L. Servilio, A. Gupta, and R. E. Lenski. 2016. Ecological speciation of bacteriophage lambda in allopatry and sympatry. *Science* 354:1301-1304.

Meyer, J. R., D. T. Dobias, J. S. Weitz, J. E. Barrick, R. T. Quick, and R. E. Lenski. 2012. Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science* 335:428-432.

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* 190:2172-2182.

Murphy, K. C. 2016. λ Recombination and Recombineering. *EcoSal Plus* 7.

Murrieta, R., S. Garcia-Luna, D. Murrieta, G. Halladay, M. Young, J. Fauver, A. Gendernalik, J. Weger-Lucarelli, C. Rückert, and G. Ebel 2021. Impact of extrinsic incubation temperature on natural selection during Zika virus infection of *Aedes aegypti* and *Aedes albopictus*. *PLoS pathogens* 17.

Padidam, M., S. Sawyer, and C. M. Fauquet. 1999. Possible emergence of new geminiviruses by frequent recombination. *Virology* 265:218-225.

Patiño-Galindo, J., I. Filip, and R. Rabadan. 2021. Global Patterns of Recombination across Human Viruses. *Mol Biol Evol* 38:2520-2531.

Payne, J. L. and A. Wagner. 2019. The causes of evolvability and their evolution. *Nat Rev Genet* 20:24-38.

Peck, K. M., C. L. Burch, M. T. Heise, and R. S. Baric. 2015. Coronavirus Host Range Expansion and Middle East Respiratory Syndrome Coronavirus Emergence: Biochemical Mechanisms and Evolutionary Perspectives. *Annu Rev Virol* 2:95-117.

Petrie, K. L., N. D. Palmer, D. T. Johnson, S. J. Medina, S. J. Yan, V. Li, A. R. Burmeister, and J. R. Meyer. 2018. Destabilizing mutations encode nongenetic variation that drives evolutionary innovation. *Science* 359:1542-1545.

Phillips, A. M., A. I. Ponomarenko, K. Chen, O. Ashenberg, J. Miao, S. M. McHugh, V. L. Butty, C. A. Whittaker, C. L. Moore, J. D. Bloom, Y. S. Lin, and M. D. Shoulders. 2018. Destabilized adaptive influenza variants critical for innate immune system escape are potentiated by host chaperones. *PLoS Biol* 16:e3000008.

Poirier, E. Z., B. C. Mounce, K. Rozen-Gagnon, P. J. Hooikaas, K. A. Stapleford, G. Moratorio, and M. Vignuzzi. 2015. Low-Fidelity Polymerases of Alphaviruses Recombine at Higher Rates To Overproduce Defective Interfering Particles. *J Virol* 90:2446-2454.

Presloid, J. B., T. F. Mohammad, A. S. Lauring, and I. S. Novella. 2016. Antigenic diversification is correlated with increased thermostability in a mammalian virus. *Virology* 496:203-214.

Priyadarsini, S. L., M. Suresh, and D. Huisingh. 2020. What can we learn from previous pandemics to reduce the frequency of emerging infectious diseases like COVID-19? *Glob Transit* 2:202-220.

Pérez-Losada, M., M. Arenas, J. C. Galán, F. Palero, and F. González-Candelas. 2015. Recombination in viruses: mechanisms, methods of study, and evolutionary consequences. *Infect Genet Evol* 30:296-307.

Raman, A. S., K. I. White, and R. Ranganathan. 2016. Origins of Allosterity and Evolvability in Proteins: A Case Study. *Cell* 166:468-480.

Rathi, P. C., K. E. Jaeger, and H. Gohlke. 2015. Structural Rigidity and Protein Thermostability in Variants of Lipase A from *Bacillus subtilis*. *PLoS One* 10:e0130289.

Rothenburg, S. and G. Brennan. 2020. Species-Specific Host-Virus Interactions: Implications for Viral Host Range and Virulence. *Trends Microbiol* 28:46-56.

Sackman, A. M., D. Reed, and D. R. Rokyta. 2015. Intergenic incompatibilities reduce fitness in hybrids of extremely closely related bacteriophages. *PeerJ* 3:e1320.

Sang, X., A. Wang, J. Ding, H. Kong, X. Gao, L. Li, T. Chai, Y. Li, K. Zhang, C. Wang, Z. Wan, G. Huang, T. Wang, N. Feng, X. Zheng, H. Wang, Y. Zhao, S. Yang, J. Qian, G. Hu, Y. Gao, and X. Xia. 2015. Adaptation of H9N2 AIV in guinea pigs enables efficient transmission by direct contact and inefficient transmission by respiratory droplets. *Sci Rep* 5:15928.

Sanjuán, R. 2010. Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. *Philos Trans R Soc Lond B Biol Sci* 365:1975-1982.

Sanjuán, R. 2012. From molecular genetics to phylodynamics: evolutionary relevance of mutation rates across viruses. *PLoS Pathog* 8:e1002685.

Sant, D. G., L. C. Woods, J. J. Barr, and M. J. McDonald. 2021. Host diversity slows bacteriophage adaptation by selecting generalists over specialists. *Nat Ecol Evol* 5:350-359.

Shi, Y., Y. Wu, W. Zhang, J. Qi, and G. F. Gao. 2014. Enabling the 'host jump': structural determinants of receptor-binding specificity in influenza A viruses. *Nat Rev Microbiol* 12:822-831.

Sikosek, T. and H. S. Chan. 2014. Biophysics of protein evolution and evolutionary protein biophysics. *J R Soc Interface* 11:20140419.

Simon-Loriere, E. and E. C. Holmes. 2011. Why do RNA viruses recombine? *Nat Rev Microbiol* 9:617-626.

Simon-Loriere, E., D. P. Martin, K. M. Weeks, and M. Negroni. 2010. RNA structures facilitate recombination-mediated gene swapping in HIV-1. *J Virol* 84:12675-12682.

Song, H., J. Qi, H. Xiao, Y. Bi, W. Zhang, Y. Xu, F. Wang, Y. Shi, and G. F. Gao. 2017a. Avian-to-Human Receptor-Binding Adaptation by Influenza A Virus Hemagglutinin H4. *Cell Rep* 20:1201-1214.

Song, W. J., J. Yu, and F. A. Tezcan. 2017b. Importance of Scaffold Flexibility/Rigidity in the Design and Directed Evolution of Artificial Metallo- β -lactamases. *J Am Chem Soc* 139:16772-16779.

Stern, A., S. Bianco, M. T. Yeh, C. Wright, K. Butcher, C. Tang, R. Nielsen, and R. Andino. 2014. Costs and benefits of mutational robustness in RNA viruses. *Cell Rep* 8:1026-1036.

Strobel, H. M., E. K. Horwitz, and J. R. Meyer. 2022. Viral protein instability enhances host-range evolvability. *PLoS Genetics* e1010030.

Studer, R. A., P. A. Christin, M. A. Williams, and C. A. Orengo. 2014. Stability-activity tradeoffs constrain the adaptive evolution of RubisCO. *Proc Natl Acad Sci U S A* 111:2223-2228.

Tenaillon, O., F. Taddei, M. Radmian, and I. Matic. 2001. Second-order selection in bacterial evolution: selection acting on mutation and recombination rates in the course of adaptation. *Res Microbiol* 152:11-16.

Tokuriki, N., F. Stricher, L. Serrano, and D. S. Tawfik. 2008. How protein stability and new functions trade off. *PLoS Comput Biol* 4:e1000002.

Tokuriki, N. and D. S. Tawfik. 2009. Protein dynamism and evolvability. *Science* 324:203-207.

Tétart, F., C. Desplats, and H. M. Krisch. 1998. Genome plasticity in the distal tail fiber locus of the T-even bacteriophage: recombination between conserved motifs swaps adhesin specificity. *J Mol Biol* 282:543-556.

Tétart, F., F. Repoila, C. Monod, and H. M. Krisch. 1996. Bacteriophage T4 host range is expanded by duplications of a small domain of the tail fiber adhesin. *J Mol Biol* 258:726-731.

V'Kovski, P., A. Kratzel, S. Steiner, H. Stalder, and V. Thiel. 2021. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol* 19:155-170.

Vihinen, M. 1987. Relationship of protein flexibility to thermostability. *Protein Eng* 1:477-480.

Visher, E., S. E. Whitefield, J. T. McCrone, W. Fitzsimmons, and A. S. Lauring. 2016. The Mutational Robustness of Influenza A Virus. *PLoS Pathog* 12:e1005856.

Wagner, A. 2005. Robustness, evolvability, and neutrality. *FEBS Lett* 579:1772-1778.

Wagner, A. 2008. Robustness and evolvability: a paradox resolved. *Proc Biol Sci* 275:91-100.

Wang, X., G. Minasov, and B. K. Shoichet. 2002. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *J Mol Biol* 320:85-95.

Wang, Y., D. Liu, W. Shi, R. Lu, W. Wang, Y. Zhao, Y. Deng, W. Zhou, H. Ren, J. Wu, G. Wu, G. F. Gao, and W. Tan. 2015. Origin and Possible Genetic Recombination of the Middle East Respiratory Syndrome Coronavirus from the First Imported Case in China: Phylogenetics and Coalescence Analysis. *mBio* 6:e01280-01215.

Weller, S. K. and J. A. Sawitzke. 2014. Recombination promoted by DNA viruses: phage λ to herpes simplex virus. *Annu Rev Microbiol* 68:237-258.

Wilkinson, D. E. and S. K. Weller. 2004. Recruitment of cellular recombination and repair proteins to sites of herpes simplex virus type 1 DNA replication is dependent on the composition of viral proteins within prereplicative sites and correlates with the induction of the DNA damage response. *J Virol* 78:4783-4796.

Worobey, M. and E. C. Holmes. 1999. Evolutionary aspects of recombination in RNA viruses. *J Gen Virol* 80 (Pt 10):2535-2543.

- Wu, X., C. Cao, Y. Xu, and X. Lu. 2004. Construction of a host range-expanded hybrid baculovirus of BmNPV and AcNPV, and knockout of cysteinase gene for more efficient expression. *Sci China C Life Sci* 47:406-415.
- Yadid, I., N. Kirshenbaum, M. Sharon, O. Dym, and D. S. Tawfik. 2010. Metamorphic proteins mediate evolutionary transitions of structure. *Proc Natl Acad Sci U S A* 107:7287-7292.
- Yang, G., C. R. Ojha, and C. J. Russell. 2021a. Relationship between hemagglutinin stability and influenza virus persistence after exposure to low pH or supraphysiological heating. *PLoS Pathog* 17:e1009910.
- Yang, T. J., P. Y. Yu, Y. C. Chang, and S. D. Hsu. 2021b. D614G mutation in the SARS-CoV-2 spike protein enhances viral fitness by desensitizing it to temperature-dependent denaturation. *J Biol Chem* 297:101238.
- Yehl, K., S. Lemire, A. C. Yang, H. Ando, M. Mimee, M. T. Torres, C. de la Fuente-Nunez, and T. K. Lu. 2019. Engineering Phage Host-Range and Suppressing Bacterial Resistance through Phage Tail Fiber Mutagenesis. *Cell* 179:459-469.e459.
- Young, C. S., G. Cachianes, P. Munz, and S. Silverstein. 1984. Replication and recombination in adenovirus-infected cells are temporally and functionally related. *J Virol* 51:571-577.
- Zhang, H., D. E. Fouts, J. DePew, and R. H. Stevens. 2013. Genetic modifications to temperate *Enterococcus faecalis* phage Ef11 that abolish the establishment of lysogeny and sensitivity to repressor, and increase host range and productivity of lytic infection. *Microbiology (Reading)* 159:1023-1035.
- Zhao, L., M. Seth-Pasricha, D. Stemate, A. Crespo-Bellido, J. Gagnon, J. Draghi, and S. Duffy. 2019. Existing Host Range Mutations Constrain Further Emergence of RNA Viruses. *J Virol* 93.
- Zheng, J., N. Guo, and A. Wagner. 2020. Selection enhances protein evolvability by increasing mutational robustness and foldability. *Science* 370.
- Zheng, J., J. L. Payne, and A. Wagner. 2019. Cryptic genetic variation accelerates evolution by opening access to diverse adaptive peaks. *Science* 365:347-353.
- Zimmermann, J., E. L. Oakman, I. F. Thorpe, X. Shi, P. Abbyad, C. L. Brooks, 3rd, S. G. Boxer, and F. E. Romesberg. 2006. Antibody evolution constrains conformational heterogeneity by tailoring protein dynamics. *Proc Natl Acad Sci U S A* 103:13722-13727.

CHAPTER 2

Conformational heterogeneity in a bacteriophage receptor binding protein opens an evolutionary pathway to expanded host range

2.1 Abstract

When proteins evolve new activity, a concomitant decrease in stability is often observed because the mutations that confer new activity can destabilize the native fold. In the conventional model of protein evolution, reduced stability is considered a purely deleterious cost of molecular innovation because unstable proteins are more prone to aggregation and more sensitive to environmental extremes. However, recent work has revealed that non-native, often unstable protein conformations play an important role in mediating evolutionary transitions, raising the question of whether instability can itself potentiate the evolution of new activity. We explored this question in a bacteriophage receptor binding protein (RBP) during host-range evolution. We studied the properties of the RBP of bacteriophage λ before and after host-range evolution and demonstrated that the evolved protein is relatively unstable and may exist in multiple conformations with unique receptor preference. This study raises the intriguing possibility that protein instability might play a previously unrecognized role in mediating host-range expansion in viruses.

2.2 Introduction

The evolution of new phenotypes at the organismal level can often be traced to functional innovation in a protein. In viruses, for example, changes in proteins that mediate host attachment may confer expanded host-range (Tétart et al. 1996; Yehl et al. 2019; Boon et al. 2020). It is

commonly observed across diverse protein types that functional innovations come with a concomitant decrease in stability (Wang et al. 2002; Studer et al. 2014). The most accepted explanation for this is that decreased stability is a deleterious side effect of mutations that generate innovations (Bloom et al. 2006), and compensatory stabilizing mutations ameliorate the cost (Wang et al. 2002; Tokuriki et al. 2008). An alternative hypothesis is that instability is not simply a side effect of innovation but instead contributes to its emergence by facilitating the ability of a single amino acid sequence to sample multiple folds, potentially fueling faster innovation (Tokuriki and Tawfik 2009; Sikosek and Chan 2014; Dellus-Gur et al. 2015)

The relationship between stability and the evolution of new functions has been understudied in virus receptor binding proteins (RBPs) compared to enzymes. A study on *vesicular stomatitis virus* showed that genetic robustness, a trait correlated with stability, did not enhance the ability of viruses to adapt to a new host, but the study did not explore a molecular mechanism (Cuevas et al. 2009). A separate study on bacteriophage λ revealed that the mutations that confer infectivity using a new receptor (Meyer et al. 2012) destabilize the viral particle (Petrie et al. 2018). But instead of simply causing generic misfolding, the instability was thought to alter the folding landscape of λ such that sometimes the peptide sequence stochastically folded into an alternative, less-stable conformation. Three lines of evidence were uncovered suggesting that evolved λ s also exhibited this type of non-genetic phenotypic heterogeneity. First, the decay rate of the evolved λ genotype decreased with time, a pattern that would be expected for a phenotypically mixed population (Russell 2021). Second, when evolved λ particles were incubated without host bacteria to replicate on, infectivity on the new receptor, OmpF, declined more rapidly than infectivity on the native receptor, LamB, suggesting the presence of an unstable subpopulation of OmpF-using particles. Third, it was shown that the subpopulation of

OmpF-binding particles could be removed from stocks of evolved λ particles by incubation with cells only expressing OmpF, leaving behind particles that preferentially bound LamB. All three results would be abnormal for an isogenic phage stock and were instead consistent with phenotypic heterogeneity. The observation that it was the unstable phenotype that appeared to preferentially interact with OmpF suggests that destabilizing the native fold was necessary for the production of the OmpF-binding conformation.

In the 2018 study all experiments were conducted using whole phage particles, and since the ancestor genotype and the evolved genotype differed only by six amino acid substitutions in the C-terminal region of the receptor binding protein (J), it was inferred that conformational heterogeneity in the J protein was the cause of phenotypic heterogeneity in the evolved λ s (Petrie et al. 2018). This inference seems reasonable, but it is also possible that the heterogeneity could arise due to interactions between J and other λ proteins and structures. In order to establish better support for the hypothesis that heterogeneity in λ had an evolutionary benefit of expanding the conformational repertoire of J, we measured the stability and conformational heterogeneity of the purified reactive domain of J before and after evolving to use OmpF.

The first step was to establish a robust protocol for purifying the reactive region of J. This posed a significant challenge because previous attempts to purify the J protein encountered difficulties with expression of even the wild type protein and relied on fusing the C-terminal J domain to maltose binding protein (MBP) in order to achieve sufficient solubility for purification (Wang et al. 1998; Wang et al. 2000; Berkane et al. 2006). Purification of the evolved OmpF⁺ version had never been attempted. We developed a protocol to purify the reactive domain of the wild type and evolved OmpF⁺ J without MBP, an achievement that allowed more sensitive biochemical assays without interference from the fused MBP molecule. Having successfully

purified both wild type and evolved proteins, we employed a thermal shift assay to measure temperature-mediated unfolding. The assay revealed striking differences in baseline foldedness and melting temperature between the wild type and evolved proteins, suggesting that the evolved protein is unstable. We then compared the empirically measured melting temperatures to those predicted from a machine learning based algorithm and uncovered a discrepancy suggestive of conformational heterogeneity. Seeking a functional link between conformational heterogeneity and the new activity of the evolved protein on OmpF, we tested the hypothesis that an isogenic population of evolved J proteins is composed of at least two subpopulations: one stable, LamB-preferring and one unstable, OmpF-preferring. Indeed, we found that a mild heat treatment reduced activity on OmpF but did not affect activity on LamB, consistent with the existence of uniquely folded forms.

2.3 Results

2.3.1 A protocol for purifying the reactive J domain without MBP

Previous efforts to purify the J protein have used two approaches. The first attempted to purify the full-length J (1132 amino acids) resulted in an insoluble, aggregated form that required treatment with the surfactant sarkosyl to improve solubility (Wang et al. 1998). A second attempt created multiple constructs by fusing the C-terminal 249, 349, and 449 residues of J to MBP, a protein that acts both as an affinity tag and improves the solubility of its fusion partner (Wang et al. 2000). This allowed expression and purification without using chemical solubilization, and the resulting protein was active on the LamB receptor (Berkane et al. 2006). Seeking to improve on these pioneering efforts to increase solubility and yield, we generated a double-tagged construct containing the C-terminal 249 amino acids of J, MBP, and a histidine tag to enable

purification using nickel affinity chromatography (Nallamsetty et al. 2005). Although the double-tagging strategy yielded highly pure, active protein suitable for some assays, the large MBP tag rendered these constructs useless for a thermal shift assay because MBP is itself a protein that interacts with the fluorescent dye, obscuring the signal from the smaller J domain. Therefore, we sought to develop a strategy to purify MBP-free J domain for the thermal shift assay. During the course of this study, the revolutionary structural prediction tool AlphaFold (Jumper et al. 2021) became publicly available (Mirdita et al. 2022) and examining the predicted J protein structure yielded a clue as to how we might increase solubility of the J domain without MBP. We observed that the C-terminal truncation (249 amino acids) consisted primarily of a region of structurally ordered beta sheets, except for an initial stretch of apparent disorder. We hypothesized that an even shorter truncation that excluded this initial region might increase solubility. Indeed, when we expressed only the C-terminal 152 amino acids of J, along with a his-tag (Figure 2.1A), we obtained soluble protein for both the wild type and evolved OmpF⁺ genotypes without the MBP tag (Figure 2.1B).

2.3.2 Assessing stability of the wild type and evolved J domains

With the purified reactive J domains, we then sought to test the hypothesis that the observed heterogeneity in evolved λ s was caused by instability and conformational heterogeneity in the reactive J domain. The first evidence that the evolved OmpF⁺ version of J is less stable than the ancestor came from the purification process itself. Although the shorter truncation allowed both the wild type and evolved OmpF⁺ J domains to be purified without MBP, the evolved OmpF⁺ produced a significantly lower yield of target protein than the wild type in the total cell extract as well as the purified fraction (Figure 2.1B). In the SDS PAGE gel, examining

the total cell extract lanes reveals that the background, non-target protein content appears lower for the evolved compared to the wild type. Overexpression of misfolded proteins can negatively impact *E. coli* growth (Hunke and Betton 2003), and one possibility is that the overexpression of the putatively unstable evolved protein caused reduced growth of the *E. coli* expression strain, leading to fewer total cells in the extracted pellet.

Seeking a more systematic and controlled method of comparing the stability of the wild type and evolved J domains, we employed a thermal shift assay (Crowther et al. 2010; Huynh and Partch 2015). Purified proteins are mixed with SYPRO orange, a fluorescent dye that binds hydrophobic residues, which are buried within the core of proteins in the folded state but become exposed during unfolding (Figure 2.1C). The temperature is slowly increased, and fluorescence is measured at 0.5 °C increments. The fluorescent signal for an initially well-folded protein resembles a sigmoidal curve, remaining low until the temperature reaches the melting point, then the signal rapidly rises as the hydrophobic core unfolds (Crowther et al. 2010; Huynh and Partch 2015). By first examining the initial fluorescence, we observed that the signal from the evolved J domain was substantially higher than the initial fluorescence of the wild type domain (Figure 2.2A), suggesting that some fraction of the molecules are unfolded at the starting temperature of the assay, 25 °C (Crowther et al. 2010), as would be expected for an unstable protein. Beyond the difference in initial fluorescence, there was also a shift in the melting temperature. The wild type J domain's melting temperature of about 60 °C, compared to about 40 °C for the evolved (Figure 2.2B). There are several phenomena that could explain the dramatic difference in melting temperature. One possibility is that some fraction of the evolved proteins begin the assay folded but melt at a lower temperature than the wild type proteins. Another possibility is that most of the evolved proteins begin the assay in an unfolded, aggregated state, and the melting peak is

observed when the aggregate melts. We did not probe this result further, and this conundrum has been previously cited as a limitation of the thermal shift assay (Crowther et al. 2010). Both outcomes would suggest that the evolved J domain is unstable relative to the ancestral, which is consistent with experiments on whole phage particles.

2.3.3 Assessing heterogeneity in the wild type and evolved J domains

We next explored whether the evolved OmpF⁺ J domain exhibited signs of conformational heterogeneity (Petrie et al. 2018). The first piece of evidence suggestive of conformational heterogeneity was obtained by comparing the empirically measured melting points of the wild type and evolved J domains with predictions based on machine learning algorithms. It is thought that proteins with conformational heterogeneity “confuse” such algorithms (Madhurima et al. 2021) because the set of published structures is biased toward stable proteins that fold into a single conformation (Mishra et al. 2019). The reasoning is that proteins prone to folding into multiple conformations might be difficult to crystallize and therefore underrepresented among solved structures. We used SCooP (Mishra et al. 2019), a bioinformatic tool that predicts melting point based on structure, to generate predictions for the wild type and evolved J domain AlphaFold structures. SCooP predicted the wild type J domain’s melting temperature with surprising accuracy (67.4 °C) but only predicted a ~3 °C lower melting temperature for the evolved J domain (64.4 °C). Because SCooP used structures that were themselves AlphaFold predictions, it may be that SCooP would have predicted a difference if AlphaFold had predicted more distinct structures (Figure 2.1A). However, AlphaFold is also trained on solved protein structures and so is likely subject to the same biases. Although not

conclusive, this intriguing result is consistent with the hypothesis that the evolved OmpF⁺ J domain exhibits conformational heterogeneity.

We then sought a test of conformational heterogeneity in the evolved J domain that would shed light on the relationship between non-native conformations and OmpF activity. In whole phage experiments, it was shown that populations of evolved λ lost infectivity on cells expressing OmpF faster than on cells expressing LamB, suggested the OmpF-binding phenotype may be less stable (Petrie et al. 2018). To test whether this was the case for the purified J domains as well, we designed a treatment that would selectively remove the least stable J molecules and measured the ability to bind LamB and OmpF receptors before and after treatment. For this assay we used the double-tagged constructs because they produced higher yields in purification. As expected, the wild type proteins had robust activity on LamB and completely lacked activity on OmpF, and that was unchanged by the heat treatment (Figure 2.3A-B). The evolved proteins, on the other hand, lost the ability to bind OmpF faster than they lost the ability to bind LamB (Figure 2.3B-C), consistent with the hypothesis that the unstable J conformations are responsible for OmpF binding. Unexpectedly, there was a slight gain in LamB binding after the heat treatment (Figure 2.3B-C). We did not delve further into this finding, but we suspect that perhaps the different conformations might interfere with each other's binding and removing a subset of conformations by heat treatment might have improved the remaining conformation's ability to bind. Together, these results are consistent with the model that evolved λ s contain J proteins of different conformations, and this heterogeneity allowed evolved λ s to use OmpF as a receptor.

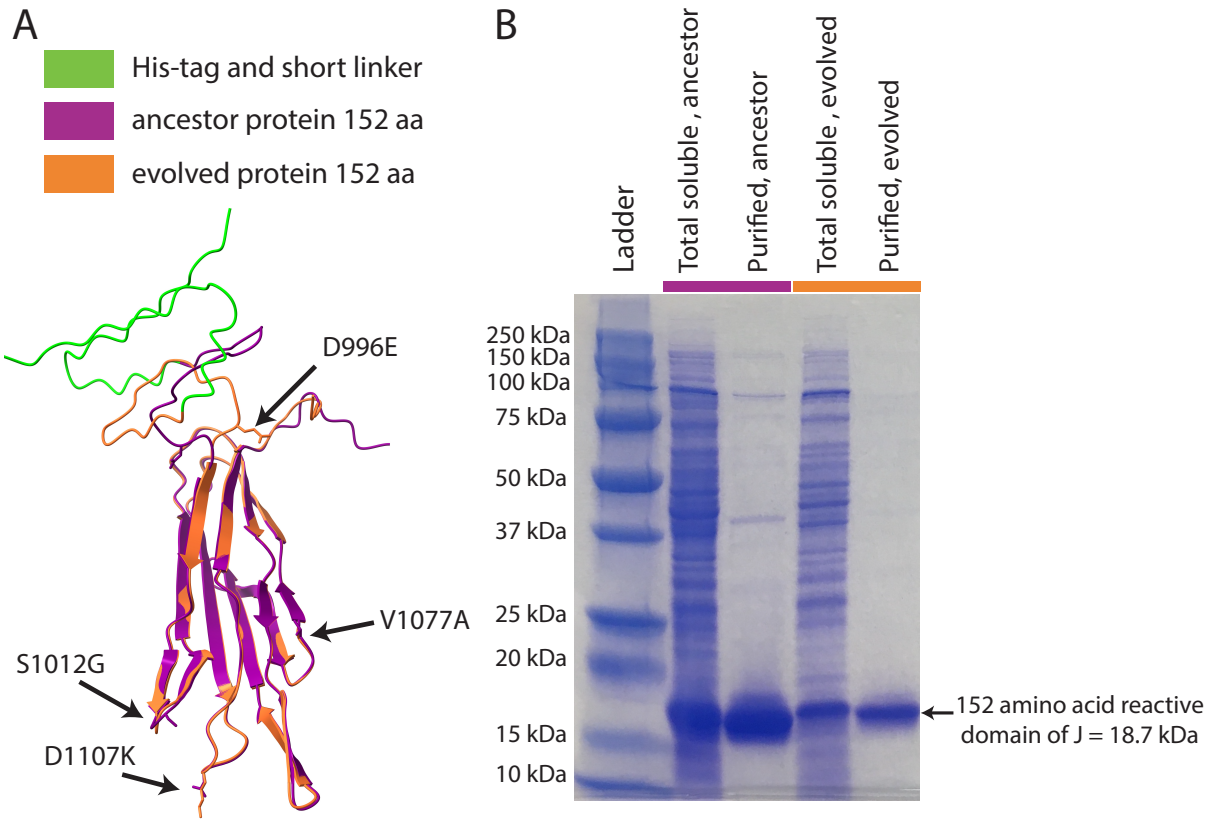


Figure 2.1 Despite nearly identical AlphaFold predicted structures, the evolved protein is dramatically less soluble. Panel A: Aligned AlphaFold modelled structures of ancestor J (fuchsia) and evolved J (orange). Models correspond to the 152 c-terminal amino acids of the full 1132 amino acid J protein. Each was purified using a his-tag attached N-terminally via a short linker sequence (green). The four amino acid changes that confer OmpF are indicated. Panel B: SDS PAGE gel showing that the evolved J produces lower soluble and purified yield than the ancestor.

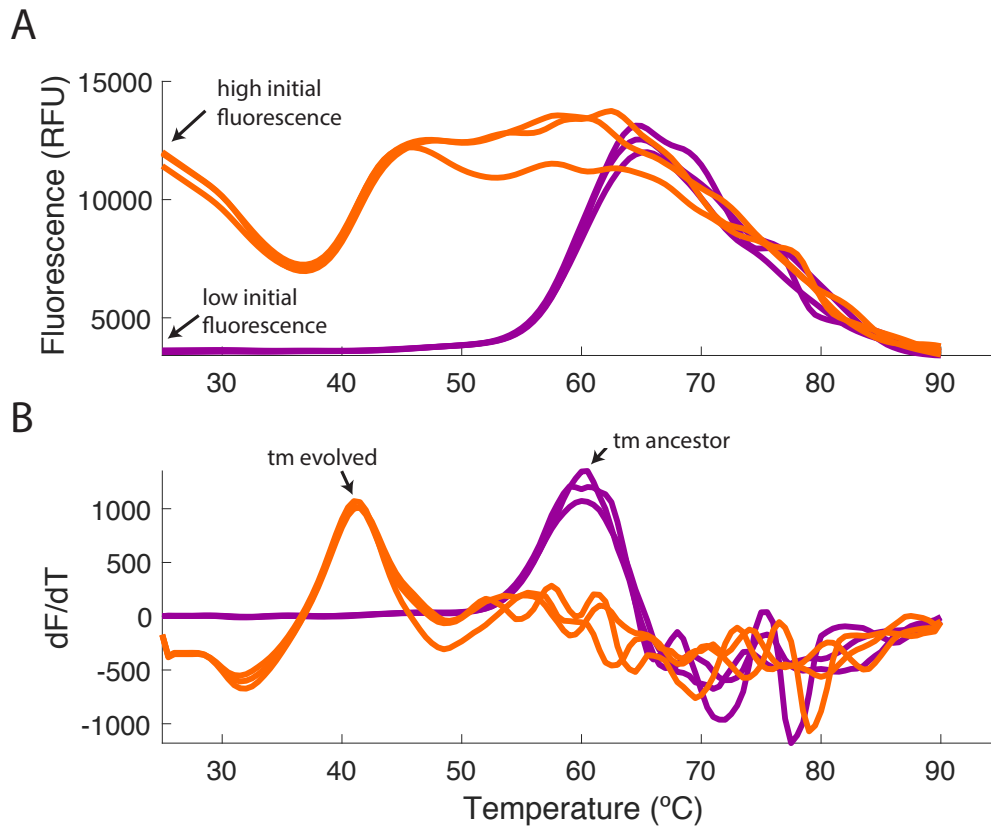


Figure 2.2 In a thermal shift assay, the evolved J has a higher initial fluorescence and lower melting point than the ancestor. Panel A: Melting curves for three replicates of each protein variant revealed substantially higher initial fluorescence for the evolved J compared to the ancestor. This indicates that some fraction of the evolved proteins have exposed hydrophobic residues even at 25 °C, the starting temperature of the assay. Panel B: The first derivative of the melting curve revealed that the evolved J also has a lower melting point than the ancestor, as indicated by the shift in the peak. Fuchsia curve = ancestor, orange curve = evolved.

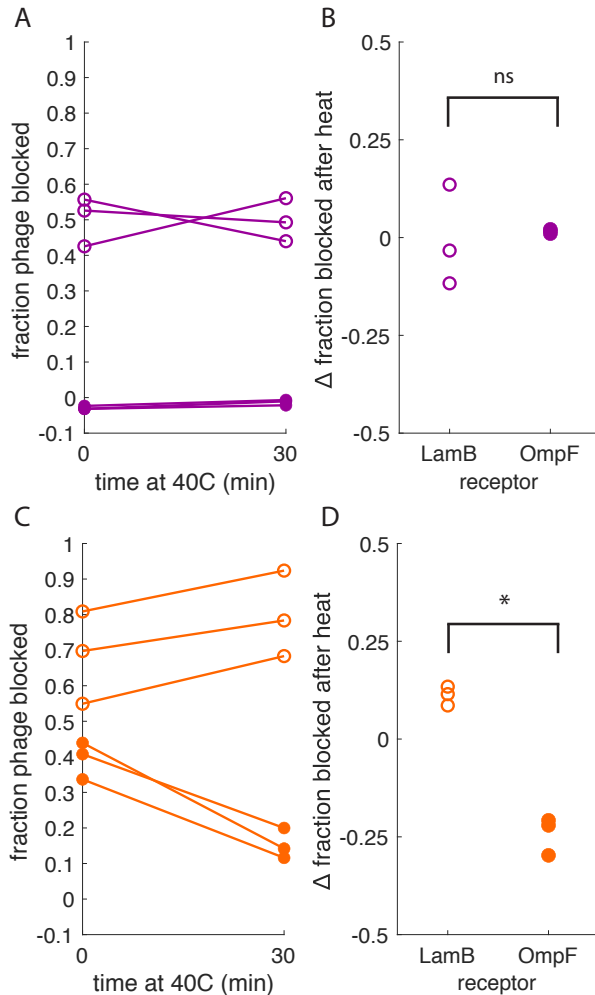


Figure 2.3 A mild heat treatment selectively reduces activity on the new receptor, while not affecting activity on the ancestral receptor. This supports the hypothesis that an alternative, less stable conformation is responsible for binding the new receptor. Activity was measured indirectly by quantifying the fraction of phage blocked by pre-incubating cells with J protein prior to phage adsorption. Higher values on the y axis correspond to higher fraction of phage blocked from adsorbing, indicating higher J protein activity. Open circles correspond to measurements on the LamB receptor, closed circles correspond to measurements on the OmpF receptor. Panel A: Activity of ancestor J protein on each receptor before and after heat treatment (heat treatment is indicated on x-axis). Note that the ancestor J protein has activity only on LamB, as expected. Panel B: Change in fraction blocked after heat treatment. The effect of the heat treatment on activity was indistinguishable between the two receptors (paired t-test, $t_{stat} = -0.2777$, $df = 2$, $p = 0.8073$). Panel C: Activity of evolved J protein on each receptor before and after heat treatment (heat treatment is indicated on x-axis). Note that the evolved J protein has activity on both LamB and OmpF, as expected. Activity of the evolved J protein on OmpF is reduced by the mild heat treatment, and activity on LamB is subtly increased. Panel D: The effect of the heat treatment on activity was significantly different between LamB and OmpF receptors (paired t-test, $t_{stat} = 10.326$, $df = 2$, $p = 0.0092$). Bonferroni adjusted significance values: ns = $p > 0.025$, * = $p < 0.025$.

2.4 Discussion

Instability is generally considered a cost of protein adaptation. We presented evidence that in some cases, protein instability potentiates adaptation by facilitating heterogeneity. A prior study on bacteriophage λ demonstrated instability and phenotypic heterogeneity among particles that had evolved to use a new receptor, OmpF. In this study, we narrowed our focus to just the reactive domain of the λ RBP (J) to evaluate whether conformational heterogeneity in J could have caused the phenotypic heterogeneity in the whole particle, thereby potentiating the evolution of OmpF use. We purified wild type and evolved J domains and measured their purified yield, baseline foldedness, melting temperature, and conformational heterogeneity. Consistent with our hypothesis, the evolved version produced a lower purified yield, as a first indication that it may be unstable and more prone to aggregation. Furthermore, the evolved J domains exhibited a higher initial fluorescence when treated with a dye that binds hydrophobic residues, suggesting a higher fraction of unfolded protein even before heat treatment. Strikingly, the melting temperature of the evolved version was ~ 20 °C lower than the wild type, a difference that was not anticipated by prediction algorithms trained on folded structures. A functional assay demonstrated that evolved J domains lose activity on the new receptor, OmpF, after heat treatment but retain full activity on the native receptor, LamB, provide further evidence of conformational heterogeneity. Our results corroborate a model in which destabilizing mutations alter J such that it produces a new conformational subpopulation with activity on OmpF, conferring expanded host-range (Petrie et al. 2018).

As structural techniques become more sensitive, it is becoming apparent that structural heterogeneity is not uncommon among proteins (Madhurima et al. 2021). The most common examples are intrinsically disordered proteins, which lack a defined three-dimensional structure,

(Oldfield and Dunker 2014), and allosteric enzymes, which undergo conformational rearrangements upon binding a ligand, (Beveridge et al. 2016). Another class, known as metamorphic proteins, can access multiple folded conformations from a single peptide sequence and their conformational state is not determined by binding (Dishman and Volkman 2018). Examples of metamorphic proteins span diverse protein types, from enzymes (Chang et al. 2015) to signaling proteins (Tuinstra et al. 2008) to scaffold proteins (Markley et al. 2013). In many animal viruses, glycoproteins that mediate membrane fusion in often undergo a pH induced conformational change (Carr and Kim 1993; Roche et al. 2007). Metamorphic proteins typically interconvert between states, but there is considerable variation in timescale, (Tuinstra et al. 2008; X and H 2008; Markley et al. 2013; Chang et al. 2015), and some may not interconvert at all due to a high energetic barrier between structures (Sinclair et al. 2022). In the case of the evolved λ J domain, our ability to separate conformational subpopulations with pulldown experiments suggests that interconversion between conformations is slow or nonexistent.

In addition to being structural curiosities, metamorphic proteins are also thought to play important roles in evolutionary transitions (Madhurima et al. 2021). By adopting multiple structures, a metamorphic protein can act as a bridge between the current and new functions (Sikosek and Chan 2014). In a study on synthetic small proteins, fold-switching sequences facilitated the formation of diverse multimeric structures with new binding activities (Yadid et al. 2010). In λ , the generalist J protein domain used in this study may represent an evolutionary bridge between a LamB-reliant ancestor and an OmpF specialist. If the metamorphic property of the evolved J domain is in fact related to instability, it might then be expected that mutations that optimize the new function on OmpF might re-stabilize J (Sikosek and Chan 2014). Indeed, prior

work demonstrated this result in whole λ particles (Meyer et al. 2016; Petrie et al. 2018), and it will be informative to test this hypothesis with the purified J domains from OmpF specialists.

A better understanding of the interaction between destabilizing mutations, protein metamorphosis, and evolvability could help shed light on the question of how proteins land on innovations among astronomically vast sequence space (Wagner 2014). The prevailing hypothesis is that neutral mutations create pathways in sequence space that allow proteins to change much of their sequence without deleterious costs, thereby placing them within a few mutations of an innovation (Wagner 2012). An alternative hypothesis is that sampling multiple phenotypes from the same genotype could allow proteins to reach new innovations more rapidly by relaxing the need for mutation (James and Tawfik 2003; Tokuriki and Tawfik 2009). If an expanded conformational repertoire can be achieved through generically destabilizing mutations rather than mutations that alter its shape in a highly specific way (Sikosek and Chan 2014), then innovations might be achieved rapidly because destabilizing mutations are common (Tokuriki et al. 2008). Some of the mutations that confer OmpF use in λ are clustered in loops on one surface of the reactive J domain, but some occur on the opposite end of the protein, suggesting their possible role as generic destabilizing mutations that pave the way for the contact surface mutations to optimize new function (Strobel et al. 2022).

Our results might also help to guide efforts aiming to employ synthetic biology and directed evolution to generate proteins with novel functions. It is commonly suggested that directed evolution experiments intentionally stabilize proteins before or during rounds of mutagenesis and selection (Socha and Tokuriki 2013; Stimple et al. 2020). While there are certainly examples of increased stability improving outcomes of directed evolution experiments (Bloom and Arnold 2009), our results demonstrate how stabilizing a protein might cause the

unintended outcome of reducing its ability to generate conformational diversity and therefore reducing innovative capacity.

Our study also serves as a cautionary tale for reliance on algorithms to predict protein properties. Algorithm based predictors represent a new frontier in protein science, and with the rise of strikingly accurate tools like AlphaFold (Jumper et al. 2021), researchers might be tempted to forego the expense and effort of pursuing empirical structural studies (Moore et al. 2022). Indeed, on the heels of the Alphafold release, several studies revealed limitations in its ability to predict disordered proteins or those with conformational heterogeneity (Ruff and Pappu 2021; Chakravarty and Porter 2022). The AlphaFold predict structures of our wild type and evolved J domains were nearly identical (Figure 2.1A), leading to nearly identical predicted melting temperatures. The prediction was only about 7 °C from the empirically determined value for our wild-type, stable protein but closer to 20 °C off for the evolved, unstable protein, consistent with the emerging trend that bioinformatic algorithms struggle with unstable, conformationally heterogeneous proteins. The question also has practical applications to global health, as it has been suggested that AlphaFold could help predict the next pandemic (Higgins 2021). Our findings are particularly relevant to this question because we observed that an unstable protein mediated the evolutionary transition of receptor use, as occurs during viral emergence. These results underscore the necessity of coupling predictions with empirical studies as well as improving predictive tools to better handle unstable proteins (Del Alamo et al. 2022).

Several limitations affected this study. One limitation was that we did not conduct structural studies on J. Even with advanced structural techniques, it is notoriously difficult to detect conformational heterogeneity in proteins (Goodchild et al. 2011; Madhurima et al. 2021). X-ray crystallography has been used to detect conformational heterogeneity, but the process of

crystallization itself likely biases structures toward a single conformation (Tuinstra et al. 2008; Dishman and Volkman 2018). Advances in cryo-EM and NMR are revealing that alternative conformations are much more common than previously recognized (Dishman and Volkman 2018). Further study of the wild type and evolved J domains using these methods will be particularly informative.

This study represents a step forward in understanding the mechanism of protein innovation that fueled a viral host-range expansion. We showed that phenotypic heterogeneity among whole λ particles that had evolved to use a new receptor in the previous study was likely caused by conformational heterogeneity in the receptor binding protein. This suggests that destabilization was not simply a deleterious cost of adaptation but instead had an independent evolutionary benefit of increasing conformational heterogeneity and potentiating OmpF⁺ evolution. The role of protein heterogeneity in mediating evolutionary transitions is a growing area of interest, and our results suggest that virus host-shifts can occur through this mechanism.

2.5 Materials and Methods

2.5.1 Media recipes for growing bacteria and phage

We prepared media used for growing bacteria and phage exactly as described in (Strobel et al. 2022).

2.5.2 Protein constructs

For each protein variant (wild type and evolved), two sets of constructs were purified. In one set, the 249 amino acid J domain was fused to maltose binding protein (MBP) coding sequence via a short linker sequence that included a 6-histidine tag. These constructs were used in the functional heterogeneity assay in Figure 2.3. To create these constructs, we used the In-

Fusion cloning kit (Takara Bio), using the pMAL-c2x plasmid as a backbone. To generate the insert (composed of the 6-histidine tag plus the 249 amino acids of J), we first cloned the 249 amino acids of J into the pET-45b plasmid backbone, which contained the 6-histidine tag (oligos to linearize pET-45b: 5' CTTGTCGTCGTCATTCGAACCGGTACC and 5' AGTCCGGATCCCAATTGGGAGCTCGTG; oligos to amplify the insert: 5' ATTGGGATCCGGACTTCAGACCACGCTGATGCC and 5' GATGACGACGACAAGATGGAGGACACGGAGGAAGG). We then amplified the 6-histidine tag + 249 amino acids of J from those constructs and cloned them into the pMAL-c2x backbone (oligos to linearize pMAL-c2x: 5' GAATTCTGAAATCCTTCCCTCGATC and 5' AAGCTTGGCACTGGCCGTC; oligos to amplify the insert: 5'- GGGAAGGATTTTCAGAATTCATGGCACATCACCACCAC and GCCAGTGCCAAGCTTTCAGACCACGCTGATGCCC). We transformed cloning products into Stellar Competent Cells (Takara Bio cat# 636767) and used colony PCR to verify the correct sequence (oligos: 5' TGGCGAAAGATCCACGTATTG and 5' AGGCGATTAAGTTGGGTAACG). We then minipreped the plasmid and transformed it into BL21DE3 cells that we had modified to be *malT*⁻. The purpose of knocking out *malT* was to prevent the binding of expressed J proteins to LamB receptors on the outer membrane after cell lysis (*malT* regulates the expression of LamB, the receptor for the J protein (Chaudhry et al. 2018).)

The second set of constructs contained only the 6-histidine tag and the c-terminal 152 c-terminal amino acids. These constructs were used in the thermal shift assay. To create these constructs, we used the Q5 site directed mutagenesis kit (NEB catalog #E0554S, with oligos: 5' AACTGTACGATAAACGGTAC and CTTGTCGTCGTCATTC) to delete the first 97

amino acids of J from constructs that had been previously generated containing the 249 c-terminal amino acids of J inserted in the pET-45b backbone. We transformed the products into NEB 5-alpha competent *E. coli* and used colony PCR to screen for successful deletions (5'GCGAAATTAATACGACTCACTATA and 5' AAGGGGTTATGCTAGTTATTG). We then minipreped the plasmid and transformed it into BL21DE3 cells that we had modified to be *malT*⁻.

Early in the project, we attempted to express a construct containing the full length J protein, 1132 amino acids, plus a 6-histidine tag, but it was completely insoluble, consistent with a previous study (Wang et al. 1998).

2.5.3 Protein expression and purification

To express proteins, we added 2 mL overnight BL21DE3 *malT*⁻ cultures containing the plasmid in 100 mL LB and 240 μ L 50 mg/ml carbenicillin for 2 hours at 37 °C, shaking at 110 rpm. We then added 100 μ L of 1M IPTG and incubated at room temperature (~22 °C), shaking at 110 rpm for 18-20 hours. Expression cultures were then pelleted by centrifugation at 3214 x g for 10 min, resuspended in 10 mL of buffer containing 20 mM Tris pH 9, 150 mM NaCl using an 18-gauge needle to reduce viscosity, and frozen at -80 for 30 min or overnight. Then samples were thawed and sonicated in three cycles of 60 seconds at 20% amplitude, with 60 seconds on ice in between each round using Fisher FB-50 sonic dismembrator with the standard 1/8" diameter microtip. Sonicates were centrifuged again at 3214 x g for 10 min at 10 °C and the soluble fraction was collected by filtering supernatants through 0.22 μ m syringe filters.

Filtered supernatants were purified using 400 μ L of resuspended Ni-NTA resin (Thermo Scientific cat# 88221), using the batch method in 15 mL falcon tubes. Resin was first centrifuged

at 500 x g for 2 min and storage fluid was removed, then resin was washed with 4 mL of 20 mM Tris pH 9, 150 mM NaCl, centrifuged again at 500 x g for 2 min and buffer removed. Then 10 mL of sonicate was added to the washed resin and incubated for 30 minutes with periodic mixing. Then the sample was centrifuged at 500 x g for 2 min and the supernatant was removed and discarded. The resin with bound protein was then washed four times using 4 mL of 20 mM Tris pH 9, 150 mM NaCl, 20 or 80 mM imidazole (20 mM for the MBP-containing constructs and 80 mM for the MBP-free constructs). After the final wash, bound proteins were eluted from the resin using 20 mM Tris pH 9, 150 mM NaCl, 380 mM imidazole and filtered through 0.22 μ m spin column filters.

2.5.4 SDS PAGE

Samples were prepared for SDS PAGE by combining 100 μ L sample with 100 μ L sample buffer (900 μ L 2x Laemmli sample buffer, BioRad cat #1610737 + 100 μ L 1M DTT), heating to 65 °C for 10 minutes, centrifuging at 16,000 x g for 10 minutes, and loading 20 μ L in a 12% TGX gel (BioRad cat #4561043).

2.5.5 Melting point determination using thermal shift assay

Purified protein samples of the MBP-free constructs were used for the thermal shift assays. Our protocol was designed following current best practices in the field (Huynh and Partch 2015; Kazlauskas et al. 2021). First, the 5000x Sypro Orange (Sigma-Aldrich catalog #S5692-50UL) was diluted to 200x in 20 mM tris pH 9, 150 mM NaCl. Then, three replicate wells of 45 μ L of protein sample and 5 μ L 200x sypro orange were prepared for each protein variant in an optically clear PCR plate, covered with a clear adhesive seal. Three replicate controls were also run using elution buffer instead of protein (20 mM tris pH 9, 150 mM NaCl,

380 mM imidazole) used instead of the protein sample. Then the sample was run on a BioRad CFX96 qPCR machine with scan mode set to FRET, under the following thermal conditions: 25 °C for fifteen minutes, followed by 0.5 °C incremental increase from 25 °C to 90 °C holding at each temperature for 30 seconds and then capturing a reading at each temperature step, followed by a final step at 25 °C for five minutes.

2.5.6 Alphafold

We generated structural predictions of our constructs using the publicly available version (Mirdita et al. 2022) of AlphaFold (Jumper et al. 2021) using all standard presets. We used Chimera to visualize and color-code predicted structures (Pettersen et al. 2004).

2.5.7 Functional heterogeneity assay

Purified MBP-containing constructs were used for this assay. Purified samples were first diluted to approximately 5 μ M and separated into three replicates of 800 μ L in Lo-Bind microcentrifuge tubes. Then, from each tube 400 μ L was immediately sampled and set on ice (the pre-heat treated sample). Samples were then incubated in a warm water bath at 40 °C for 30 minutes, and another 400 μ L was sampled and put on ice. The heated samples were filtered through 0.22 μ m spin column filters to remove any large aggregates.

Then, we measured the activity of each construct on each receptor, before and after heat treatment. We measured activity on receptors as a function of their ability to block receptors from being bound by whole phage particles. To measure this, we first incubated 100 μ L of 5 μ M protein with 10 uL overnight bacteria culture ($\sim 10^7$ cells) expressing only LamB or only OmpF (Keio collection knockouts) (Baba et al. 2006) allowing the proteins to adsorb to the receptors on

the cell surface. Then, we added whole phage particles and allowed the phage to adsorb to receptors not blocked by protein. Tubes were then placed on ice and centrifuged at 16,000 x g for one minute, and 50 μL of the supernatant (containing unbound phage) was plated with 100 μL of wild type *E. coli* (Keio collection parental strain BW25113) (Baba et al. 2006). For the whole phage particles, we chose a receptor generalist genotype of λ that can infect through both LamB and OmpF (Meyer 2012). EvoC phage were induced from a lysogenic prophage integrated into the HWEC106 genome by heat shock. Lysogens were grown up at 37 °C in LB, then 140 μL was inoculated into 4 mL LBM9 and 40 μL MgSO₄, grown at 30 °C for two hours, heat shocked at 42 °C for fifteen minutes, then incubated at 37 °C for 90 minutes. The lysate was filtered through a 0.22 μm syringe filter and diluted in 9 mL M9 minimal media containing no sugar source, supplemented with 90 μL MgSO₄.

We also included three replicate controls containing phage with LB media instead of cells and 20 mM tris pH 9, 150 mM NaCl, 380 mM imidazole buffer instead of protein, in order to calculate the total number of phage particles. Additionally, to capture the fraction of phage that adsorb in the absence of protein, we included three replicate controls containing phage with LamB-only cells and 20 mM tris pH 9, 150 mM NaCl, 380 mM imidazole buffer instead of protein, as well as three analogous replicates with OmpF-only cells. Only about 1-2% of total phage remained un-adsorbed to either cell type. To calculate the fraction of phage blocked by protein, we subtracted the number of unbound phage in the buffer treatment from the number of unbound phage in the protein treatment and divided the result by the total number of phage initially added.

To compare the effect of heat treatment on different receptors, we used a paired t-test. Because the variances were unequal for measurements on the ancestor protein, an unequal variance t-test was also performed and the significance did not change.

2.6 Acknowledgements

We thank Dr. Katherine Petrie for her valuable help and advice with this project. We thank Dr. Randolph Hampton, Dr. Sonya Neal, and Dr. Matthew Flagg for generously offering their expert advice and laboratory space for this project. Without their help, this project would not have been possible.

Chapter 2, in full, is coauthored with Meyer, Justin R. The dissertation author was the primary investigator and author of this paper.

2.7 References

Baba, T., T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. Datsenko, M. Tomita, B. Wanner, and H. Mori 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular systems biology* 2.

Berkane, E., F. Orlik, J. Stegmeier, A. Charbit, M. Winterhalter, and R. Benz. 2006. Interaction of bacteriophage lambda with its cell surface receptor: an in vitro study of binding of the viral tail protein gpJ to LamB (Maltoporin). *Biochemistry* 45.

Beveridge, R., L. Migas, K. Payne, S. Nigel, L. David, and P. Barran. 2016. Mass spectrometry locates local and allosteric conformational changes that occur on cofactor binding. *Nature Communications* 7:1-9.

Bloom, J. D. and F. H. Arnold. 2009. In the light of directed evolution: pathways of adaptive protein evolution. *Proc Natl Acad Sci U S A* 106 Suppl 1:9995-10000.

Bloom, J. D., S. T. Labthavikul, C. R. Otey, and F. H. Arnold. 2006. Protein stability promotes evolvability. *Proc Natl Acad Sci U S A* 103:5869-5874.

Boon, M., D. Holtappels, C. Lood, V. van Noort, and R. Lavigne. 2020. Host Range Expansion of *Pseudomonas* Virus LUZ7 Is Driven by a Conserved Tail Fiber Mutation. *PHAGE* 1:87-90.

Carr, C. and P. Kim. 1993. A spring-loaded mechanism for the conformational change of influenza hemagglutinin. *Cell* 73:823-832.

Chakravarty, D. and L. Porter. 2022. AlphaFold2 fails to predict protein fold switching.

Chang, Y., Y. Cohen, C. Phong, W. Myers, Y. Kim, R. Tseng, J. Lin, L. Zhang, J. Boyd, Y. Lee, S. Kang, D. Lee, S. Li, R. Britt, M. Rust, S. Golden, and A. A LiWang. 2015. A protein fold switch joins the circadian oscillator to clock output in cyanobacteria. 349.

Chaudhry, W., M. Pleška, N. Shah, H. Weiss, I. McCall, J. Meyer, A. Gupta, C. Guet, and B. Levin. 2018. Leaky resistance and the conditions for the existence of lytic bacteriophage. *PLoS biology* 16.

Crowther, G., P. He , P. Rodenbough, A. Thomas, K. Kovzun, D. Leibly, J. Bhandari , L. Castaneda, W. Hol , M. Gelb, A. Napuli, and V. Van , WC. 2010. Use of thermal melt curves to assess the quality of enzyme preparations. *Analytical biochemistry* 399.

Cuevas, J. M., A. Moya, and R. Sanjuán. 2009. A genetic background with low mutational robustness is associated with increased adaptability to a novel host in an RNA virus. *J Evol Biol* 22:2041-2048.

Del Alamo, D., D. Sala, H. Mchaourab, and J. Meiler. 2022. Sampling alternative conformational states of transporters and receptors with AlphaFold2. *eLife* 11.

Dellus-Gur, E., M. Elias , E. Caselli, F. Prati, M. Salverda, J. de Visser, J. Fraser, and D. Tawfik. 2015. Negative Epistasis and Evolvability in TEM-1 β -Lactamase--The Thin Line between an Enzyme's Conformational Freedom and Disorder. *Journal of molecular biology* 427.

Dishman, A. and B. Volkman. 2018. Unfolding the Mysteries of Protein Metamorphosis. *ACS chemical biology* 13.

Goodchild, S., P. Curmi , and L. Brown. 2011. Structural gymnastics of multifunctional metamorphic proteins. *Biophysical reviews* 3.

Higgins, M. 2021. Can We AlphaFold Our Way Out of the Next Pandemic? *Journal of molecular biology* 433.

Hunke, S. and J. Betton. 2003. Temperature effect on inclusion body formation and stress response in the periplasm of *Escherichia coli*. *Molecular microbiology* 50.

Huynh, K. and C. Partch. 2015. Analysis of protein stability and ligand interactions by thermal shift assay. *Current protocols in protein science* 79.

James, L. and D. Tawfik. 2003. Conformational diversity and protein evolution--a 60-year-old hypothesis revisited. *Trends in biochemical sciences* 28.

Jumper, J., R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Židek, A. Potapenko, A. Bridgland, C. Meyer, S. Kohl, A. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596.

Kazlauskas, E., V. Petrauskas, V. Paketyrytė, and D. Matulis. 2021. Standard operating procedure for fluorescent thermal shift assay (FTSA) for determination of protein-ligand binding and protein stability. *European biophysics journal* : EBJ 50.

Madhurima, K., B. Nandi, and A. Sekhar. 2021. Metamorphic proteins: the Janus proteins of structural biology. *Open biology* 11.

Markley, J., J. Kim, Z. Dai, J. Bothe, K. Cai, R. Frederick, and M. Tonelli. 2013. Metamorphic protein IscU alternates conformations in the course of its role as the scaffold protein for iron-sulfur cluster biosynthesis and delivery. *FEBS letters* 587.

Meyer, J. R., D. T. Dobias, S. J. Medina, L. Servilio, A. Gupta, and R. E. Lenski. 2016. Ecological speciation of bacteriophage lambda in allopatry and sympatry. *Science* 354:1301-1304.

Meyer, J. R., D. T. Dobias, J. S. Weitz, J. E. Barrick, R. T. Quick, and R. E. Lenski. 2012. Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science* 335:428-432.

Mirdita, M., K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger. 2022. ColabFold - Making protein folding accessible to all.

Mishra, S., L. Looger, and L. Porter 2019. Inaccurate secondary structure predictions often indicate protein fold switching. *Protein science : a publication of the Protein Society* 28.

Moore, P., W. Hendrickson, R. Henderson, and A. Brunger. 2022. The protein-folding problem: Not yet solved. *Science (New York, N.Y.)* 375.

Nallamsetty, S., B. Austin, K. Penrose, and D. Waugh 2005. Gateway vectors for the production of combinatorially-tagged His6-MBP fusion proteins in the cytoplasm and periplasm of *Escherichia coli*. *Protein science : a publication of the Protein Society* 14.

Oldfield, C. and A. Dunker. 2014. Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. <http://dx.doi.org/10.1146/annurev-biochem-072711-164947> 83:553-584.

Petrie, K. L., N. D. Palmer, D. T. Johnson, S. J. Medina, S. J. Yan, V. Li, A. R. Burmeister, and J. R. Meyer. 2018. Destabilizing mutations encode nongenetic variation that drives evolutionary innovation. *Science* 359:1542-1545.

Pettersen, E., T. Goddard, C. Huang, G. Couch, D. Greenblatt, E. Meng, and T. Ferrin. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* 25.

Roche, S., F. Rey, Y. Gaudin, and S. Bressanelli. 2007. Structure of the prefusion form of the vesicular stomatitis virus glycoprotein G. *Science (New York, N.Y.)* 315.

Ruff, K. and R. Pappu. 2021. AlphaFold and Implications for Intrinsically Disordered Proteins. *Journal of molecular biology* 433.

Russell, C. J. 2021. Hemagglutinin Stability and Its Impact on Influenza A Virus Infectivity, Pathogenicity, and Transmissibility in Avians, Mice, Swine, Seals, Ferrets, and Humans. *Viruses* 13.

Sikosek, T. and H. S. Chan. 2014. Biophysics of protein evolution and evolutionary protein biophysics. *J R Soc Interface* 11:20140419.

Sinclair, J. F., M. M. Ziegler, and T. O. Baldwin. 2022. Kinetic partitioning during protein folding yields multiple native states. *Nature Structural Biology* 1:320-326.

Socha, R. and N. Tokuriki 2013. Modulating protein stability - directed evolution strategies for improved protein function. *The FEBS journal* 280.

Stimple, S., M. Smith, and P. Tessier 2020. Directed evolution methods for overcoming trade-offs between protein activity and stability. *AIChE journal. American Institute of Chemical Engineers* 66.

Strobel, H., E. Horwitz, and J. Meyer 2022. Viral protein instability enhances host-range evolvability. *PLoS genetics* 18.

Studer, R. A., P. A. Christin, M. A. Williams, and C. A. Orengo. 2014. Stability-activity tradeoffs constrain the adaptive evolution of RubisCO. *Proc Natl Acad Sci U S A* 111:2223-2228.

Tokuriki, N., F. Stricher, L. Serrano, and D. S. Tawfik. 2008. How protein stability and new functions trade off. *PLoS Comput Biol* 4:e1000002.

Tokuriki, N. and D. S. Tawfik. 2009. Protein dynamism and evolvability. *Science* 324:203-207.

Tuinstra, R., F. Peterson, S. Kutlesa, E. Elgin, M. Kron, and B. Volkman. 2008. Interconversion between two unrelated protein folds in the lymphotactin native state. *Proceedings of the National Academy of Sciences* 105:5057-5062.

Tétart, F., F. Repoila, C. Monod, and H. Krisch. 1996. Bacteriophage T4 host range is expanded by duplications of a small domain of the tail fiber adhesin. *Journal of molecular biology* 258.

Wagner, A. 2012. The role of robustness in phenotypic adaptation and innovation. *Proceedings. Biological sciences* 279.

Wagner, A. 2014. *Arrival of the Fittest*. Current, New York, New York.

Wang, J., M. Hofnung, and A. Charbit. 2000. The C-terminal portion of the tail fiber protein of bacteriophage lambda is responsible for binding to LamB, its receptor at the surface of *Escherichia coli* K-12. *Journal of bacteriology* 182.

Wang, J., V. Michel, M. Hofnung, and A. Charbit. 1998. Cloning of the J gene of bacteriophage lambda, expression and solubilization of the J protein: first in vitro studies on the interactions between J and LamB, its cell surface receptor. *Research in microbiology* 149.

Wang, X., G. Minasov, and B. K. Shoichet. 2002. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *J Mol Biol* 320:85-95.

X, L. and Y. H. 2008. Protein metamorphosis: the two-state behavior of Mad2. *Structure* (London, England : 1993) 16.

Yadid, I., N. Kirshenbaum, M. Sharon, O. Dym, and D. S. Tawfik. 2010. Metamorphic proteins mediate evolutionary transitions of structure. *Proc Natl Acad Sci U S A* 107:7287-7292.

Yehl, K., S. Lemire, A. C. Yang, H. Ando, M. Mimee, M. T. Torres, C. de la Fuente-Nunez, and T. K. Lu. 2019. Engineering Phage Host-Range and Suppressing Bacterial Resistance through Phage Tail Fiber Mutagenesis. *Cell* 179:459-469.e459.

CHAPTER 3

Viral protein instability enhances host-range evolvability

3.1 Abstract

Viruses are highly evolvable, but what traits endow this property? The high mutation rates of viruses certainly play a role, but factors that act above the genetic code, like protein thermostability, are also expected to contribute. We studied how the thermostability of a model virus, bacteriophage λ , affects its ability to evolve to use a new receptor, a key evolutionary transition that can cause host-range evolution. Using directed evolution and synthetic biology techniques we generated a library of host-recognition protein variants with altered stabilities and then tested their capacity to evolve to use a new receptor. Variants fell within three stability classes: stable, unstable, and catastrophically unstable. The most evolvable were the two unstable variants, whereas seven of eight stable variants were significantly less evolvable, and the two catastrophically unstable variants could not grow. The slowly evolving stable variants were delayed because they required an additional destabilizing mutation. These results are particularly noteworthy because they contradict a widely supported contention that thermostabilizing mutations enhance evolvability of proteins by increasing mutational robustness. Our work suggests that the relationship between thermostability and evolvability is more complex than previously thought, provides evidence for a new molecular model of host-range expansion evolution, and identifies instability as a potential predictor of viral host-range evolution.

3.2 Introduction

The evolvability of life is evident in its remarkable diversity of forms and persistence through time. While life may be inherently evolvable, it is thought that evolvability is a

malleable trait itself subject to evolutionary tinkering (Kirschner and Gerhart 1998). A number of mechanisms have been proposed to enhance evolvability, centering on the increased capacity to produce novel phenotypes (Payne and Wagner 2019). The most straightforward is evolving a higher mutation rate which allows populations to explore phenotypic variation that results from genetic changes (Stern and Andino 2016). However, this mechanism is constrained by concomitant increases in mutation load that can slow adaptation (Sprouffske et al. 2018). There are other qualities of biological systems that can facilitate phenotypic novelty, although these too can have conflicting effects. This manuscript focuses on one such trait, protein thermostability, which is the propensity to resist misfolding when heated. High thermostability can promote evolvability by buffering protein folding against destabilization from mutations, a property called mutational robustness (Wagner 2005), thus allowing genomes to accumulate more mutations and increasing the chance of uncovering the mutations for an innovation (Bloom et al. 2006; Fasan et al. 2008; McBride et al. 2008; Bloom and Arnold 2009; Thyagarajan and Bloom 2014). However, high thermostability can also limit evolvability because high heat tolerance is often achieved by increasing conformational rigidity (Vihinen 1987; Jaenicke and Böhm 1998; Smirnova and Kaback 2003; Besenmatter et al. 2007; Razvi and Scholtz 2009), precluding the formation of non-native conformers that sometimes support new functions (Yadid et al. 2010; Raman et al. 2016). Indeed, the conformational flexibility of unstable proteins can be essential for evolvability by allowing non-native conformers, protein molecules that fold into conformations other than the ground-state conformation, to explore promiscuous functions (Tokuriki and Tawfik 2009; Sikosek and Chan 2014). While the most widely accepted view is that thermostability promotes faster protein evolution, most scholars recognize that the

relationship between thermostability and evolvability is more complex and that certain types of proteins may be more or less sensitive to either mechanism.

As the world experiences the second year of the COVID-19 Pandemic, the attention of many scientists has turned to the problem of predicting which viral strains are most likely to emerge. Presumably, more evolvable viruses are more likely to gain the mutations and innovations necessary to shift species. Fortunately, the ease with which viruses evolve under laboratory conditions allows for direct experimental tests of the role thermostability plays in evolvability. Such experiments have been conducted by culturing different viral variants with potentially different evolvabilities in parallel under identical conditions, and quantifying evolvability as the viruses' ability to adapt to imposed challenges. So far, this method has revealed opposite results: When genotypes of phage $\phi 6$ with high and low robustness were evolved to cope with heat stress, the more robust variant adapted faster (Ogbunugafor et al. 2009). However, a separate study challenged vesicular stomatitis virus with replicating on a novel cell type and reported the opposite pattern: less robust variants were more evolvable (Cuevas et al. 2009). These findings suggest that while robustness is one mechanism that can enhance viral evolvability, there are other mechanisms associated with low robustness that may override it.

One particular class of viral proteins that are relevant for host-shift evolution are host-recognition proteins (Hueffer et al. 2003; Shi et al. 2014; Lu et al. 2015) , and thus knowing the determinants of their evolvability is particularly important. We used bacteriophage λ as a model system to test how thermostability impacts the evolution of receptor recognition. Through coevolution with *E. coli* in the laboratory, λ can evolve from a specialist able to bind only the ancestral receptor, LamB, to a generalist by gaining the ability to bind a second receptor, OmpF

(Meyer et al. 2012). The mutations that produced expanded receptor recognition occurred in the *J* gene, which encodes λ 's host-recognition protein, and were also shown to reduce λ 's thermostability, as measured by the rate of decrease in infectious titer (Petrie et al. 2018). Additional experiments revealed that genetically homogeneous populations of the evolved generalist produced a subpopulation of phage particles that lost infectivity faster than the population as a whole, and the unstable subpopulation also possessed enhanced ability to use OmpF (Petrie et al. 2018). This system exemplifies how our current understanding of protein evolvability can lead to divergent predictions: it could be reasoned that destabilization promoted OmpF⁺ evolution by increasing the conformational flexibility of the protein, allowing it to occupy unstable, non-native, yet functionally innovative conformers, but it is equally reasonable to predict that because the innovative mutations were destabilizing, they would have been more likely to evolve in a more stable, robust background. We hypothesize that, in the case of λ 's receptor use evolution, protein instability was not simply a cost for attaining the OmpF⁺ mutations, but instead it actually enhanced evolvability by enabling the formation of phenotypically diverse subpopulations of particles with latent new binding abilities.

3.3 Results

3.3.1 Naturally evolved thermostabilizing mutations reduced host-range expansion evolvability

We initiated our evaluation of the thermostability-evolvability relationship by comparing the propensity to evolve OmpF⁺ among three OmpF⁻ precursor strains: one that is unstable and two naturally evolved thermostable derivatives. To generate the unstable precursor strain, we started with a well-studied unstable OmpF⁺ λ generalist ("7-mut") (Petrie et al. 2018) and engineered out a single mutation (N1107K) that is known to be critical for OmpF function

(Maddamsetti et al. 2018). The resulting “6-mut” genotype was unable to use OmpF and remained as unstable as “7-mut” (Figure 3.1A), with both less stable than the ancestral λ genotype (Figure 3.1A). Next, we selected the resulting OmpF⁻ “6-mut” for increased thermostability without selecting for receptor function (see materials and methods). We found two mutations (T987A and F1122L) that each independently enhanced 6-mut thermostability (Figure 3.1B), restoring thermostability to that of the ancestral λ (Figure 3.1B). Intriguingly, the stabilizing mutations bracket the region of J that contained most host-range mutations in the original study documenting the OmpF innovation (Meyer et al. 2012). Amino acid 987 lies at the very beginning of that “reactive region” but was never mutated in the original study, whereas amino acid 1122 lies at the very end and was mutated in a single isolate from the original study (Meyer et al. 2012). That mutation, I1122F, was one of the seven mutations in the 7-mut genotype (Petrie et al. 2018) that was used to initiate this study.

To test the effect of stabilization on λ evolvability, we performed evolutionary replay experiments in which the unstable precursor and the two thermostable derivatives were cultured for hundreds of generations. The experimental design (Figure 3.2A) mirrored the original coevolution experiment in which OmpF function first evolved in the laboratory (Meyer et al. 2012), except that in half of the replicates, we coevolved phage with a host genotype isolated from later in the original experiment, in an effort to correct for the time-shift between phage and bacteria caused by starting with the partially evolved 6-mut. This time-shifted host had a duplication in the *malT* gene, a regulator of LamB expression, causing LamB to only be expressed in rare mutants that had reverted the duplication (Chaudhry et al. 2018). All three genotypes were capable of evolving OmpF use, but the two thermostable variants required more time to evolve (6–10 days for the stable genotypes compared to 1–2 days for the unstable

genotype), and most replicate populations of the stable genotypes never evolved during the ten-day experiment (Figure 3.2B). This result was despite the fact that most replicate populations of stabilized genotypes attained titers equal to or higher than those of the unstable 6-mut (Figure 3.A.1).

Sequencing revealed an explanation for why unstable genotypes evolved OmpF⁺ faster than thermostable genotypes: a single amino acid substitution allowed the unstable genotype to evolve OmpF⁺, while thermostable genotypes required two substitutions (Figure 3.2C). It was not surprising that a single mutation could restore OmpF⁺ in the unstable 6-mut background, since 6-mut was created by reverting a single mutation from an OmpF⁺ genotype. Unexpectedly though, only three of the six 6-mut replicates restored the same amino acid that was reverted (I1107), while the other three replicates obtained a different mutation (S1011R or S1049R), suggesting the role N1107K plays in OmpF function can be accomplished through other mutations (Figure 3.2C). Each thermostable genotype that evolved OmpF⁺ also gained either N1107K, S1011R, or S1049R (Figure 3.2C). Parallel evolution of these four mutations across unstable and thermostable backgrounds suggests that this set of mutations may produce a similar phenotypic effect in J that is required for activity on OmpF, regardless of the thermostability of the genotype in which it occurs. Curiously, another set of mutations (V966L, S970Y, and L1122F) only evolved in thermostable backgrounds and were never detected in the unstable background (Figure 3.2C). Each instance of OmpF⁺ in a thermostable background required one mutation from each set (Figure 3.2C).

3.3.2 Engineering a library of thermostability variants by manipulating a single amino acid

The two naturally selected thermostable λ genotypes provided notable cases in contradiction to the hypothesis that thermostability promotes evolvability and instead pointed to instability as a trait that potentiates functional innovation. However, this pattern was observed in only a small set of variants, and we next sought to test the hypotheses across a larger set of λ variants. To generate additional variants, we focused on codon 987 in J, the site of one of the naturally selected thermostabilizing mutations (T987A). We chose to focus on 987 instead of 1122 because 1122 lies in the C-terminal region in the J protein that is typically associated with receptor binding (Wang et al. 2000) and OmpF use evolution (Meyer et al. 2012) and was already mutated in the 7-mut, relative to the λ ancestor. We reasoned that further study should focus on variation at amino acid 987 because physical separation from the region previously implicated in host-range evolution might reduce the risk of confounding effects. We used Multiplexed Automated Genome Engineering with coselection (Wang et al. 2012) to create a library of amino acid variants at this position in the 6-mut genotypic background. From this procedure we were able to generate nine of the remaining 18 possible amino acid variants at position 987 (New amino acids L, C, S, G, K, R, Y, I, and P).

We hoped to generate a gradient of stability variants to assess the precise relationship between thermostability and evolvability; however, we were only able to engineer three levels of thermostability: stable, unstable, and catastrophically unstable. Six variants were as stable as ancestral λ (paired t-tests, $n = 3$ per genotype; T987S: $p = 0.697$; T987C: $p = 0.575$; T987Y: $p = 0.412$; T987R: $p = 0.448$; T987G: $p = 0.503$; T987K: $p = 0.575$, Bonferroni corrected α value = 0.0016) and only one variant was as unstable as 6-mut (T987L: $p = 0.510$, Figure 3.3A). One of the remaining variants, T987P, failed to produce any phage upon lysogen induction, and a

second variant, T987I, produced a culture with $\sim 10^{-4}$ of the normal 6-mut titer. Upon sequencing *J* from plaques that formed from the T987I lysogen, we discovered that all viable descendants had an additional mutation, indicating that the T987I variant can only produce infectious particles if it gains a “rescue” mutation while the phage genome replicates in the lysogenic state. The two rescue mutations identified were A1077V and F1122L. Although we are not able to measure the stability of the T987I variant without the rescue mutations, we can infer that the mechanism of rescue is likely stabilization, since F1122L is stabilizing in the 6-mut genotype background, and A1077V is a reversion of one of the six mutations that the stable λ ancestor evolved *en route* to the unstable 6-mut. We concluded that T987I and T987P might be catastrophically unstable, with J proteins that completely fail to fold, so we excluded these genotypes from further experiments.

3.3.3 Among the engineered library, most thermostable variants had reduced evolvability

We then measured the evolvability to gain OmpF function for each newly engineered variant in the library. Based on the results from the naturally selected thermostable variants, we predicted that T987L would be the most evolvable of the engineered variants because it had the lowest stability while still remaining viable. To test our prediction, we evolved six replicate populations of each engineered genotype, as well as six replicate populations of the 6-mut as a control, using the same experimental conditions and host strain as the initial experiment. As predicted, T987L was the most evolvable engineered genotype, with all six replicates using OmpF after just one day of evolution (Figure 3.3B). Among the six thermostable variants the fraction of replicate populations to evolve and number of days required differed widely among variants (Figure 3.3B). T987G never evolved OmpF⁺, while T987S, T987K, T987R, and T987Y

were able to evolve OmpF⁺ but fewer replicate populations evolved and required longer than the 6-mut and T987L (Figure 3.3B). A notable exception to the overall pattern was T987C, which was the only thermostable variant with high evolvability, requiring just one or two days of evolution (Figure 3.3B).

As with the initial experiment, we sequenced a single clone from each isolate that evolved OmpF⁺ on the first day that plaques were visible on the LamB knockout lawn. And consistent with the initial experiment, the time to evolve OmpF⁺ was explained by the number of mutations required. Each of the 6-mut control populations required a single mutation for OmpF⁺, from the same set of three mutations (S1011R, S1049R, and N1107K) as the initial experiment, except that one population evolved OmpF⁺ by a single mutation (T1053K) that was not previously detected as a N1107K substitute (Figure 3.3C). And T987L, the only unstable engineered variant, also evolved OmpF⁺ in all six replicates with a single mutation, either N1107K or one of the same substitutes (Figure 3.3C). Among the stable engineered variants, T987S, T987K, T987R, and T987Y required two mutations for OmpF⁺, while the variant that evolved more quickly (T987C) required only one (Figure 3.3C). We will return to potential explanations for this outlier in a subsequent section, but for now we will discuss the structural and functional roles of the two sets of mutations.

3.3.4 J structural prediction provides insight into mutation function

We wondered whether the pattern of mutation observed in the coevolution experiments might shed light on a structural mechanism behind J evolution. Perhaps evolution to use OmpF involves manipulating both the conformational flexibility of J, in addition to altering the very specific residues that make direct contact with OmpF. Recall that all variants, regardless of

thermostability, require either N1107K or a substitute mutation, whereas thermostable variants required an additional mutation, typically toward the N-terminus of the protein. We predicted that N1107K and substitutes are the OmpF contacting residues, and presumably would lie clustered on one surface of J. Under this model, the additional mutations, required only by the thermostable variants, destabilize J, and thus increase conformational flexibility. We predicted that these mutations would lie embedded within the protein. Because the J protein structure remains unsolved experimentally, we used AlphaFold, a new technology that uses machine learning to predict protein structures with remarkable accuracy (Jumper et al. 2021), to predict the structure of the reactive region of the 6-mut genotype (predictions for the ancestor and 6-mut were nearly identical, Figure 3.A.2). We modelled the protein multiple times by varying which segments of the protein to include. In the end, we found that the segment with the highest confidence score and for which all the mutations being studied were included, was a 173 amino acid segment at the C-terminal end. The problem with predicting a larger portion of the protein is that from ~amino acid 780 to 960, the model confidence is very low (40–60%) compared to the terminal region (80–90%). The J region we chose is known through biochemical assays to be the portion that binds the receptor (Wang et al. 2000), it encompasses the majority of host-range altering mutations previously reported (Meyer et al. 2012), and has been shown to have an elevated rate of evolution in nature (Maddamsetti et al. 2018). As we expected, N1107K and its substitutes all lie in loops at one surface of the protein, while the thermostabilizing mutations and putative destabilizing mutations all lie at the other end of a series of beta sheets (Figure 3.4). It is notable that this spatial separation is not simply a reflection of proximity in the peptide sequence but are likely clustered because of 3D position-function correlations. For example, S1011R and

N1107K are proximate in the folded structure despite being 96 residues apart, and the same is true for T987A and F1122L (135 residues apart).

3.3.5 Reconstructing J molecular evolution

To experimentally test this two-step model of protein functional evolution, we reconstructed the evolutionary sequence that started from four different genetic backgrounds and ended with an OmpF⁺ λ (Figure 3.2C: pop. B initiated from 6-mut, pop. C initiated from 6-mut T987A, pop. E initiated from 6-mut F1122L; and Figure 3.3C: pop. E initiated from 6-mut T987S) and measured how stability changed. We included the wild type pathway from 6-mut to 7-mut via N1107K for comparison, as we did not expect this pathway to be associated with a change in stability, and indeed there was no effect (Figure 3.5). In line with our prediction, the putative destabilizing mutations (S987L, L1122F, and S970Y) were in fact destabilizing in the background in which they occurred, although the destabilizing effect was stronger in the 6-mut F1122L and 6-mut T987S backgrounds compared to the 6-mut T987A background (Figure 3.5). Notably, the destabilizing mutation that occurred in the 6-mut F1122L background was an exact reversion of the stabilizing mutation (L1122F, Figure 3.5). As expected, the addition of the putative contact surface mutation did not alter viral particle stability in any reconstructed background (Figure 3.5).

3.3.6 A novel molecular mechanism for host-range evolvability

We propose a multiple phase model of J evolution consistent with the sum of our findings. In the first phase, ancestral J gains destabilizing mutations that alter its folding dynamics such that a portion of the expressed proteins fold into an alternative, non-native

conformer (Figure 3.1A, ancestor → 6-mut). The next phase of mutations alters the binding surface to facilitate specific surface amino acid interactions between J and OmpF (Figure 3.1A, 6-mut → 7-mut). The catch with this model is that the second phase mutation only works if it occurs in an unstable protein that expresses some fraction of a non-native conformer that places key amino acid residues in proximity to binding partners on the OmpF molecule. If the stabilizing mutations increase protein rigidity and suppress the formation of non-native conformers, then even if the N1107K mutation occurs, the residue would not be in the correct position to interact with its partner on OmpF. To test this, we engineered the surface mutation N1107K into all variants and tested for OmpF activity. As expected, both unstable backgrounds and the outlier T987C became OmpF⁺, whereas none of the remaining stable backgrounds yielded strong OmpF use (S1 Text, Figure 3.A.3, Table 3.A.1). Counter to our expectations, though, editing in N1107K into two of the stable backgrounds, T987A and T987Y, yielded very weak activity on OmpF. The edited genotypes were unable to produce plaques on a lawn of *lamB*⁻ cells; however, they produced faint clearing at high phage concentrations spotted on *lamB*⁻ lawns (S1 Text and Figure 3.A.3). We verified that faint clearings on plates corresponded to weak growth on *lamB*⁻ cells in liquid culture (S1 Text and Figure 3.A.4). This result suggests that the steps in the molecular model are not as discrete as anticipated, yet the path to strong OmpF use does involve both destabilizing mutations and binding surface alterations.

3.3.7 T987C: The thermostable, yet evolvable variant

While our model was able to correctly predict the evolvability of six engineered strains, it failed to predict the high evolvability of T987C, the only variant to be both highly thermostable

and highly evolvable. To first verify that this outlier did not result from additional mutations outside *J* that might have occurred during library creation and affected OmpF⁺ evolution, we generated an independently engineered T987C variant and re-measured its evolvability with 12 replicate populations, along with 12 replicate populations of 6-mut as a control. All 12 populations of T987C evolved OmpF⁺ on the first day (Figure 3.A.5), consistent with the initial experiment, even slightly faster, ruling out this possibility. At present, our best hypothesis to reconcile T987C with our model is that there may be rare genotypes that are thermostable but produce sufficient copies of the non-native, OmpF-binding *J* to complete infection, bypassing an extra mutation that yields non-native conformers via destabilization. If a destabilized *J* protein is not strictly required for evolving OmpF-binding, as suggested by the T987C variant, this raises the question of whether there may be mutational pathways available to the ancestral λ genotype that do not include destabilizing mutations before receiving the surface contact mutation. Such pathways, if they exist, should be highly favored because intermediate genotypes would not pay the cost of low thermostability. One explanation for why λ might have originally evolved via destabilization despite the cost of instability is that the simultaneous optimization of evolvability and thermostability instigates a trade-off with other traits that determine fitness. Previous studies on phages have shown fitness tradeoffs between thermostability and growth rate (Dessau et al. 2012; Singhal et al. 2017) that could close off pathways that include evolving stabilizing mutations in competitive environments, making variants like 6-mut T987C accessible via engineering, but not through adaptive evolution. To test this idea, we measured the growth rates of all variants on the same host as was used in the evolution experiment, and T987C had the lowest average growth rate of any of the variants, significantly lower than that of its 6-mut

progenitor (Figure 3.6). This result suggests that the two-phase model of molecular innovation outlined above emerged from a combination of biophysical and adaptive constraints.

3.3.8 Evaluating alternative models of protein evolution

Our hypothesis for why some variants required two mutations to evolve OmpF⁺ is that they required an extra destabilizing mutation to restore evolvability. An alternative possibility is that the destabilizing effect is coincidental, and the true function of the extra mutation is to compensate for fitness costs incurred by the stabilizing mutations in the 6-mut background. Recall that this is unlikely since most of the less evolvable stabilized variants had growth rates greater than or equal to that of 6-mut (Figure 3.6).

Along similar lines of thinking, perhaps the extra mutation is required to ameliorate a genetic incompatibility between the stabilizing mutations and the mutations found in the contact surface. This could occur if stabilizing mutations and surface mutations are beneficial alone but produce a non-functional protein when they co-occur in the same genotype. To rule out this possibility, we measured productivity of lysogens for the naturally selected thermostable variants (T987A, F1122L) with and without N1107K and found no difference in titer (Figure 3.A.6) indicating that there is not a genetic incompatibility (Meyer et al. 2016).

Another possibility is that the stabilizing mutations interfere specifically with the very last step in OmpF⁺ evolution (the N1107K mutation), but had they been introduced in a background further removed from OmpF⁺, they would have enhanced evolution. To test this hypothesis, we generated a new background, a 5-mut created by reverting another 7-mut mutation by changing the amino acid at codon 1012 from G back to S. We then edited in the T987A stabilizing mutation and confirmed that it does stabilize the 5-mut as expected (Figure

3.7A). To measure evolvability, we ran a 10-day coevolution experiment and found that all six populations of the unmodified 5-mut genotype evolved in the first two days of the experiment, while only one out of six populations of the stabilized 5-mut evolved OmpF⁺, requiring 5 days (Figure 3.7B). Under the conventional model of protein stability and evolvability, stabilizing the J protein should have increased the J protein's capacity to search for adaptive mutations. This effect should have been enhanced when starting further from the innovation; however, we found the opposite. The difference in evolvability between the unmodified 5-mut and the stabilized 5-mut was equal, if not slightly greater, than the difference in evolvability between the unmodified 6-mut and the stabilized 6-mut.

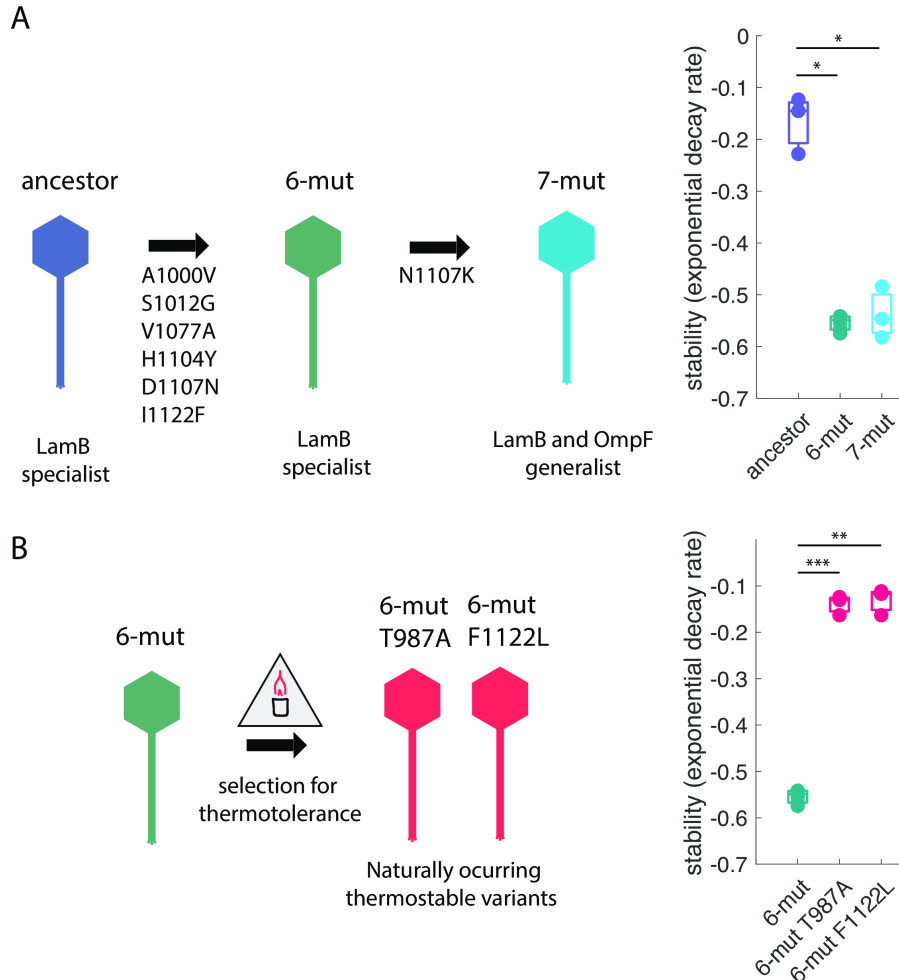


Figure 3.1 - Experimental system and selection for thermostability. A: To generate our experimental system, we began with an unstable OmpF⁺ genotype (7-mut) and edited out a critical mutation (N1107K), creating an OmpF⁻ genotype that remained unstable and was on the verge of evolving OmpF⁺. Stability plots: n = 6 replicates for 6-mut, n = 3 replicates for ancestor, and n = 3 replicates for 7-mut. Comparisons of 6-mut and 7-mut to ancestor were made using paired t-tests corrected for multiple comparisons by Bonferroni method (ancestral λ to 6-mut: $p = 8.55 \times 10^{-4}$; ancestor to 7-mut: $p = 9.80 \times 10^{-4}$). Bonferroni corrected significance thresholds: ns: $p > 0.0024$, *: $p < 0.0024$, **: $p < 0.00024$, ***: $p < 2.4 \times 10^{-5}$. B: We then selected the 6-mut for enhanced thermostability, generating two naturally evolved thermostable genotypes: T987A and F1122L. Stability plots: n = 6 replicates for 6-mut, n = 3 replicates for T987A, and n = 3 replicates for F1122L. Comparisons of 6-mut and 7-mut to ancestor were made using paired t-tests corrected for multiple comparisons by Bonferroni method (6-mut to 6-mut T987A: $p = 1.40 \times 10^{-5}$; 6-mut to 6-mut F1122L: $p = 2.73 \times 10^{-5}$). Bonferroni corrected significance thresholds: ns: $p > 0.0024$, *: $p < 0.0024$, **: $p < 0.00024$, ***: $p < 2.4 \times 10^{-5}$.

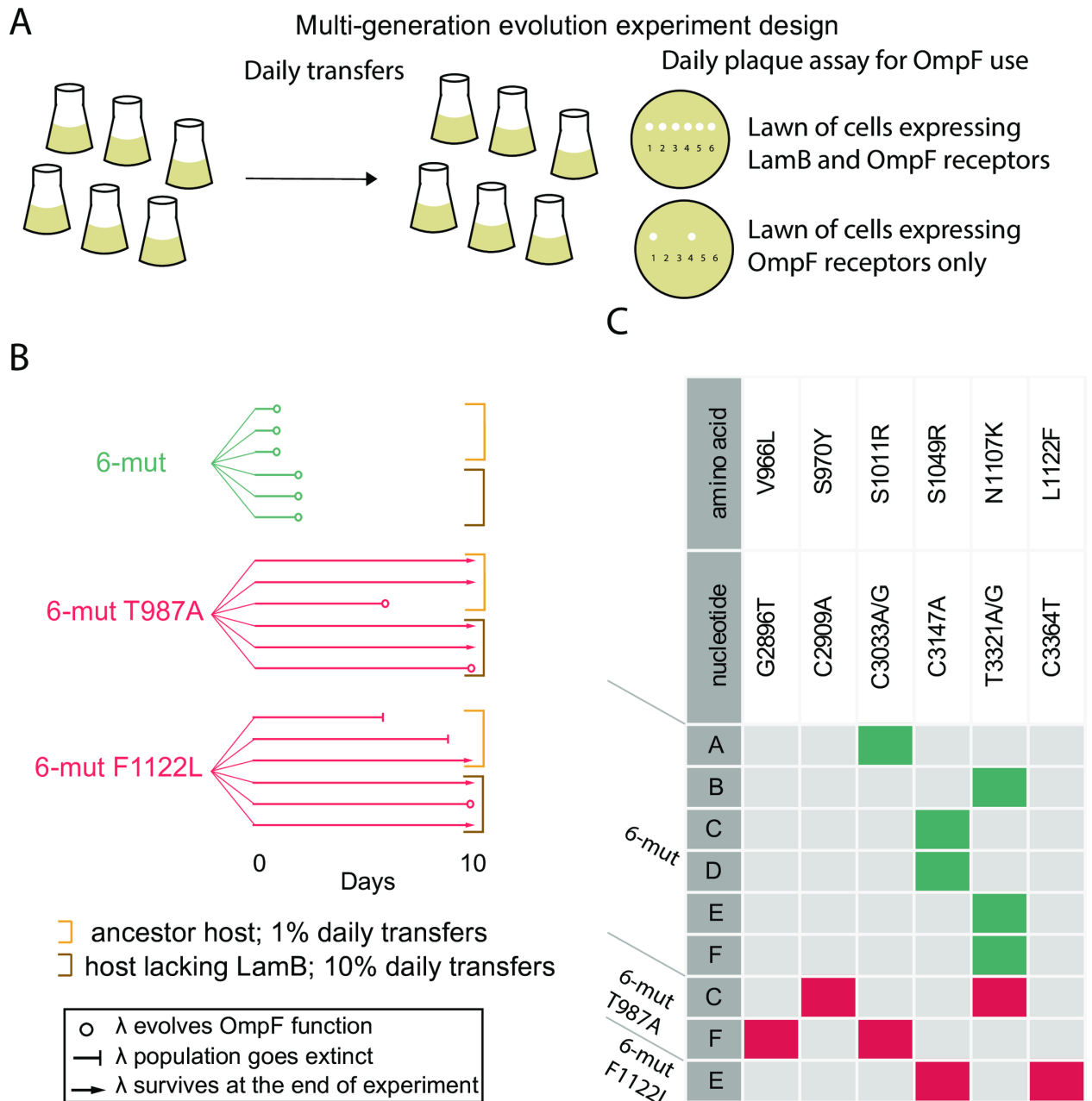


Figure 3.2 - Evolution experiment with naturally occurring thermostable genotypes. A: A ten-day evolution experiment was performed to assess OmpF-use evolvability. B: Evolutionary trajectories of six replicate populations of each starting genotype are denoted by six parallel lines. When λ evolves the innovation or goes extinct is indicated by symbols indicated in legend. Brackets surround replicates from two different ways of running the coevolution, which did not appear to impact whether OmpF⁺ evolved. C: A single plaque from each replicate population was sequenced on the first day OmpF⁺ was detected. Mutations are indicated along the top. Boxes with colored fill indicate that the amino acid change occurred in an isolate. The fill color indicates the stability of the genotypic background in which the mutation evolved (teal = unstable, red = thermostable.) Population IDs are indicated by letters.

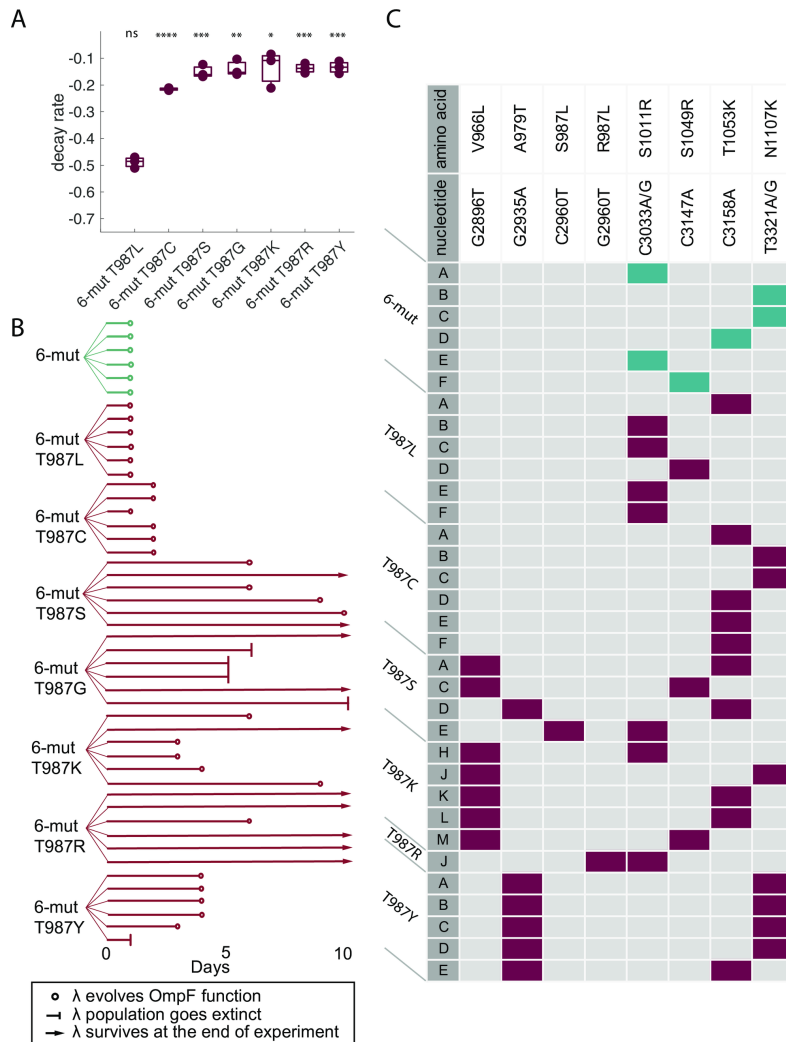


Figure 3.3 - Evolution experiment with engineered library genotypes. We generated additional variants of 6-mut differing only at the amino acid at position 987 in J. A: Most variants were more thermostable than 6-mut, but T987L remained as unstable as 6-mut. N = 3 replicates per genotype. Comparisons to 6-mut decay rate were made using paired t-tests corrected for multiple comparisons by the Bonferroni method (6-mut T987L: $p = 0.510$; 6-mut T987C: $p = 7.58 \times 10^{-8}$, 6-mut T987S: $p = 1.02 \times 10^{-5}$, 6-mut T987G: $p = 1.26 \times 10^{-4}$, 6-mut T987K: $p = 8.55 \times 10^{-4}$, 6-mut T987R: 2.96×10^{-6} , 6-mut T987Y: $p = 7.26 \times 10^{-6}$.) B: We measured the evolvability of OmpF⁺ in each variant using a nearly identical evolution experiment as in Figure 3.2A and 2B. C: A single plaque from each replicate population was sequenced on the first day OmpF⁺ was detected. Mutations are indicated along the top. Boxes with colored fill indicate that the amino acid change occurred in an isolate. The fill color indicates the stability of the genotypic background in which the mutation evolved (teal = 6-mut, dark red = engineered codon 987 variants.) Population IDs are indicated by letters. Asterisks along the lines indicate significant differences in decay rate between the genotypes connected by the line (Bonferroni adjusted significance: ns: $P > 0.05$, *: $P < 0.05$, **: $P < 0.01$, ***: $P < 0.001$, ****: $P < 0.0001$).

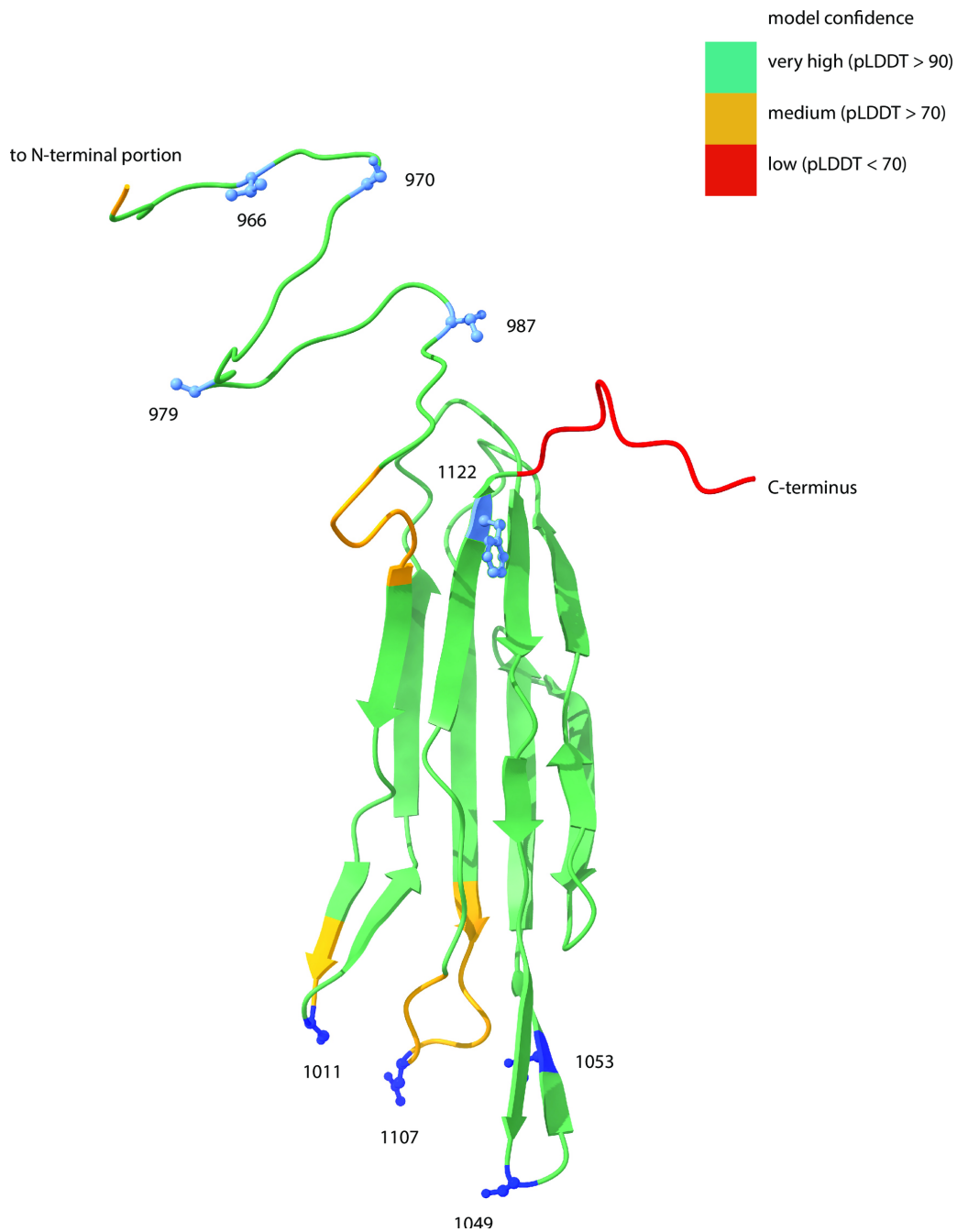
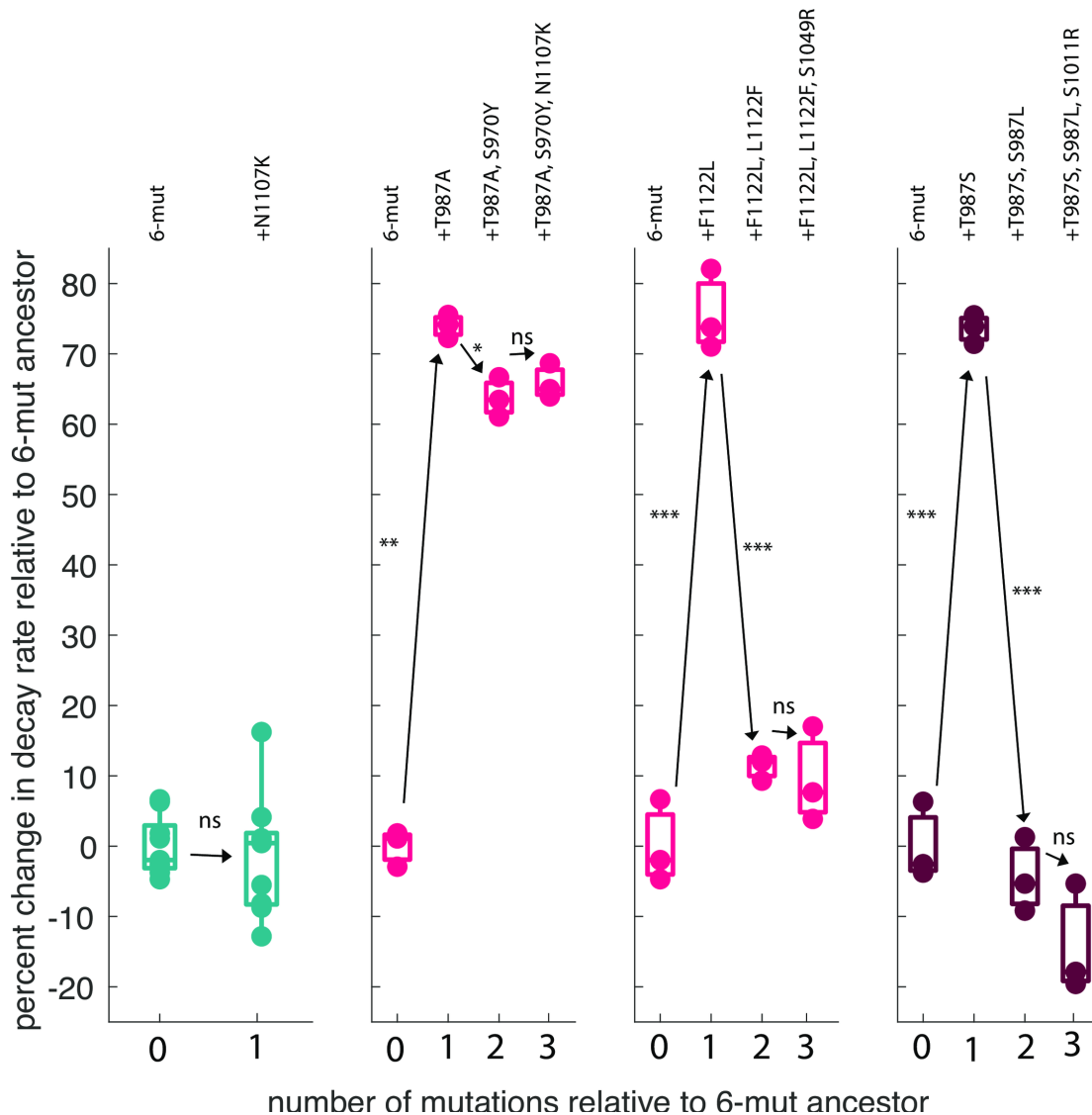


Figure 3.4 - Mapping thermostabilizing and putative destabilizing mutations on a structural prediction of J's reactive region. We used AlphaFold (Jumper et al. 2021) to predict the structure of the reactive region of J (amino acids 960–1132). This structure corresponds to the 6-mut genotype, but the 6-mut prediction was nearly identical to the ancestor prediction (Figure 3.A.2). We then mapped the surface binding mutations (dark blue residues) and the thermostabilizing and destabilizing mutations (light blue residues) onto the structure. Coloration of the backbone indicates model confidence.



number of mutations relative to 6-mut ancestor

Figure 3.5 -Trajectories of evolution of OmpF⁺ in unmodified and stabilized backgrounds. Evolutionary trajectories of selected isolates from the replay experiment, reconstructed with genomic engineering. In the 6-mut (teal), a single mutation led to an OmpF⁺ genotype. In three stabilized backgrounds (naturally evolved = red, engineered = dark red), an additional destabilizing mutation was required as a steppingstone to the OmpF⁺ genotype. Asterisks along the lines indicate significant differences in decay rate between the genotypes connected by the line, as compared by paired t-tests corrected for multiple comparisons using the Bonferroni method. First panel: 6-mut vs. 6-mut N1107K: $p = 0.678$; Second panel: 6-mut vs. 6-mut T987A: $p = 1.88 \times 10^{-6}$; 6-mut T987A vs. 6-mut T987A S970Y: $p = 0.0055$; 6-mut T987A S970Y vs. 6-mut T987A S970Y N1107K: $p = 0.3818$; Third panel: 6-mut vs 6-mut F1122L: $p = 9.25 \times 10^{-5}$; 6-mut F1122L vs. 6-mut F1122L L122F: $p = 5.13 \times 10^{-5}$; 6-mut F1122L L122F vs. 6-mut F1122L L122F S1049R: $p = 0.667$; Fourth panel: 6-mut vs. 6-mut T987S: $p = 2.65 \times 10^{-5}$; 6-mut T987S vs. 6-mut T987S T987L: $p = 1.83 \times 10^{-5}$; 6-mut T987S T987L vs. 6-mut T987S T987L S1011R: $p = 0.143$. (Bonferroni adjusted significance: *: $p < 0.0167$, **: $p < 0.00167$, ***: $p < 0.000167$, ****: $p < 1.67 \times 10^{-5}$).

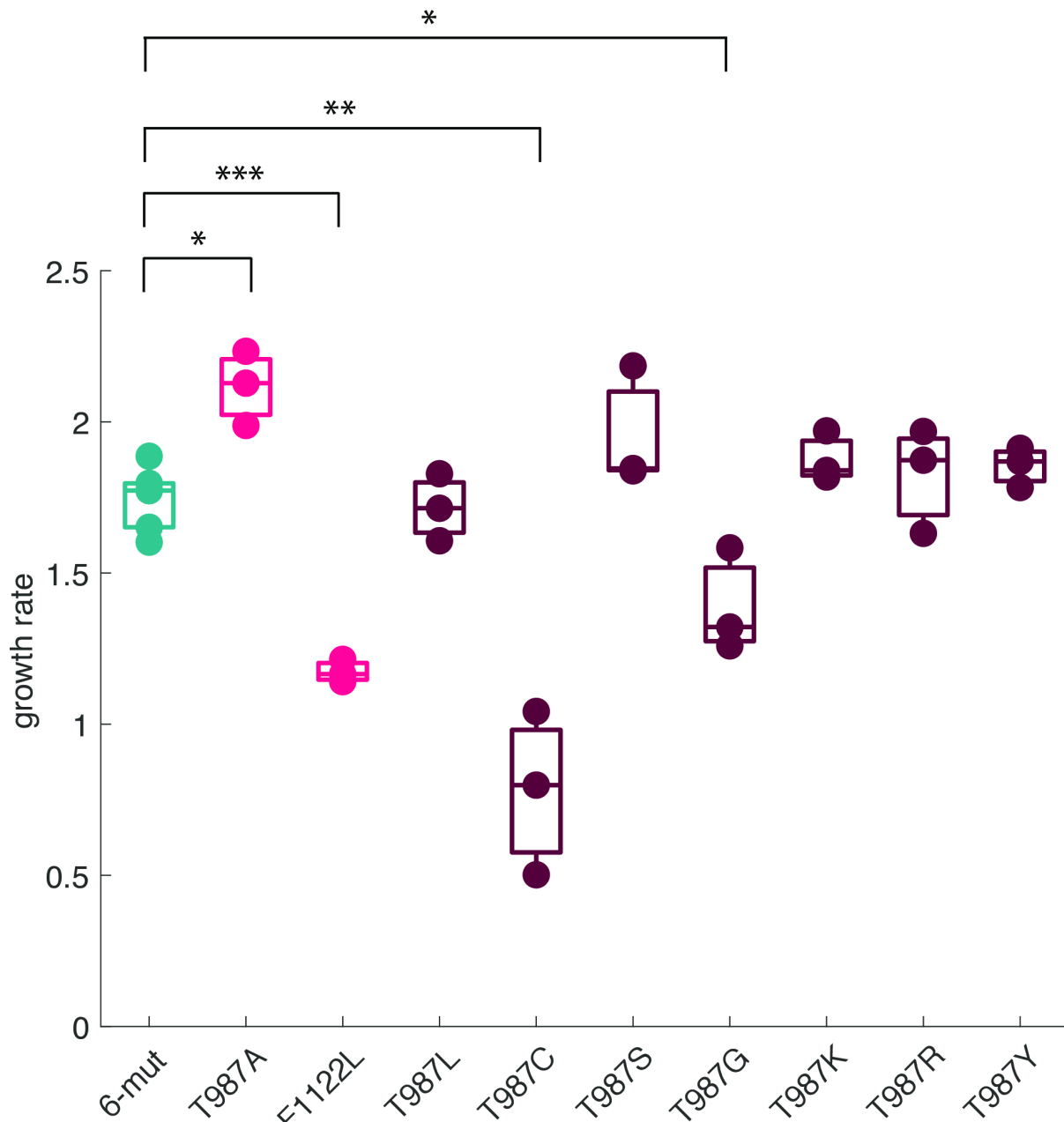


Figure 3.6 - Growth rates of naturally occurring thermostable genotypes and engineered library genotypes. Growth rates on REL606 in M9 glucose + MgSO₄ media over four hours. Comparisons to 6-mut growth rate were made using paired t-tests corrected for multiple comparisons using the Bonferroni method, N = 3 per genotype. Significantly higher: T987A: p = 0.002; significantly lower: F1122L: p = 4.12x10⁻⁵; T987G: p = 0.0052; no difference: T987L: p = 0.695; T987S: p = 0.068; T987K: p = 0.105; T987R: p = 0.419; T987Y: tstat = p = 0.150 Bonferroni adjusted significance: ns: P > 0.0056, *: P < 0.0056, **: P < 0.00056, ***: P < 0.000056.

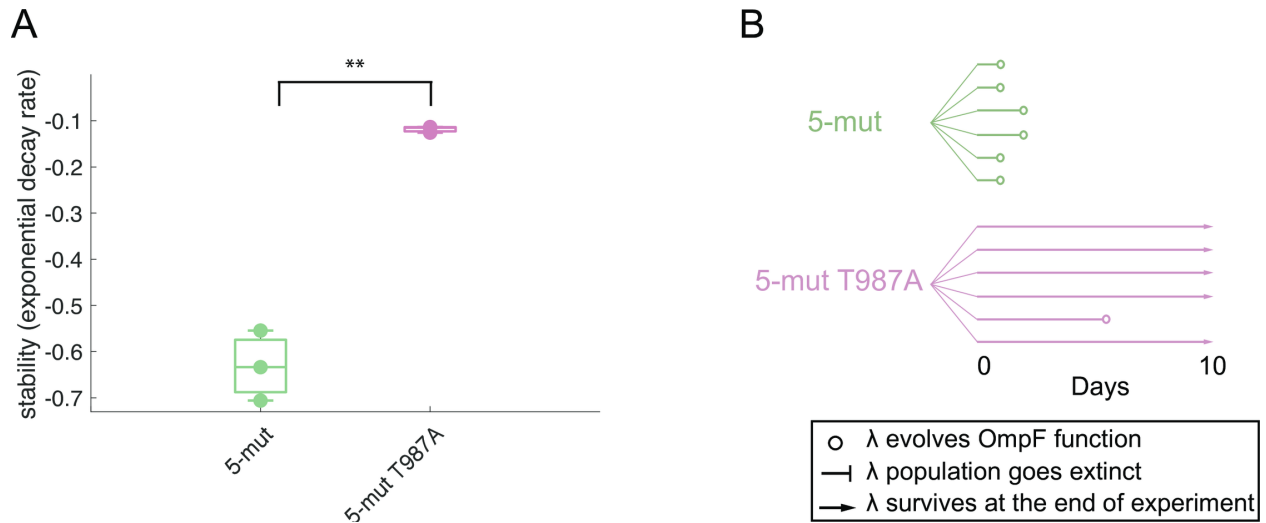


Figure 3.7 - Stability and evolvability of 5-mut and 5-mut T987A stable variant. A: In a background that is further away from evolving OmpF⁺ (two mutations reverted compared to one mutation in 6-mut), the T987A mutation is still stabilizing. B: In the 5-mut background, T987A reduces evolvability to an even greater extent. Two sample t-test for unequal variance, N = 3 per genotype, **: P = 0.0069.

3.4 Discussion

We found that, on average, increasing the thermostability of λ 's host-recognition protein resulted in the loss of evolvability, which was restored when thermostable genotypes evolved destabilizing mutations. This pattern is the opposite of the prediction that could be made based on the preponderance of current literature—that inserting thermostabilizing mutations should enhance evolvability. This current consensus is rooted in an observed trade-off between thermostability and function in proteins: mutations that confer new functions often destabilize the native fold as a side effect (Fasan et al. 2008; Studer et al. 2014). The stability-function trade-off has presented a particular challenge for directed evolution, which requires multiple rounds of mutagenesis and selection (Stimple et al. 2020). One solution that is thought to be relatively drawback-free has been to select for naturally occurring stabilizing mutations or deliberately engineer known stabilizing mutations into enzymes either before or between cycles of directed evolution (Bloom and Arnold 2009; Socha and Tokuriki 2013; Stimple et al. 2020). The robustness conferred by extra thermostability preserves the native fold and function while allowing for new functions to be explored.

In the case of J, mutations that led to gain-of-function on the OmpF receptor were destabilizing, indicating that the J protein is also subject to the stability-function trade-off. However, inserting thermostabilizing mutations into an unstable J genotype on its way to evolving expanded host-range actually slowed or completely stalled evolution. This indicates that J discovered an alternative solution to the stability-function trade-off. Based on our results and prior work (Petrie et al. 2018), we suggest that unstable genotypes circumvent the trade-off by producing multiple conformers from a single peptide sequence, at least one of which retains preferential binding to the ancestral LamB receptor (Petrie et al. 2018). If a genotype produces a

high fraction of thermostable, LamB-binding particles, the less stable, non-ground state conformers would be released from the selective pressure to maintain LamB function and free to optimize latent binding interactions with new partners. We propose that it is specifically the lack of structural rigidity in the 6-mut genotype and variant T987L that enhanced the evolvability of those genotypes.

Why might some proteins solve the stability-function trade-off with robustness and others with conformational flexibility, as in the case of λ ? The lack of a universal pattern suggests that the relationship between thermostability and evolvability is multifaceted. We can only speculate on the true nature of this relationship, but our leading idea is that each protein exists on a spectrum of thermostability, and extremes at either end can inhibit evolvability. Indeed, we found that one variant from our engineered library, T987I, was likely too unstable to fold, requiring a rescue thermostabilizing mutation to produce infectious particles. Unfortunately, none of the variants fell between T987I and T987L in thermostability, so we cannot precisely identify the inflection point where instability is so severe that evolvability is inhibited. Similarly, it is also unclear at what point increasing thermostability might begin to inhibit evolvability. Additionally, prior work on this question has focused primarily on enzyme proteins, and it is not entirely surprising that the paradigm that emerged from those studies might not apply to viral attachment proteins, especially if viral proteins generally have different properties from cellular proteins (Tokuriki et al. 2009). Given these unknowns, it is not surprising that different studies have reported both positive and negative relationships between thermostability and evolvability. Even within a single protein, we did not find a universal pattern, as evidenced by the clear exception of the T987C variant. The balance between robustness and conformational flexibility may also depend upon the extent of structural change required to achieve novel functionality (i.e.

localized amino acid substitutions vs. conformation changes). It is possible that for some proteins, single amino acid substitutions that subtly alter active sites or binding surfaces may be sufficient to confer new functions, whereas in other proteins, more global shape changes are required. We suspect that J's evolution from OmpF⁻ to OmpF⁺ falls into the latter category, and single mutations such as N1107K, do not confer OmpF⁺ unless they occur in a background with conformational flexibility. Under this model, for most of the thermostable variants, the increased structural rigidity caused by stabilizing mutations (Rathi et al. 2015) would stifle evolvability by locking J into a single conformation. This interpretation requires the caveat that we did not perform structural studies on J, but the predicted structure of J's reactive region is consistent with our model.

While thermostability explained much of the variation in evolvability, it was not fully predictive. Two mutations, T987L and T987C, differed greatly in thermostability (Figure 3.3A), yet had nearly identical evolvability, both evolving OmpF⁺ in all populations in 1–2 days with one mutation (Figure 3.3B and 3C). Similarly, among a cluster of engineered variants that had indistinguishable thermostabilities (T987S, T987G, T987K, T987R, T987Y, Figure 3.3A), we observed a breadth of responses in evolvability, as measured by the number of days required to evolve OmpF⁺, the fraction of populations that eventually evolved, and the number of unique 2-mutation pathways taken during evolution (Figure 3.3B and 3C). There are two insights that can be drawn from these observations. First, the ability to evolve OmpF⁺ did not hinge on the specific residue at position 987; instead, two residues besides the wild type could be interchanged without impacting evolvability. And second, viruses differing only at a single residue in a single protein were surprisingly variable in their capacity for host-range expansion, and differences in key traits like thermostability and growth rate were only partially predictive.

That our library variants were genetically identical except for the amino acid at a single site was both a strength and a limitation of our study. This design allowed us to pinpoint and precisely manipulate the genetic basis for stability and host-range evolvability in λ ; however, focusing in on a single site might result in non-independent effects, and conclusions drawn from a single protein might reflect idiosyncrasies of its particular evolutionary history.

In this study, we evolved engineered viruses under highly controlled laboratory conditions to examine how thermostability affects host-range evolvability, allowing us to precisely manipulate our variable of interest. Such a design cannot, and was not intended to, recapitulate the complexity of nature. However, prior work indicates that the evolution of λ 's J gene in the lab may be uniquely informative for understanding evolution in nature due to a remarkable parallelism between sites under selection in the laboratory and in natural populations (Maddamsetti et al. 2018). The entire reactive region of J, encompassing the vast majority of mutations that evolve in during OmpF gain-of-function experiments, was shown to be a hotspot for molecular evolution in natural J homologs, and amino acids 1012 and 1107 had particularly high rates of evolution (Maddamsetti et al. 2018). Whether this parallelism between laboratory and nature extends to stabilizing mutations is unknown, but our simple selection identified mutations that affected stability at a physiologically relevant temperature (37°C), and it is plausible that the same mutations might aid virus survival under natural conditions. That said, our results are based upon a relatively small number of genotypes in a single experimental system, and additional work will be required to assess the generality of the pattern and its use in predicting host-range expansions.

Understanding the predictors of evolvability has important applications, such as providing insights into the characteristics that potentiate viral host shifts. There has been

considerable interest in conducting surveillance on viral populations with the goal of detecting variants with emergence potential before they first spill over into human populations (Flanagan et al. 2012). This type of surveillance would be aided by information about the characteristics of viruses that influence the likelihood of host-range evolution. Our results were derived from a study of bacteria-infecting virus rather than an animal virus, and further work is warranted to investigate the role of viral particle instability in animal virus host-range evolvability. Viral instability can be rapidly assessed in the laboratory, or if only genome sequences are available for given viruses, there are bioinformatic tools capable of predicting protein stability (Pucci et al. 2017). This work was completed during the 2020–2021 coronavirus pandemic, raising the question of whether instability in SARS-CoV-2 host recognition protein (S) aided its transition to humans. At the time of publication this had not been tested; however, a mutation that evolved during the pandemic stabilizes the SARS-CoV-2 S protein (Zhang et al. 2020) and also increases viral titer (Li and Luo 2020), suggesting that the variant that jumped to humans was relatively unstable.

3.5 Materials and methods

3.5.1 Media

We used the following recipes to prepare media: LB Lennox broth: 20 g LB Lennox powder per liter of water. LBM9: 20 g tryptone, 10 g yeast extract per liter of water supplemented with 47.7 mM disodium phosphate heptahydrate, 22.0 mM potassium phosphate monobasic, 18.7 mM ammonium chloride, 8.6 mM sodium chloride, 0.2 mM calcium chloride and 10 mM magnesium sulfate. M9 glucose: 47.7 mM disodium phosphate heptahydrate, 22.0 mM potassium phosphate monobasic, 18.7 mM ammonium chloride, 8.6 mM sodium chloride, 0.2 mM calcium chloride, 10 mM magnesium sulfate, 5.55 mM glucose, 7.54 μ M thiamine, and

0.02% LB. M9 minimal glycerol: 47.7 mM disodium phosphate heptahydrate, 22.0 mM potassium phosphate monobasic, 18.7 mM ammonium chloride, 8.6 mM sodium chloride, 0.2 mM calcium chloride, 10 mM magnesium sulfate, 7.54 uM thiamine, 17.1 mM glycerol, 1.22 mM deoxygalactose, and 0.02% LB. LB agar: 10 g tryptone, 5 g yeast extract, 5 g sodium chloride 16 g agar per liter of water. MacConkey agar galactose: 40 g MacConkey base and 10 g galactose per 1 liter of water. Soft agar: 10 g tryptone, 1 g yeast extract, 8 g sodium chloride 7 g agar per 1 liter of water, supplemented with a final concentration of 2 mM calcium chloride 5.55 mM glucose, and 10 mM magnesium sulfate.

3.5.2 λ Phage strains

The phage used in this study were all derived from the 7-mut λ strain described in (Petrie et al. 2018). The 7-mut is a derivative of the cI857 λ lysogen integrated into the HWEC106 genome. HWEC106 has two features that allow genome editing: a deletion in the *mutS* mismatch repair gene and the pKD46 λ red recombineering plasmid (Datsenko and Wanner 2000). Plasmids were maintained by addition of 50 μ g/mL ampicillin to lysogen cultures. Phage particles were produced from lysogen stocks as described in the section “induction of lysogens by heat shock.”

3.5.3 Host *E. coli* strains

For the coevolution replay, we used REL606 and its *malT*⁻ derivative, LR01 (Chaudhry et al. 2018), which was isolated from a previous coevolution experiment (Meyer et al. 2012) and has a 25-base duplication in *malT*, which results in the loss of LamB expression except in mutants that revert the duplication. For density measurements during the replay, we used the

Keio knockout collection parental wild type strain BW25113 (Baba et al. 2006). For detection of OmpF⁺ mutants during the replay, we used the *lamB*⁻ strain (JW3996) from the Keio collection (Baba et al. 2006). For all other experiments, we used Keio knockout collection strains (wild type BW25113 and *lamB*⁻JW3996) for all culturing and plating. Throughout this manuscript we refer to BW25113 as “WT”.

3.5.4 Sanger sequencing

We sequenced the region of the *J* gene known to interact with the LamB receptor (approximately position 2650 to 3399 of 3399 total bp) to identify stabilizing mutations, confirm genetic edits, and identify mutations in genotypes that evolved to be OmpF⁺ in the replay experiment. We submitted unpurified PCR products (primers: Forward 5' CCT GCG GGC GGT TTT GTC ATT TA; Reverse 5' CGC ATC GTT CAC CTC TCA CT) to Genewiz La Jolla, CA, for sequencing with the reverse primer. We aligned sequences using Unipro UGENE v1.31.1 (Okonechnikov et al. 2012).

3.5.5 MAGE to produce 6-mut lysogen

We used MAGE (Multiplexed Automated Genome Engineering) (Wang et al. 2009; Wang and Church 2011) to edit a single reversion into the *J* gene of the 7-mut OmpF⁺ lysogen. We reversed the mutation at position 3321 from A back to the ancestral nucleotide, T. We call this new genotype ‘6-mut’.

3.5.6 Induction of lysogens by heat shock

The night before induction, we grew lysogens in LB at 30°C shaking overnight. The next morning, we inoculated 140 μL overnight culture into 4 mL LBM9 supplemented with 40 μL MgSO_4 , grew at 30°C shaking for 2 hours, heat shocked at 42°C for 15 minutes, then incubated at 37°C shaking until lysates became clear (90 minutes).

3.5.7 Selection for thermostabilizing mutations

We generated 6-mut phage by inducing the 6-mut lysogen, as described in the section ‘Induction of lysogens by heat shock’. The lysate was filtered through a 0.22 μM syringe filter, diluted in 0.8% NaCl, and plated with 100 μL WT cells infused in 4 mL soft agar. We then picked 6 plaques from the overlay plate and incubated them with 100 μL WT in 4 mL LBM9 and 40 μL MgSO_4 at 37°C shaking overnight. The next morning, plaque cultures were centrifuged and filtered through 0.22 μM syringe filters to remove cells. Phage were then incubated at 55°C, a temperature much higher than their normal growth temperature of 37°C, for 6 hours. After incubation, 3.3 mL from each phage stock were plated with 250 mL WT infused in 10 mL soft agar. The phage that were still able to form plaques after the 55°C incubation were selected as candidate thermostable mutants and were further screened for enhanced stability as follows. Plaques were picked into 0.8% NaCl and incubated with 100 μL WT cells in 4 mL LBM9 and 40 μL MgSO_4 at 37°C shaking overnight. The next morning, we added 100 μL chloroform to each culture and centrifuged at 357 \times g for 20 min. As a secondary screen to verify the increased heat tolerance of the putative thermostable mutants, we incubated phage 58°C for 1 hour and plated densities before and after incubation. Out of four candidates, one exhibited increased thermostability in the secondary screen and was isolated for sequencing. We identified a single

base pair substitution at position 2959 (codon 987) of *J* resulting in an amino acid change from threonine to alanine.

We identified a second stabilizing mutation using a very similar protocol with minor procedural variations. Before heat selection, plaques were grown up for 7 hours instead of overnight, and after heat selection, 4 mL were plated instead of 3.3 mL. Survivors of heat selection were picked into NaCl and re-plated on WT lawns, then a plaque was picked and grown up for 6 hours for the secondary stability screen (whereas for the isolation of the first stabilizing mutation, the plaques were directly grown up instead of being re-plated first). The secondary screen for stability was conducted at 56°C for 1 hour (instead of 58°C for 1 hour). Using this slightly modified protocol we identified the stabilizing mutation at position 3364 in *J* (codon 1122) changing the amino acid from phenylalanine to leucine.

3.5.8 CoS-MAGE to insert stabilizing mutations into 6-mut lysogen

To rule out the possibility that mutations outside *J* could be driving the increased stability, we edited each stabilizing mutation into a lysogenic strain of 6-mut λ that had not undergone the selection for increased thermostability. We designed oligos for each stabilizing mutation and used CoS-MAGE (Wang et al. 2012), a genomic editing tool, to insert each mutation into the 6-mut lysogen. To increase the likelihood of picking successful colonies, we co-transformed with an oligo containing a nonsense mutation in the *galK* gene (*galK*⁻ oligo, Table 3.A.2). Successful recombination with this oligo resulted in a color change when cells were plated on MacConkey agar plates containing galactose as the only sugar source. The oligos for stabilizing mutations and the *galK* oligos were both used at a final concentration of 5 μ M.

We also attempted to edit in a control mutation in codon 987 that we predicted would be neutral with respect to stability. We intended the control mutation to be conservative with respect to the chemistry of the amino acid side chain, so we chose threonine to serine (T987S) because both have polar, uncharged side chains. However, we found that this mutation unexpectedly had a stabilizing effect, similar to the T987A mutation. We proceeded with all three mutations; two uncovered through selection and the third through intentional engineering.

3.5.9 CoS-MAGE to create a variant library at codon 987

To generate additional λ 's varying at the amino acid at position 987 in the J protein, we used an oligo pool containing random amino acids at nucleotide positions 2959, 2960, 2961, the codon for amino acid 987. We then carried out CoS-MAGE (Wang et al. 2012) with our oligo pool plus the *galK*⁻oligo (Table S2) by the same protocol as described in the previous section, using the 6-mut lysogen as the recipient strain. To enrich the population for successful recombinants, we selected for conversion of the *galK* state by growing up the recovered populations in M9 minimal glycerol media containing deoxygalactose as the selective agent. After enriching the population, we plated cultures on MacConkey agar plates containing galactose and picked white colonies (nonfunctional *galK*) for sequencing. Approximately one third of the colonies had mutations at codon 987, and we grew these up overnight to create lysogen stocks. All library strains were stored in the lysogenic state at -80°C, and lysogens were induced to create fresh phage stocks for all further experiments.

3.5.10 Decay assays

Decay assays were performed identically for all genotypes, and three replicates were included for each genotype. On the day of the assay, phage were induced from lysogens as described in the section ‘Induction of lysogens by heat shock’ with three replicates per genotype. Lysates were then filtered through 0.22 μM syringe filters. Filtered lysates were immediately serially diluted in LBM9 and incubated in glass culture tubes in a warm water bath at 37°C for 6 hours, quantifying surviving phage at $t = 0$ (just before starting the incubation) and $t = 6$ hours. Measurements were taken from repeated sampling of the same tube. Quantification was done by overlay plating 4 mL molten soft agar with 100 uL of WT bacteria + 10 or 100 uL phage lysate over an LB plate.

To measure stability, we calculated overall exponential decay rates from $t = 0$ to $t = 6$ hours (d_{06}). To compare decay rates among genotypes, we used one-way ANOVAs followed by post-hoc Tukey HSD test ($n = 3$ for each genotype).

We were not able to include all genotypes in a single experiment on the same day, so we carried out experiments in a blocked design such that each set of measurements included 6-mut replicates, as a way of gauging how much measurements varied across days. For the majority of blocks, no day effects were detected; however, there were a few cases where the measurements for 6-mut stability slightly deviated. To ensure that day-effects were not driving significance in comparisons between genotypes, statistical comparisons of each genotype to 6-mut were made only between replicates measured on the same day (paired t-tests corrected by the Bonferroni method). For comparisons to ancestral λ , it was not possible to constrain comparisons to same-day measurements because ancestral λ was only measured on one day. To address this, we computed relative decay rates for each genotype by dividing the raw decay rate by the 6-mut

decay rate measured on the same day. We then ran another set of comparisons between the ancestral λ and all other genotypes using the relative decay rates, and the same pairs were significant as with the raw decay rates, even with the stringent Bonferroni adjusted α value of 0.0016. Since the day effects were small compared to the differences between variants, we do not believe that the day-effects substantively affected our interpretation of the results. Both raw and relative decay rates are reported in Table 3.A.3.

3.5.11 Coevolution replay experiment

For the experiment including the naturally evolved stabilizing mutations and the engineered variant T987S:

On the day preceding the first day of the experiment, we induced phage from each genotype from their corresponding lysogens, as described in the section ‘Induction of lysogens by heat shock’. We then filtered lysates through 0.22 μM syringe filters and quantified lysate densities by serially diluting in the replay experiment media, M9 glucose, and spotting 2 μL on a lawn of WT cells. The next day, designated ‘Day 1’ of the replay, we inoculated 6 flasks per starting genotype with 9.7 mL M9glucose media, 100 μL MgSO₄, 100 μL host bacteria, and phage. Each genotype was diluted to equalize the number of initial particles across genotypes. We added approximately 10^6 phage to each initial flask, which corresponded to 50 μL of 6-mut and 5 μL of 6-mut T987A, 6-mut T987S, and 6-mut F1122L. Exact numbers of phage added were 1.8×10^6 for 6-mut, 4.9×10^6 for 6-mut T987A, 4.6×10^6 for 6-mut T987S, and 2.5×10^6 for 6-mut F1122L. For each phage genotype half of the replicates (3 out of 6 flasks) were initiated with REL606 wild type hosts and the other half were initiated with LRO1 hosts (a *malt*⁻ derivative of REL606). Flasks were incubated for 24 hours at 37°C shaking. On day 1 of the experiment, we

verified that no populations started the experiment as $OmpF^+$ by plating $\sim 6x$ the volume of phage added to experimental flasks with $lamB^-$ bacteria. No $OmpF^+$ plaques were detected. Daily transfers were made (every 24 hours) from the previous day flask to a new flask with fresh M9 glucose and $MgSO_4$. Because the $malT^-$ is a more challenging host for phage to grow on compared to REL606, we transferred 10% of flask contents for $malT^-$ flasks and 1% of flask contents for REL606 flasks. After completing each daily transfer, we sampled 1 mL from the previous day's flask and processed the samples as follows. To each 1 mL sample we added 40 μL chloroform and centrifuged at $3,214\times g$ for 10 minutes. Then we spotted 2 μL phage from each sample onto a lawn of $lamB^-$ cells to test for $OmpF^+$. We also serially diluted phage from each sample in M9 glucose and spotted 2 μL of each dilution on a lawn of WT cells to estimate the total number of phage in each flask. To save samples for future analysis, we made a frozen stock of each population by adding 40 μL 80% glycerol to 200 μL sample. When $OmpF^+$ phage were detected on the $lamB^-$ plates, we isolated a single plaque and sequenced the reactive region of the J gene to identify mutations.

For the experiments including the remaining engineered variants:

We measured evolvability of the engineered library variants using co-evolution replay experiments very similar to the one described in the previous section. Due to the number of variants in the library, we performed two separate 10-day experiments (experiment 1: 6-mut, T987L, T987G, T987A, T987C, T987Y; experiment 2: 6-mut, T987A, T987R, T987K). To gauge whether results of the two experiments could be directly compared, we included six replicates of 6-mut and six replicates of T987A in each experiment. In both cases the outcomes were identical or similar between experiments (6-mut: experiment 1: 6/6 replicates evolved, all six required 1 day, all six required 1 mutation; experiment 2: 6/6 replicates evolved, all six

required 1 day, all required 1 mutation; T987A: experiment 1: 6/6 replicates evolved, average 4.2 days, all required 2 mutations, experiment 2: 4/6 replicates evolved, average 5.25 days, all required 2 mutations). The higher variation in T987A compared to 6-mut is expected given increased opportunities for stochasticity when two mutations are required instead of just one. This indicates that results are consistent and repeatable between experiments. For all variants in both experiments, between 9.6×10^5 and 3.0×10^7 phage were added to each experimental flask, consistent with the experiment that included naturally evolved thermostable genotypes. All other elements of the experimental design were kept as identical to the initial experiment as possible, with a few minor exceptions: (1) REL606 was used as the host in all six replicates, as opposed to the three MalT⁻/ three REL606 design used for the initial experiment; (2) we carried out 1% transfers each day, consistent with the REL606 replicates from the initial experiment; (3) due to the larger size of the later experiment, we did not measure density of phage on WT each day, but instead spotted 5 μ L of undiluted chloroformed lysate from each flask on a WT lawn to determine the presence or absence of phage.

3.5.12 Bioinformatic prediction of J structure

We used the publicly available version (Mirdita et al. 2022) of AlphaFold (Jumper et al. 2021) to predict the structure of a truncated version of J containing the 173 most C-terminal amino acids, for both the ancestor and the 6-mut. We used all recommended presets in the publicly available version, and sequence coverage and model confidence are shown in Figure 3.A.2. We used Chimera (Pettersen et al. 2004) to visualize structures and create Figs 4 and S2.

3.5.13 Reconstruction of evolutionary pathways from coevolution replay

When replay populations evolved OmpF⁺, we sequenced the binding region of *J* from a single plaque on the first day OmpF⁺ was detected. We chose to further explore one population from each background and reconstruct the genotypes that evolved by engineering the mutations into a lysogenic λ . The purpose of reconstructing genotypes of interest as lysogens was to (1) ensure stable long-term storage of these genotypes as lysogens and (2) ensure that subsequent decay assays would be performed on freshly produced phage particles generated by triggering lysis by heat shock rather than growing up overnight stocks, which may contain phage particles of varying ages. The oligos used to engineer these strains can be found in Table 3.A.2. Decay assays were performed in the same way described in the section “decay assays”. Due to the relatively low-throughput nature of our decay assay, we performed these decay assays in a blocked design, with three blocks measured on three different days. Each block consisted of the stabilized starting genotype, as well as derivatives that evolved during the experiment. In each block, we also measured 6-mut and 7-mut. Since 6-mut and 7-mut were measured on three different days, we were able to gauge how much variation in decay rates was due to our blocked design. We found that there were no significant differences among the decay rates for 6-mut measured on different days (one-way ANOVA, $n = 3$ per day per genotype, $F[2,6] = 0.32$, $P = 0.7391$), and the same was true for 7-mut (one-way ANOVA, $n = 3$ per day per genotype, $F[2,6] = 4.12$, $P = 0.0748$) indicating that there is very little variation across days. This allowed us to pool 6-mut and 7-mut decay rates across days, and we used a t-test on pooled 6-mut vs. pooled 7-mut ($n = 9$ per genotype) to measure the effect of the N1107K mutation in the 6-mut background. To compare among the genotypes that evolved in stabilized backgrounds (6-mut + stabilizing mutation; 6-mut + stabilizing mutation + putative destabilizing mutation; 6-mut + stabilizing

mutation + putative destabilizing mutation + final gain-of-function mutation), we measured the five genotypes in each panel of Figure 3.5 on the same day and used paired t-tests to compare each genotype to its immediate ancestor, using only measurements taken on the same day. Significance was corrected for multiple testing using the Bonferroni method.

3.5.14 Editing N1107K into all variant backgrounds

As an additional verification of our model, we engineered N1107K into all variants and measured OmpF use. For some backgrounds (6-mut, T987A, F1122L, T987L and T987C) we edited N1107K in using the same CoS-MAGE technique described in the section “CoS MAGE to insert stabilizing mutations into 6-mut lysogen”, and then picked colonies and verified the change by sequencing. For the remaining engineered library variants (T987S, T987G, T987K, T987R, T987Y), we modified our CoS-MAGE protocol to increase throughput such that we could check whether N1107K confers OmpF⁺ to each variant without picking colonies and sequencing individually for each variant (S1 Text). Regardless of whether the edited genotypes were produced in the low or high throughput method, we tested for OmpF use in edited genotypes first by plating dilution series of induced lysogens on *lamB*⁻ lawns. As a second measure of OmpF use, we also calculated growth rates on *lamB*⁻ cells in liquid culture for a subset of variants, which allowed more sensitive detection of weak OmpF activity (S1 Text).

3.5.15 Additional verification of plaque-based detection of OmpF⁺ with a liquid-based growth assay in populations from replay experiment

Because we observed that the 7-mut T987A and T987Y thermostable mutant plaqued poorly on *lamB*⁻ cells yet grew measurably on *lamB*⁻ cells in liquid culture, we wondered if any

populations in the coevolution replay experiments had become OmpF⁺ without being detected by the plaque assay. To address this possibility, we measured growth on *lamB*⁻ in liquid among all six replicates of a subset of naturally evolved and engineered genotypes (6-mut, T987A, F1122L, and T987S) at day 6 of co-evolution. We combined 1 mL from each experimental flask with 40 μ L chloroform to kill the coevolving bacteria and centrifuged the tubes at 21,000 \times g for 1 minute. Then we inoculated 100 μ L from each population into culture tubes containing 4 mL LBM9, 40 μ L MgSO₄, and 100 μ L *lamB*⁻ cells. Tubes were incubated at 37°C shaking for 12.5 hours. After incubation, cells were treated with chloroform (40 μ L chloroform to 1 mL sample). Quantification of phage was assessed before and after incubation by serially diluting chloroformed phage in M9 glucose and spotting 2 μ L on a lawn of WT cells. The assay did not reveal any populations that grew on *lamB*⁻ in liquid but were not detected by the plate-based assays. Growth rates are reported in Table 3.A.4.

3.5.16 Growth rate on REL606 in M9 glucose media

To assess the growth rate of each variant in the conditions used for the evolution experiment, we induced three replicates of each variant using the same procedure as described in the section Induction of lysogens by heat shock, filtered through a 0.22 μ M filter, and diluted each in M9 glucose + MgSO₄ depending on initial titer upon induction, such that $\sim 10^4$ phage were added to each flask containing 10 mL M9 glucose, 100 μ L MgSO₄. Flasks were mixed well, and phage titers were measured by plating with WT cells infused in soft agar. Then, we added $\sim 10^8$ REL606 cells to each flask and incubated at 37 C shaking for 4 hours. After incubation, samples were taken from each flask, filtered through 0.22 μ M filters, and phage titers were re-measured by diluting in M9 glucose + MgSO₄ and plating with WT cells infused in soft

agar. We chose to measure growth over 4 hours rather than 24 hours (the length of time between transfers in the evolution experiment) to reduce the possibility that genotypes would evolve mutations during the growth experiment. T987A, T987L, T987S, T987G, T987K, T987R, and T987Y were measured alongside 6-mut in a single experiment. F1122L and T987C were measured on a different day alongside three additional replicates of 6-mut, and no day effects were detected. Paired t-tests were used to compare growth rates of variants to 6-mut and corrected for multiple comparisons using the Bonferroni method ($\alpha = 0.0056$).

3.5.17 Assays for productivity in naturally evolved thermostable variants containing stabilizing mutations and N1107K

We used the same procedure as described in the section “CoS MAGE to insert stabilizing mutations into 6-mut lysogen” to edit N1107K into 6-mut and the two naturally evolved thermostable genotype (T987A, and F1122L). To test the productivity of these genotypes, we induced phage as described in the section ‘Induction of lysogens by heat shock’. We then filtered lysates through 0.22 μM syringe filters, serially diluted in LBM9, and spotted 2 μL of each dilution on a lawn of *WT* cells. Statistical comparisons were made between titers produced by genotypes with and without N1107K using unequal variance t-tests and corrected for multiple comparisons using the Bonferroni method ($\alpha = 0.025$).

3.5.18 Testing the effect of the T987A stabilizing mutation in 5-mut, a background further removed from OmpF⁺

We used the same procedure as described in the section “CoS MAGE to insert stabilizing mutations into 6-mut lysogen” to revert another one of the 7-mut mutations (changing the amino acid at codon 1012 from G back to S) to create a 5-mut. We then used CoS MAGE again to insert the stabilizing mutation T987A into the 5-mut. To assess whether T987A is stabilizing in this background, we used the same procedure described in the section “Decay Assays” to measure decay rates. The decay rates of 5-mut and 5-mut T987A were compared with an unequal variance 2-sample T-test. To assess evolvability, we used the same procedure described in the section “Coevolution replay experiment” to conduct a 10-day experiment with six replicate flasks each for 5-mut and 5-mut T987A. All six flasks were inoculated with an initial population of 10^6 phage particles, and all six flasks contained the REL606 host bacteria. As with the other coevolution experiments, we transferred 1% of the flask contents (phage and host bacteria) every 24 hours and plated daily on WT and *LamB*- hosts.

3.5.19 Software for statistics and plotting

All statistics and plotting were carried out in Matlab (version R2019b). Statistical methods are detailed in materials and methods subsections above. All t-tests were two-tailed. Data were checked for normality and homogeneity of variance prior to performing statistical tests.

3.6 Acknowledgments

We would like to thank Sarah Medina for help in the laboratory. We would also like to thank Katie Petrie, Morgan Mouchka, Animesh Gupta, and Josh Borin for valuable discussion of the manuscript.

Chapter 3, in full, is a reprint of the material as it appears in Strobel HM, Horwitz EK, Meyer JR (2022). Viral protein instability enhances host-range evolvability. PLoS Genet 18: e1010030. The dissertation author was the primary researcher and author of this paper.

3.A Appendix

3.A.1 High throughput Cos-MAGE to edit N1107K into engineered backgrounds

Initially, we edited N1107K into only a subset of all variant backgrounds. To do this, we used a low throughput method of Cos-MAGE, requiring isolation and sequencing individual lysogen clones and inducing phage production. We later sought to confirm the effect of N1107K in all backgrounds, so we designed a modified Cos-MAGE procedure to increase throughput and reduce cost of searching for clones that received the mutation. To do this, we conducted two cycles of CoS-MAGE in each remaining background using the appropriate *galK* oligo as well as the N1107K oligo, with three replicates per background. Following an overnight recovery step, cultures were induced using the same procedure as described in the section Induction of lysogens by heat shock. Lysates were filtered, diluted in LBM9, and spotted on two different lawns: one containing WT cells and the other containing *lamB*⁻ cells. The titer on WT lawns provided the total number of phage particles in the lysate, and the titer on *lamB*⁻ cells provided the number of phage that had been converted to OmpF⁺ after MAGE with the N1107K oligo. We included 6-mut as a positive control and observed a 10-20% conversion rate. As an additional control to verify that the MAGE process was working in each variant, we computed the conversion rate of

the *galK* selectable marker as a positive control and observed a conversion rate of 3-12%. Given the high density of lysogen cells prior to induction ($>10^7$) it is unlikely that our method failed to detect conversions to $OmpF^+$. In Table 3.A.1, we present the conversion rates both of the *galK* selectable marker and of the phage genotypes to $OmpF^+$. When no $OmpF^+$ conversions were detected, we calculated an upper bound for the conversion rate if a single plaque had been detected.

3.A.2 Liquid assay to detect weak growth on OmpF after editing in N1107K

Upon receiving the N1107K mutation, most backgrounds either produced obvious plaques when serial dilutions were spotted on *lamB*⁻ lawns (6-mut, T987L, and T987C) or produced no visible effect when spotted on *lamB*⁻ (F1122L, T987S, T987G, T987K, and T987R). However, two backgrounds, T987A and T987Y, failed to produce individual plaques in a dilution series, but produced turbid clearings when high concentrations of phage were spotted. We hypothesized that perhaps these two backgrounds gained the ability to infect using *OmpF* at very low levels, resulting in some killing of cells but at a rate too low to produce plaques. To test this, we designed an assay to more sensitively measure growth rate on *lamB*⁻ cells. For this assay we chose representative variants that produced good plaques (6-mut + N1107K), no plaques but turbid clearings (T987A + N1107K), and no visible effect (F1122L + N1107K) and inoculated three replicate tubes with 10 μ L of filtered lysate and 100 μ L of *lamB*⁻ cells to 4 mL LBM9 supplemented with 40 μ L $MgSO_4$ and incubated at 37 $^{\circ}C$ shaking for 14.5 hours. Phage growth was quantified by plating on the permissive host (WT) before and after phage were incubated with *lamB*⁻ cells. We quantified pre-growth densities by diluting phage in LBM9 (before cells were added) and spotting 2 μ L of each dilution a WT lawn. Post-growth densities of phage were obtained by chloroforming cultures (to remove cells) and diluting phage in LBM9

and spotting 2 μ L of each dilution on a WT lawn. We found that, as predicted, genotypes that produce no visible effect on plates do not measurably grow on *lamB*⁻ in liquid culture, whereas genotypes that produce turbid clearings grow at a measurable but significantly reduced rate compared to genotypes that plaque well on *lamB*⁻ lawns (Figure 3.A.4). We compared growth rates among genotypes using paired t-tests corrected for multiple comparisons using the Bonferroni method.

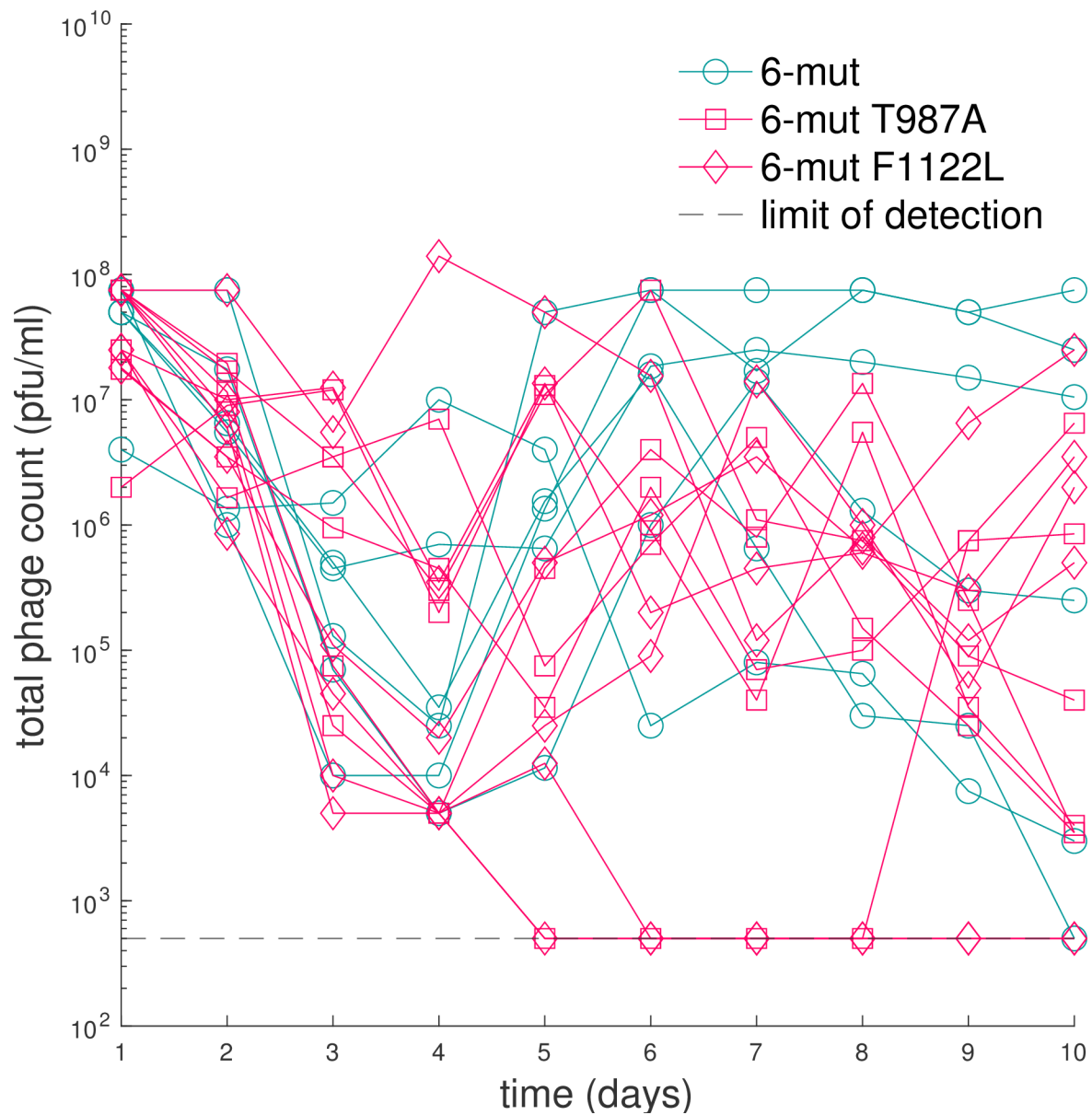


Figure 3.A.1 - Phage titer during evolution experiment on naturally occurring thermostable genotypes. Phage titer measured on WT cells on each day of the evolution experiment on the naturally evolved thermostable variants (Figure 3.2).

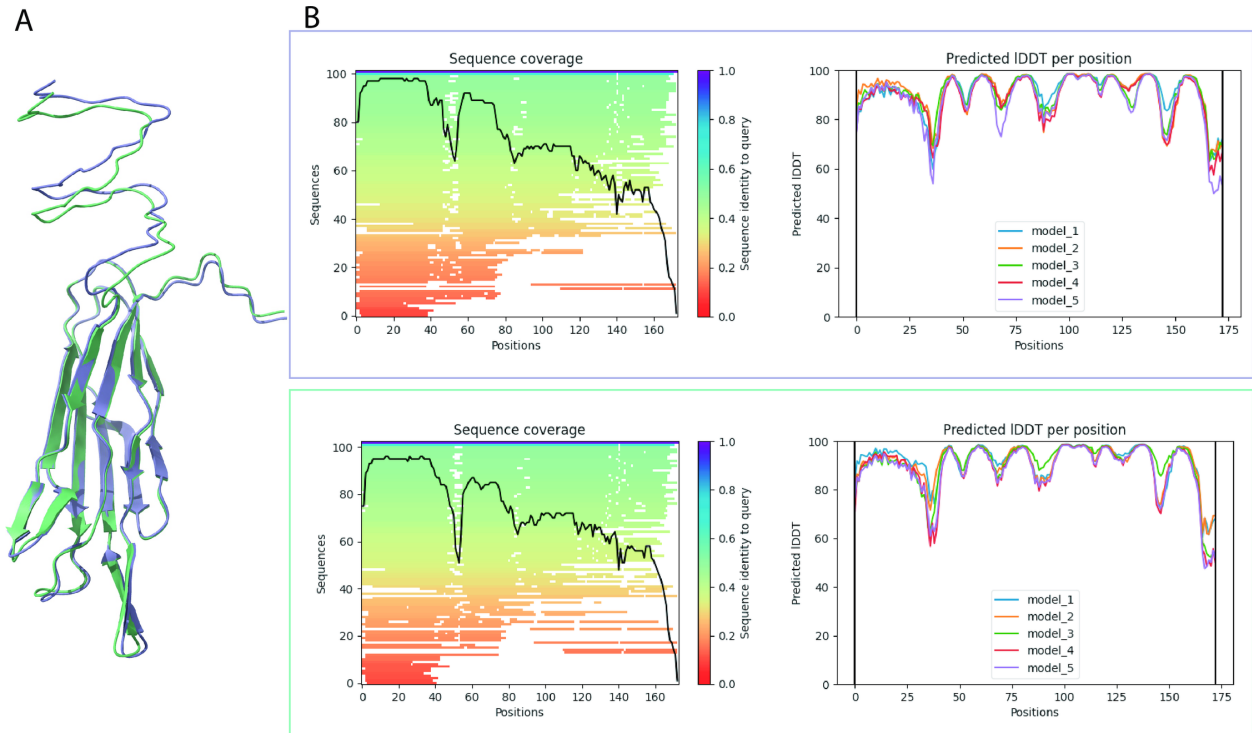


Figure 3.A.2 - Aligned AlphaFold predictions for ancestor and 6-mut J proteins. A: aligned AlphaFold predictions for ancestor (purple) and 6-mut (teal) J protein reactive region (amino acids 960–1132). B: Coverage of multiple sequence alignment (MSA) used to make structural prediction and predicted IDDT (model confidence out of 100) at each position.

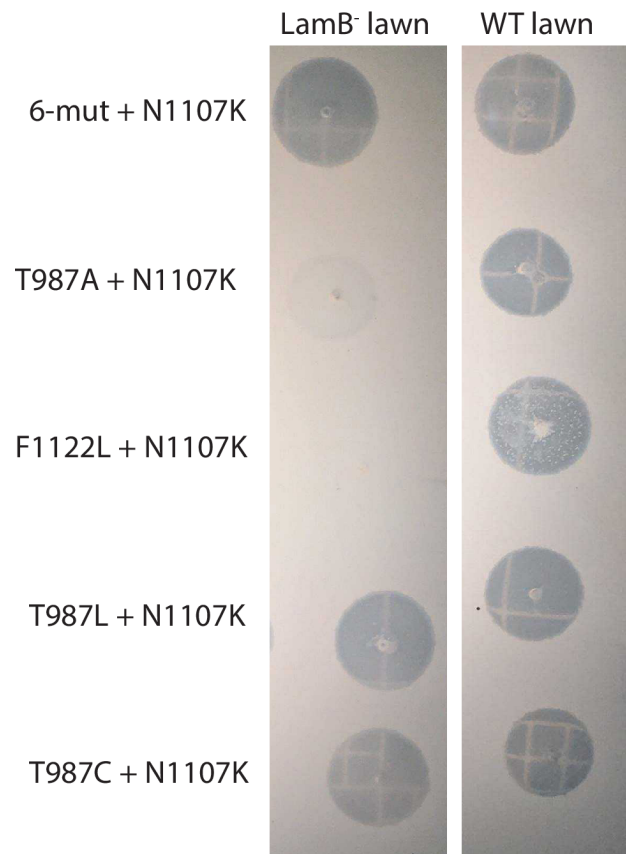


Figure 3.A.3 - Editing N1107K into all variant backgrounds.

To verify that most of the stable genotypes required an additional mutation along with N1107K for OmpF⁺, we edited N1107K back into all genotypes in the library (see appendix) and tested their ability to plaque on *LamB*⁻. N1107K conferred full OmpF⁺ on three genotypes (6-mut, T987L, and T987C), and these were all detected during the evolution experiment. Two variants, T987A and T987Y, were able to form very turbid clearings when spotted on *LamB*⁻, but because these genotypes were not able to form individual plaques we did not detect them during the evolution experiment. Consistent with their poor plaquing, genotypes with partial OmpF-use grew at a dramatically lower rate compared to genotypes with full OmpF-use (Figure 3.A.4). All genotypes were spotted on WT lawns as a positive control to indicate that plaquing effect was specific to OmpF-use and not indicative of generic viability.

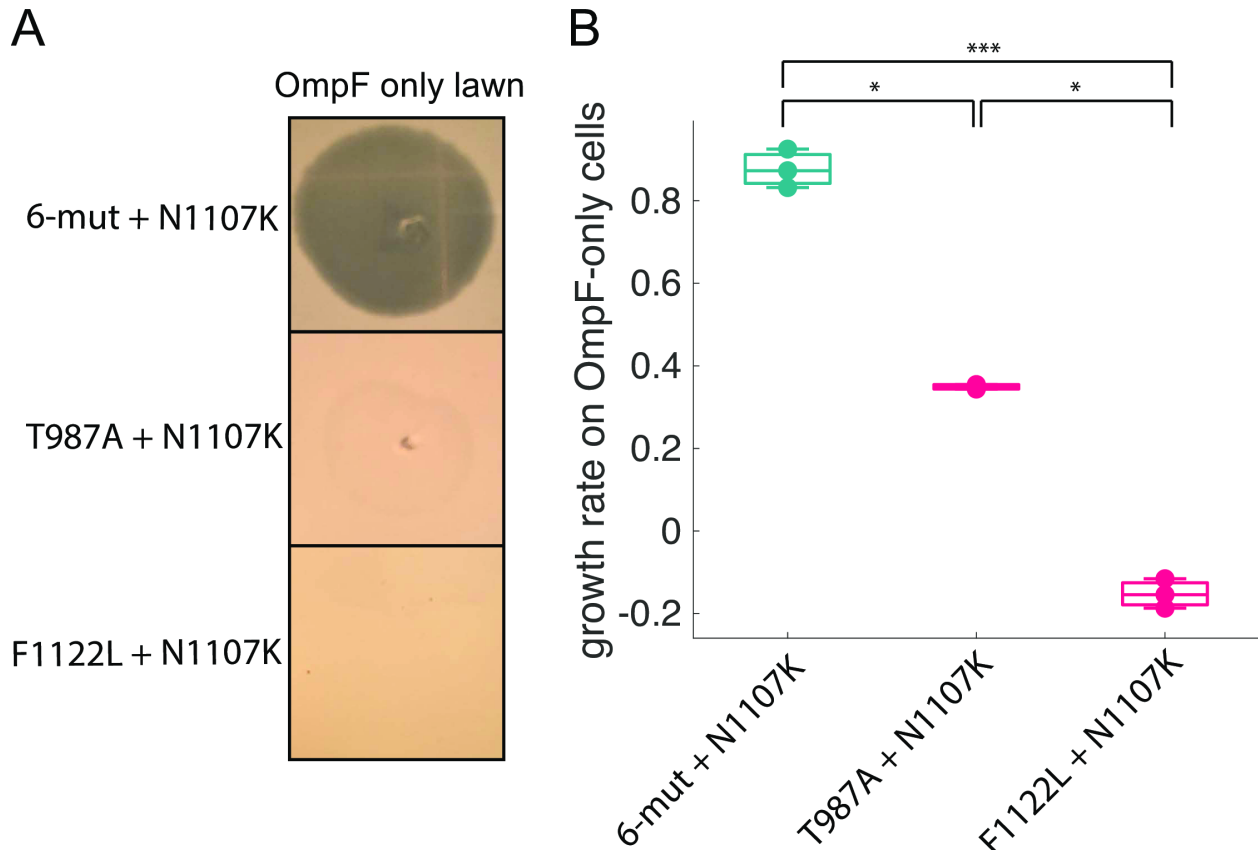


Figure 3.A.4 - Turbid clearings on plates correspond to weak growth in liquid.

A: Plate photo of each genotype spotted on a lawn of cells that express OmpF only. B: Growth rate of three replicates of each genotype on cells that express OmpF but not LamB ($n = 3$ for 6-mut N1107K and F1122L N1107K, $n = 2$ for T987A N1107K; paired t-tests corrected for multiple comparisons using Bonferroni method; 6-mut + N1107K vs. 6-mut T987A + N1107K: $p = 6.38 \times 10^{-4}$; 6-mut + N1107K vs. 6-mut F1122L + N1107K: $p = 3.38 \times 10^{-4}$; 6-mut + N1107K vs. 6-mut F1122L + N1107K: $p = 1.06 \times 10^{-6}$. Bonferroni adjusted significance: ns: > 0.0167 , *: $p < 0.00167$, **: $p < 1.67 \times 10^{-4}$, *** $p < 1.67 \times 10^{-5}$, ****: $p < 1.67 \times 10^{-6}$.)

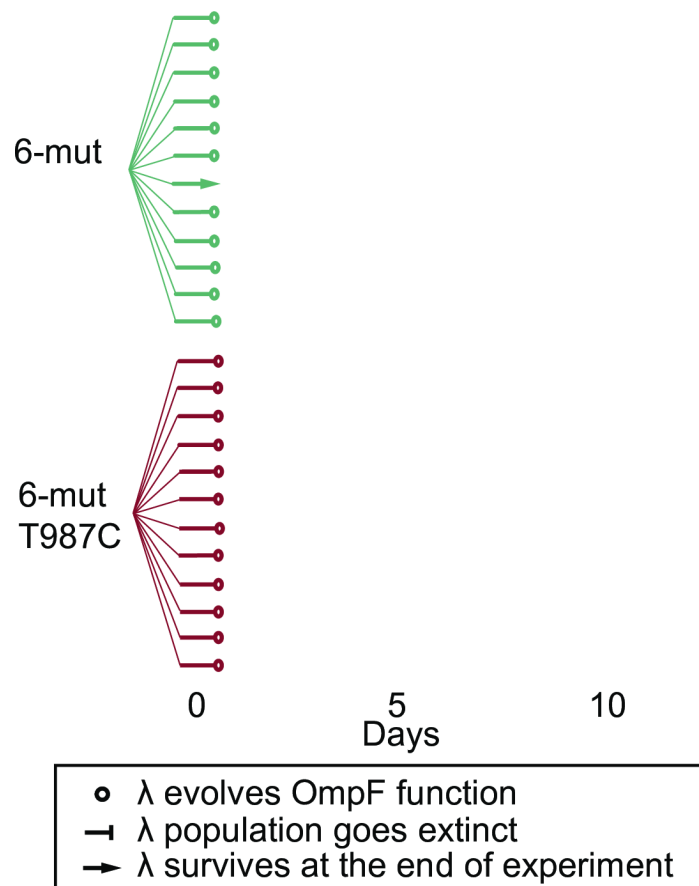


Figure 3.A.5 - Independently created T987C lysogen.

To rule out the possibility that T987C had acquired mutations outside of J that helped it evolve OmpF⁺ faster, we re-engineered a new lysogen with an oligo specifically designed to produce T987C. We then measured its evolvability in 12 replicate populations, as well as 12 replicate populations of the 6-mut as a control. All 12 T987C replicate populations evolved OmpF⁺ in one day, compared to eleven of twelve 6-mut replicate populations, confirming that these genotypes have nearly identical evolvabilities under the conditions used in our evolution experiments.

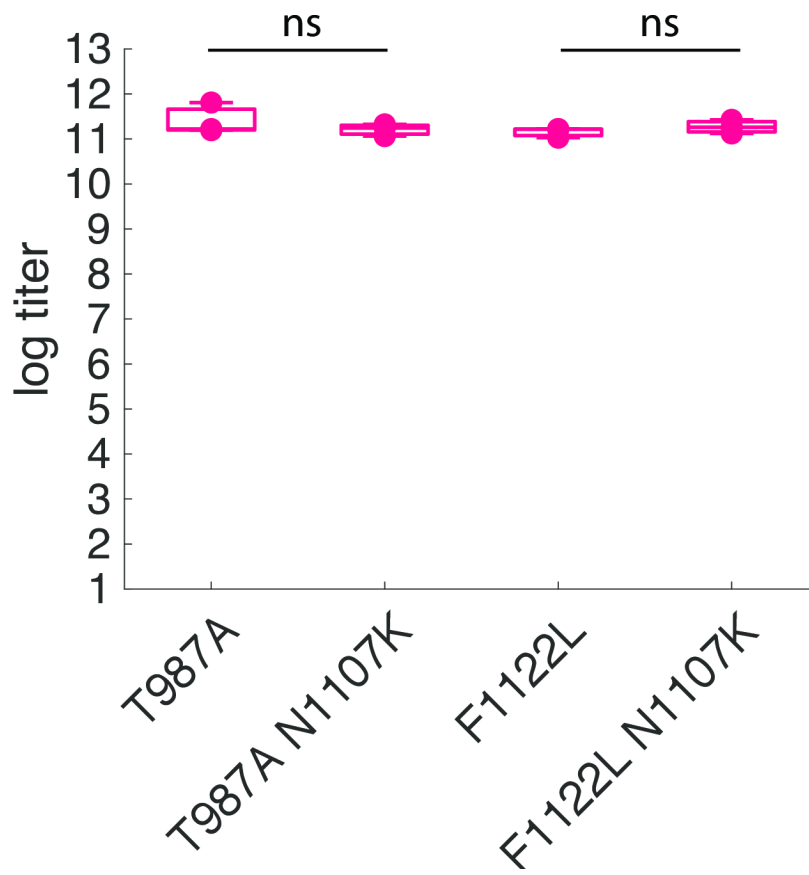


Figure 3.A.6 - Productivity of stable genotypes with and without N1107K. Productivity (log viable titer) of each genotype immediately after lysogen induction, with and without N1107K mutation. Statistical comparisons were made using two t-tests and corrected for multiple comparisons using the Bonferroni method, N = 3 per genotype. 6-mut T987A vs. 6-mut T987A N1107K: $p = 0.4376$, 6-mut F1122L vs. 6-mut F1122L N1107K: $p = 0.3508$. Bonferroni adjusted significance: ns: $p > 0.025$.

Table 3.A.1 - High throughput MAGE to engineer N1107K into genotypes that either did not evolve OmpF⁺ in the evolution experiment (T987G) or required two mutations to evolve OmpF⁺ (T987S, T987K, T987R, T987Y).

6-mut was included as a positive control because it is expected to convert to OmpF⁺ upon receiving the N1107K mutation. As an additional positive control to ensure that the MAGE process was working for each variant, we used an oligo mix containing N1107K as well as the appropriate *galk* conversion oligo and computed the fraction of cells that were successfully converted from *galk*⁺ to *galk*⁻ or vice versa. We computed the conversion rate to OmpF⁺ after MAGE by inducing the lysogens and plating dilutions of each lysate on two lawns: one with cells expressing only OmpF, and another with WT cells. We then divided the number of plaques on the OmpF only lawn by the total number of plaques on the WT lawn to get the conversion rate. In T987S, T987K, T987R, and T987Y we did not observe any OmpF⁺ plaques, so we computed an upper bound as the conversion rate if a single OmpF⁺ plaque had been observed. For T987A, F1122L, T987L and T987C, we separately engineered in N1107K and isolated and sequenced individual clones rather than using the high throughput method described here.

genotype	positive control (<i>galk</i>) conversion rate	conversion rate to OmpF ⁺ by N1107K oligo
6-mut	6.20E-02	2.00E-01
6-mut	9.04E-02	1.85E-01
6-mut	6.54E-02	9.38E-02
T987S	9.49E-02	<3.20E-07
T987S	1.07E-01	<3.60E-07
T987S	1.20E-01	<6.40E-07
T987G	4.96E-02	<8.00E-04
T987G	2.86E-02	<1.30E-03
T987G	5.36E-02	<2.70E-03
T987K	5.01E-02	<1.10E-06
T987K	5.31E-02	<1.10E-06
T987K	5.04E-02	<4.00E-05
T987R	4.93E-02	<1.60E-06
T987R	6.24E-02	<3.20E-06
T987R	6.90E-02	<3.20E-06
T987Y	3.48E-02	<1.10E-06
T987Y	3.10E-02	<4.00E-07
T987Y	3.47E-02	<6.40E-07

Table 3.A.2 - MAGE oligos.

Asterisks indicate phosphorothioated bond and underlined letters indicate the mutated bases.

mutation	background	Oligo sequence
A3321T (K1107N)	7-mut	C*A*T*C*GCTGGCAAACGTATACGGCGGAAT <u>A</u> TTTG CCGAATACCGTGTGGACGTAAGCGTGAACGTCAGGA TCACGTTTCCCCGACCCGCTG
A2959G (T987A)	6-mut and 7- mut	C*G*C*C*ACCTTTACAATGTCCCCGACGATTTTTTCC GCCCTCAGCG <u>C</u> ACCGTTTATCGTACAGTTTTTCAGCTA TCGTCACATTACTGAGCG
T3364C (F1122L)	6-mut and 7- mut	T*C*T*G*TAACACACTCAGACCACGCTGATGCCAG CGCCTGTTTCTTAAGCACCATAACCTGCACATCGCTG GCAAACGTATACGGCGGAAT
A2959T (T987S)	6-mut	C*G*C*C*ACCTTTACAATGTCCCCGACGATTTTTTCC GCCCTCAGCG <u>A</u> ACCGTTTATCGTACAGTTTTTCAGCTA TCGTCACATTACTGAGCG
C2909A (S970Y)	6-mut, 6-mut T987A, and 7- mut	T*T*A*T*CGTACAGTTTTTCAGCTATCGTCACATTACT GAGCGTCCCGT <u>A</u> GTTTCGCATTCACACTGCCACTGAT ATCCGCATTTTTAGCGGTCA
S987L (C2960T)	6-mut T987S	C*G*C*C*ACCTTTACAATGTCCCCGACGATTTTTTCC GCCCTCAGC <u>AA</u> ACCGTTTATCGTACAGTTTTTCAGCTA TCGTCACATTACTGAGCGT
S1011R (C3033G)	6-mut S987L	A*T*C*G*GTCACGGTGACAGTACGGGTACCTGACGG CCAGTCCACACC <u>C</u> TTTCACGCTGGCGCGGAAAAGC CGCGCTCGCCACCTTTACAAT
S1049R (C3147A)	6-mut	G*T*T*C*ATCAGTACTTTCAGATAACACATCGAATA CGTTGTCCTGCC <u>T</u> CTGACAGTACGCTTACTTCCGCGA AACGTCAGCGGAAGCACCAC
galK t435a (premature stop codon)	Any CI857 derivative	G*C*T*T*CACTGGAAGTCGCGGTTCGGAACCGTATTG CAGCAGCTTTA <u>A</u> CATCTGCCGCTGGACGGCGCACAA ATCGCGCTTAACGGTCAGGAA
Codon 987 library	6-mut	C*G*C*C*ACCTTTACAATGTCCCCGACGATTTTTTCC GCCCTCAG <u>NNN</u> ACCGTTTATCGTACAGTTTTTCAGCTA TCGTCACATTACTGAGCG

Table 3.A.3 - Unadjusted and adjusted decay rates for engineered library variants.

genotype	date collected	unadjusted decay rate	adjusted decay rate
6-mut A	63021	-0.567	
6-mut B	63021	-0.554	
6-mut C	63021	-0.542	
ancestral λ A	63021	-0.145	0.262
ancestral λ B	63021	-0.123	0.222
ancestral λ C	63021	-0.228	0.411
6-mut A	82420	-0.575	
6-mut B	82420	-0.541	
6-mut C	82420	-0.543	
7-mut A	82420	-0.582	1.053
7-mut B	82420	-0.547	0.989
7-mut C	82420	-0.484	0.875
6-mut T987A A	82420	-0.163	0.295
6-mut T987A B	82420	-0.124	0.224
6-mut T987A C	82420	-0.13	0.235
6-mut F1122L A	82420	-0.117	0.212
6-mut F1122L B	82420	-0.163	0.295
6-mut F1122L C	82420	-0.112	0.202
6-mut	60521	-0.541	
6-mut	60521	-0.544	
6-mut	60521	-0.537	
T987S	60521	-0.123	0.228
T987S	60521	-0.168	0.311
T987S	60521	-0.161	0.298
T987C	60521	-0.211	0.39
T987C	60521	-0.215	0.398
T987C	60521	-0.22	0.407
T987Y	60521	-0.111	0.205
T987Y	60521	-0.157	0.291
T987Y	60521	-0.134	0.247
T987R	60521	-0.155	0.287
T987R	60521	-0.138	0.255
T987R	60521	-0.119	0.22
6-mut	21821	-0.487	
6-mut	21821	-0.446	
6-mut	21821	-0.492	
T987L	21821	-0.485	1.022
T987L	21821	-0.47	0.989
T987L	21821	-0.511	1.075
T987G	21821	-0.16	0.336
T987G	21821	-0.103	0.218
T987G	21821	-0.153	0.321
6-mut	33121	-0.502	
6-mut	33121	-0.482	
6-mut	33121	-0.482	
T987K	33121	-0.211	0.433
T987K	33121	-0.085	0.174
T987K	33121	-0.108	0.222

Table 3.A.4 - λ growth rates on *lamB*⁻ hosts measured in liquid culture on D6 of the replay experiment.

Positive growth rates are bold. Growth rates could not be calculated (“NA”) for some populations (6-mut T987A 2 and 6-mut F1122L 1 and 2) because no viable phage were detected at day 6 (i.e. the population had gone extinct), or because phage decayed to zero during the overnight growth period on *lamB*⁻ (6-mut T987S 4). Growth rates for each population were measured in a single replicate.

starting genotype	population	growth rate
6-mut	1	0.79
6-mut	2	0.04
6-mut	3	0.66
6-mut	4	0.60
6-mut	5	0.86
6-mut	6	1.38
6-mut T987A	1	-0.04
6-mut T987A	2	NA
6-mut T987A	3	0.95
6-mut T987A	4	-0.08
6-mut T987A	5	-0.18
6-mut T987A	6	-0.03
6-mut T987S	1	0.09
6-mut T987S	2	-0.15
6-mut T987S	3	-0.16
6-mut T987S	4	NA
6-mut T987S	5	0.04
6-mut T987S	6	-0.03
6-mut F1122L	1	NA
6-mut F1122L	2	NA
6-mut F1122L	3	-0.06
6-mut F1122L	4	-0.09
6-mut F1122L	5	-0.10
6-mut F1122L	6	-0.07

3.7 References

- Baba, T., T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. Datsenko, M. Tomita, B. Wanner, and H. Mori 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular systems biology* 2.
- Besenmatter, W., P. Kast, and D. Hilvert. 2007. Relative tolerance of mesostable and thermostable protein homologs to extensive mutation. *Proteins* 66:500-506.
- Bloom, J. D. and F. H. Arnold. 2009. In the light of directed evolution: pathways of adaptive protein evolution. *Proc Natl Acad Sci U S A* 106 Suppl 1:9995-10000.
- Bloom, J. D., S. T. Labthavikul, C. R. Otey, and F. H. Arnold. 2006. Protein stability promotes evolvability. *Proc Natl Acad Sci U S A* 103:5869-5874.
- Chaudhry, W., M. Pleška, N. Shah, H. Weiss, I. McCall, J. Meyer, A. Gupta, C. Guet, and B. Levin. 2018. Leaky resistance and the conditions for the existence of lytic bacteriophage. *PLoS biology* 16.
- Cuevas, J. M., A. Moya, and R. Sanjuán. 2009. A genetic background with low mutational robustness is associated with increased adaptability to a novel host in an RNA virus. *J Evol Biol* 22:2041-2048.
- Datsenko, K. and B. Wanner. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America* 97.
- Dessau, M., D. Goldhill, R. McBride, P. Turner, and Y. Modis. 2012. Selective pressure causes an RNA virus to trade reproductive fitness for increased structural and thermal stability of a viral enzyme. *PLoS genetics* 8.
- Fasan, R., Y. Meharena, C. Snow, T. Poulos, and F. Arnold 2008. Evolutionary history of a specialized p450 propane monooxygenase. *Journal of molecular biology* 383.
- Flanagan, M., C. Parrish, S. Cobey, G. Glass, R. Bush, and T. Leighton. 2012. Anticipating the species jump: surveillance for emerging viral threats. *Zoonoses and public health* 59.
- Hueffer, K., J. Parker, W. Weichert, R. Geisel, J. Sgro, and C. Parrish 2003. The natural host range shift and subsequent evolution of canine parvovirus resulted from virus-specific binding to the canine transferrin receptor. *Journal of virology* 77.
- Jaenicke, R. and G. Böhm. 1998. The stability of proteins in extreme environments. *Current opinion in structural biology* 8.

Jumper, J., R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Židek, A. Potapenko, A. Bridgland, C. Meyer, S. Kohl, A. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596.

Kirschner, M. and J. Gerhart. 1998. Evolvability. *Proceedings of the National Academy of Sciences of the United States of America* 95.

Li, J. and L. Luo. 2020. Nurturing Undergraduate Researchers in Biomedical Sciences. *Cell* 182.

Lu, G., Q. Wang, and G. Gao 2015. Bat-to-human: spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond. *Trends in microbiology* 23.

Maddamsetti, R., D. T. Johnson, S. J. Spielman, K. L. Petrie, D. S. Marks, and J. R. Meyer. 2018. Gain-of-function experiments with bacteriophage lambda uncover residues under diversifying selection in nature. *Evolution* 72:2234-2243.

McBride, R. C., C. B. Ogbunugafor, and P. E. Turner. 2008. Robustness promotes evolvability of thermotolerance in an RNA virus. *BMC Evol Biol* 8:231.

Meyer, J. R., D. T. Dobias, S. J. Medina, L. Servilio, A. Gupta, and R. E. Lenski. 2016. Ecological speciation of bacteriophage lambda in allopatry and sympatry. *Science* 354:1301-1304.

Meyer, J. R., D. T. Dobias, J. S. Weitz, J. E. Barrick, R. T. Quick, and R. E. Lenski. 2012. Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science* 335:428-432.

Mirdita, M., K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger. 2022. ColabFold - Making protein folding accessible to all.

Ogbunugafor, C., R. McBride, and P. Turner. 2009. Predicting virus evolution: the relationship between genetic robustness and evolvability of thermotolerance. *Cold Spring Harbor symposia on quantitative biology* 74.

Okonechnikov, K., O. Golosova, and M. Fursov. 2012. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics (Oxford, England)* 28.

Payne, J. L. and A. Wagner. 2019. The causes of evolvability and their evolution. *Nat Rev Genet* 20:24-38.

- Petrie, K. L., N. D. Palmer, D. T. Johnson, S. J. Medina, S. J. Yan, V. Li, A. R. Burmeister, and J. R. Meyer. 2018. Destabilizing mutations encode nongenetic variation that drives evolutionary innovation. *Science* 359:1542-1545.
- Pettersen, E., T. Goddard, C. Huang, G. Couch, D. Greenblatt, E. Meng, and T. Ferrin. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* 25.
- Pucci, F., J. Kwasigroch, and M. Rooman. 2017. SCooP: an accurate and fast predictor of protein stability curves as a function of temperature. *Bioinformatics (Oxford, England)* 33.
- Raman, A. S., K. I. White, and R. Ranganathan. 2016. Origins of Allostery and Evolvability in Proteins: A Case Study. *Cell* 166:468-480.
- Rathi, P. C., K. E. Jaeger, and H. Gohlke. 2015. Structural Rigidity and Protein Thermostability in Variants of Lipase A from *Bacillus subtilis*. *PLoS One* 10:e0130289.
- Razvi, A. and M. Scholtz. 2009. Lessons in stability from thermophilic proteins. *Protein Science* 15:1569–1578.
- Shi, Y., Y. Wu, W. Zhang, J. Qi, and G. F. Gao. 2014. Enabling the 'host jump': structural determinants of receptor-binding specificity in influenza A viruses. *Nat Rev Microbiol* 12:822-831.
- Sikosek, T. and H. S. Chan. 2014. Biophysics of protein evolution and evolutionary protein biophysics. *J R Soc Interface* 11:20140419.
- Singhal, S., G. Leon, CM, S. Whang, E. McClure, H. Busch, and B. Kerr 2017. Adaptations of an RNA virus to increasing thermal stress. *PloS one* 12.
- Smirnova, I. and H. Kaback. 2003. A mutation in the lactose permease of *Escherichia coli* that decreases conformational flexibility and increases protein stability. *Biochemistry* 42.
- Socha, R. and N. Tokuriki 2013. Modulating protein stability - directed evolution strategies for improved protein function. *The FEBS journal* 280.
- Sprouffske, K., J. Aguilar-Rodríguez, P. Sniegowski, and A. Wagner 2018. High mutation rates limit evolutionary adaptation in *Escherichia coli*. *PLoS genetics* 14.
- Stern, A. and R. Andino 2016. Viral Evolution: it is all about mutations. . Pp. 233-240 in M. Katze, G. Law, M. Korth, and N. N, eds. *Viral Pathogenesis* (3rd edition). New York.
- Stimple, S., M. Smith, and P. Tessier 2020. Directed evolution methods for overcoming trade-offs between protein activity and stability. *AIChE journal. American Institute of Chemical Engineers* 66.

Studer, R. A., P. A. Christin, M. A. Williams, and C. A. Orengo. 2014. Stability-activity tradeoffs constrain the adaptive evolution of RubisCO. *Proc Natl Acad Sci U S A* 111:2223-2228.

Thyagarajan, B. and J. Bloom. 2014. The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. *eLife* 3.

Tokuriki, N., C. Oldfield, V. Uversky, I. Berezovsky, and D. Tawfik 2009. Do viral proteins possess unique biophysical features? *Trends in biochemical sciences* 34.

Tokuriki, N. and D. S. Tawfik. 2009. Protein dynamism and evolvability. *Science* 324:203-207.

Vihinen, M. 1987. Relationship of protein flexibility to thermostability. *Protein Eng* 1:477-480.

Wagner, A. 2005. Robustness, evolvability, and neutrality. *FEBS Lett* 579:1772-1778.

Wang, H. and G. Church. 2011. Multiplexed genome engineering and genotyping methods applications for synthetic biology and metabolic engineering. *Methods in enzymology* 498.

Wang, H., F. Isaacs, P. Carr, Z. Sun, G. Xu, C. Forest, and G. Church. 2009. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460.

Wang, H., H. Kim, L. Cong, J. Jeong, D. Bang, and G. Church. 2012. Genome-scale promoter engineering by coselection MAGE. *Nature methods* 9.

Wang, J., M. Hofnung, and A. Charbit. 2000. The C-terminal portion of the tail fiber protein of bacteriophage lambda is responsible for binding to LamB, its receptor at the surface of *Escherichia coli* K-12. *Journal of bacteriology* 182.

Yadid, I., N. Kirshenbaum, M. Sharon, O. Dym, and D. S. Tawfik. 2010. Metamorphic proteins mediate evolutionary transitions of structure. *Proc Natl Acad Sci U S A* 107:7287-7292.

Zhang, L., C. Jackson, H. Mou, A. Ojha, H. Peng, B. Quinlan, E. Rangarajan, A. Pan, A. Vanderheiden, M. Suthar, W. Li, T. Izard, C. Rader, M. Farzan, and H. Choe 2020. SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nature communications* 11.

CHAPTER 4

Evolvability is an important trait in the selection of bacteriophages for therapeutic use

4.1 Abstract

The number of multidrug resistant strains of bacteria is increasing rapidly, while the discovery of new antibiotics has stagnated. Subsequently, interest in bacteriophages as anti-bacterial therapeutics has surged, in part, because there is near limitless diversity of phages to harness. While this diversity provides opportunity, it also creates the dilemma of having to decide which criteria to use to select phages. Two traits previously proposed are thermostability and reproductive rate. Here we show that focusing on phage's current abilities may be shortsighted if maximizing traits like stability and reproduction limits future evolution. We studied the ability of three phages to suppress bacteria. Each phage differed by only one amino acid at site 987 in their host-recognition protein (J), yet they varied significantly in their stability, growth rate, and evolvability. We uncovered a three-way tradeoff such that each phage maximized only two of the three traits. The most suppressive phage was an evolvable, fast-reproducing variant, supporting the importance of evolvability and reproduction rate. We tested whether these traits were interdependent but found that each had individual effects on suppression. Seeking to test the environmental contingency of our results, we altered the experiment to reflect more challenging conditions inside a patient's body. The stable, fast-reproducing phage was most suppressive in the short term, but the fast-reproducing, evolvable phage was more suppressive in the long term. Our results highlight the importance of evolvability, an often-overlooked trait in phage therapy, and underscore the need to consider long-term dynamics when testing phage for therapeutic use.

4.2 Introduction

The growing threat of antibiotic resistant bacteria has generated interest in the use of alternative therapeutics, such as bacteriophage (Kutter et al. 2010; Gordillo and Barr 2019). Bacteriophages are natural predators of bacteria and have been successfully used to treat multidrug resistant bacterial infections (Broncano-Lavado et al. 2021). Phages are ubiquitous in nearly every environment and are more genetically diverse than any other taxonomic group. This means that there are nearly endless varieties to harness (Zablocki et al. 2015; Jurczak-Kurek et al. 2016; Batinovic et al. 2019), whereas antibiotic drugs are far fewer and new discoveries are rare (Donadio et al. 2010; Brown and Wright 2016). To understand phage's enormous potential, there are currently more than 20,000 bacteriophages isolated and stored in the Actinobacteriophage Database, and those are the phages that infect just a single phylum of bacteria (<https://phagesdb.org/>). This incredible diversity is beneficial for providing practitioners with many options for treatment, but it also necessitates the development of criteria for selecting the best phage from among numerous options.

Currently, researchers select phages by first evaluating infectivity on the pathogen of interest. Infectivity is measured by testing whether phage lyse bacteria in liquid culture or by plating phage on a lawn of the pathogenic bacteria and measuring plaquing ability (Hyman 2019). If multiple candidate phages all plaque, then additional phage traits might be considered. Current literature recommends stability as one such trait (Hejnowicz et al. 2014; Casey et al. 2018), and mathematical models predict that stable phages will be more suppressive than unstable phages (Bull et al. 2019). There is also interest in deliberately enhancing stability in therapeutic phages via directed evolution (Favor et al. 2020). Another trait that is thought to be desirable in therapeutic phages is a high reproductive rate (Casey et al. 2018). The reasoning is

intuitive: phages able to rapidly generate a large population should be more effective at arresting bacterial growth (Bull et al. 2019). Despite the intuitive appeal of both stability and reproductive rate, the importance of these traits in determining bacterial suppression has not been demonstrated empirically.

A third trait that has been relatively overlooked as a criterion for therapeutic phages is evolvability (Bono et al. 2021), defined as the capacity for adaptive evolution. The reason antibiotics and even phage treatments become obsolete is that bacteria evolve resistance (Labrie et al. 2010; Blair et al. 2015). However, unlike antibiotics, phages have an endogenous algorithm – evolution by natural selection – to overcome resistance by gaining counter defenses (Bull et al. 2019). Ideally, phage therapeutics would be able to evolve counter measures to resistance during treatment without any human intervention. Phages have been shown to diverge in their ability to evolve counter-defenses, and so favoring evolvable phages during therapeutic selection might yield a more powerful treatment (Casey et al. 2018). One mechanism by which bacteria evolve resistance to phage is by mutating, deleting, or downregulating the expression of cell surface molecules that phage use as receptors during the first stage of infection (Charbit et al. 1984; Rakhuba et al. 2010; Laanto et al. 2012; Høyland-Kroghsbo et al. 2013). Under certain conditions, phage can overcome this perturbation by evolving to use a new receptor and retain suppression (Meyer et al. 2012; Borin et al. 2021). Therefore, phages with enhanced ability to evolve to use new receptors and expand their host-range should be more effective at suppressing bacteria, especially over the long term when bacteria have the opportunity to evolve resistance.

Ideally, of course, researchers should select phages that are stable, fast-reproducing, and evolvable. However, tradeoffs constrain the simultaneous optimization of multiple traits (Goldhill and Turner 2014; Edwards et al. 2021). Indeed, there is a well-known tradeoff

between stability and reproduction that has been demonstrated in RNA viruses (Dessau et al. 2012; Singhal et al. 2017). Additionally, there may be a tradeoff between high levels of stability and evolvability, as demonstrated in bacteriophage λ (Strobel et al. 2022). If tradeoffs are common in nature, then it may be difficult to identify naturally occurring phages with multiple favorable traits, and it will be helpful to understand which traits to prioritize when selecting phage. Furthermore, with the development of more sophisticated genetic engineering technologies, researchers are increasingly attempting to design phage with desirable characteristics (Favor et al. 2020). Such attempts could easily be derailed without a solid understanding of how tradeoffs constrain trait optimization.

We capitalized on the wealth of knowledge on bacteriophage λ evolution and molecular biology (Meyer et al. 2012; Casjens and Hendrix 2015) to generate an ideal system in which to test the importance of the three phage traits for bacterial suppression. We studied three nearly identical λ genotypes from a previously created library of variants differing only by a single amino acid in the receptor binding protein J (Strobel et al. 2022). The initial library contained nine variants, and none exhibited optimal values for all three traits. Perhaps surprisingly, given their high genotypic similarity, they exhibited striking phenotypic differences across our three traits of interest, and we chose three genotypes that had each optimized a different pair of traits (Fig. 1).

To test which traits enhance suppression, we performed a ten-day experiment in which each phage genotype was incubated with permissive *E. coli* bacteria in 24-hour cycles, with daily transfers to fresh media. As a metric of bacterial suppression, we monitored bacterial population size across all ten days. The study was designed such that each trait was optimized in two of the three phages, so if a single trait strongly influenced suppression, the bacteria would be

significantly lower in two of the three populations. However, bacteria were only significantly suppressed in one treatment, suggesting two traits worked together to make the difference. Upon finding this, we tested whether the two traits had independent effects on suppression. To achieve this, we used genotypes from the previous study, in which a key OmpF-conferring mutation was edited into the fast-evolvable and stable-evolvable backgrounds, effectively collapsing their variation to a single trait: reproductive rate (Strobel et al. 2022). These new genotypes were named “fast-evolved” and “stable-evolved”. Finally, we tested the environmental contingency of our first result, aiming to adjust the experimental conditions to reflect the challenging environment of the human body.

4.3 Materials and Methods

4.3.1 Media

We prepared media exactly as described in the previous study (Strobel et al. 2022).

4.3.2 Phage Strains

The fast-evolvable, fast-stable, and stable-evolvable λ genotypes used in this study were derived from lysogenic λ phage strains of cI857 that were part of a library of phage variants created for a previous study (Strobel et al. 2022). For ease of engineering, the initial library was created by editing a lysogenic λ prophage integrated into the *E. coli* chromosome. This presented an obstacle to our suppression experiments; however, because the phage still contained a functional cI repressor gene, enabling them to re-integrate into the *E. coli* host chromosome and confer high levels of resistance. To ameliorate this, we first designed a method to switch each genotype to from lysogenic to obligate lytic. We generated an oligo that would introduce two

stop codons into the early part of the cI gene to render it nonfunctional and used one cycle of MAGE (Wang et al. 2009; Wang and Church 2011) to introduce it into each desired prophage genome. We used a slightly modified MAGE protocol than previously described. First, we grew up each lysogen overnight at 30 °C with carbenicillin. Then, we inoculated three replicate tubes of 3 mL LB with 100 μ L of overnight culture and 4 μ L of carbenicillin and incubated for 1 hour at 30 °C. Then, we added 20 μ L of 1M arabinose to each tube and incubated for another hour at 30 °C. Aliquots of 1 mL were then pelleted and washed three times with ice cold, sterile, nanopure water to remove media and salt residue. Then, oligos were added at a final concentration of 5 μ M and cells were electroporated. Finally, after electroporation cells were recovered in fresh media and antibiotic overnight. The stop codons were inserted with an oligo containing two different stop codons at separate locations:

C*G*C*A*CGGTGTTAGATATTTATCCCTTGCGGTGATAGATTTAACGTATGTGAACA
AA AAAGTAACCATTAACACAAGAGCAGCTTGAGGAC. Then, cells were recovered overnight, and in cells that successfully integrated the cI knockout oligo, subsequent cell divisions diluted out the cI protein to the point that λ initiated its lytic cycle. Since the cI gene was knocked out, the resulting phage could not re-integrate into the E. coli chromosome.

To generate the fast-evolved and stable-evolved genotypes, we began with the lysogenic strains that had been previously edited to contain the key N1107K mutation that confers OmpF-use (Strobel et al. 2022). To make these genotypes amenable to our suppression experiments, we generated obligately lytic versions using our cI knockout method described above.

4.3.3 Host E. coli Strains

For the suppression experiments, we used the *E. coli* strain REL606 (Jeong et al. 2009) as the host bacterium. For detection of phages that evolved to use the OmpF receptor during the suppression experiments, we used the *LamB*⁻ strain (JW3996) from the Keio collection (Baba et al. 2006). The wild type parent of the Keio collection (BW25113, referred to as “WT” throughout this manuscript) was used for estimating phage titer during the suppression, growth rate, and stability assays.

4.3.4 Sanger sequencing

We sequenced the reactive region of the *J* gene (approximately position 2650 to 3399 of 3399 total bp) to identify mutations in genotypes that evolved to be OmpF⁺ in the coevolution experiment. We also sequenced the *cI* gene to verify successful knockouts.

For sequencing the *J* gene we used primers: Forward 5' CCT GCG GGC GGT TTT GTC ATT TA; Reverse 5' CGC ATC GTT CAC CTC TCA CT and sent unpurified PCR products to Genewiz La Jolla, CA, for sequencing with the reverse primer.

For sequencing the *CI* gene use primers: Forward 5' CGA CCA GAA CACCTT GCC 3'; Reverse 5' CCC TTG CGG TGA TAG ATT TAA CG 3' and sent unpurified PCR products to Genewiz La Jolla, CA, for sequencing with both forward and reverse primer. All sequences were aligned to the appropriate reference using Unipro UGENE v1.31.1 (Okonechnikov et al. 2012).

4.3.5 Stability, reproductive rate, and evolvability measurements

Stability and net reproductive rates for the fast-stable, fast-evolvable, and stable-evolvable genotypes were measured in a previous study (Strobel et al. 2022). In the previous

study, stability was measured as the rate at which phage lost infectivity in media alone (i.e. no host cells), and net reproductive rate (called simply “growth rate” in Strobel et al. 2022) was measured as the rate of increase in phage titer when incubated with permissive host cells. The latter is actually a combined rate of reproduction and decay, so for the current study we obtained just the rate at which phage reproduce (i.e. factoring out decay) by subtracting the decay rate from the combined rate (i.e. reproductive rate = net reproductive rate – decay). Evolvability was measured as the ability of the phage to evolve to use the non-native OmpF receptor. Several measures of evolvability were previously reported: the number of replicate populations that evolved to use OmpF, the average number of days required to evolve OmpF, or the number of mutations required to use OmpF. All metrics were correlated and so here we only report the evolutionary path length to gaining OmpF use. Fewer mutations, or a shorter evolutionary path, corresponds to higher evolvability (Kirschner and Gerhart 1998) (Figure 4.1).

For the genotypes that we edited to receive the N1107K mutation (fast-evolved and stable-evolved), we measured net reproductive rates of the fast-evolved and stable-evolved immediately after generating the obligate lytic versions. Following the protocol in Strobel *et al.* 2022, we picked a single plaque of each genotype into 100 μ L of M9 Glucose and then divided the volume of 100 μ L into the three replicate 50ml flasks with M9 Glucose and 0.01M MgSO₄. We then measured initial phage titers by diluting the flask contents and plating with WT cells infused in soft agar. Then, we added $\sim 10^8$ REL606 cells to each flask and incubated at 37 C shaking for 4 hours. After incubation, samples were taken from each flask, filtered through 0.22 μ M filters to remove bacteria, and phage titers were re-measured by diluting in M9 glucose + MgSO₄ and plating with WT cells infused in soft agar. We chose to measure growth over 4 hours rather than 24 hours (the length of time between transfers in the evolution experiment) to

reduce the possibility that genotypes would evolve mutations during the growth experiment. We did not measure decay rates of the fast-evolved and stable-evolved genotypes, so we report net reproductive rates for all four genotypes in Figure 4.A.3.

Because we measured the net reproductive rates of the fast-evolvable and fast-stable in the previous study (lysogenic versions) and here (obligate lytic versions), we were able to assess whether knocking out the *cI* gene to make the obligate lytic versions altered net reproductive rate. We did not find a significant effect (2-sample t-test, $n = 3$ per genotype; Fast-Evolvable: $t\text{-stat} = -0.110$, $df = 4$, $p = 0.918$; Stable-Evolvable: $t\text{-stat} = 0.566$, $df = 4$, $p = 0.602$).

4.3.6 Bacterial Suppression Experiments

To determine which phage genotypes would best suppress REL606, we inoculated six 50-mL flasks per phage genotype with 10 mL modified M9-Glucose and 0.01M MgSO₄ and 10^6 bacterial cells and approximately $10^5 - 10^6$ phage particles (exact values reported in Table 4.A.2). Flasks were incubated at 37 °C, shaking at 120 rpm. After 24 hours, 100 μ L of each community was transferred into new flasks with 10 mL of fresh media. Flasks were passaged for 10 days for the first two suppression experiments (Figure 4.2 and Figure 4.3) and for 6 days in the third suppression experiment (Figure 4.4). Each day, 1 mL aliquots were removed to estimate bacterial and phage densities, as well as to freeze communities with glycerol for later analysis (40 μ L 80% glycerol per 200 μ L sample). To assess bacterial titers, aliquots were diluted in M9-Glucose and spot plated on LB agar. For phages, 1-mL aliquots were centrifuged (1 min at $15,000 \times g$) to pellet cells. Then, supernatants were serially diluted in M9-Glucose, and 2- μ L aliquots were spotted on a lawn of REL606 infused in soft agar to obtain phage titers.

4.3.7 Bioinformatic prediction of J structure

We used the publicly available version (Mirdita et al. 2022) of AlphaFold (Jumper et al. 2021) to predict the structure of the reactive region of the wild type J domain containing the 173 most C-terminal amino acids. We used Chimera to create Figure 4.1A (Pettersen et al. 2004).

4.3.8 Statistical Tests and Plotting

We used RStudio to create Figures 4.2, 4.3, 4.4, 4.A.1, 4.A.2, and 4.A.4 and carry out statistical tests in this manuscript. We used a two-sided Wilcoxon Rank Sum test and α value of 0.05 for statistical significance between replicates of pairs of genotypes on a single day. We used MATLAB to create Figure 4.1B and 4.A.3. For statistical comparisons of trait values, we used two-sample T-tests after verifying equal variances.

4.4 Results

Of the three phage genotypes, the fast-evolvable proved to be most suppressive (Figure 4.2A, 4.A.1). There are statistical differences in the bacterial densities between the fast-evolvable and fast-stable genotype on days 2, 3, 4, 5, and 6 and the fast-evolvable and fast-stable genotype on days 2, 4, 5, and 6 (Table 4.A.3) (Figure 4.2B-C). The other two genotypes, fast-stable and stable-evolvable, appear to be equally poor at suppressing and are not statistically different from each other on any day (Table 4.A.3) (Figure 4.2D). Although there are timepoints where there are not significant differences in suppression, at least one replicate of this genotype was the most suppressive on all ten days of the experiment (Figure 4.2B, 4.2C). Because two of the three genotypes suppressed so poorly, it is difficult to draw conclusions about all three traits, but it is at least clear that stability is not of paramount importance for suppression. It appears that the reproductive rate of the phage and its evolvability may both facilitate bacterial suppression.

This result led to a new question: does a fast reproductive rate have its own, independent effect on suppression by increasing the number of phage particles relative to the number of bacteria, or is increased reproduction just a second way of enhancing evolvability by increasing the number of generations and opportunities to evolve counter defenses? If the reproductive rate is just a function of how many opportunities the phage has to overcome resistance mutations, then perhaps the low reproductive rate of the stable-evolvable genotype prevented it from achieving sufficient replications to generate the mutations necessary to adapt to the resistant bacteria by acquiring activity on OmpF. To parse out the independent effect of reproductive rate, we edited in a mutation (N1107K) that conferred the ability to use OmpF in both the fast-evolvable and stable-evolvable genotypes (Strobel et al. 2022). By doing so, we hoped to remove evolvability from the equation by artificially making both genotypes ‘evolved’ with respect to OmpF use. We then repeated the suppression experiment with the two evolved genotypes, and the fast-evolved genotype was substantially more suppressive than the stable-evolved (Table 4.A.3, Figure 4.3.) In fact, not only did the stable-evolved genotype not suppress the bacteria, it did not grow fast enough to keep up with the dilution from the daily transfer, and every replicate of this genotype lost phage entirely after day two (Figure 4.A.2). Because the stable-evolved genotype was even worse at suppressing than the stable-evolvable (i.e., before receiving the OmpF⁺ granting mutation), we measured the growth rates of the evolved versions of each genotype to verify that N1107K did not introduce an unexpected fitness cost in the stable-evolvable background. It did not, although it did cause a significant gain in growth rate (combined reproductive rate + decay rate) in the fast-evolvable background (Figure 4.A.3). Together, these results indicate that reproductive rate has an independent effect on suppression, and evolvability on its own does not confer high suppression.

Thus far, these results suggest that evolvability and reproductive rate are important criteria for suppression, but stability did not appear to predict suppression. One possible explanation for this result is that the laboratory environment of our experiments is artificially permissive of instability, compared to the more challenging environment where therapeutic phage would need to be deployed, such as inside a patient's body. In our experiment, phage had access to a homogeneous population of rapidly growing hosts. Inside a patient's body, by contrast, phage would need to survive in a spatially complex environment in which their optimal host would be intermixed with numerous other microbes from the host's microbiome. In that environment, stability might be a better predictor of suppression. To evaluate this hypothesis, we repeated the suppression experiment exactly as was done in Figure 4.2, except we decreased the amount of glucose, a limiting resource of the bacteria, by ten-fold. This lowered the carrying capacity of the bacteria in the flask, thereby decreasing the density of host cells and increasing the amount of time that phages spent in the external environment between hosts, undergoing decay. We hypothesized that under these more challenging conditions, the fast-evolvable phage might not be more suppressive than the fast-stable phage. Consistent with expectations, the fast-stable phage is the most suppressive after one day (Figure 4.4, Figure 4.A.4, Table 4.A.3). However, by day three two thirds of the replicates of the fast-evolvable phage were the most suppressive, and by the end of the experiment it was all six replicates (Figure 4.4A-C, Table 4.A.3). This finding suggests that evolvability can have a stronger effect than stability for determining suppression over the long term bacteria over the long term, even in the more challenging condition where stability should be favored.

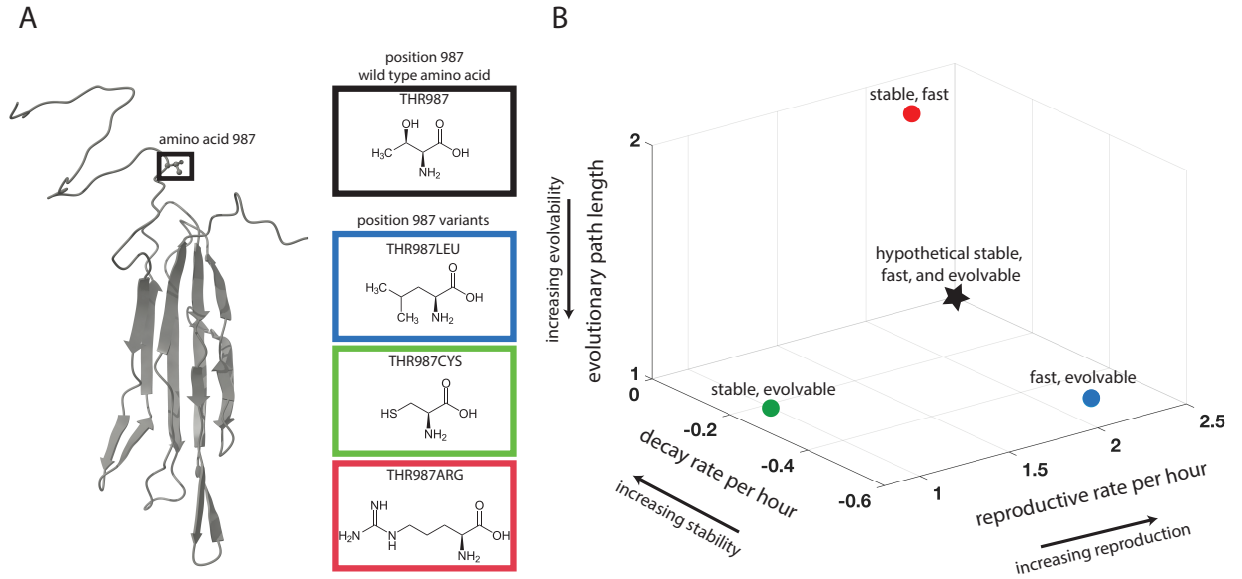


Figure 4.1 - A trio of closely related phage genotypes demonstrate a three-way trade-off between stability, reproduction, and evolvability. Panel A: AlphaFold prediction of the domain of the λ receptor binding protein that determines host range. The three λ genotypes in this study were identical except for a single amino acid difference in this domain (black box). Insets show the wild type amino acid and three variant amino acids. Panel B: Three-dimensional plot of phage trait values. Stability is measured by decay rate, the rate at which phage lose infectivity in an environment lacking hosts. Reproduction is measured by reproductive rate, the rate at which phage replicate on their host bacteria, adjusted to account for the phage lost to decay. Evolutionary path length is the number of mutations required to infect through the non-native receptor. It is the inverse of evolvability because a genotype requiring fewer mutations to achieve a new function is more evolvable. The position on the graph corresponding to optimality of all three traits is indicated by the star in the lower back corner. No phage was able to optimize all three traits.

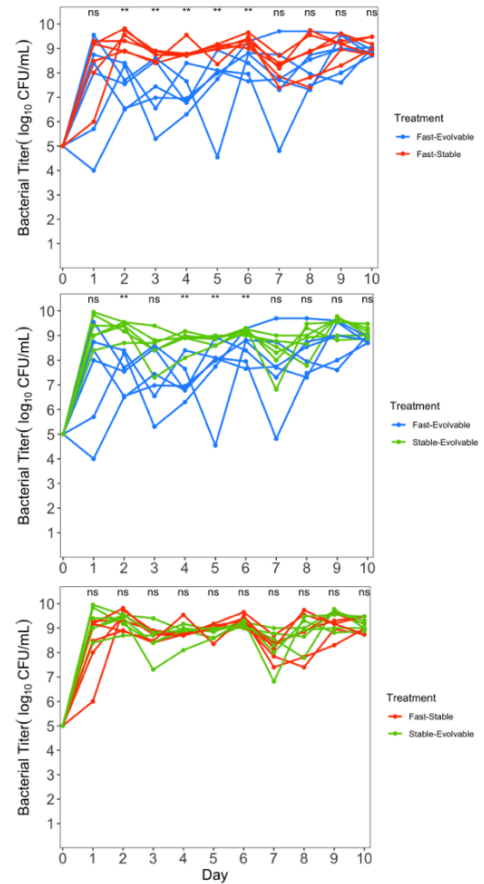
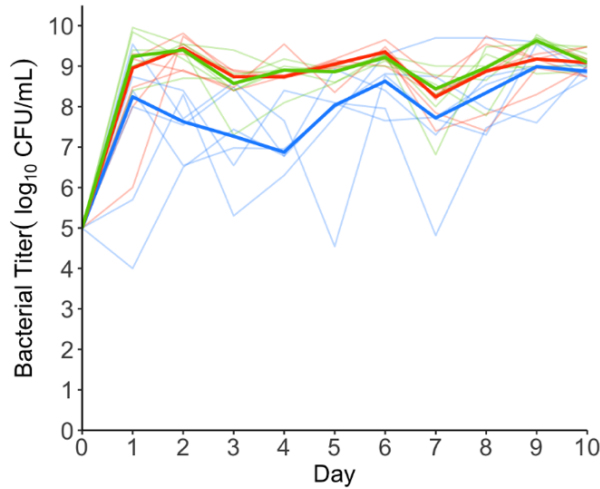


Figure 4.2 - λ suppression of bacteria monitored daily for 10 days. Each line corresponds to the bacterial titer in a single replicate flask population. Six replicate flasks were initiated for each phage genotype. Across all genotypes, the same bacterial strain was used to initiate the flask and approximately the same ratio of phage and bacteria were added. Panel A: All three phage genotypes are shown together. Median lines for bacterial population replicates A-F suppressed by a given genotype are shown in bold, and individual populations are shown by translucent lines. Panel B-D: pairs of genotypes are shown for ease of visualizing differences. Statistical differences are present on days 2-6 between the Fast-Evolvable replicates and Fast-Stable replicates, days 2, 4, 5, and 6 between the Fast-Evolvable replicates and Stable-Evolvable replicates, and statistical differences were not detected between the Fast-Stable replicates and Stable-Evolvable treatments at any time (Table 4.A.2). A Wilcoxon Ranked Sum Test was used to test statistical significance between levels of suppression between two phage genotypes.

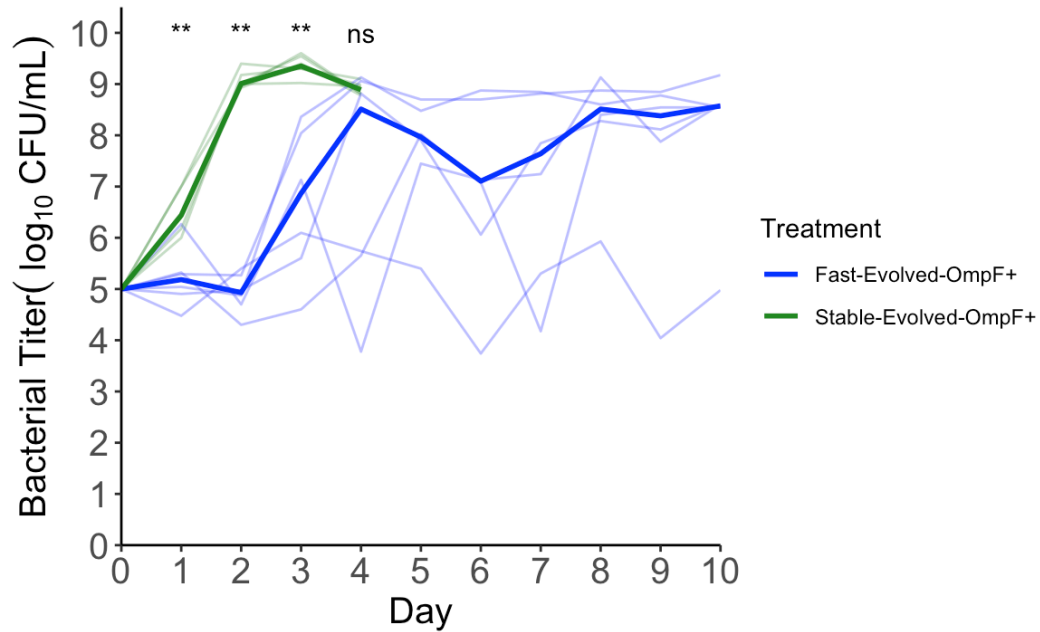


Figure 4.3 - Bacterial population dynamics exposed to two λ genotypes differing only in their reproductive rates. Both genotypes were engineered to contain a key mutation that conferred activity on the OmpF receptor, allowing infection of bacteria that are resistant to OmpF⁻ phage. Six replicate flask populations were initiated for each phage genotype, with the same bacterial strain. Panel A: Median lines for bacterial population replicates A-F of a given genotype are shown in bold, and individual populations are shown by translucent lines. Each translucent line corresponds to the bacterial titer in a single replicate flask population. The stable, slow reproducing phage (green) poorly suppressed the bacteria and the phages passed below our limit of detection after the first day. We discontinued the stable-evolved replicates after three days of no phage detection. Statistical differences between replicate populations are present on days 1, 2, and 3 (Table 4.A.2). A Wilcoxon Ranked Sum Test was used to test statistical significance between levels of suppression between the two phage genotypes.

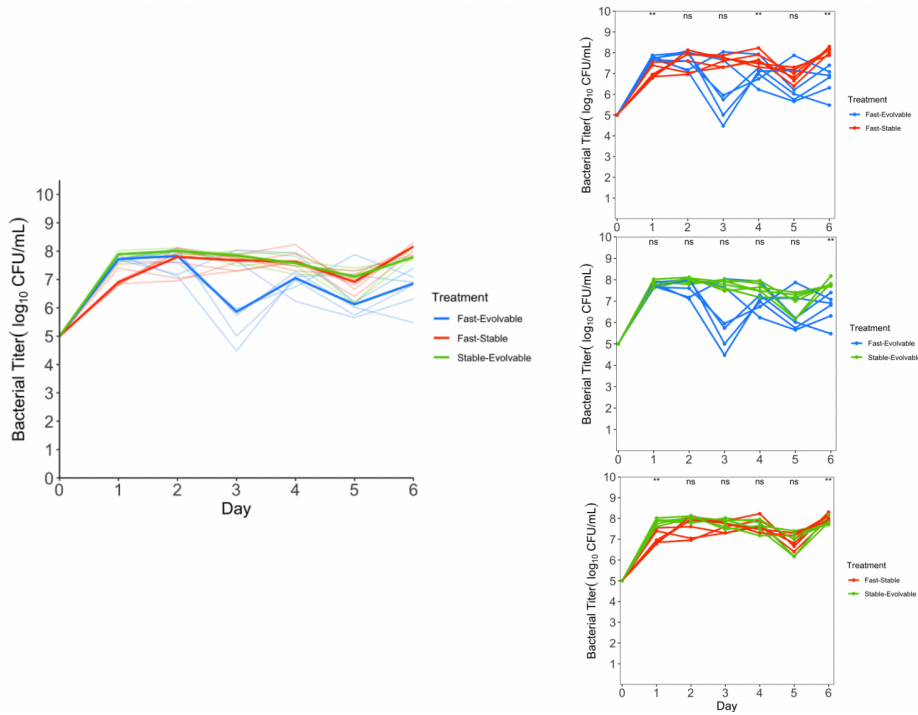


Figure 4.4 - Environmental contingency of bacterial suppression dynamics. The conditions of this experiment were identical to those of Figure 4.2, except that the available glucose was reduced by ten-fold, limiting the maximum potential bacterial carrying capacity throughout the duration of the experiment. Panel A: All three phage genotypes are shown together. Median lines for bacterial population replicates A-F suppressed by a given genotype are shown in bold, and individual populations are shown by translucent lines. Panel B-D: pairs of genotypes are shown for ease of visualizing differences. Statistical differences are present on days 1, 4, and 6 between the Fast-Evolvable replicates and Fast-Stable replicates, day 6 between the Fast-Evolvable replicates and Stable-Evolvable replicates, and days 1 and 6 between the Fast-Stable replicates and Stable-Evolvable replicates (Table 4.A.2). A Wilcoxon Ranked Sum Test was used to test statistical significance between levels of suppression between two phage genotypes.

4.5 Discussion

The aim of this study was to evaluate which phage trait is most predictive of bacterial suppression. We hypothesized that evolvability would be most predictive, and therefore the fast-evolvable and stable-evolvable genotypes would be most suppressive. The first suppression experiment with all three genotypes revealed that a single phage, the fast-evolvable genotype, was suppressive, while the fast-stable and stable-evolvable genotypes were not. This somewhat unexpected result validated the importance of evolvability over stability but also suggested that a fast reproductive rate, in addition to evolvability, is critical for suppression. In a follow-up experiment, growth rate had its own independent effect after controlling for evolvability. Finally, having shown that stability was the least predictive of suppression, we asked whether our experimental design favored that outcome by using conditions that are artificially permissive to unstable genotypes. To alter our experiment to better reflect the more challenging environment with fewer permissive hosts, we repeated the first suppression experiment using only 10% of the glucose, limiting the bacterial carrying capacity of the flask. Initially, the fast-stable phage was most suppressive, validating the hypothesis that decreasing host availability penalized unstable phage. However, by the end of the experiment the fast-stable phage had become non-suppressive, while the fast-evolvable phage had become suppressive. We confirmed that the gain in suppression was due to the fast-growing evolvable phage achieving robust growth on OmpF much earlier than the other genotypes. These findings emphasize the importance of considering not only the current traits of therapeutic phages, but also their capacity for adaptive evolution. The reversal of suppression ability across a multi-day timescale suggests the importance of measuring suppression over an entire course of treatment since resistance and counter-resistance traits only take days to evolve.

Although the need to consider the future evolutionary potential (i.e. evolvability) has not yet permeated *in vivo* phage therapy practice, numerous studies have demonstrated that providing an evolutionary advantage to phage generates more suppression. In a technique called ‘phage training’, an evolutionary leg-up is given by coevolving a candidate phage with the target bacteria, allowing an evolutionary arms race to play out, and then isolating an evolved phage strain for use against the naïve bacteria (Laanto et al. 2020; Borin et al. 2021). Another strategy aims to enhance the evolution of more suppressive phage by increasing opportunities for recombination among different strains, allowing beneficial mutations to be shuffled into a single, highly suppressive genotype (Burrowes et al. 2019). Our study raises the possibility that there is substantial genetic variation in evolvability, even among closely related genotypes. Relatively simple laboratory experiments on existing collections of phages might identify genotypes with unusually high evolutionary potential.

With the rise of genetic engineering technology, there is increasing interest in designing phage with the traits deemed desirable for suppression. Our study sheds light on some possible pitfalls that might frustrate such attempts. In the phage variant library codon 987 was targeted because it had mutated during selection for increased stability (Strobel et al. 2022). Seeking additional variants with a range of stabilities, different amino acids were edited in at codon 987. The assumption was that altering the amino acid at codon 987 affected only stability and evolvability but might not have obvious pleiotropic effects on other traits. This appeared to be true for most variants; however, the 987CYS (the stable-evolvable genotype used in this study) had a markedly reduced reproductive rate. This result demonstrates that a genotype that seems to “break” one tradeoff (stability vs. evolvability) might pay a cost in another trait (reproductive rate). Had multiple axes of variation not been considered, this seemingly tradeoff-breaking phage

might have appeared to be an ideal therapeutic phage, but the current study demonstrated the opposite result. Better outcomes might be achieved from ‘bioprospecting’ naturally occurring phages or using directed evolution to evolve enhanced genotypes, rather than attempting to design optimal genotypes, because selection should penalize genotypes with low fitness.

There are several limitations to the current study. First, we examined only three phage genotypes, and their genomes were identical except for a single codon. The evolutionary history and idiosyncrasies of this experimental system could prevent our results from being applied generally across the vast diversity of phages. In λ , there is a known tradeoff between stability and evolvability (Strobel et al. 2022) that might not exist in other phages or might be less pronounced. In other phages, different combinations of traits might produce tradeoffs, and our study provides a general framework for understanding how tradeoffs limit the optimization of therapeutic phages. Another limitation is that we studied suppression dynamics in flasks under controlled laboratory conditions and used a single phage and a single bacterial host. These conditions are unlike the environment in which therapeutic phages would be deployed. The human gut, for example, where λ -like phages might be used, is replete with myriad host-associated microbes in addition to the target, many of which would likely be unavailable as prey for the therapeutic phage (Lozupone et al. 2012). We began exploring this dimension with our limited glucose environment, and our results were robust to the perturbation, but we did not test the effect of non-host microbes. Phage may also encounter extremes of temperature, pH, and chemicals in vivo that would penalize unstable genotypes (Blazanin et al. 2022). Despite these differences between λ 's natural environment and laboratory conditions, comparisons between laboratory and natural populations of λ revealed that the sites that receive mutations that drive host-range expansion in the laboratory are present in natural populations, suggesting that

laboratory studies are informative for understanding natural dynamics (Maddamsetti et al. 2018). Prior work specific to phage therapy suggests that despite differences between in vitro and in vivo conditions, similar evolutionary dynamics can play out in vitro and in vivo, validating the use of in vitro experiments to inform clinical practices (Castledine et al. 2022).

This work demonstrates the important role of understanding evolutionary biology in phage therapy. Although phage have properties that are similar to chemical therapeutics, it is critical to remember that they are biological entities and have their own ability to propagate and evolve. When using directed evolution to create phage therapeutics, phage must be selected to evolve the properties that enable them to suppress bacteria while keeping in mind that evolving or engineering one desirable phenotypic characteristic may come at the cost of sacrificing another characteristic due to tradeoffs. Tradeoffs are difficult to avoid in biological systems, so evaluating traits in the context of tradeoffs is necessary. Considering evolutionary potential alongside conventional attributes like stability and reproductive rate allowed us to predict the genotype that best suppressed bacterial population across a multi-day timeframe.

4.6 Acknowledgements

We thank Joshua Borin for advice on experiments and feedback on the manuscript.

Chapter 4 in full, is a reprint of the material currently being prepared for publication. Horwitz EK, Strobel HM, Meyer JR (2022). Evolvability is a key trait in the selection of bacteriophages for therapeutic use. The dissertation author was one of the primary investigators and the second author of this paper.

4.A Appendix

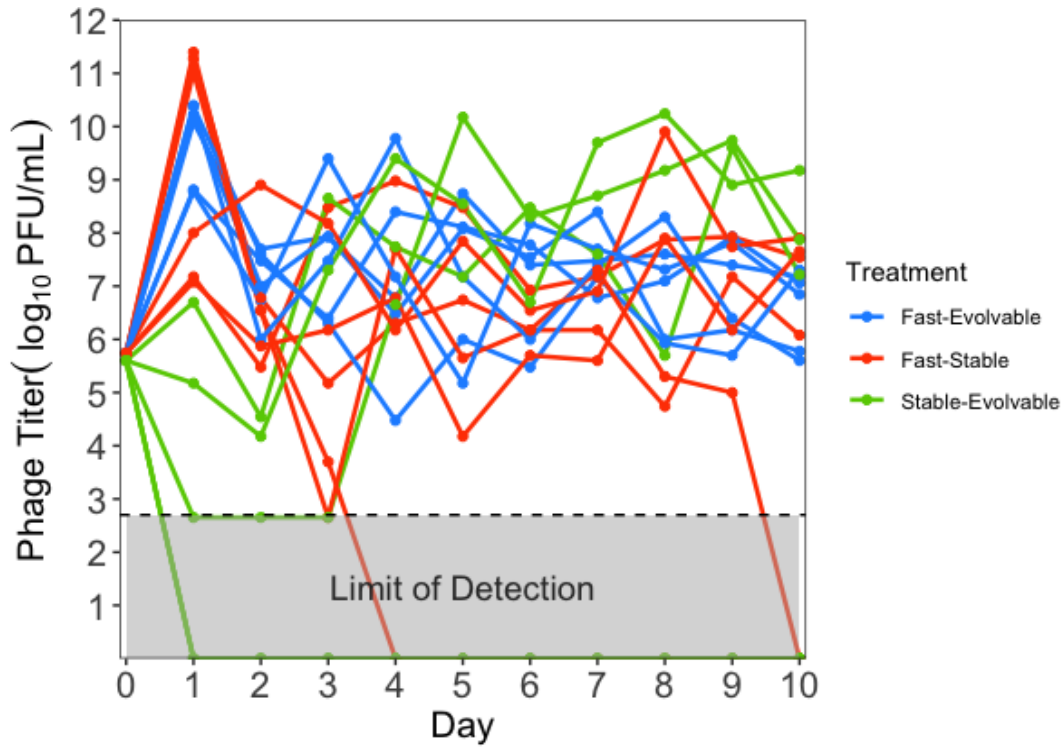


Figure 4.A.1 - Phage Titer from Figure 4.2
Phage densities were measured each day for all replicates of all 3 genotypes.

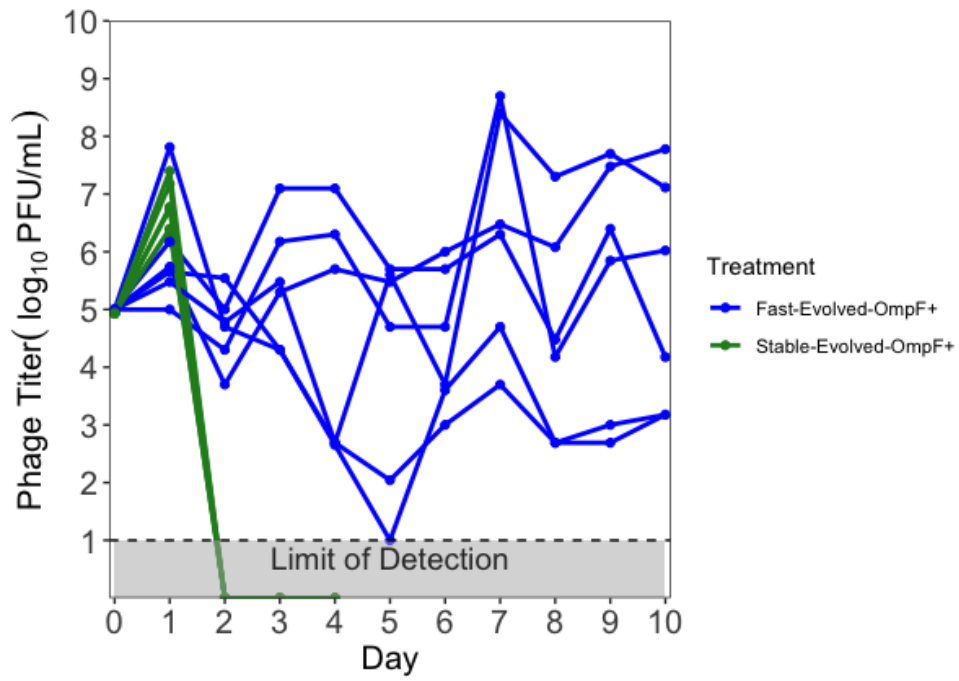


Figure 4.A.2 - Phage Titer from Figure 4.3
 Phage densities were measured each day for all replicates of both genotypes.

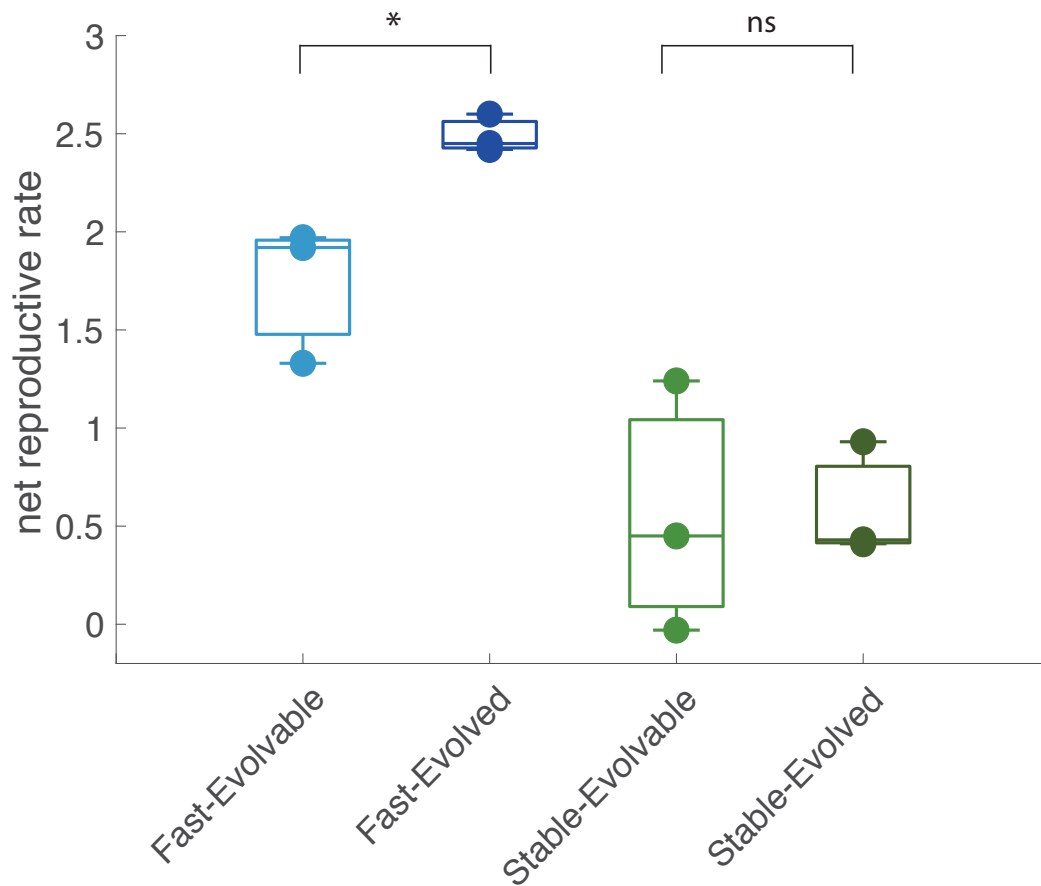


Figure 4.A.3 - Boxplots of net reproductive rate (reproductive rate + decay rate) of the fast reproducing, unstable and fast reproducing, stable genotypes without the final mutation (“evolvable”) and after receiving the final mutation via engineering (“evolved”). These genotypes were used in the suppression experiment from Figure 4.3. The net reproductive rate of Fast-Evolved was slightly higher than and Fast-Evolvable (2-sample T-test, $n = 3$ per genotype; $t\text{-stat} = -3.523$, $df = 4$, $P = 0.0244$), whereas the net reproductive rates of Stable-Evolvable and Stable-Evolved were indistinguishable (2-sample T-test, $n = 3$ per genotype; $t\text{-stat} = -0.09$, $df = 4$, $P = 0.933$).

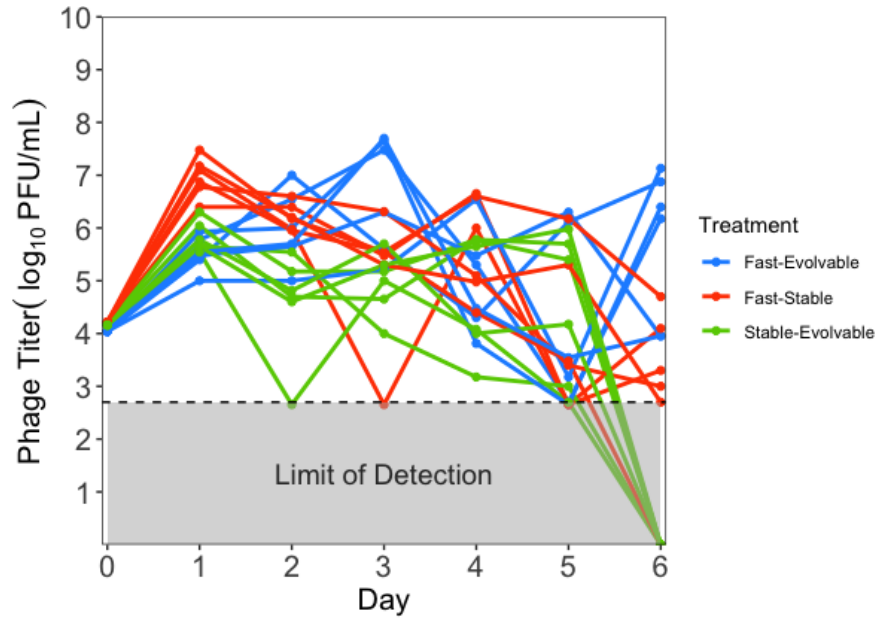


Figure 4.A.4 - Phage Titer from Figure 4.4
 Phage densities were measured each day for all replicates of each 3 genotypes.

Table 4.A.1 - Phage trait values used to create Figure 4.1A. Decay rates and evolutionary path length were first reported in Strobel, Horwitz, and Meyer 2022. Reproductive rates were computed from previously reported net growth rates and decay rates from the same paper.

genotype	net reproductive rate per hour	decay rate per hour	reproductive rate per hour	distance to adaptation (# mutations)
THR 987LEU (Fast-Evolvible)	1.606601368	-0.485	2.091601368	1
	1.828305097	-0.47	2.298305097	
	1.714115496	-0.511	2.225115496	
THR987ARG (Fast-Stable)	1.631190757	-0.155	1.786190757	2
	1.873172219	-0.138	2.011172219	
	1.968209044	-0.119	2.087209044	
THR987CYS (Stable-Evolvible)	0.798291867	-0.211	1.009291867	1
	1.041989095	-0.215	1.256989095	
	0.50178022	-0.22	0.72178022	

Table 4.A.2 - Phage titers used to initiate suppression experiments.

	Suppression Exp. 1 (Figure 4.2)	Suppression Exp. 2 (Figure 4.3)	Suppression Exp. 3 (Figure 4.4)
Phage Genotype	Phage added to flask replicates	Phage added to flask replicates	Phage added to flask replicates
LEU	5.50E+05	1.00E+06	1.10E+05
ARG	5.50E+05		1.65E+05
CYS	4.00E+05	8.50E+05	1.45E+05

Table 4.A.3 Statistics from Figure 4.2, 4.3, and 4.4

Figure 4.2			
comparison	day	W	P-Value
fast-evolvable vs. fast-stable	1	12.5	0.4217
	2	0	0.002165
	3	3.5	0.02447
	4	0	0.004847
	5	1	0.004329
	6	2	0.008658
	7	14	0.5887
	8	12	0.3939
	9	14.5	0.6298
	10	10	0.2273
fast-evolvable vs. stable-evolvable	1	5.5	0.05382
	2	0	0.004922
	3	7	0.09123
	4	1	0.007687
	5	5	0.04113
	6	4	0.03035
	7	13	0.4848
	8	10	0.2403
	9	8	0.1262
	10	7	0.09155
fast-stable vs. stable-evolvable	1	10	0.2281
	2	21	0.6863
	3	23	0.4673
	4	11	0.2928
	5	28.5	0.1087
	6	26	0.2248
	7	17	0.5211
	8	13.5	0.9372
	9	17	0.2281
	10	10	0.7457
Figure 4.3			
comparison	day	W	P-Value
fast-evolved vs. stable-evolved	1	2	0.01291
	2	0	0.002165
	3	0	0.004998
	4	13	0.468

Table 4.A.3 Statistics from Figure 4.2, 4.3, and 4.4 (continued)

Figure 4.4 comparison	day	W	P-Value
fast-evolvable vs. fast-stable	1	36	0.004922
	2	21.5	0.6304
	3	9	0.1727
	4	5	0.04113
	5	10	0.2403
	6	0	0.004998
fast-evolvable vs. stable-evolvable	1	9	0.1712
	2	10.5	0.2615
	3	8	0.132
	4	6	0.06508
	5	11	0.3095
	6	0	0.004998
fast-stable vs. stable-evolvable	1	0	0.004998
	2	10	0.2403
	3	9	0.1712
	4	20.5	0.7466
	5	16.5	0.8721
	6	32	0.02919

4.7 References

- Baba, T., T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. Datsenko, M. Tomita, B. Wanner, and H. Mori 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular systems biology* 2.
- Batinovic, S., F. Wassef, S. Knowler, D. Rice, C. Stanton, J. Rose, J. Tucci, T. Nittami, A. Vinh, G. Drummond, C. Sobey, H. Chan, R. Seviour, S. Petrovski, and A. Franks 2019. Bacteriophages in Natural and Artificial Environments. *Pathogens (Basel, Switzerland)* 8.
- Blair, J., M. Webber, A. Baylay, D. Ogbolu, and L. Piddock 2015. Molecular mechanisms of antibiotic resistance. *Nature reviews. Microbiology* 13.
- Blazanin, M., W. Lam, E. Vasen, B. Chan, and P. Turner 2022. Decay and damage of therapeutic phage OMKO1 by environmental stressors. *PloS one* 17.
- Bono, L., S. Mao, R. Done, K. Okamoto, B. Chan, and P. Turner. 2021. Advancing phage therapy through the lens of virus host-breadth and emergence potential. *Advances in virus research* 111.
- Borin, J. M., S. Avrani, J. E. Barrick, K. L. Petrie, and J. R. Meyer. 2021. Coevolutionary phage training leads to greater bacterial suppression and delays the evolution of phage resistance. *Proc Natl Acad Sci U S A* 118.
- Broncano-Lavado, A., G. Santamaría-Corral, J. Esteban, and M. García-Quintanilla. 2021. Advances in Bacteriophage Therapy against Relevant MultiDrug-Resistant Pathogens. *Antibiotics (Basel, Switzerland)* 10.
- Brown, E. and G. Wright 2016. Antibacterial drug discovery in the resistance era. *Nature* 529.
- Bull, J., B. Levin, and I. Molineux 2019. Promises and Pitfalls of In Vivo Evolution to Improve Phage Therapy. *Viruses* 11.
- Burrowes, B. H., I. J. Molineux, and J. A. Fraclick. 2019. Directed in Vitro Evolution of Therapeutic Bacteriophages: The Appelmans Protocol. *Viruses* 11.
- Casey, E., S. van, D., and J. Mahony. 2018. In Vitro Characteristics of Phages to Guide 'Real Life' Phage Therapy Suitability. *Viruses* 10.
- Casjens, S. R. and R. W. Hendrix. 2015. Bacteriophage lambda: Early pioneer and still relevant. *Virology* 479-480:310-330.
- Castledine, M., D. Padfield, P. Sierocinski, P. Soria, J. A. Hughes, L. Mäkinen, V. Friman, J. Pirnay, M. Merabishvili, D. de Vos, and A. Buckling. 2022. Parallel evolution of

- Pseudomonas aeruginosa* phage resistance and virulence loss in response to phage treatment in vivo and in vitro. *eLife* 11.
- Charbit , A., J. Clement , and M. Hofnung. 1984. Further sequence analysis of the phage lambda receptor site. Possible implications for the organization of the lamB protein in *Escherichia coli* K12. *Journal of molecular biology* 175.
- Dessau , M., D. Goldhill , R. McBride , P. Turner, and Y. Modis. 2012. Selective pressure causes an RNA virus to trade reproductive fitness for increased structural and thermal stability of a viral enzyme. *PLoS genetics* 8.
- Donadio, S., S. Maffioli , P. Monciardini , M. Sosio , and D. Jabes. 2010. Antibiotic discovery in the twenty-first century: current trends and future perspectives. *The Journal of antibiotics* 63.
- Edwards , K., G. Steward , and C. Schvarcz 2021. Making sense of virus size and the tradeoffs shaping viral fitness. *Ecology letters* 24.
- Favor, A., C. Llanos, M. Youngblut , and J. Bardales 2020. Optimizing bacteriophage engineering through an accelerated evolution platform. *Scientific reports* 10.
- Goldhill , D. and P. Turner 2014. The evolution of life history trade-offs in viruses. *Current opinion in virology* 8.
- Gordillo, A., FL and J. Barr. 2019. Phage Therapy in the Postantibiotic Era. *Clinical microbiology reviews* 32.
- Hejnowicz, M., U. Gągała, B. Weber-Dąbrowska, G. Węgrzyn, and M. Dadlez. 2014. The First Step to Bacteriophage Therapy - How to Choose the Correct Phage in J. Borysowski, R. Międzybrodzki, and A. Górski eds. *Phage Therapy: Current Research and Applications* Caister Academic Press, United Kingdom.
- Hyman, P. 2019. Phages for Phage Therapy: Isolation, Characterization, and Host Range Breadth. *Pharmaceuticals* (Basel, Switzerland) 12.
- Høyland-Kroghsbo, N., R. Maerkedahl , and S. Svenningsen 2013. A quorum-sensing-induced bacteriophage defense mechanism. *mBio* 4.
- Jeong , H., V. Barbe , C. Lee , D. Vallenet , D. Yu , S. Choi , A. Couloux, W. Lee S, S. Yoon , L. Cattolico, C. Hur, H. Park, B. Ségurens , S. Kim , T. Oh , R. Lenski , F. Studier , P. Daegelen, and J. Kim 2009. Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). *Journal of molecular biology* 394.
- Jumper, J., R. Evans, A. Pritzel , T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool , R. Bates, A. Židek, A. Potapenko , A. Bridgland , C. Meyer , S. Kohl , A. Ballard , A. Cowie , B. Romera-Paredes , S. Nikolov , R. Jain, J. Adler, T. Back , S. Petersen, D.

- Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska , T. Berghammer, S. Bodenstein, D. Silver , O. Vinyals , A. Senior , K. Kavukcuoglu , P. Kohli , and D. Hassabis 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596.
- Jurczak-Kurek, A., T. Gąsior, B. Nejman-Faleńczyk , S. Bloch , A. Dydecka , G. Topka , A. Necel, M. Jakubowska-Deredas , M. Narajczyk , M. Richert , A. Mieszkowska , B. Wróbel , G. Węgrzyn , and A. Węgrzyn 2016. Biodiversity of bacteriophages: morphological and biological properties of a large group of phages isolated from urban sewage. *Scientific reports* 6.
- Kirschner, M. and J. Gerhart. 1998. Evolvability. *Proceedings of the National Academy of Sciences of the United States of America* 95.
- Kutter, E., D. De Vos , G. Gvasalia , Z. Alavidze, L. Gogokhia, S. Kuhl , and S. Abedon. 2010. Phage therapy in clinical practice: treatment of human infections. *Current pharmaceutical biotechnology* 11.
- Laanto, E., J. Bamford, J. Laakso, and R. Sundberg L. 2012. Phage-driven loss of virulence in a fish pathogenic bacterium. *PloS one* 7.
- Laanto , E., K. Mäkelä, V. Hoikkala, J. Ravantti , and L. Sundberg 2020. Adapting a Phage to Combat Phage Resistance. *Antibiotics (Basel, Switzerland)* 9.
- Labrie , S., J. Samson , and S. Moineau. 2010. Bacteriophage resistance mechanisms. *Nature reviews. Microbiology* 8.
- Lozupone , C., J. Stombaugh , J. Gordon , J. Jansson , and R. Knight 2012. Diversity, stability and resilience of the human gut microbiota. *Nature* 489.
- Maddamsetti, R., D. T. Johnson, S. J. Spielman, K. L. Petrie, D. S. Marks, and J. R. Meyer. 2018. Gain-of-function experiments with bacteriophage lambda uncover residues under diversifying selection in nature. *Evolution* 72:2234-2243.
- Meyer, J. R., D. T. Dobias, J. S. Weitz, J. E. Barrick, R. T. Quick, and R. E. Lenski. 2012. Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science* 335:428-432.
- Mirdita, M., K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger. 2022. ColabFold - Making protein folding accessible to all.
- Okonechnikov, K., O. Golosova , and M. Fursov. 2012. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics (Oxford, England)* 28.
- Pettersen, E., T. Goddard, C. Huang, G. Couch, D. Greenblatt, E. Meng, and T. Ferrin. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* 25.

- Rakhuba, D., E. Kolomiets , E. Dey , and G. Novik 2010. Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell. Polish journal of microbiology 59.
- Singhal , S., G. Leon, CM, S. Whang, E. McClure, H. Busch, and B. Kerr 2017. Adaptations of an RNA virus to increasing thermal stress. PloS one 12.
- Strobel, H., E. Horwitz, and J. Meyer 2022. Viral protein instability enhances host-range evolvability. PLoS genetics 18.
- Wang , H. and G. Church. 2011. Multiplexed genome engineering and genotyping methods applications for synthetic biology and metabolic engineering. Methods in enzymology 498.
- Wang , H., F. Isaacs, P. Carr , Z. Sun , G. Xu , C. Forest , and G. Church. 2009. Programming cells by multiplex genome engineering and accelerated evolution. Nature 460.
- Zablocki, O., E. Adriaenssens , and D. Cowan 2015. Diversity and Ecology of Viruses in Hyperarid Desert Soils. Applied and environmental microbiology 82.

CONCLUSIONS AND FUTURE DIRECTIONS

This work focused on understanding whether viral traits affect the propensity to evolve expanded host range. Mutation rate, recombination, and stability are three intrinsic traits that are thought to influence host-range evolvability. In Chapter 1, I reviewed the current evidence for the role of each trait in enhancing viral evolvability. In my doctoral research, I focused on stability, the trait that is least understood in the context of viral evolvability. It has been proposed that viruses possessing stable proteins might be most evolvable (Ogbunugafor et al. 2009) because their proteins can tolerate more genetic change (Wagner 2012). However, another hypothesis is that viruses with unstable proteins might have more capacity for phenotypic heterogeneity (Tokuriki and Tawfik 2009) and therefore more easily acquire the new functions necessary to infect new hosts.

To approach this question, I used a well-studied host-range expansion of bacteriophage λ as a model system. In Chapter 2, I purified the receptor binding protein of bacteriophage λ before and after evolution to use a new receptor. Consistent with the heterogeneity hypothesis, the evolved proteins were unstable and contained distinguishable subpopulations of particles with distinct preferences for the wild type and new receptors.

In Chapter 3, I then examined the relationship between stability and evolvability among a library of closely related λ variants differing in stability of the receptor binding protein. We found that overall, the variants with higher stabilities evolved slower and required an extra mutation. The exact mutation differed between genotypes, but in all cases, it had a destabilizing effect, consistent with the hypothesis that instability can promote host-range expansion evolvability.

In Chapter 4, I identified a single genotype that was stable and evolvable. At first, this genotype appeared to have broken a stability-evolvability tradeoff that constrained the other genotypes. However, we then discovered that this genotype had a substantially reduced reproductive rate, indicating a three-way tradeoff between stability, evolvability, and reproduction. This finding illustrated how tradeoffs can prevent organisms from simultaneously optimizing suites of characteristics and led us to ask whether there are key traits that have a stronger influence than others in determining the outcome of inter-species evolutionary interactions. We found that phage with high evolvability and reproductive rate were more effective at suppressing populations of their host bacteria, even though they paid a cost of low stability.

Although increasing stability has enhanced evolvability in other, non-virus proteins (Bloom et al. 2006; Bloom and Arnold 2009), I showed that it inhibited evolvability in phage λ by constricting the conformational diversity of its receptor binding protein. These orthogonal results might call into question whether stability is an informative trait in predicting virus evolvability. In Chapter 1, we suggest a potential solution to these conflicting results: perhaps each protein has an optimal stability at which it is most evolvable, and extremes in either direction decrease evolvability. Testing this hypothesis will require studying a larger number of variants than I examined in the present work. Efforts to discover and sample virus genetic diversity, as well as rapidly improving genetic engineering tools, will facilitate this work.

References

- Bloom, J. D. and F. H. Arnold. 2009. In the light of directed evolution: pathways of adaptive protein evolution. *Proc Natl Acad Sci U S A* 106 Suppl 1:9995-10000.
- Bloom, J. D., S. T. Labthavikul, C. R. Otey, and F. H. Arnold. 2006. Protein stability promotes evolvability. *Proc Natl Acad Sci U S A* 103:5869-5874.
- Ogbunugafor, C., R. McBride, and P. Turner. 2009. Predicting virus evolution: the relationship between genetic robustness and evolvability of thermotolerance. *Cold Spring Harbor symposia on quantitative biology* 74.
- Tokuriki, N. and D. S. Tawfik. 2009. Protein dynamism and evolvability. *Science* 324:203-207.
- Wagner, A. 2012. The role of robustness in phenotypic adaptation and innovation. *Proceedings. Biological sciences* 279.