# Hypermutation and Adaptation of Experimentally Evolved Marine *Vibrio* Bacteria

by

Sean Aidan Clarke

B.A., University of Texas at Austin (2000)
B.S.M.E., University of Texas at Austin (2000)



Submitted to the Department of Biological Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Biological Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2013



Author .....
Department of Biological Engineering
May 23, 2013

Certified by .....
Eric J. Alm
Associate Professor of Biological Engineering
Thesis Supervisor

Accepted by .....
Forest M. White
Associate Professor of Biological Engineering
Chair, Graduate Program Committee

Thesis Committee

Professor Edward F. DeLong
Chair, Thesis Committee
Morton and Claire Goulder Family Professor in Environmental Systems
Biological Engineering and Civil & Environmental Engineering

Professor Eric J. Alm
Thesis Supervisor
Karl Van Tassel (1925) Career Development Associate Professor
Biological Engineering

Professor Christopher J. Marx
Associate Professor of Biology
Department of Organismic and Evolutionary Biology
Harvard University

**Hypermutation and Adaptation of Experimentally Evolved Marine *Vibrio* Bacteria**

by

Sean Aidan Clarke

Submitted to the Department of Biological Engineering
on May 23, 2013, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Biological Engineering

## Abstract

Environmental bacteria display tremendous genetic diversity, but we are still learning how this diversity arises and relates to their wide range of habitats. Investigating how bacteria adapt helps us understand their contributions to environmental processes and informs forward engineering of bacteria for industrial applications. Experimental evolution is a powerful approach, with microbes especially, but it has mostly been applied to model organisms and metabolic functions.

In the work here, we investigated the possibility, degree, and variability of adaptation of an environmental *Vibrio* strain by applying a little-used selection method appropriate to a relevant condition, salinity. We successfully isolated mutants with higher salt tolerance by selecting on salt gradient plates. Resequencing the genomes of the evolved strains revealed unprecedented hypermutation in three of nine parallel lineages. These mutator lines arose independently, and each of them accumulated more than 1500 single-base mutations. By comparison, there are only 302 single-base differences between the ancestor strain and another strain isolated in the wild. Hypermutation was associated with a deletion resulting from improper prophage excision. Members of this family of prophages are found in other proteobacteria, including well-studied human pathogens, from very different environments.

Mutators are known to arise spontaneously in wild and clinical bacteria, but the extent of their adaptive contribution is unknown. We have preliminary evidence that this mechanism of evolution could be relevant in the environment, where horizontal gene transfer and mobile elements play known, significant roles in bacterial evolution.

Thesis Supervisor: Eric J. Alm
Title: Associate Professor of Biological Engineering

## Acknowledgments

So many people have supported me in my years at MIT. I would like to recognize and thank many of them here.

My committee, for professional encouragement and empathy:
Ed, for steady reinforcement and insightful suggestions.
Chris, for compassion, encouragement, and sharing his love of experimental evolution.
Eric, for being ever-exuberant and positive with his support, and for showing me how to unite Borges and biological engineering.

In Parsons Lab, a uniquely interdisciplinary, exciting, and friendly place to work:

The original Almstars: Jesse, Lawrence, Sonia
The rest of the Alm lab, past and present: Arne, Yonatan, Eli, Sarah, Ali, Albert, Inês, Matt, Mark, Chris, Greg, Ilana, Paco, Manu, Spence, Sü, Scott
The Parsons pyramid of power and support: Jim, Sheila, Vicki
The Polz/Thompson/DeLong/Chisholm/Alm power pentagon of environmental microbiology folk who go out of their way to be helpful and share resources and knowledge.
phage buddies: Kathryn, Katya, Simon
other fans of environmental bacteria: Michael, Alison, Otto, Fatima, Diana, Jessie
Jeff, Ben, Dave, Amy, Matt, Mike, Bec, Gaj, Libusha, Melissa
the 212/208, quads

BE:
I have also been fortunate to be part of the supportive and innovative Biological Engineering community at MIT.

Irrepressible, inventive friends: JB, Marcio, BE 2005, ex-members of the Endy lab, Nate T., Sarah K.
Tremendously encouraging faculty and staff: Doug Lauffenburger, Dalia Fares, Drew Endy, Natalie Kuldell, Jaime Goldstein

Around and outside MIT:

My parents, John and Donna, for nurturing my love of science and all other learning
Val
Dan and Jess
Drs. Lisa, Dan, Allie
Steph
Prof. Alan Grossman, for frank and patient discussion of how to finish a Ph.D.
Random Hall: Nina and Chris, 4 years of Randomites
yoga friends, for keeping me balanced: Checka, Whitney, Anna, Bruce, Johanna
Carissa Y., for nudges and feedback on thesis writing
my UROP, Brian Ross
Intramural softball and air pistol teammates

# Contents


**1 Introduction** **15**

1.1 Introduction . . . 15

1.2 Outline . . . 16

**2 Experimental evolution of salt tolerance: Phenotypic adaptation** **17**

2.1 Introduction . . . 17

2.1.1 Experimental evolution and bacteria–advantages and prior work's limited scope . . . 17

2.1.2 Vibrios are a rising model of evolutionary ecology . . . 18

2.1.3 Lack of forward evolution studies for salinity, a relevant environmental factor . . . 18

2.1.4 Drawbacks to typical experimental evolution methods for this case . . . 18

2.1.5 Gradient plates . . . 18

2.1.6 Outputs . . . 19

2.2 Results/Discussion . . . 19

2.2.1 Gradient plates and selection scheme methodology . . . 19

2.2.2 Mutant colonies can be selected on gradient plates . . . 21

2.2.3 Magnitude: Mutant strains can improve salt tolerance 50% over ancestor 21

2.2.4 Parallelism: Improvement varies across replicate lines . . . 22

2.3 Discussion . . . 22

2.3.1 Patterns of phenotypic adaptation . . . 22

2.3.2 Extending prior niche work . . . 22

2.3.3 Salt and natural Vibrios . . . 22

2.3.4 Fast adaptation in the wild . . . 23

2.3.5 Advantages of gradients . . . 23

2.4 Conclusions . . . 23

2.5 Methods . . . 24

2.5.1 Bacterial strains . . . 24

2.5.2 Cultures and growth conditions . . . 24

2.5.3 Gradient plates . . . 24

2.5.4 Selection cycles . . . 25

2.5.5 Growth in variable salt levels . . . 25

**3 Experimental evolution of salt tolerance: Genome resequencing** **27**

3.1 Introduction . . . 27

3.1.1 Experimental evolution and resequencing . . . 27

3.1.2 Parallel molecular evolution common in past studies . . . 28

3.1.3 Salinity tolerance mutations unknown and may be broader, search is enabled by WGS . . . 28
3.1.4 Mechanisms of adaptive mutation vary, with roles for mobile DNA, loss of function, regulation . . . 28
3.1.5 Applications of re-sequencing information . . . 29
3.2 Results . . . 29
3.2.1 Genome sequencing of forty-five mutant strains . . . 29
3.2.2 Three lines are mutators and far more single base mutations were observed than expected . . . 30
3.2.3 Parallelism at the locus level exists, but is not significant . . . 32
3.2.4 Non-synonymous and synonymous mutations are both common . . . 34
3.2.5 Mutators share an exclusive deletion feature . . . 36
3.2.6 Mutations in each round vary widely within nonmutator lines . . . 38
3.2.7 Transition mutations greatly predominate, especially in rounds with more mutations . . . 39
3.2.8 Mutation data support a model using a mix of low and high transition probabilities . . . 40
3.2.9 Salt tolerance does not correlate with SNP totals . . . 40
3.3 Discussion . . . 42
3.3.1 Mutators . . . 42
3.3.2 Parallel mutations . . . 42
3.3.3 Salt tolerance mutations . . . 43
3.4 Conclusion . . . 43
3.5 Methods . . . 44
3.5.1 DNA Extraction, Library Preparation, Sequencing . . . 44
3.5.2 Sequencing data analysis . . . 44
3.5.3 Large deletions . . . 44
3.5.4 Probabilistic models of transition/transversion and non-synonymous/synonymous ratios . . . 45
3.5.5 Strain-specific sequencing barcodes . . . 45

**4 Prophage excision causes hypermutation: Mechanism and implications 47**
4.1 Introduction . . . 47
4.1.1 Phage conversion of host bacteria . . . 47
4.1.2 Bacterial mutators . . . 48
4.2 Results . . . 49
4.2.1 All the mutator lines show specific excision of a prophage . . . 49
4.2.2 Excision of a prophage causes frameshift that inactivates *mutS* . . . 49
4.2.3 Alternative recombination homology matches the frameshift . . . 50
4.2.4 PCR assay strategy detects presence and excision of prophage . . . 52
4.2.5 The same PCR assay always detects a subpopulation of wild-type cells with excised prophage . . . 52
4.2.6 The *attB* PCR product from wild-type has same scarred sequence as mutators . . . 53
4.2.7 Pure colonies with prophage excision can be isolated by colony PCR . . . 53
4.2.8 Colonies with the excision have a distinct morphology . . . 54
4.2.9 No circular episome is detected . . . 56
4.2.10 Extracellular particles are detectable in wild-type cultures . . . 56

4.2.11 Related prophages are in genomes of other Vibrios . . . 56
4.2.12 Related prophages are in other beta- and gamma-proteobacteria . . . 58
4.3 Discussion . . . 60
4.3.1 Excision mechanics . . . 60
4.3.2 Genomic context . . . 60
4.3.3 Larger phenomenon of prophage disruption of host genes . . . 60
4.3.4 Model of potential relevance of this mechanism to bacterial evolution . 61
4.3.5 Further examination of *mutS* as integration site . . . 61
4.3.6 Broad presence of this element . . . 62
4.4 Conclusions . . . 62
4.5 Methods . . . 63
4.5.1 PCR assays . . . 63
4.5.2 Database searches for related prophage . . . 63
4.5.3 *V. splendidus* 12B01 contig 25 prophage gene contents . . . 63

**5 Future work** **65**
5.1 Mutators and *mutS* prophage
(work continued from Chapter 4) . . . 65
5.1.1 Can naive strains be infected with the *mutS* prophage element? . . . 65
5.1.2 Do naive strains represent a different wild-type/negative control? . . . 66
5.1.3 What is the fate of the excised prophage DNA?
What are the particles observed by fluorescence and their contents? . . 66
5.1.4 What are the baseline frequency and modulating factors of excision? . 67
5.1.5 Bioinformatic study of the *mutS* prophage family
found in many sequenced proteobacteria . . . 67
5.1.6 What are the effects of phage interruption of *mutS*? . . . 68
5.2 Genomics of salt-tolerant mutants
(work continued from Chapter 3) . . . 68
5.2.1 Did we discover salt-tolerance mutations? . . . 68
5.2.2 Testing mutator survival . . . 69
5.2.3 Detecting tradeoffs . . . 69
5.2.4 Further study of the cause of variability in the number of SNPs/round
in the nonmutators. . . . 70
5.2.5 Closing the genome with *de novo* assembly of existing reads . . . 70
5.2.6 Small insertions and deletions . . . 70
5.3 Selection method, Salt tolerance phenotype
(work continued from Chapter 2) . . . 70
5.3.1 Comparing selective methods . . . 70
5.3.2 What are the dynamics and limits of salt adaptation? . . . 71
5.4 Conclusion . . . 72

# List of Figures


2-1 Serial selection involved plating on gradient plates, colony isolation, freezing and sequencing, with growth in liquid medium to grow the required cells between steps. . . . . 20

2-2 Five serial rounds on each of nine lines all began from the same ancestor strain, *Vibrio splendidus* 12B01. Lines are named as Line.Round or LinehsRound, e.g., 1hs5 or 1.5, 8hs2 or 8.2. Line 0, not pictured, went through five rounds of selection as a pilot study. The other lines, 1-8, were evolved by the same method. Lines 1-8 were replicates initiated from one wild-type culture, and they entered each selection round simultaneously. . . . . 20

2-3 Maximum salt tolerances of mutants after five selection rounds are improved and variable after five selection rounds. . . . . 21

3-1 Parallel lineages evolved from a common ancestor fall into two categories, mutators and nonmutators. Number of SNPs vs. selection round for each of nine lines. . . . . 31

3-2 Number of SNPs vs. round, scale adjusted to resolve non-mutators . . . . . 32

3-3 Genes with SNPs in multiple nonmutator lineages. Sites are categorized by non-synonymous, synonymous or in the 5' 100bp. . . . . 33

3-4 dN/dS vs. SNPs for each round. Rounds with zero SNPs are not plotted, and in cases when there were non-synonymous mutations but no synonymous mutations, a pseudocount of one synonymous mutation was added. The ratio of non-synonymous to synonymous sites in 12B01 is 2.9, used to normalize the ratio of mutational counts to calculate dN/dS. . . . . 35

3-5 Mutator strains show a drop in coverage when sequencing reads are aligned to the reference, indicating a deletion from 44890-78138 in contig 25. The spikes in coverage within the deletion area are due to artifactual mismapped reads that correspond to homologs elsewhere in the genome. . . . . 37

3-6 Nonmutator lines experience widely varying numbers of SNP mutations in different rounds, even within the same line. . . . . 38

3-7 Transition frequencies are elevated, especially in rounds with more mutations. These mutants have mutations with a different signature than those between 12B01 and related natural strains. . . . . 39

3-8 Maximum salt tolerance of mutant strains after five rounds of selection does not correlate with accumulated SNPs. These endpoints are only for the 8 parallel lines. Line 0 was not tested. . . . . 41

4-1 The number of SNPs added in the prior selection round is substantially higher for strains with a deletion in contig 25. Points are just spread out uniformly for visibility, so there is no significance to displacement on the x-axis. SNP totals for strains without the deletion are sorted from highest to lowest to show their maximum next to the minimum for strains with the deletion. . . . 49
4-2 Alignment of the amino acid sequences for 12B01 with and without prophage shows how the post-excision *mutS* regains the host starting amino acids, but will now code for a different amino acid sequence. MutS translation in close relatives and the putative uninfected 12B01 begins with a GTG start codon. . 50
4-3 The prophage homology with the host allows it to recapitulate a coding start of the gene. An alternate homology allows improper excision, with the excised DNA removing a 2 bp scar segment from the host. Top: Arrows indicating genes are not to scale. . . . . 51
4-4 Primers flanking attachment sites amplify PCR products that detect the insertion or excision state of the prophage. . . . . 52
4-5 PCR on 12B01 wild-type (ancestor of our mutants) shows detectable excision of prophage (*attB*). Signal is smaller than from expected amplicons of *attL* and *attR* junctions. . . . . 53
4-6 PCR on single colony of 12B01 wild-type (ancestor of our mutants) shows detectable excision of prophage (*attB* amplicon) and absence of integrated prophage (no amplicons for *attL*, *attR*, region inside prophage). No circular or linear prophage DNA is detected. . . . . 54
4-7 Colonies of 12B01 with excised prophage have a distinct starburst morphology. 55
4-8 Colonies of 12B01 with excised prophage differ in morphology from wild-type (with prophage). These spots were transferred from another plate of single colonies in order to space them apart from each other, so these did not originate from single cells, but from a small pipet tip transfer. . . . . 55
4-9 Vibrios with the *mutS* prophage (circled) are mostly in one clade, but are not monophyletic. This tree was constructed by Sonia Timberlake from alignment of the common core genome. . . . . 57

# List of Tables


3.1 Categories of SNPs from 8 parallel lines after 5 selection rounds, based on coding effect and transition/transversion. . . . . 31
3.2 Mutators (lines 0,2,6) share an exclusive deletion, found among the few large (>100 bp) deletions, listed below, found in the 9 final strains after 5 rounds of selection. . . . . 36
3.3 A mix of parameters (20/80, fourth line in table) for transition/transversion ratio is more probable than any individual parameter. Parameters and likelihoods for some alternate probabilistic models of transition data. . . . . 40

4.1 Wide variation in sizes of the related prophages integrated into close relatives of 12B01. . . . . 57
4.2 A high-confidence, partial list of bacterial strains containing an integrase adjacent to a disrupted host *mutS*. . . . . 59
4.3 Predicted phage tail structural genes from contig 25 prophage. . . . . 60
4.4 Primers used to amplify *attL*, *attR*, *attB*, *attP* sites in *V. splendidus* 12B01. . 63

# Chapter 1

# Introduction

## 1.1 Introduction

When this thesis began, it was not yet routine to sequence multiple bacterial genomes for one project. That capacity is with us now, nearly taken for granted. Part of the essential experience of working on a recent thesis on microbes has been seeing those technologies develop. With the arrival of deeper and broader means of interrogating genome sequences and the accompanying rafts of data, it is even clearer that we still have more work ahead in interpreting and attributing biological meaning to the sequences.

A main pleasure of this thesis has been the freedom and rewards in exploring the data generated essentially by one experiment. Much other work went on in parallel, but that seed work was evolving one strain of bacteria to live in higher salinity. We selected for mutants nine parallel times over five steps, testing history and repeatability. The phenomena observed in these bacteria redirected our thoughts about salt and microbial ecology into two other very unsettled areas of research, mutators and bacteriophage. These unexpected shifts were among the most rewarding aspects of the project, leading us to subjects where new biological entities and mechanisms are still being discovered all the time.

This thesis began with techniques invented around the same time as antibiotics, when our ideas of bacteria were mostly as pathogens of humans and of the plants and animals we raise. We now study many more bacteria, not just ones adapted to life in the laboratory after decades since their isolation. Until not long ago, the idea of even finding bacteria and viruses in the ocean was without credence. We now know that bacteria play an essential role in our environment, and that they have shaped it into the one we know. Now we also have tools to investigate their full genetic contents, changes and histories.

Rapid whole-genome sequencing of bacterial genomes, a 21st-century technique, allows us to discover, compare and interpret the presence and functions of existing wild microbes. There is also growing use of experimental evolution by many methods, one of which we used to find salt tolerant mutants. Evolution in the laboratory plus genome resequencing help us interpret the process and results of adaptation. At the same time, applications are appearing. Synthetic biology, metabolic engineering, and applications of bacteria as industrial tools are maturing fields, releasing products with transformative effects on how compounds are made and energy is harvested.

Parts of this thesis could have been done 60 years ago, but other parts, not until 6 years ago. The combination, applied to still developing environmental and evolutionary questions, is much more powerful than any arbitrary dating of methods would indicate.

Despite all the leaps in technology and our growing understanding of bacteria in the environment (*their* environment, largely), these areas still feel full of open questions and active research: environmental microbiology, experimental evolution with genomic resequencing, mutators, mobile elements, and horizontal transfer. We have learned quite a lot, but each time we look, there are good reasons for looking deeper or at an organism or phenomenon previously unknown.

## 1.2 Outline

This thesis has three central chapters that characterize and investigate the same set of bacterial mutants, evolved in the lab. Each chapter carries its own introductory material to the relevant questions. They reflect a progression in time and in deepening understanding of the results in the previous chapter. The final chapter lays out questions still open and work that merits further pursuit.

**Chapter 2** describes the experimental evolution of salt-tolerant mutants and the resulting phenotype of the mutants. We selected for salt tolerance by plating on gradients of salinity, with the aims of learning about the adaptability of this phenotype. To investigate the repeatability and limits of this process, the experiment involved parallel replicates and serial, repeated rounds of selection. We also planned from the start to resequence these mutants in order to learn about the genetic mechanisms of adaptation to salinity.

**Chapter 3** describes the process and results of resequencing of the 45 genomes from the isolated mutants. These genomes represent the genetic changes that occurred during the selection process. We describe the characteristics of the observed mutations and the discovery of a hypermutation phenotype in multiple independent lines.

**Chapter 4** describes the mechanism of this hypermutation phenotype, which involves the excision of a prophage. We also discuss further attempts to characterize the behavior of this mobile element and its implications for Vibrios and other bacteria, including environmentally and clinically important genera.

**Chapter 5** covers open questions and future work proposed to follow from all three chapters. A better understanding of the prophage's relation to the host and evolutionary importance seem very feasible. Further bioinformatic study of this family of prophages could especially give us insight on its effect on the evolution of other proteobacteria in many ecologies, including pathogenic ones.

# Chapter 2

# Experimental evolution of salt tolerance: Phenotypic adaptation

Environmental bacteria impact global ecology and are adaptable to valuable biotechnological applications. The relationship between the genetic diversity and habitats of bacterial strains is highly studied, but it is unknown how a particular strain adapts to changes in environmental conditions. Bacteria are ideal subjects for experimental evolution, though the field is limited in scope to a few organisms and traits. Further, the choice of selection method affects the variety and types of mutants isolated. Here we describe the novel application of a gradient plate method and how it results in salt-tolerant mutants descended from *Vibrio splendidus*, a prevalent marine bacterium. By plating cultures on salinity gradient plates, we rapidly isolated mutants that tolerate higher salinity than the ancestor. After five selection rounds in series, some mutant strains can grow in 1.5M NaCl, compared to a maximum of 1.0M in the ancestor and the average ocean salinity of 0.5M. The improvement in phenotype differs across parallel evolutionary lineages, which recommends our method of finding any mutation that increases salt tolerance, as opposed to methods that select for only a few mutations of greatest effect on growth. This work helps us understand the potential effects of environmental change and suggests a method for forward engineering bacteria.

## 2.1 Introduction

### 2.1.1 Experimental evolution and bacteria–advantages and prior work's limited scope

Experimental evolution enables us to test evolutionary theories and isolate organisms more adapted to chosen selection conditions [67, 119]. In the laboratory, we can witness the process of evolution at a detailed and controlled scale not possible with field samples and phylogenetic inference. These studies can yield interesting outputs for science and engineering–a detailed mechanistic understanding of evolution and improved strains for biotechnological applications.

Experimental evolution with bacteria has several distinct advantages that have been well-reviewed in the literature [10, 40, 44, 79]. Established techniques for growing, observing, counting, modifying, sequencing and archiving bacteria make them nearly ideal for looking at evolutionary events in detail on a laboratory timescale. We can control the composition and number of starting populations, stop and restart the process, watch adaptation as it

happens, and modify conditions during the experiment. These are huge advantages in the difficult study of evolution.

A major limitation of prior experimental evolution work is that only a few microbial species have been examined. Studies have been dominated by a few model organisms, especially *E. coli* and *Pseudomonas* [45, 66, 75, 95], which are often laboratory-adapted after decades in culture. Given the great genetic and physiological diversity among bacteria, these findings may not hold for other strains.

### 2.1.2 Vibrios are a rising model of evolutionary ecology

Vibrios have become a model system for studying environmental adaptation and genetic diversity [32, 59, 104, 122, 123]. Their evolution and ecology have been studied by combinations of phylogenomics, field studies of environmental distribution, and phenotypic assays on hundreds or thousands of cultured strains, but their potential for adaptation has not been tested experimentally. Many different named species of Vibrios are plentiful in the marine environment, with *V. splendidus* often a significant proportion [59, 70, 114, 122]. Environmental conditions, including salinity, have been observed to affect their species distribution [64, 97, 137] and growth characteristics in the laboratory [65, 85, 108, 110, 134].

### 2.1.3 Lack of forward evolution studies for salinity, a relevant environmental factor

Though there are many studies of salt tolerance and of correlation of abundance with salinity, it is unknown if or how a particular strain can adapt to higher salinity. Since salt is a relevant factor for this genus from a marine habitat, environmental microbiologists, including our lab, have made measurements of tolerance and interaction with other environmental factors [4, 80, 85, 110, 134]. Other genera (in addition to the obvious obligate halophiles) have also been characterized and categorized for their salt tolerance [17, 130]. Despite all this evaluation of current phenotype, bacteria have not been evolved experimentally to test improvement of this phenotype and learn the genetic basis of adaptation. We can address this potential for improvement by using experimental evolution.

### 2.1.4 Drawbacks to typical experimental evolution methods for this case

The most common techniques used in experimental evolution, chemostats [42, 88] and serial transfer in batch culture [75], have limitations in speed and the types of mutations favored. These systems select for the mutations that increase fitness, which is growth rate under the selection condition. Advantageous mutations take time to sweep through a population depending on its size, and clonal interference slows the rise of adaptive mutations [36, 91]. Parallel or repeatable changes in phenotype often result, perhaps in part from this bias [28, 31, 45, 99]. It is unclear if this trend of parallelism implies that only a limited quantity of adaptive mutations exists, or that typical selection methods are limited in accessing a fuller spectrum of possibilities. Instead of growth rate, we wanted to focus on salinity tolerance, finding any mutants that would tolerate a new salinity regime.

### 2.1.5 Gradient plates

To focus on salinity tolerance and find a variety of mutations, we picked an atypical selection method, gradient agar plates. These plates were invented for a similar purpose–to select for

resistant mutants to new antibiotics when that phenomenon was first being understood 60 years ago [19, 115]. Gradient plates have been commonly used for measurement of tolerance ranges for salt and many other diffusible chemicals [120], but gradient methods to investigate adaptation are uncommon. Selecting for individual colonies from plates minimizes the effect of clonal interference (Figure 2-1). If there are alternate paths to salt adaptation, multiple gradient plates with parallel lineages will potentially generate a broader set of mutants than other methods would isolate.

### 2.1.6 Outputs

The outputs of this experiment were an evaluation of this selection method and a set of improved strains that can be evaluated phenotypically and genetically to investigate their limits, consistency, trajectory and mechanism of adaptation.

Besides greater scientific understanding of the adaptive process, results of selection experiments also have valuable application now that engineered bacteria are tools for chemical production, sensing, remediation, and other industrial needs. One goal is engineered strains that will survive in the wild for bioremediation or other *in situ* tasks. Beginning with a native strain from a target environment might improve its survival and effectiveness there as an engineered strain, as when local microflora were shown to be mostly responsible for cleaning up the Gulf oil spill [55]. In another scenario, a new version of a strain with improved tolerance to a controllable environmental factor would allow selection for that strain in an open or mixed culture bioreactor, or an engineered organism could operate in a regime less susceptible to contamination. Both the resulting improved strains and knowledge of how best to improve bacteria will be helpful–indicating particular genetic targets for engineering and tailored methods for strain improvement.

In the work here, we investigate the possibility, degree, and variability of adaptation of an environmental Vibrio strain by applying a little-used selection method appropriate to a relevant condition, salinity.

## 2.2 Results/Discussion

### 2.2.1 Gradient plates and selection scheme methodology

We applied a simple gradient method to evolve mutant strains of *V. splendidus* with higher salinity tolerance than their ancestor, wild-type 12B01. Our selection scheme (Figure 2-1) was designed with the goal of isolating a set of mutations that confer salt tolerance and producing a collection of strains adapted in parallel and series.

In order to test for convergence and quantitatively assess variability, we studied parallel replicate lines beginning from the same ancestor. In addition, to isolate more mutants and study the adaptability by multiple mutations, we used rounds of selection in series (Figure 2-2).

The selection scheme is included here for context and because the gradient plate method is not typical. The results of mutant isolation follow.

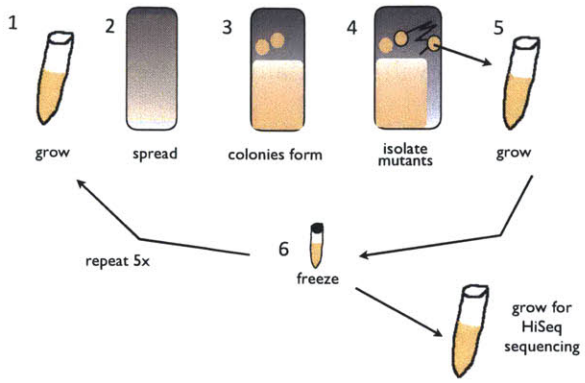


Figure 2-1: Serial selection involved plating on gradient plates, colony isolation, freezing and sequencing, with growth in liquid medium to grow the required cells between steps.

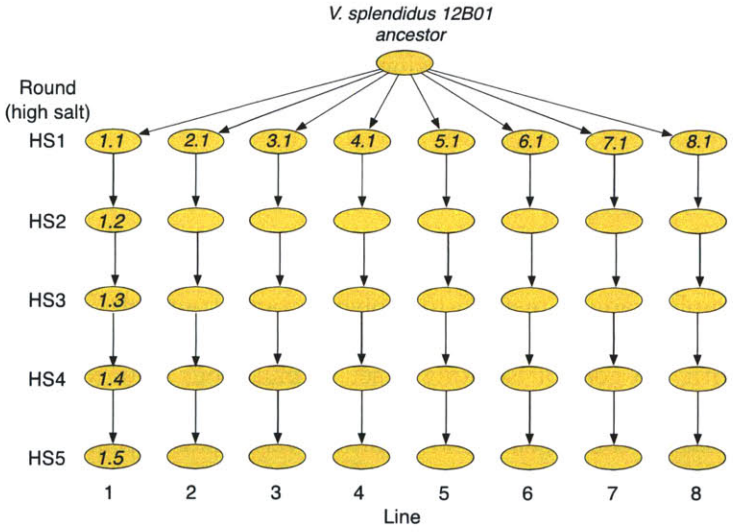


Figure 2-2: Five serial rounds on each of nine lines all began from the same ancestor strain, *Vibrio splendidus* 12B01. Lines are named as Line.Round or LinehsRound, e.g., 1hs5 or 1.5, 8hs2 or 8.2. Line 0, not pictured, went through five rounds of selection as a pilot study. The other lines, 1-8, were evolved by the same method. Lines 1-8 were replicates initiated from one wild-type culture, and they entered each selection round simultaneously.

### 2.2.2 Mutant colonies can be selected on gradient plates

The gradient selection successfully isolated mutants with higher salt tolerance. On all gradient plates, colonies of mutants grew in the area of higher salinity beyond the confluent front where most cells could grow. We thus achieved immediate results of adapted strains by sampling from naturally occurring mutations.

We quickly find mutants with improved salt tolerance, in just the amount of time they need to grow into a colony. These colonies grow at a salt level higher than the tolerable level for the rest of the population. This is an advantage over other methods, since mutations need time to rise to fixation in a chemostat or batch population, and fixation takes a long time if the improvement in growth rate is small. Another drawback to a chemostat or batch method for this application is that the salt concentration regime would need to be chosen *a priori*, whereas our method inherently reveals both the upper tolerance limit and mutants that surpass that limit. Since selection is isolating from among cells just in a small zone of the plate around the limit of the bulk population, the effective population size is much smaller than the number of cells plated.

### 2.2.3 Magnitude: Mutant strains can improve salt tolerance 50% over ancestor

Serial repetition of the selection process produced further improvement in each strain's maximum salt tolerance (Figure 2-3). After five rounds of serial selection, evolved strains were able to tolerate and grow in 1.2-1.5M NaCl, up from 1.0M in the ancestor.

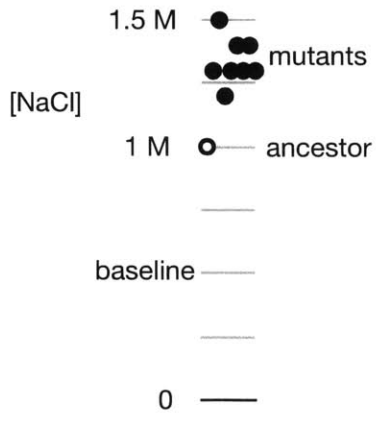


Figure 2-3: Maximum salt tolerances of mutants after five selection rounds are improved and variable after five selection rounds.

We stopped selecting after 5 rounds due to sequencing limitations, but even further improvement could still be possible. So far, each new round has always produced more tolerant mutants growing beyond the confluent boundary.

### 2.2.4 Parallelism: Improvement varies across replicate lines

Parallel or convergent evolution in fitness improvement, gene expression, and protein levels is a common result of experimental evolution [28, 60]. We used an original methodology intended to enable discovery of more varied mutations. Strains did not all improve to the same degree. Variation in final tolerance can be seen in Figure 2-3.

## 2.3 Discussion

### 2.3.1 Patterns of phenotypic adaptation

Studies of protein (not genome) evolution have often revealed very few adaptive paths [133], and other experimental evolution work has shown repeatable parallel changes in phenotype [60, 95]. Here we found divergent phenotypic results of selection among parallel lineages. The selection scheme isolates any mutation that improves fitness, regardless of the selective advantage, so varying levels of improvement are perhaps not surprising, despite the literature on limited trajectories. Two other features of salt tolerance support variable adaptive outcomes. First, this trait is general, not corresponding to one pathway or enzyme. Secondly, it is continuously variable. We hypothesize that improvements are thus not limited to a single gene locus or discrete set of alleles. These results agree with work in a different context of bacterial selection that found greater diversity as adaptation increased [111, 126].

A process with analogous goals has recently been carried out to evolve a thermotolerant *E. coli* [14]. Like in our experiment, the goal was to find mutants that could survive conditions lethal to the ancestor. In their case, they used a gradient in time, gradually increasing the temperature over 8 months. By contrast, we used a gradient in space, for one month. Both methods were successful at isolating improved strains, and both should find further applications. That both studies included whole-genome sequencing is heartening, since knowing the changes is the best way to enable forward strain engineering.

### 2.3.2 Extending prior niche work

The natural variability and divergence among related strains of Vibrio was studied in earlier work in the lab and correlated with their phylogeny [80]. In that work, the six strains of *V. splendidus* characterized had maximum salt tolerances of 0.99-1.13M NaCl. A strain of *V. parahaemolyticus* isolated from dried sardines had a maximum of 1.55M, in line with our improved strains (1.5M). Our evolution experiment thus isolated strains with salinity tolerance both well outside the normal range for their species and comparable to a wild strain in a severe environment.

Rather than just classifying strains based on their original salt tolerance, our approach here begins with a single strain and asks how adaptable that strain is to new levels of salinity. Here we artificially constructed a new clade of 40 evolved strains descended from a common ancestor, and we could see changes in tolerance range even on this short timescale. In this work, we forced rapid evolution by making selection the dominant force, while the genetic diversity in the wild strains can result from other sources like recombination and drift.

### 2.3.3 Salt and natural Vibrios

Vibrio lifestyles incorporate varying environmental conditions, and response to stress conditions, including salinity, is important in their association, pathogenic or symbiotic, with

eukaryotes [85, 134]. Some pathogenic strains infect humans and our food, so understanding environmental effects on Vibrios could have economic and epidemiological benefits [69, 70, 117, 121].

Halophilic Vibrios, such as *V. costicola*, tolerate salinities much greater than typical marine levels [46, 129]. Can we improve a non-halophilic Vibrio to resemble these halophiles? Perhaps pleiotropy or tradeoffs prevent non-halophiles from adapting toward halophily. Alternately, the number of required adaptive mutations could be too great to sample in lab population sizes and timescales.

### 2.3.4 Fast adaptation in the wild

We have shown that evolution can take place quickly along a biologically fundamental and ecologically ever-present axis, and without new DNA, only with new SNPs, deletions and rearrangements that alter one starting genome. Vibrios, particularly *V. splendidus*, seem to change ecology rapidly [59, 114], but the genetic changes accompanying these shifts are not understood. Flexible genes and recombination can change genomes on fast time scales and have a proven role in environmental adaptation [104, 122]. Those forces are absent from our single-strain study.

While horizontal transfer and recombination are known drivers of bacterial evolution, here we have evolved just one strain isolated from lateral sources of variation, so we can just examine the adaptive potential of mutation during vertical descent. We simply do not know *a priori* how much improvement of a strain is possible, especially without HGT. Here we are discovering the magnitude and repeatability of improvements made just by mutation of a single genome.

### 2.3.5 Advantages of gradients

Our plate-based assay is expected to return mutants that can grow where ancestor could not, no matter how slowly (within an arbitrary observation interval, which is limited mostly by the loss of the gradient by lateral diffusion and the potential for resumed bacterial growth). Selection on plates should work even if there is a tradeoff between growth and tolerance.

Thinking genetically, our method may also reveal a different class or number of mutations from those found by gradual methods. More gradual methods may not reveal as many different beneficial mutations, as competition within the population will select for those with greatest fitness benefit. The quantity and character of the genetic changes will be addressed in the following chapter.

## 2.4 Conclusions

By novel application of a classic method, we investigated the adaptation of an isolated strain of a species, *Vibrio splendidus*, that is a new model for microbial ecology. This adaptation was to salinity, an environmental condition where the limits on bacterial growth are well studied, but the capacity to change those limits is not.

First, we found that the salt tolerance phenotype of this strain can change, and dramatically. Here, wild-type cells were selected to tolerate much higher salinity levels than they originally tolerate and normally inhabit. This was not a guarantee–populations of bacteria often encounter a lethal condition they cannot overcome by mutation. A particular genotype might be limited in adaptive range, or adaptation might require multiple mutations or

foreign DNA that are inaccessible in the scope of the experiment. Colonies grew beyond the bulk of the culture and showed measurable improvements in tolerance, so there was enough variation by mutation to permit adaptation. Our selection method allowed those mutations to appear no matter how much of an improvement in tolerance they conferred and without favoring increases in growth rate, two possible differences from other methods.

Secondly, the improvement in salt tolerance varied across parallel lines. This result runs counter to the many examples of parallelism seen in prior experimental evolution. We chose our gradient plate method in order to discover a pool of more varied mutations than other methods might yield. We can infer some greater variety simply from the phenotypic variation we observe. Different mutations could result in the same tolerance gain, but differences in phenotype should indicate different genetic solutions.

Available high-throughput sequencing means we can now follow these evolution experiments at the genomic level. The following chapter addresses the quantity and sites of the genetic variation that arose and accompanied the phenotypic changes just described.

## 2.5 Methods

### 2.5.1 Bacterial strains

*V. splendidus* strain 12B01 was isolated from seawater collected at the Plum Island Estuary Long-term Ecological Research site (http://ecosystems.mbl.edu/pie/data/est/EST.htm) [122].

45 mutant strains of *V. splendidus* 12B01 are designated HS 0.1-8.5, where the first digit is the line and the second digit the selection round. We also used the nomenclature 1hs5 (where hs is high salinity). Line 0 was a pilot line selected for five rounds. Lines 1-8 were later begun from the same ancestor as line 0, but were selected simultaneously for five rounds in parallel.

### 2.5.2 Cultures and growth conditions

Liquid cultures were grown at 20°C with gentle shaking (190-200 rpm). Cultures on solid media were grown at 20-22°C.

Salinity-adjusted medium (LB-S) was prepared following the protocol for lysogeny broth (LB) medium with the NaCl concentration adjusted as required. The pH of LB-S was adjusted to 8.0 using NaOH pellets and 1N liquid, and the concentration of agar was 15 g/L. Liquid and agar media were both autoclaved.

### 2.5.3 Gradient plates

Salinity gradients were poured in disposable, non-vented 241 x 241 x 20 mm bio-assay polystyrene dishes (Nunc) (technique after [19, 80, 90, 115]). Salinity gradients were established by diffusion between two adjacent medium layers using the following procedure: The dish was tilted by raising one side by 5 mm and a wedge-shaped high salinity base layer was poured, using 170 mL LB-S agar medium containing 2.2 M NaCl. After solidification, the plate was leveled and the top layer was poured with 170 mL solid LB-S medium without additional NaCl (0M). The gradient plate then equilibrated for 48 hours to ensure diffusion-driven formation of the salinity gradient along one axis.

### 2.5.4 Selection cycles

Cultures were grown overnight in liquid LB-S and spread over the two outer thirds of the plate in the direction of the salinity gradient, leaving a central alley. Plates were incubated at 20°C for 24 up to 48 hours, until colonies had grown large enough to transfer. Colonies growing beyond the front of bulk cell growth were restreaked at the same salinity position in the central alley to ensure salt tolerance and prevent carryover of dormant, non-salt tolerant cells. Two of these streaks from each gradient plate were streaked to colony purify on separate LB-S 10cm culture dishes. Two streaks were taken from every round in case one failed to grow at some point in the isolation or selection process. These alternates were denoted by a ".1" or ".2" following the strain and round designation.

The salinity of the non-gradient plates in the process depended on the round according to the following: round 1: 0.7M NaCl, round 2: 0.9, round 3: 1.0, round 4: 1.1, round 5: 1.2M. These increases were intended to keep a floor of selection pressure and prevent contamination or reversion. We did not directly measure the concentration on the gradient where the colonies were isolated.

Single colonies from the individual plates were used to inoculate liquid cultures of the same salinity. After overnight growth, these cultures were combined with glycerol (autoclaved or 0.2$\mu$m-filtered) to make 15-25% v/v glycerol stocks in cryotubes, which were frozen at -80°C. Replicate glycerol tubes were also made as backups. To begin the next round, a small inoculum from frozen glycerol stock, usually from streak 1 was used to start a liquid culture that would begin the next selection round on a gradient plate.

### 2.5.5 Growth in variable salt levels

Overnight cultures in standard 0.5M LB-S were used to inoculate 1:100 into 1mL volumes of varying salinity in 2mL deep-well plates using 3 wells per strain as replicates. Maximum salinity tolerated was determined as the level of the wells with detectable growth after 48 hours of incubation at 20°C.

# Chapter 3

# Experimental evolution of salt tolerance: Genome resequencing

Through experimental evolution, we have obtained salt-tolerant mutants in phenotype, but we have questions about the genetic changes that occurred during the process. We expect the inventory and characterization of these mutations to reveal answers to these questions:

What types of mutations were selected?–SNPs, indels, large genomic deletions, rearrangements.

Are mutations affecting broad regulators of gene expression more common than single amino acid substitutions in enzymes?

Are changes in loci with relevance to salinity, such as in pathways relating to osmolytes, ion pumps or outer membrane structures?

Can adaptive benefit be attributed to the observed mutations by signatures of selection, such as parallelism and prevalence of non-synonymous mutations?

Does a general trait like salt result in parallel genetic changes across replicate lines?

We used whole-genome resequencing to find the mutations experienced by *Vibrio splendidus* strains during salt selection and generate hypotheses about adaptation, with surprising results.

## 3.1 Introduction

### 3.1.1 Experimental evolution and resequencing

The advent of feasible whole-genome sequencing of bacteria has dramatically increased the power of evolution experiments by revealing *all* genetic changes in the mutants of interest [128]. Past studies have been limited in genetic analysis, relying on candidate gene sequencing or non-sequencing approaches–methods with lower resolution, scope or accuracy. Now, entire genomes can be checked for changes at the nucleotide level by high-throughput resequencing, in experimental evolution with whole genome sequencing (EE-WGS) [18, 40]. Studies, especially in mutation, are still reconciling results from work on specific loci and whole genomes [41].

With this increase in sequencing power, we can eschew dependence on what is in public genome sequence databases and move beyond the handful of model organisms to choose subjects not for their historical importance, but for their interesting physiology or ecology [73, 92, 104]. Until recently, a few model organisms and organisms with shorter genomes

had been the only references in sequencing databases. A 2012 review of EE-WGS still only listed *E. coli*, *Myxococcus xanthus*, and *Pseudomonas fluorescens* as study organisms [40]. Obtaining an initial reference sequence is now much easier. A lab can even sequence the ancestral genome *de novo* and resequence evolved genomes as part of the same project, even in the same lane of a sequencer.

Experiments in microbial evolution have taken advantage of the new sequencing capacity to extend valuable experimental features like replication, control, and varied conditions [54]. Experimental evolution followed by re-sequencing allows us to study the dynamics and spectra of molecular sequence evolution and the genetic causes of phenotypic change.

### 3.1.2 Parallel molecular evolution common in past studies

Parallel genetic changes at varying levels have been found in multiple or even all replicate evolved strains [9, 20]. This phenomenon is dependent on the selection method, the feature of interest, and the underlying fitness distribution of mutations. In a case where beneficial single mutations are competing under selection, then the one with the greatest fitness benefit will dominate and eventually fix in the population.

Studies of single proteins have indicated only a few molecular paths to more adaptive alleles out of the vast permutations possible. A fitter beta-lactamase allele differing by 5 point mutations has 120 possible mutational trajectories, but experiments showed only a few were likely, implying that protein evolution may be reproducible and even predictable [133]. In other work on selection for thermostability in a "weak link" enzyme, less than 1% of possible missense mutations were observed [33]. These studies each focused on one locus, but does this constrained, potentially reproducible adaptation apply across genomic evolution?

### 3.1.3 Salinity tolerance mutations unknown and may be broader, search is enabled by WGS

Literature studies of bacterial experimental evolution are dominated by metabolic targets, such as adaptation to limitation or substitution of a nutrient [73, 86, 125, 131]. A few environmental conditions like temperature and pH have been used, but most often for investigating patterns of adaptation and tradeoffs by studying just phenotypic improvements, not molecular origins of adaptation [11, 12, 124] . A literature search found no reports of adaptive mutations for bacterial salt tolerance found by experimental evolution.

For a phenotype that broadly affects cellular processes, we need genomic-scale information. Proteomes in halophiles typically have broad changes in amino acid signatures, showing preferences for acidity and against hydrophobicity [129]. Such broad shifts in many proteins may be inaccessible on laboratory timescales. Some salinity tolerance candidate genes, such as those coding for ion pumps and proteins in osmolyte-related pathways, have been identified by studies of physiology, expression data following salt shock or adaptation, and by comparison with halophiles [56, 77, 84], but the nature and sites of potential genetic adaptations are really unknown.

### 3.1.4 Mechanisms of adaptive mutation vary, with roles for mobile DNA, loss of function, regulation

Studies of adaptation have also indicated mobile elements, loss of gene function, and changes in regulation as mechanisms of adaptation, besides single SNPs that might alter one amino

acid [62]. Regulatory changes have been reported that affect expression convergently but through a wide spectrum of mutation types [28], and in another study, 24 genes exhibit parallel expression changes in three clonal groups [60]. In a complementary case, growth phenotypes converged, but with different gene expression states, so modulation of regulation can allow different adaptive paths to the same outcome [45].

On a larger scale, gene gain and loss is a strong feature of even closely-related Vibrio strains. Co-occurrent strains differing by only 1% in 16S sequence differ in genome size by up to 20% [122], and more than 80 sequenced Vibrio strains show substantial flexible gene content (Sonia Timberlake, personal communication). In our cultures of a single strain, we may still observe the gene loss side of that process, which can itself be adaptive [74].

We expected few mutations *a priori* with this selection method. There are still only a few studies that apply whole-genome resequencing to mutants, but past experimental evolution in the literature usually identifies just a few mutations, nearly all non-synonymous [9, 128]. Essential to this field has been a mutation rate that is high enough to produce adaptive solutions, but not so high that there are many neutral or deleterious hitchhikers that obscure the truly adaptive changes. Further reviews of experimental evolution with resequencing are in [30, 40, 79])

### 3.1.5 Applications of re-sequencing information

Sequencing the results of experimental evolution can help us understand past or potential evolution in the wild. In a study conducted on viruses, which have much smaller genomes, the mutations in evolved viral strains recapitulated genetic differences seen in natural evolution when compared with a related virus [20]. Lab evolution can thus supplement comparative sequencing approaches and computed signatures of selection to identify relevant loci for an environment or speciation, especially in a model system of ecological and genetic diversity like Vibrio [59, 104, 122]. Discovering Vibrios' adaptive mutations to salinity, a relevant environmental condition for these marine bacteria, may help us better understand the genetic diversity we see among Vibrio strains in the wild.

Even in model organisms, a large proportion of genes remain hypothetical or poorly annotated (see NCBI Genbank). Selection experiments may reveal the function or relevance of the genes that mutate during selection. This method is a positive counterpart to methods that assign relevance by loss of function, such as knockout libraries [8] or transposon sequencing approaches [127].

The adaptive mutations and relevant loci identified and categorized by selection experiments can also benefit forward, human-directed engineering of useful strains of bacteria. This approach extends the directed gene evolution methods common in metabolic engineering to the level of the organism, optimizing its growth or activity in a particular environment [5, 112]. Bacteria customized at the whole-organism level (strain engineering vs. gene engineering) have vast potential applications and already appearing in academia and industry [24, 61, 68].

## 3.2 Results

### 3.2.1 Genome sequencing of forty-five mutant strains

We resequenced the collection of evolved strains of *Vibrio splendidus* 12B01 described in the previous chapter in order to discover the genetic changes that occurred during the

experiment. Forty-five whole genomes were resequenced with Illumina HiSeq sequencing. These represent 5 selection rounds for one pilot lineage, line 0, and eight replicate lines, 1-8, begun simultaneously in parallel. Strains are denoted as line.round, so the five selection rounds in series for line 1 are strains 1.1 to 1.5. When aligned to a reference genome of the ancestor, mean coverage of the 5,596,386 bp genome was greater than 25X (SD=10) for the strain with the lowest coverage among the 40 parallel strains. Typical mean coverage for strains was 30-40X.

### 3.2.2 Three lines are mutators and far more single base mutations were observed than expected

After resequencing the strains that resulted from selection for salt tolerance, we aligned the reads to the reference genome of the common ancestor, 12B01. The most striking result is the sheer number of single nucleotide mutations accumulated in the evolved lines (Figure 3-1). In order to show the trajectories and SNP totals of the nonmutator lineages, Figure 3-2 presents the same data with a smaller y-axis range. The SNPs tallied here are only those that appear in all subsequent rounds. This criterion ensures that SNPs are fixed and less likely to be sequencing errors or artifacts of growing up cultures for sequencing. The mutator lines, 0, 2 and 6, had 1589, 1539, and 1802 mutations, respectively, after five selection rounds. By comparison, the nonmutators (1,3,4,5,6,7,8) had 126, 88, 140, 73, 37, and 325 SNPs.

Three of nine lineages became mutators during evolution from a common ancestor. These mutators accumulated more than 1500 SNPs. All lines of evolved 12B01 accumulated far more mutations than expected, and 3 of 9 have substantially more SNPs than the other 6. We define two categories of evolved strains and lines from this experiment–mutators and non-mutators. The term "mutator" is not well-defined in the literature, simply indicating strains with higher than typical mutation rates when compared with a wild-type or related strain [83].

By comparison, a very closely-related (identical at 16S) wild isolate, 12F01, has only 302 SNP differences from 12B01.

We estimate that each round corresponds to 100-150 generations. There are three colony isolations ( 30 generations each) and growth in liquid culture twice for about 15 generations each time. This small number of generations makes the SNP totals more striking. 1500 mutations in 600 generations for a typical mutator yields 2.5 mutations/generation, compared with 0.12 mutations/generation in a mutation accumulation experiment with a mismatch-repair-deficient strain [72].

As a summary, Table 3.1 describes the SNPs accumulated in each line after five rounds of selection.

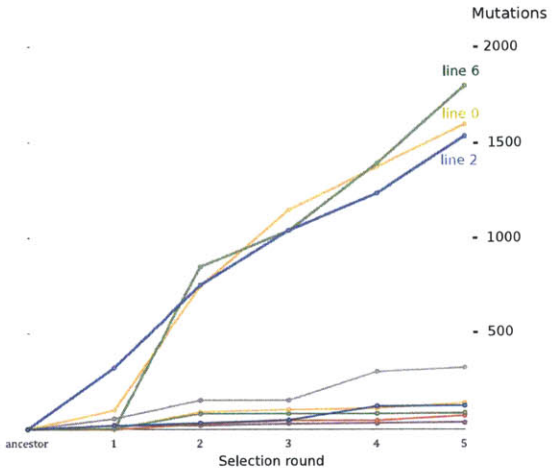


Figure 3-1: Parallel lineages evolved from a common ancestor fall into two categories, mutators and nonmutators. Number of SNPs vs. selection round for each of nine lines.

Table 3.1: Categories of SNPs from 8 parallel lines after 5 selection rounds, based on coding effect and transition/transversion.

| line | dN | dS | noncoding | ti | tv | total SNPs |
|---|---|---|---|---|---|---|
| 1 | 82 | 18 | 26 | 121 | 5 | 126 |
| 2 | 960 | 380 | 199 | 1483 | 56 | 1539 |
| 3 | 52 | 22 | 14 | 82 | 6 | 88 |
| 4 | 76 | 31 | 33 | 128 | 12 | 140 |
| 5 | 36 | 21 | 16 | 67 | 6 | 73 |
| 6 | 1075 | 500 | 227 | 1728 | 74 | 1802 |
| 7 | 23 | 5 | 9 | 35 | 2 | 37 |
| 8 | 189 | 88 | 48 | 311 | 14 | 325 |

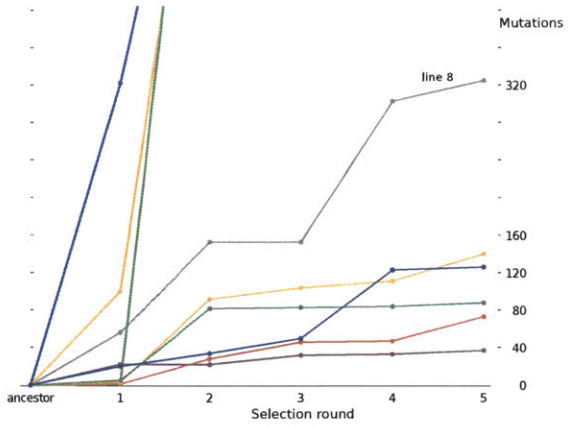


Figure 3-2: Number of SNPs vs. round, scale adjusted to resolve non-mutators

### 3.2.3 Parallelism at the locus level exists, but is not significant

After identifying SNPs for each strain, we compared them across lines to look for common sites or genes, which could indicate selection. We began with the nonmutators, since they have fewer mutations to consider, and a lower probability of overlap. Of 634 total loci affected, 49 (7.7%) were seen in multiple lines. This includes synonymous mutations, often assumed to be neutral, but which may be adaptive due to codon bias [1].

Despite some common loci, we do not find enrichment of common sites or loci among the nonmutators. Some loci are mutated in two or more lines, but not more than expected by a Poisson distribution. Since the data do not diverge from a random model, we cannot attribute significance to the common loci. Comparing the mutations across parallel lines does not recapitulate the parallelism seen in many other evolution experiments. We are using quite a different system from those that report more parallelism.

One base has changed identically in all 8 parallel lines (in the gene ZP_00991008.1 flagellar motor protein, C to T, A284V), but it seems more likely to be due to a mutation in the common ancestor before selection, but after the reference genome was sequenced (or an error in the reference), rather than convergent evolution.

The mutator lines 2 and 6 have three SNP sites in common (below). The expected value would only be 1 if SNPs were Poisson distributed, and that would be a conservative model of *any* change at that site. Here we see the *same* base change made in both lines in all three cases, convergence which may be evidence of selection. The functional implications of these mutations are unknown. The synonymous substitution might not seem to be selective, but it does change the codon usage from a common one to one used 7x less often (in *E. coli*), and it is a transversion, rare in this dataset.

```
originated nt aa contig|site and annotation
2.1,6.4 A>G I661V NZ_AAMR01000021.1|21193
condesin_subunit_B
2.1,6.4 G>T 478dS NZ_AAMR01000032.1|59145
DNA-directed_RNA_polymerase_beta_subunit
2.2,6.4 T>C H53R NZ_AAMR01000034.1|872
phosphoadenosine_phosphosulfate_reductase
```

| total | dN | dS | 5' | gene |
|---|---|---|---|---|
| | 6 | 6 | 0 | 0 12B01__flagellar_motor_protein__2680374__1 |
| | 4 | 4 | 0 | 0 12B01__DNA_gyrase_subunit_B__2682261__1 |
| | 3 | 3 | 0 | 0 12B01__ATPase_domain_protein__2680223__1 |
| 5 in 3L | 3 | 2 | 0 | 12B01__hypothetical_protein__2678229__1 |
| | 3 | 3 | 0 | 0 12B01__omega-3_polyunsaturated_fatty_acid_synthase_PfaC__2680096__1 includes a stop |
| | 3 | 2 | 0 | 1 12B01__GTPase_subunit_of_restriction_endonuclease-like__2680774__1 |
| | 3 | 2 | 1 | 0 12B01__hypothetical_protein__2679805__1 |
| 3 in2L | 2 | 1 | 0 | 12B01__ABC_transporter,_ATP-binding_protein__2677882__1 |
| | 2 | 2 | 0 | 0 12B01__ABC_transporter,_ATP-binding_protein__2681092__1 |
| 3 in2L | 2 | 1 | 0 | 12B01__ClpB_protein__2681732__1 |
| | 2 | 2 | 0 | 0 12B01__Hypothetical_nitrate_reductase_large_subunit__2678114__1 |
| | 2 | 2 | 0 | 0 12B01__hypothetical_protein__2677996__1 |
| | 2 | 2 | 0 | 0 12B01__hypothetical_protein__2678097__1 |
| | 2 | 2 | 0 | 0 12B01__hypothetical_protein__2678765__1 |
| | 2 | 2 | 0 | 0 12B01__hypothetical_protein__2679276__1 |
| | 2 | 2 | 0 | 0 12B01__hypothetical_protein__2680654__1 |
| | 2 | 2 | 0 | 0 12B01__hypothetical_protein__2681336__1 |
| | 2 | 2 | 0 | 0 12B01__hypothetical_protein__2681408__1 |
| | 2 | 2 | 0 | 0 12B01__hypothetical_protein__2681820__1 |
| | 2 | 2 | 0 | 0 12B01__hypothetical_protein__2682686__1 |
| | 2 | 2 | 0 | 0 12B01__NrfD_protein__2681228__1 |
| | 2 | 2 | 0 | 0 12B01__pantoate--beta-alanine_ligase__2682141__1 |
| | 2 | 2 | 0 | 0 12B01__prolyl_endopeptidase__2679131__1 |
| | 2 | 2 | 0 | 0 12B01__putative_transcriptional_regulator,_LysR_family__2679517__1 |
| 3 in2L | 2 | 1 | 0 | 12B01__TPR-repeat-containing_protein__2682607__1 |
| | 3 | 1 | 1 | 1 12B01__hypothetical_protein__2681877__1 |
| | 2 | 1 | 1 | 0 12B01__2-oxoglutarate_dehydrogenase_complex,_dehydrogenase_(E1)_component__2681612__1 |
| | 2 | 1 | 1 | 0 12B01__AcrB/AcrD/AcrF_family_protein__2679907__1 |
| | 2 | 1 | 1 | 0 12B01__hypothetical_protein__2677718__1 |
| | 2 | 1 | 1 | 0 12B01__hypothetical_protein__2678223__1 |
| | 2 | 1 | 1 | 0 12B01__hypothetical_protein__2678226__1 |
| | 2 | 1 | 0 | 1 12B01__hypothetical_protein__2678246__1 |
| | 2 | 1 | 1 | 0 12B01__hypothetical_protein__2680226__1 |
| | 2 | 1 | 0 | 1 12B01__hypothetical_protein__2680645__1 |
| | 2 | 1 | 1 | 0 12B01__putative_ABC_transporter,_permease_protein__2681064__1 |
| | 2 | 1 | 1 | 0 12B01__putative_glycine_betaine_ABC_transporter_ATP-binding_protein__2677903__1 |
| | 2 | 1 | 1 | 0 12B01__putative_sodium/myo-inositol_cotransporter__2679604__1 |
| | 2 | 1 | 0 | 1 12B01__putative_transposase__2678784__1 |
| | 2 | 1 | 0 | 1 12B01__RTX_toxin_transporter__2678857__1 |
| | 2 | 1 | 0 | 1 12B01__Thiol-disulfide_isomerase__2677629__1 |
| | 2 | 0 | 1 | 1 12B01__2,4-dienoyl-CoA_reductase__2681766__1 |
| | 2 | 0 | 2 | 0 12B01__ATP-dependent_protease__2680973__1 |
| | 2 | 0 | 1 | 1 12B01__C4-dicarboxylate_transporter_family_protein,_DctQ_subunit__2678482__1 |
| | 2 | 0 | 2 | 0 12B01__membrane-bound_metallopeptidase__2679266__1 |
| | 2 | 0 | 1 | 1 12B01__putative_phage_gene__2681467__1 |
| | 2 | 0 | 1 | 1 12B01__putative_protein-export_membrane_protein_SecF__2678599__1 |
| | 2 | 0 | 2 | 0 12B01__Soluble_lytic_murein_transglycosylase__2681722__1 |
| | 2 | 0 | 1 | 1 12B01__von_Willebrand_factor_type_A_domain_protein__2678398__1 |

Figure 3-3: Genes with SNPs in multiple nonmutator lineages. Sites are categorized by non-synonymous, synonymous or in the 5' 100bp.

At the locus level, mutators 2 and 6 have 426 loci in common, which is fewer than a Poisson distribution would suggest is possible by chance, given the number of mutated loci in each line. Hence, that set of loci is not particularly informative without some prior hypothesis about convergence or categories of loci.

This result repeats and explains the lack of parallelism in phenotype discussed in the previous chapter. Our findings diverge from many past studies that showed repeatable evolution, but there are plausible explanations. We speculate that the coupling of high mutation rate with population bottlenecks fixed neutral or deleterious mutations that might have been purged by selection during growth in a larger population. The signal from these random, unselected-upon changes overwhelm any possibly parallel adaptive changes. Also, there may actually be a wider range of adaptive mutations for salt tolerance than for traits where fitness effects are concentrated in one gene.

Whole genome resequencing of mutants was intended to generate hypotheses about adaptive mutations for salt tolerance. There are too many changes to deduce immediately which were adaptive, but these sequences do provide some hypothetical sites of adaptive changes. Allele swaps in cultured cells could evaluate the adaptive benefit of particular putatively relevant SNPs.

### 3.2.4 Non-synonymous and synonymous mutations are both common

In order to look for signals of selection or mutational bias in our strains, we analyzed the ratio of non-synonymous to synonymous mutations. When corrected for the likelihood of each change, the ratio of non-synonymous mutations to synonymous ones, dN/dS, indicates a near-random distribution (Figure 3-4). This signature dominated by random mutation makes estimating the adaptive fraction of the mutations difficult.

An average dN/dS = 0.79 (95% CI = 0.77-0.81) best fits the data from all 40 strains, according to a simple binomial probabilistic model. By comparison, the SNPs that separate 12B01 from its close sister strain 12F01 have a dN/dS of 0.32. In previous experimental evolution, positive selection is so dominant that there are few to no synonymous mutations [9] in resequenced nonmutator lines.

Low dN/dS ratios indicate purifying selection, ratios near one are random, and higher ratios can indicate positive selection. Our data is dominated by the random signal, which may contain some positive selection. We are sampling the distribution of mutations before purifying selection has had much time to act, since the rate of mutation is so high compared with the time between bottleneck steps.

This result makes the large number of mutations tolerated in the mutators even more impressive, since most are non-synonymous changes, which are predicted to be deleterious more often than synonymous or non-protein-coding SNPs. This signature dominated by random changes is uncommon in experimental evolution, which usually isolates few neutral (synonymous or non-coding) mutations. Usually these neutral changes are maintained due to hitchhiking with a positively selected allele elsewhere on the same chromosomal background. In our selections, picking single colonies fixes (sometimes multiple) mutations while minimizing purifying selection, which causes the dN/dS ratio to resemble a random sample of sites.

With so many more changes than even between natural relatives, it was natural to check that this was not simply contamination by another Vibrio isolate in the lab. If our experiment had been contaminated by a natural strain, the strain-specific differences would have a lower dN/dS (as with 12F01), given the purifying selection over a much-

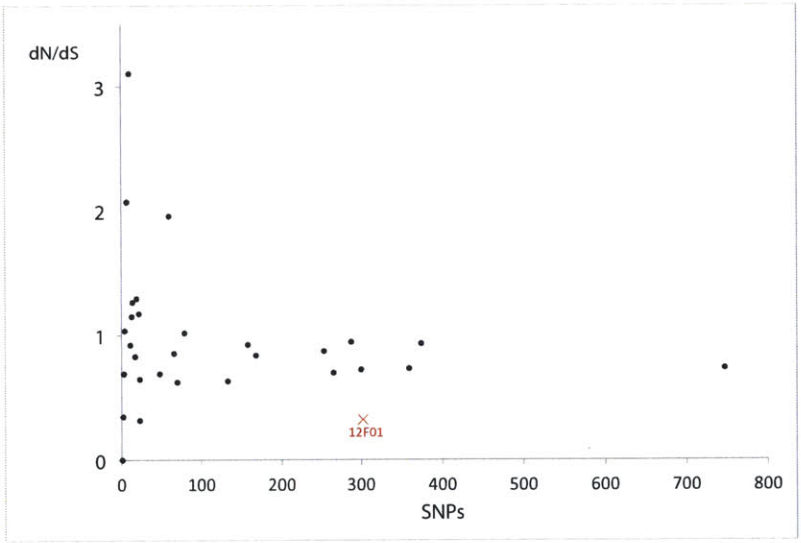


Figure 3-4: dN/dS vs. SNPs for each round. Rounds with zero SNPs are not plotted, and in cases when there were non-synonymous mutations but no synonymous mutations, a pseudo-count of one synonymous mutation was added. The ratio of non-synonymous to synonymous sites in 12B01 is 2.9, used to normalize the ratio of mutational counts to calculate dN/dS.

longer evolutionary time since divergence. Instead, the ratio is higher, including many non-synonymous mutations, indicating that our strains are true recent descendants, diverged just during our experiment.

We know from the strains' phenotypic improvement (see Chapter 2) that our selection process must have fixed adaptive alleles that confer high salt tolerance, so we presume that strains gain at least one adaptive mutation per round (which may be another form of mutation, not necessarily a SNP). The signal of that positive selection is swamped by other changes, neutral or deleterious.

### 3.2.5 Mutators share an exclusive deletion feature

Deletion of an specific and identical region is the common genetic feature unique to the mutator strains. In trying to find the cause of the two levels of mutation, I looked for common features exclusive to the mutators. I first looked in the SNP analysis itself to find mutator loci (e.g., genes associated with DNA repair, replication or metabolism) that were mutated, the most common explanation [82, 93, 106]. No mutator locus had common SNPs in the mutators, so we turned to analysis of deleted regions. Regions greater than 100bp are presented in Table 3.2. A deletion of approximately 26kb in contig 25 was only found in the mutators (0,2,6) and in none of the nonmutators (Figure 3-5).

Table 3.2: Mutators (lines 0,2,6) share an exclusive deletion, found among the few large (>100 bp) deletions, listed below, found in the 9 final strains after 5 rounds of selection.

| genes | length | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|---|
| contig 25 phage integrase restriction-modification system, regulatory proteins etc. | 26 kb | + | | + | | | | + | | |
| contig 18 recombinase, DNA polymerase, response regulator, etc. | 17 kb | + | + | | | | | | + | |
| ABC-type uncharacterized transport system, periplasmic component | 148 b | | | | | | + | | | |

All three lineages with elevated SNP numbers had this same deletion feature, and it was never seen among the strains with a lower mutation rate. This perfect association has a significance of p=0.0119 (Fisher's exact test, two-tailed). The deletion can even be linked in time with the mutation increases. In lineage 6, the genome sequence of 6.1 is whole and has just a few SNPs, but the deletion appears in 6.2, coincident with the rise in mutation rate in round 2 (800+ mutations).

One round in another lineage, number 5, seems to have become a mutator on the way to sequencing, since strain 5.1 has many mutations that were not found in later rounds of the same line. These mutations were not included in the earlier analysis because we required that SNPs be consistent through subsequent rounds, in order to discount mutational artifacts

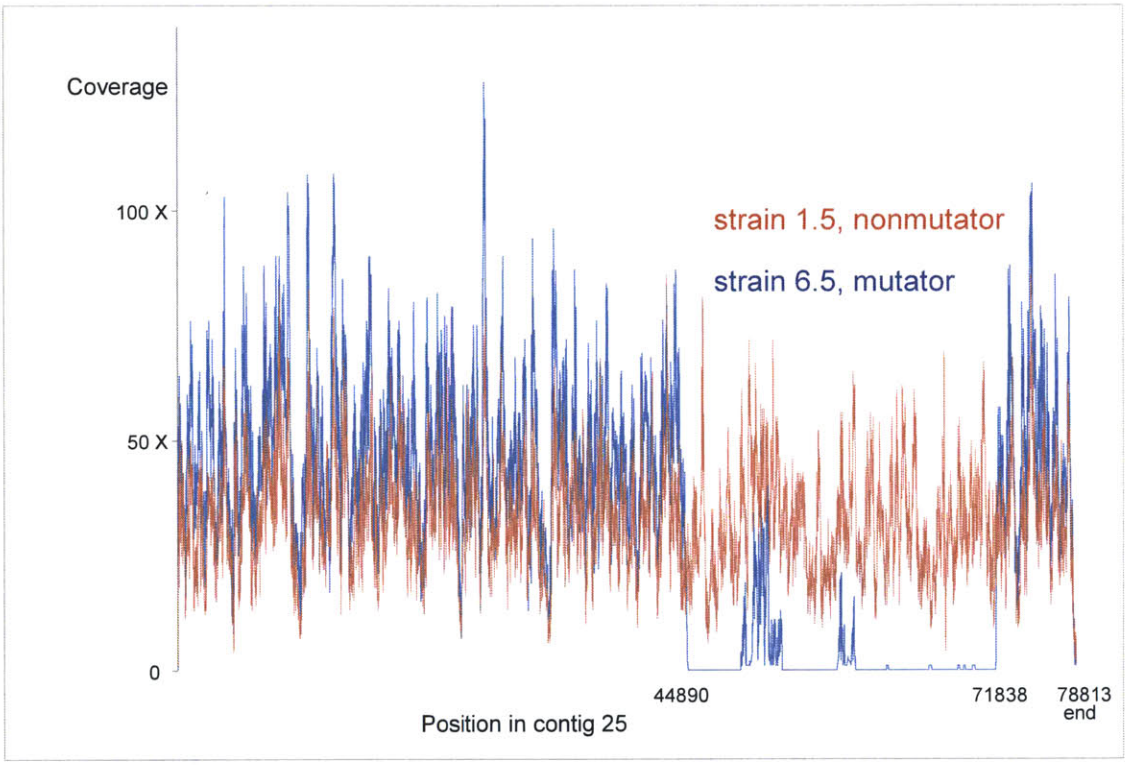


Figure 3-5: Mutator strains show a drop in coverage when sequencing reads are aligned to the reference, indicating a deletion from 44890-78138 in contig 25. The spikes in coverage within the deletion area are due to artifactual mismapped reads that correspond to homologs elsewhere in the genome.

due to growing strains for sequencing. 5.1's large number of mutations is accompanied by the same deletion seen in the strains that ended as mutators, independently confirming the association between deletion and high mutation rate. Since the mutations sequenced in 5.1 were not seen in later rounds of this line, we do not believe there was a reversion of the deletion, which would have kept the mutations accumulated in the first round. Rather, the deletion and mutations happened on a branching on the way to sequencing due to the necessary plating and outgrowth steps for DNA extraction. In support of this hypothesis, 5.1 had only about 130 SNPs, less than any increment in the mutator lines.

It is important to note that the genomic alignment and analysis of mutations can need attention and curation. This full deletion was not obvious, since paralogs and assembly artifacts lent spurious coverage to regions within the true deletion (Figure 3-5). Realigning sequence reads from the mutators to the putative spliced sequence revealed correct, continuous alignments, and the deletion was confirmed by PCR and sequencing (details are in Chapter 4).

### 3.2.6 Mutations in each round vary widely within nonmutator lines

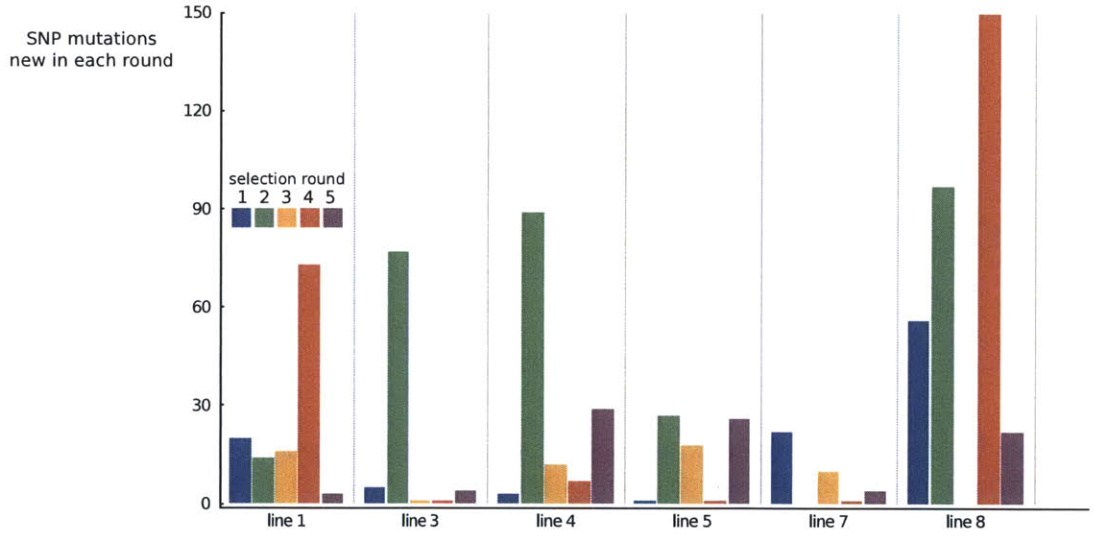


Figure 3-6: Nonmutator lines experience widely varying numbers of SNP mutations in different rounds, even within the same line.

Plotting the mutation results for individual selection rounds shows variability between rounds, even for the same lineage (Figure 3-6). Each nonmutator line has one or more rounds in which there were substantially more mutations. This pattern did not occur in any particular round in the progression, nor was the trend fixed after the increase. Subsequent rounds often returned to the prior low mutation levels. The cause of this transient mutation is unknown, though there is some intriguing related data in the next section on the mutational signatures.

### 3.2.7 Transition mutations greatly predominate, especially in rounds with more mutations

Deficiencies in the enzymes that copy and repair DNA result in signature mutational spectra that are due to the biophysical mechanisms of these enzymes. A striking feature of our SNP data is a predominance of transition mutations over transversions (Figure 3-7). This bias guides the search for a molecular cause of the hypermutation.

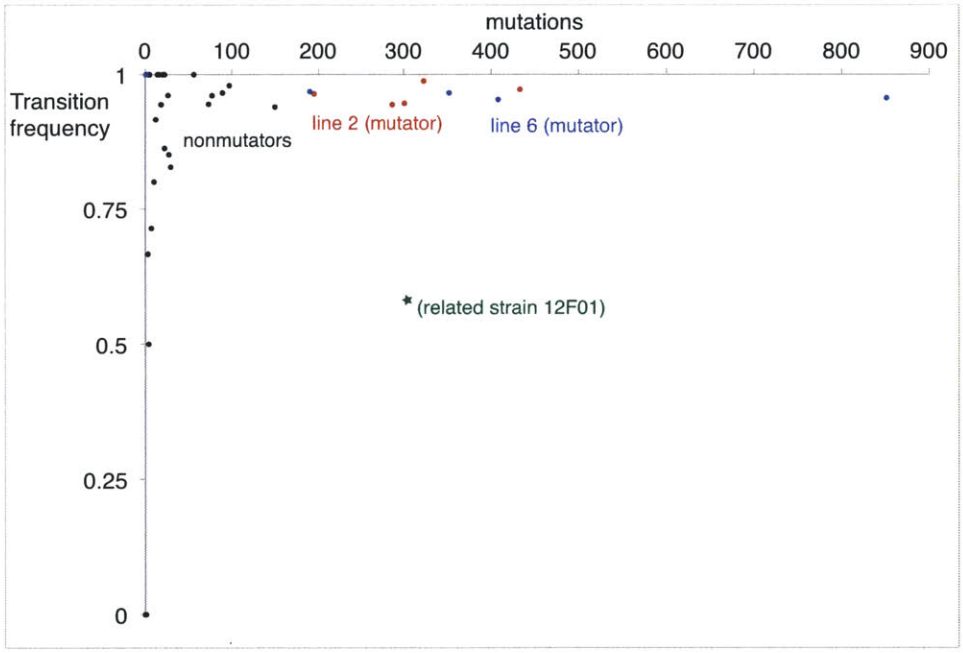


Figure 3-7: Transition frequencies are elevated, especially in rounds with more mutations. These mutants have mutations with a different signature than those between 12B01 and related natural strains.

The data are plotted as transition frequency (transitions/total mutations), because some rounds have no transversions at all, making the usual metric of transition/transversion ratio (ti/tv) undefined. By comparison, SNPs between 12B01 and its close relative *V. splendidus* 12F01 have ti/tv = 1.4 (a transition frequency of 0.58 on the plot), closer to literature estimates of ti/tv = 2-2.5 (0.67-0.71) [48]. These ratios reflect a slight biochemical tendency to transitions, but are lower than almost all of our data points (Figure 3-7). In our lines, transitions are favored much more.

### 3.2.8 Mutation data support a model using a mix of low and high transition probabilities

All of the nonmutators have some rounds in which they experience substantially more mutations, and those rounds share a signature of transition bias with the mutators (Figure 3-7), which leads us to speculate that the two phenomena might be related. To test this idea, we built a mathematical model to explain the nonmutator transition frequency. The results are in Table 3.3. Details of the model are in the Methods section at the end of the chapter.

A binomial model with a mix of two probabilities that a mutation is a transition explained the data better than a model with only one transition probability. The mix was of ti/tv = 2.5 and 20.5, at likelihoods of 0.2 and 0.8, respectively. Even with a single parameter for the ratio, the model that best fit the data has a ti/tv of 16.5, far greater than the typically reported values for bacteria. The mixed ratios represent an underlying, low level of mutation that has a typical ti/tv ratio, combined with additional mutations that carry a signature of very high ti/tv. The better fit of this model suggests that transient hypermutation with the same distinctively high ti/tv ratio may even be occurring in the nonmutators.

| ti/tv | ti frequency | p(model \| data) | normalized likelihood |
|---|---|---|---|
| 2.5 (literature) | 0.714 | 2.0E-79 | 1.5E-59 |
| 20.5 (mutators) | 0.95 | 1.4E-21 | 0.11 |
| 50/50 mix of above | | 8.7E-22 | 0.06 |
| 20/80 mix of above | | 1.3E-20 | 1 |
| 16.5 | 0.94 | 3.6E-21 | 0.27 |

Table 3.3: A mix of parameters (20/80, fourth line in table) for transition/transversion ratio is more probable than any individual parameter. Parameters and likelihoods for some alternate probabilistic models of transition data.

### 3.2.9 Salt tolerance does not correlate with SNP totals

Does improvement correlate with the number of mutations? We measured the maximum growth of the strains after five rounds before knowing their genotypes. Now we can address the question of whether strains with more mutations are more adapted. Figure 3-8 compares genotype and phenotype data for the final strains from the eight parallel lines.

These data show no strong correlation between mutation totals and salt tolerance. One mutator, strain 2.5, did outperform other strains, but the other mutator from the parallel lines, strain 6.5, is undistinguished in salt tolerance. Nor is there a reverse trend, so the mutators have not paid an obvious penalty in this fitness dimension due to their excess of mutations. This finding agrees with our understanding that the SNP totals are dominated by random, not adaptive, mutations. It is widely believed that most mutations are neutral or harmful, so strains with more mutations might be expected to have accumulated a fitness disadvantage. The mutators may have disadvantages in other dimensions, but their thousands of mutations seem largely neutral with respect to salinity tolerance.

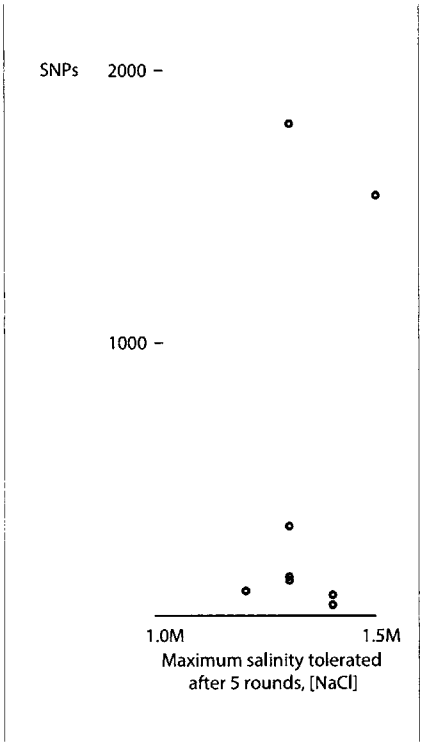


Figure 3-8: Maximum salt tolerance of mutant strains after five rounds of selection does not correlate with accumulated SNPs. These endpoints are only for the 8 parallel lines. Line 0 was not tested.

## 3.3 Discussion

### 3.3.1 Mutators

While mutators are not uncommon in long-term evolution experiments, ours are exceedingly mutation-prone [105]. These results are exceptional in comparison with prior reports. We estimate that these strains passed through 100-200 generations per round, a possible total of 1000 generations. Even known mutator strains subjected to mutation accumulation yielded fewer SNPs [47]. Foster, and colleagues sequenced 34 *mutL- E. coli* lines after 12,750 total generations (all lines combined) and identified 1625 total mutations, roughly equivalent to the total from just one of our mutators. Resequenced lines from Lenski's long-term *E. coli* lines had 45 SNPs after 20,000 generations [9]. Though their resequenced mutator lineage arose after 20,000 generations, at 40,000 total generations, it had only 627 SNPs total. In a 2000 generation experiment with 115 parallel *E. coli* populations, an average of 6.9 SNPs/line, excluding one mutator line (with 73 mutations) was identified [118].

In experimental settings, mutator strains have arisen by diverse mechanisms, usually when replication and repair loci are victims of IS interruption or deleterious SNPs. These mutator loci are often mismatch repair genes, such as *mutS* and *mutL*. The Lenski mutator's SNPs had a transversion bias due to a frameshift in *mutT*. Different mutator mechanisms often produce distinctive mutational spectra, which helps determine the genetic cause of the mutator phenotype. There is also a precedent for transient, stress-induced hypermutation. The SOS response produces a transient, reversible rise in mutation, biased to transversions [96].

Besides studies that incidentally produce mutators during the experiment, some studies intentionally use mutator strains during experimental evolution to study their behavior [47, 49, 50, 72]. Few of these strains, however, have been resequenced completely. Often the mutators are identified phenomenologically and quantified by the inaccurate assay of frequency of antibiotic resistance, not by the actual number of SNPs accumulated [102].

The significance of mutators goes beyond just these controlled studies. Mutators are regularly detectable in wild and clinical samples, and are thought to have serious medical consequences via accelerated host adaptation and antibiotic resistance[52], e.g., *Burkholderia* spp. in cystic fibrosis [89], and uropathogenic *E. coli* [39].

Enrichment of mutators by hard selection has been predicted by theory and seen in experiments [51, 71, 87, 105]. Theory holds that mutators may produce beneficial mutations more quickly than nonmutators can, leading to selection of the mutator allele (initially adaptively neutral) on the coattails of the beneficial mutations. This idea implies a balance between the rate of mutator production and the rate of beneficial allele production, so mutators must be above a threshold in the population in order to succeed. Multiple rounds or environments that favor multiple mutations are also thought to enrich for mutators [87, 116]. Though both are features of our system, we saw mutators even in the first round. Learning the causal mechanism of the mutator phenotype and the rate at which mutators arise, covered in the next chapter, should help us predict the likelihood of selecting mutators.

### 3.3.2 Parallel mutations

In light of earlier molecular studies that reveal constrained paths of evolution, we examined the reproducibility of mutations in our sequences. Our experiments select on an environmental axis affecting many parts of the organism and, by extension, involve a larger portion

of the genome. Also, our isolation of colonies are bottlenecks that make the effective population size smaller and reduce competitive exclusion, which should reveal a more diverse set of adaptive mutations. With several steps of hard selection, in retrospect, the experiment resembles a mutation accumulation assay. Based on literature mutation rates, it was expected that just one mutation (or perhaps a few) that conferred salt adaptation would be fixed in the strains in each round. We did not induce mutation extrinsically, but only selected on the existing variation in the liquid culture population that was plated on the gradients. The sequencing results involved surprisingly high totals, seemingly dominated by random mutations.

### 3.3.3 Salt tolerance mutations

Attributing adaptive advantage to any single mutation is difficult with so many changes and little significant parallelism. Since osmolytes are one of the ways that nonhalophiles resist salt stress, we had a prior expectation that these could be among the likely adaptive loci for salt. In strain 3.1, there were only five SNPs, and one was -65 of the start of glycine betaine/L-proline transport ATP-binding protein ProV ZP_00992752.1. Finding this functional annotation among a low number of SNPs makes this change a strong candidate for an adaptive mutation to higher salinity. Six genes related to glycine betaine production or transport (out of 19 in the genome) experienced mutations in the nonmutators, and these would be good initial candidates to test for benefit individually in a wild-type background. The mechanisms of tolerance gain from these changes are unclear and likely variable.

Experimental evolution has often led to changes in regulation, which can dramatically shift phenotype with one or a few genetic changes [28, 31, 131]. With limited gene annotations and no expression data for these strains, we cannot say if such mutations in transcription factors or other regulators may be partially responsible for salt adaptation. As a precedent, adaptive changes in regulation were identified in a study selecting for salinity tolerance in yeast [3].

We hypothesize that the variation seen in experimental evolution on suspected environmental traits of importance (nutrients, temperature, salinity, pH) may correspond to the observed natural diversity on an allele or locus level. If so, we could assign putative adaptive benefits to variation in loci or SNPs found in wild strains. This remains to be explored, since there were too many new alleles to know yet which are adaptive. Putative loci from each method, phylogenomics of wild strains and resequencing experimentally evolved strains, may enrich the search for adaptive loci by the other method.

## 3.4 Conclusion

Given the scope and resolution afforded by next-generation genome re-sequencing, we have observed the genetic results of selection for salt tolerance. We saw far more genetic changes than expected, and three of the nine lines have distinctly more mutations than the other six. The SNPs show a marked bias for transition mutations, indicating a possible molecular mechanism for hypermutation. The ratio of non-synonymous to synonymous changes looks close to random, obscuring the signal of positive selection and ruling out contamination. Only the mutator lines underwent a deletion of a 26 kb region containing prophage-related genes. The mechanism distinguishing the exceptional mutators is the subject of the next chapter.

## 3.5 Methods

### 3.5.1 DNA Extraction, Library Preparation, Sequencing

Strains were grown overnight in 5 mL LBS (per L: 10 g tryptone, 5 g yeast extract, 29.2 g NaCl (0.5M)), pH 8, in 14mL culture tubes, at 20°C, with shaking at 200 rpm. DNA was extracted with the Qiagen Blood and Tissue kit, using the protocol for Gram-negative bacteria. DNA was fragmented by sonication: 30 second on/off cycles for 18 minutes total on M setting, ice bath replenished every 6 minutes. Sequencing libraries were prepared with Illumina barcoding primers at MIT's BioMicroCenter. The five strains in line 0 were prepared by Arne Materna and sequenced on an earlier Illumina GAII lane at the BioMicroCenter. Libraries for 40 genomes were multiplexed and sequenced in one Illumina HiSeq lane at BioMicroCenter, resulting in approximately 10E8 40 bp, paired-end reads. The barcode was read after the first genomic read and before the second. Average coverage was approximately 40x, with the least represented strains at 27x.

### 3.5.2 Sequencing data analysis

data flow:
Split reads by strain-specific barcodes into files for each of the 40 strains.
In Galaxy bioinformatics portal:
FASTQ Groomer [15] Aligned reads with bowtie (1.X.X) to corrected reference sequence of wild-type *Vibrio splendidus* 12B01 NZ_AAMR01000001.1-119.1 (119 contigs).
The NCBI AAMR sequence (obtained by Sanger sequencing) was corrected at 23-25 single base sites by Sonia Timberlake using next-generation resequencing.
Converted BAM to SAM.
Ran SNP Finder in Galaxy (SAMtools), which finds sites with variant reads.
Used minimum of 80 percent variant reads to detect mostly fixed SNPs with tolerance for errors.
Saved local files of SNP sites for all 40 strains.
Compiled a list of fakes by finding ambiguous sites, mismapped homologs like tRNA and ribosomal elements, contig ends, ends of deleted regions. All these produce mapping artifacts that are not true variation.
Subtracted these fakes from list of sites.
Added consensus base based on SNP reads.
A Perl script from Sonia Timberlake used gene annotations and SNP sites and classified each change as a transition or transversion, determined the genetic locus and whether the change was in the gene coding sequence or 100bp upstream, and if coding, whether synonymous or non-synonymous.

### 3.5.3 Large deletions

To find large deletions, a script searched for jumps of >100 bp in covered positions in pileup files of assemblies to reference contigs.

### 3.5.4 Probabilistic models of transition/transversion and non-synonymous/synonymous ratios

In order to calculate the transition frequency that best explains our data, we use a binomial model to calculate the likelihood of each data point (e.g., x transitions out of y total SNPs) given a predicted probability, $p_{ti}$, for a SNP to be a transition. (An analogous model was used for non-synonymous/synonymous mutations.)

$$\Pr(\text{binomial(k transitions, n total SNPs, } p_{ti})) = \binom{n}{k} {p_{ti}}^k (1 - p_{ti})^{n-k}$$

The probability of the model given the data is the product of the probability for each data point:

$$\text{p(model with } p_{ti} | \text{data from j lines)} = \prod_i^j \binom{n}{k} {p_{ti}}^k (1 - p_{ti})^{n-k}$$

For the model of multiple transition frequencies, the transition probability is itself a probability distribution function, so the probability for a data point is:

$$\sum Pr(k, n, p_{ti}) * Pr(p_{ti})$$

We have kept to functions of simple, discrete scenarios. For example, for the mix of scenarios with these probabilities:

$$\text{P} = \{\Pr(p_{ti} = 0.2) = 0.5 \ , \ \Pr(p_{ti} = 0.9) = 0.5\}$$

The overall probability of the data would be:

$$\Pr(\text{k,n,P}) = \Pr(\text{k,n,0.2})\text{*}0.5 + \Pr(\text{k,n,0.9})\text{*}0.5$$

The models can be compared by the ratio of their probabilities given the data, a likelihood ratio. We normalize to the most likely model to express relative likelihoods and determine 95% confidence intervals. In our calculations, we have not adjusted the likelihood of the mixed model for the additional parameter of another possible transition probability.

### 3.5.5 Strain-specific sequencing barcodes

strain barcode
1.1 GAACCT
1.2 ACACGA
1.3 TGGACT
1.4 CATGCC
1.5 CTTGTG
2.1 ACCTGA
2.2 TGGATG
2.3 CTCTTG
2.4 GTCATG
2.5 TGTGAC
3.1 ACACTG
3.2 GAACAC
3.3 CTCTAC
3.4 GAGCGT

3.5 ACACCT
4.1 TGTGTG
4.2 GAGAAC
4.3 CTCTGA
4.4 ACGATT
4.5 GAACTG
5.1 ACCTAC
5.2 TGTGCT
5.3 CTTGGA
5.4 CTATAA
5.5 CTCTCT
6.1 CTTGAC
6.2 GAACGA
6.3 TGTGGA
6.4 AGCCGG
6.5 ACACAC
7.1 GAGAGA
7.2 CTTGCT
7.3 GAGATG
7.4 TCTGCA
7.5 TGGAGA
8.1 ACCTTG
8.2 ACCTCT
8.3 TGGAAC
8.4 TGATAC
8.5 GAGACT

# Chapter 4

# Prophage excision causes hypermutation: Mechanism and implications

Mutator strains and mobile elements are both drivers of diversity and evolution in bacteria. Many environmental and clinical species contain subpopulations of mutators, but the specific mechanism of elevated mutation rate is usually unknown. The potential for mtuators to rise to significant population levels is also unknown. Experiments selecting for salt tolerance in *Vibrio splendidus* isolated three independent lineages of mutators out of nine lines. These mutator strains share a previously unreported hypermutation mechanism. Here we present the discovery of a mobile element that integrates into the 5' coding sequence of *mutS*, a critical mismatch repair gene. The prophage provides a new start codon in-frame with the rest of the host protein. Upon excision, this element leaves a scarred version of *mutS* that confers hypermutation in *Vibrio splendidus*, a common marine bacterium. These mutators occur naturally at low frequency even under non-selective conditions. We used computational sequence comparison and manual curation to identify similar mobile elements in other previously sequenced bacteria, including pathogens and industrially relevant bacteria. Related elements are found in many sequenced beta- and gamma-proteobacteria, including *E. coli*, *Yersinia* spp., and *Burkholderia* spp., so this is not a strain-specific phenomenon. Elements like this one may play a central role in the hypermutation previously observed in clinical strains.

## 4.1 Introduction

### 4.1.1 Phage conversion of host bacteria

Bacteriophage can greatly alter host phenotype and lifestyle by integrating into a bacterial chromosome and providing new genetic content or disrupting host systems. Since bacteria are key to health, chemical cycles and ecology, this process of phage conversion of bacteria has dramatic effects. Examples of new functions contributed by prophage include toxins and virulence factors that render commensal bacteria pathogenic, as well as antibiotic resistance cassettes [132]. Prophage insertion and excision can be a regulated process, known best from the classic lambda phage, but these genomic alterations can themselves regulate host genes if those genes are disrupted by prophage.

There are already documented examples of prophages regulating host phenotype and under regulation of host conditions. A prophage in *Listeria monocytogenes* excises and activates genes required for the bacterium to escape phagosomes and infect its own mammalian host [94]. This excision even happens in a regulated response to the bacterium's host environment. In *Streptococcus pyogenes*, a prophage interrupts mutL expression, inactivating mismatch repair and rendering the bacteria a temporary mutator [101, 102]. Again, there is control of the process by growth state–in this case, stationary phase. More examples are sure to be found as sequencing efforts continue and more attention is paid to phage, prophage and other mobile elements.

The relationship between phage and host, involving coevolution, evasion and innovation, is often discussed as an arms race [113, 135]. There may be other, more subtle or cooperative dimensions to this relationship [21]. Lytic phage or lysogenized phage that excise to become lytic are not advantageous to the immediate host, but these examples of stable or regulated prophage may actually provide benefit to the host.

### 4.1.2 Bacterial mutators

Mutators have been observed regularly both in the environment and in the clinic. Environmental studies of mutation rates in wild bacterial samples regularly find a mutator phenotype in a low percentage of isolates [71, 78, 81]. The influence of these mutators on bacterial populations and adaptation is well-explored by theory and some experiments, but their relevance to evolution in the wild is not well understood. Since bacteria profoundly affect life and chemical cycles on the planet, the role of mutator strains in their evolution could have important consequences [35].

One of the clearest examples of the effects of mutators is in the clinic. Pathogenic bacteria that adapt to resist antibiotic treatment cause prolonged illness, death and great expense. Mutators may play a significant role in this phenomenon [53, 63]. *Pseudomonas* mutators are enriched in cystic fibrosis and contribute to antibiotic resistance and other adaptations in the host [89].

Though mutators are regularly observed in wild and experimental populations [13, 39, 78, 83], the cause is only known in a small fraction of cases. In parallel lines from the Lenski long-term *E. coli* evolution study, one population had an insertion of a single base in *mutS*, and in two populations, mutators formed independently when changes in repeat length disrupted *mutL* [106]. Whether these changes are revertible is also an important question, since mutators are thought to have dim long-term prospects [35]. As mentioned above, a prophage-based system for changing mutation rate based on gene inactivation is known, and it is particularly intriguing for its reversibility.

Hypermutation and mobile elements can both help bacteria adapt to environmental stresses. These mechanisms combine in the genetic feature we have discovered–a mobile element resembling a prophage integrates within the coding sequence of *mutS*, a DNA repair gene. The prophage replaces the upstream region, start codon, and first 10-20 amino acids of *mutS*, replacing the host control of the gene.

We have preliminary evidence that suggests a switchable mutator genotype, which may be accessible even in the absence of salt stress. The hypermutation may also be transient, since the mechanism is possibly reversible.

## 4.2 Results

### 4.2.1 All the mutator lines show specific excision of a prophage

The three lines that ended our selection with over 1500 SNPs all had a deletion of a prophage region in contig 25. Absence of this region was strongly associated with high SNPs in rounds. The deletion was the same in all three mutators.

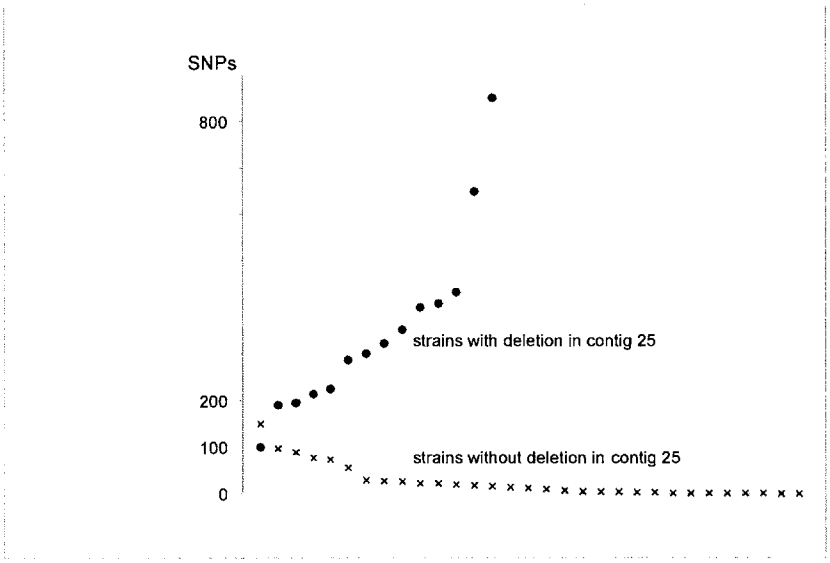


Figure 4-1: The number of SNPs added in the prior selection round is substantially higher for strains with a deletion in contig 25. Points are just spread out uniformly for visibility, so there is no significance to displacement on the x-axis. SNP totals for strains without the deletion are sorted from highest to lowest to show their maximum next to the minimum for strains with the deletion.

### 4.2.2 Excision of a prophage causes frameshift that inactivates *mutS*

Improper excision of a prophage causes a frameshift in *mutS*. This shift throws most of the protein sequence out of frame and introduces a premature stop codon. The truncated and garbled MutS lacks its functional domains for DNA mismatch recognition and repair.

```
12B01 wild-type \textit{mutS} (containing prophage)
M  Q  K  E  T  T  K  N  K  D  I  K  A  A  H  T  P  M  M  Q  Q  Y  L  K
ATGCAAAAAGAAACCACTAAAAACAAAGACATAAAAGCAGCGCATACACCTATGATGCAGCAGTACCTAAAA
-------------------------------------------------- deleted
```

```
infected mutS   MQKETTKNKDIKAAHTPMMQQYLKLKAENPEILLF...          (861 aa)
uninfected mutS         MKADQKHTPMMQQYLKLKAENPEILLF...          (853 aa)
post-excision mutS      MKADQKHTPDALVCNNIKNNNLKQPKLKNHTFTYTNQKS* (41 aa)
```

Figure 4-2: Alignment of the amino acid sequences for 12B01 with and without prophage shows how the post-excision *mutS* regains the host starting amino acids, but will now code for a different amino acid sequence. MutS translation in close relatives and the putative uninfected 12B01 begins with a GTG start codon.

MutS deficiency leads to a characteristic mutation spectrum dominated by transitions [100, 107]. We had already observed a strong transition bias in the mutations, lending further support to this deletion as the cause of hypermutation.

### 4.2.3 Alternative recombination homology matches the frameshift

While analyzing the boundaries of this deletion, we noticed that the sequences flanking the deletion had sequence homology to each other, forming *attL* and *attR* direct repeat regions typical of prophage elements (Figure 4-3). Further, there were two registers of homology, one in-frame for coding MutS, and one offset by two base pairs. We used NCBI BLAST [2] to compare with uninfected relatives and found the putative pre-insertion sequence of the region upstream and at the beginning of the *mutS* open reading frame. However, the sequence deletion in the evolved strains did not neatly restore this pre-insertion sequence, but rather recombined using the offset region of homology, causing an additional two base pairs to be lost in the deletion.

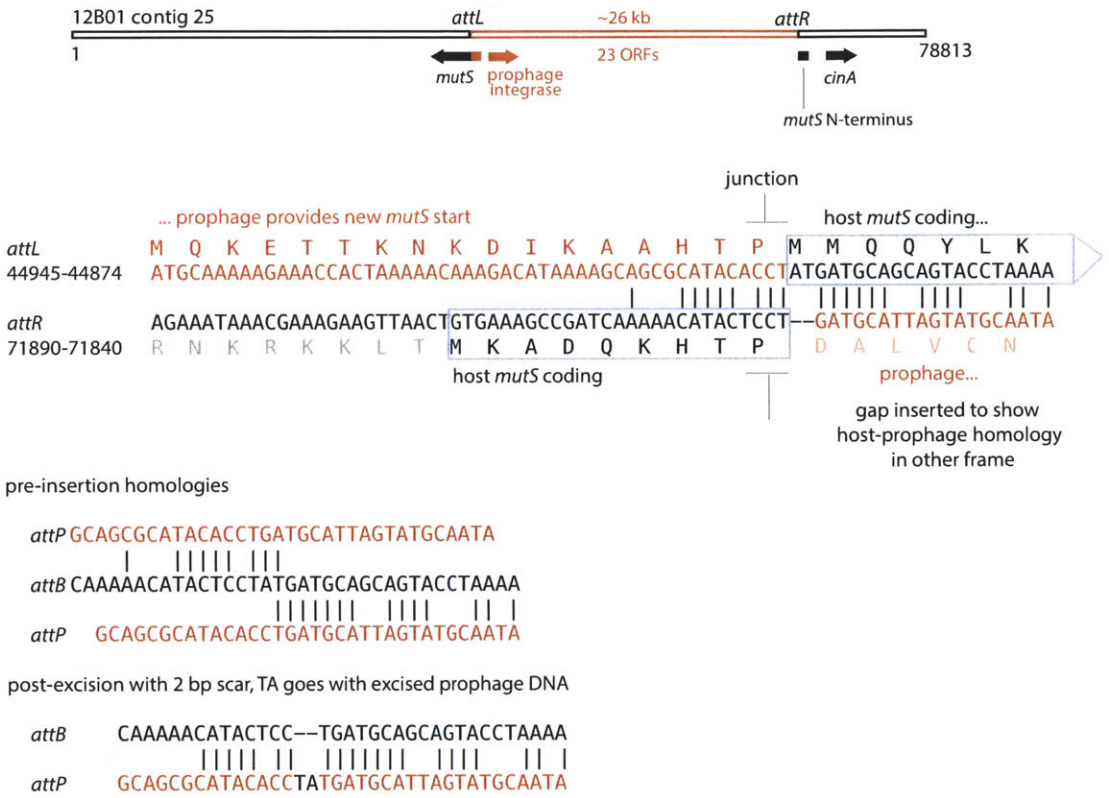


Figure 4-3: The prophage homology with the host allows it to recapitulate a coding start of the gene. An alternate homology allows improper excision, with the excised DNA removing a 2 bp scar segment from the host. Top: Arrows indicating genes are not to scale.

### 4.2.4 PCR assay strategy detects presence and excision of prophage

Since gene annotations and homologous repeats in the genome indicated that a prophage was responsible for the identical independent deletions in our mutators, we expected predictable and perhaps repeatable boundaries to the deletion. In order to track the prophage, we designed a set of PCR primers to target amplicons that would tell us if the prophage was present in the genome, excised, and if a circular extrachromosomal version was present (Figure 4-4).

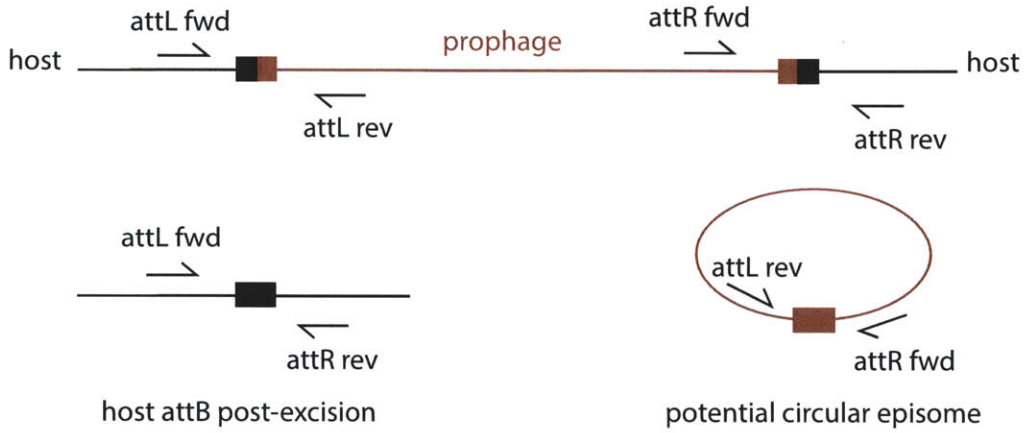


Figure 4-4: Primers flanking attachment sites amplify PCR products that detect the insertion or excision state of the prophage.

product: primers
*attL*: *attL* forward + *attL* reverse
*attR*: *attR* forward + *attR* reverse
*attB*: *attL* forward + *attR* reverse
*attP*: *attL* reverse + *attR* forward

### 4.2.5 The same PCR assay always detects a subpopulation of wild-type cells with excised prophage

In a subpopulation of wild-type ancestor from glycerol, the *attB* host-host junction amplifies. The *attL* and *attR* amplicons are present due to the intact prophage-host junctions in most cells, but there is detectable signal of excision as well. Critically, this template DNA was extracted from cultures grown at baseline salinity, comparable to seawater (LB-S 0.5M), not under salt stress or selection. High salinity is thus not a requirement for excision of the prophage region.

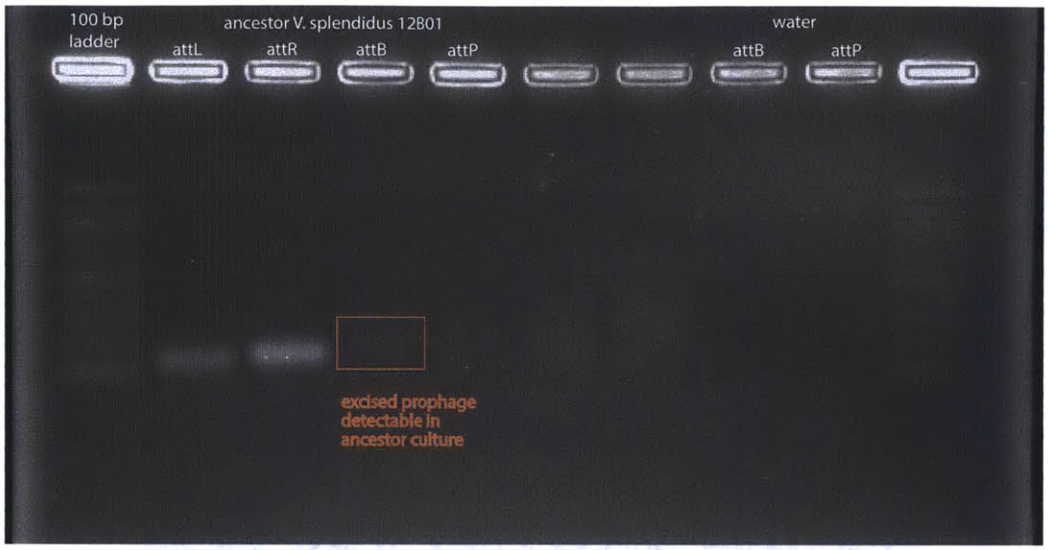


Figure 4-5: PCR on 12B01 wild-type (ancestor of our mutants) shows detectable excision of prophage (*attB*). Signal is smaller than from expected amplicons of *attL* and *attR* junctions.

### 4.2.6 The *attB* PCR product from wild-type has same scarred sequence as mutators

Though the band on the gel was faint, we were able to sequence the *attB* amplicon from DNA extracted from wild-type 12B01 culture. The sequence matched the *attB* of the mutator lines, post-deletion. We have sequenced several of these from different cultures and colonies of wild-type cells and never amplified an *attB* product whose sequence matches the putative uninfected sequence. Excision seems always to take place with this scar, which is the molecular mechanism of *mutS* inactivation.

evolved mutator *attB*, from whole genome sequence

```
GTGAAAGCCGATCAAAAACATACTCCTGATGCAGCAGTACCTAAAACTAAAA
||||||||||||||||||||||||||||||||||||||||||||||||||||
GTGAAAGCCGATCAAAAACATACTCCTGATGCAGCAGTACCTAAAACTAAAA
```

12B01 ancestor culture *attB*, from PCR product (above, and again below with gap to show alignment)

```
GTGAAAGCCGATCAAAAACATACTCCT--GATGCAGCAGTACCTAAAACTAAAA
|||||||||||||||||||||||||||  ||||||||||||||||||||||||
GTGAAAGCCGATCAAAAACATACTCCTATGATGCAGCAGTACCTAAAACTAAAA
```

putative pre-infection 12B01 *attB*, taken from *V. splendidus* 1S-124

### 4.2.7 Pure colonies with prophage excision can be isolated by colony PCR

We saw evidence of excision in the form of faint PCR bands for the *attB* product amplified from DNA extracts of wild-type cultures grown in baseline salinity (LB-S 0.5M) (Figure 4-5). We sequenced those PCR products and found they matched the junction sequence

from the mutators, having the same two base pair deletion (previous section). To rule out PCR artifact, we set out to isolate cells that were responsible for this signal. We diluted a sample of the wild-type 12B01 glycerol stock and plated it immediately. These colonies grew from single cells that had been frozen in the glycerol. We then ran PCR *attL* and *attB* assays on 96 colonies. 95 of 96 were positive for *attL*, but colony 94 (c94) was negative for *attL* and clearly positive for *attB* (Figure 4-6). We sequenced the c94 *attB* product, and its sequence matched the mutator as well.

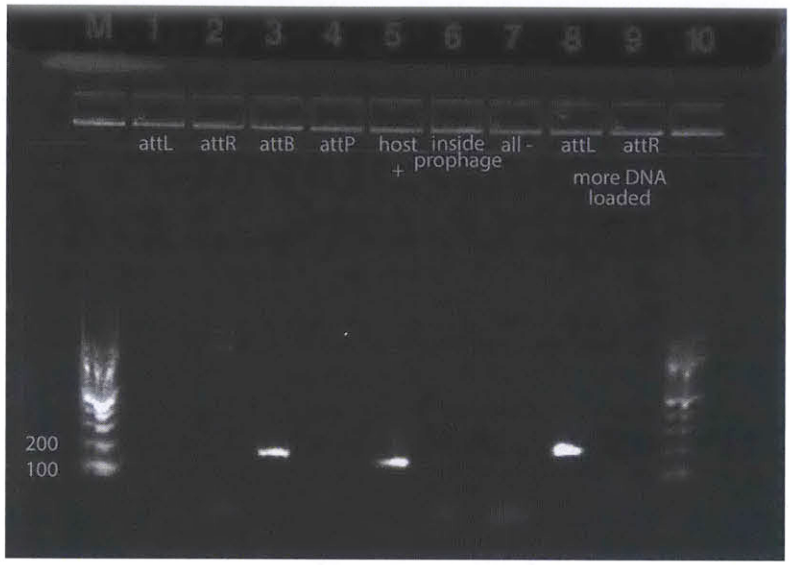


Figure 4-6: PCR on single colony of 12B01 wild-type (ancestor of our mutants) shows detectable excision of prophage (*attB* amplicon) and absence of integrated prophage (no amplicons for *attL*, *attR*, region inside prophage). No circular or linear prophage DNA is detected.

Below is the sequence of *attB* PCR product from colony 94 for the region around the integration site in the host. These sequences were 100% identical for the entire *attB* PCR product (178 bp). Colony 94, grown from a cell already present in the wild-type 12B01 glycerol stock, shows the same excision as the mutators. Hence, a subpopulation of wild-type cells carries this excision, the mutator-associated genotype.

```
evolved mutator \textit{attB}, from whole genome sequence
GTGAAAGCCGATCAAAAACATACTCCTGATGCAGCAGTACCTAAAACTAAAA
||||||||||||||||||||||||||||||||||||||||||||||||||||
GTGAAAGCCGATCAAAAACATACTCCTGATGCAGCAGTACCTAAAACTAAAA
c94 colony from glycerol stock, \textit{attB} PCR product
```

### 4.2.8 Colonies with the excision have a distinct morphology

When assaying colonies for excision, we extracted DNA from part of each colony for PCR, but also respotted some cells from each of the colonies on a new plate, kept as an index. After successfully detecting total loss of the element in one of the colonies, we noticed a distinct morphology in the supercolony grown from the transferred cells. Photographs of this starburst morphology are below (Figures 4-7 and 4-8). Even though the initial observation

was not in colonies grown from a single cell, this morphology does breed true. All descendant colonies present this morphotype.

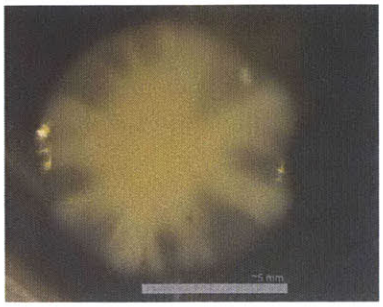

Figure 4-7: Colonies of 12B01 with excised prophage have a distinct starburst morphology.

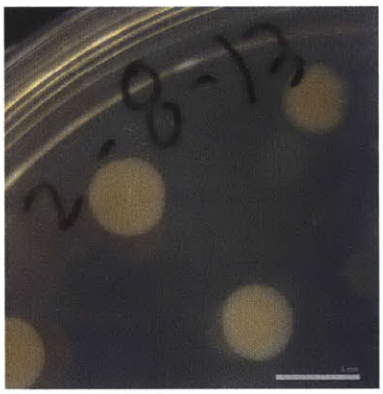

Figure 4-8: Colonies of 12B01 with excised prophage differ in morphology from wild-type (with prophage). These spots were transferred from another plate of single colonies in order to space them apart from each other, so these did not originate from single cells, but from a small pipet tip transfer.

Once we discovered this phenotype, we were able to run the same colony screen in reverse, beginning by plating colonies, screening by eye for the distinct starburst morphology, and successfully confirming the excision in that colony by PCR. The morphology is not easily visible with the naked eye until the colonies grow to about 5 mm in diameter.

We had observed that occasional wild-type colonies have one or a few clearer notches, similar to a partial, less dramatic version of the starburst, so we compared cells picked from inside the notch and outside the notch. Tests for the excision trait found no excision in either group, so that morphology must be tied to some other change. We believe the appearance of the colonies is not inherently due to the deletion, but to secondary effects of the hypermutation phenotype. Higher mutation rates turn rare morphological mutation events into common features even in a population the size of a single colony.

In the initial colony PCR above, we found one colony pure for the excision out of 96 total colonies (and then noticed its morphotype). In screening for the distinct morphology, we found the starburst morphology in 1/54 of plated colonies and confirmed the loss of the prophage in that one. Hence, there are cells carrying the excision genotype present at a detectable level of about 1-2% even within the wild-type stock.

### 4.2.9 No circular episome is detected

Though other cases of prophage excision involve circular episomes [94, 102], we have been unable to detect a circular episome by PCR in any strains. In cells lacking *attL* and *attR* amplification, indicating excision, we have also failed to amplify a region in the middle of the prophage element away from the junctions, which would indicate the presence of extrachromosomal linear prophage DNA, though exhaustive tests have not been made with this method.

### 4.2.10 Extracellular particles are detectable in wild-type cultures

In trying to learn the fate of the excised DNA, we filtered media from cultures of cells containing the prophage after growth to stationary phase. We used SYBR DNA staining and epifluorescence microscopy, a method used to detect free phage virions in liquid, and we detected particles consistent with the punctate appearance of virions. Attempts to isolate these particles and extract DNA have so far failed. We suggest future work to investigate these particles and their contents in Chapter 5.

### 4.2.11 Related prophages are in genomes of other Vibrios

Eighty closely-related Vibrio isolates have been whole-genome sequenced [32]. Searching for the conserved *mutS* and typically adjacent *cinA* gene reveals the presence of an intervening element in six closely-related genomes, mostly found in a clade of *Vibrio splendidus*. However, the strains with the feature are not monophyletic, according to alignment of the shared core genome. A tree of the prophage-containing strains is in Figure 4-9.

A strain not in the *V. splendidus* clade, *Enterovibrio norvegicus* FF-162, has a similar element within *mutS*, but it is lower in sequence identity, implying more distant relation. These inserted regions vary in length and gene content among the Vibrio strains, consistent with the commonly dynamic and mosaic character of phage and other mobile elements [23, 136]. Whether the element has been transmitted vertically from the common ancestor of the clade and lost in strains that lack it or whether it has infected multiple times and transferred laterally is difficult to determine. Prophage are usually under relaxed selection in the host genome, and they experience common gene deletions and interruptions [22, 23, 25, 57], so it is unclear if the differences between these elements have happened while in the current host during vertical descent, or in between infection events. From a preliminary phylogenetic analysis, the integrases are very well conserved, with few polymorphic sites, and the patterns of element size and gene content are not obviously informative.

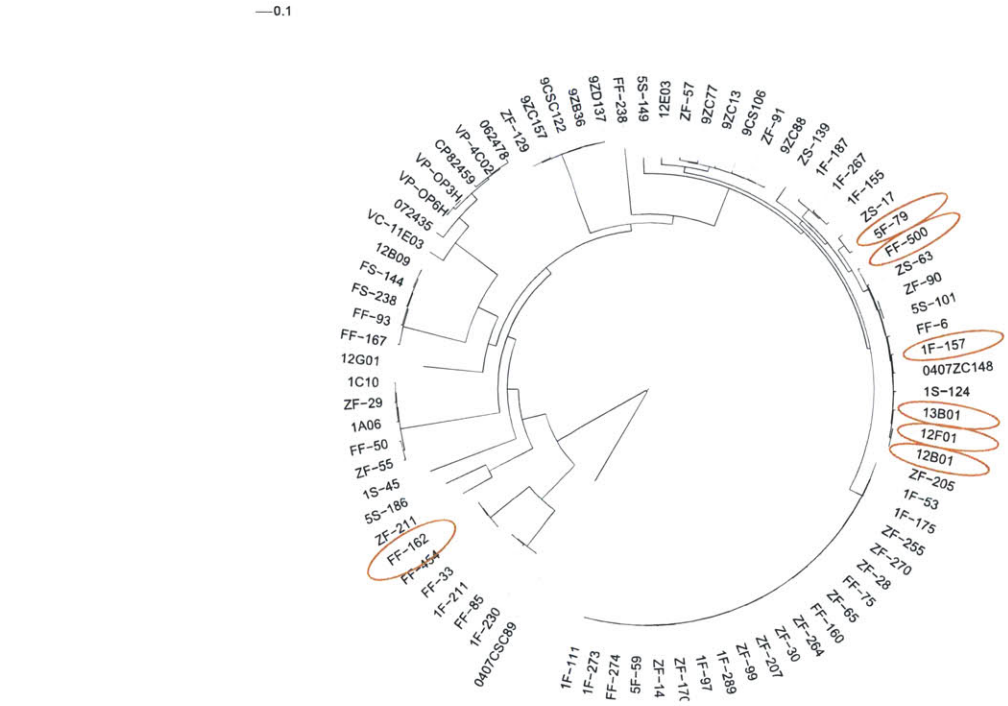


Figure 4-9: Vibrios with the *mutS* prophage (circled) are mostly in one clade, but are not monophyletic. This tree was constructed by Sonia Timberlake from alignment of the common core genome.

Table 4.1: Wide variation in sizes of the related prophages integrated into close relatives of 12B01.

| Strain | Length (kb) |
|---|---|
| *Vibrio splendidus* 12B01 | 27 |
| *Vibrio splendidus* 12F01 | 7 (split over 2 contigs) |
| *Vibrio splendidus* 13B01 | 46? (3 contigs) |
| *Vibrio splendidus* 1F-157 | 20 |
| *Vibrio splendidus* FF-500 | 23 |
| *Vibrio tasmaniensis* 5F-79 | 24 (2 contigs) |

### 4.2.12 Related prophages are in other beta- and gamma-proteobacteria

Since nearly every bacterial genome contains *mutS*, but not necessarily the prophage, we used the prophage integrase from 12B01 as a hook to search for related elements in sequence databases. Nearly all related integrases (some annotated as hydrolases) were located adjacent to *mutS*. Further searches for regions homologous to the insertion site revealed that these prophage elements are also inserted *within mutS* and also provide a new coding start sequence. The orphaned 5' upstream and N-terminal coding region of *mutS* can often be found at the other end of the prophage. A list of strains is in Table 4.2. When genome sequences of closely related bacteria are available, the beginning of *mutS*, the preceding gene and intergenic region can be used to determine the bounds and homology of the integration in the same way that we used uninfected *V. splendidus* sequences to infer the integration site in 12B01. The bounds of the prophages are extremely rarely annotated, though manual sequence comparison and curation reveal them. The ACLAME database of mobile elements only contained one of the prophages from strains listed below, that of *Burkholderia thailandensis* E264 [76].

Hits to *mutS* and an adjacent integrase were also found in metagenomic samples, including an "uncultured bacterium from groundwater metagenome" from Rifle, Colorado [138] and the well-known Global Ocean Survey [98]. Metagenomic assemblies may be too short to find the full prophage element in sampled genomes. Nonetheless, the discovery of related integrases adjacent to *mutS* broadens our knowledge of this family of elements and their hosts by directly linking the phylogenetic information from a host gene and a prophage gene.

Table 4.2: A high-confidence, partial list of bacterial strains containing an integrase adjacent to a disrupted host *mutS*.

| |
|---|
| *Acinetobacter baumannii* str. SDF |
| *Alcanivorax* sp. W11-5 contig6 |
| *Azoarcus* sp. BH72 |
| *Brenneria* sp. EniD312 |
| *Burkholderia pseudomallei* 112 |
| *Burkholderia pseudomallei* MSHR346 |
| *Burkholderia thailandensis* Bt4 |
| *Burkholderia thailandensis* E264 |
| *Enterobacter cloacae subsp. dissolvens* SP1 C813contig7 |
| *Erwinia amylovora* CFBP 1430 |
| *Erwinia carotovora subsp. atroseptica* |
| *Erwinia tasmaniensis* strain ET1/99 |
| *Escherichia coli* 536 |
| *Escherichia coli* F11 |
| *Haemophilus haemolyticus* M21621 014 |
| *Klebsiella pneumoniae* KCTC 2242 CP002910.1 |
| *Klebsiella pneumoniae subsp. pneumoniae* DSM 30104 AJJI01000014.1 |
| *Klebsiella pneumoniae subsp. pneumoniae* Ecl8 contig chr CANH01000031.1 |
| *Klebsiella pneumoniae subsp. pneumoniae* LCT-KP214 AJHE01000108.1 |
| *Klebsiella pneumoniae subsp. pneumoniae* WGLW1 AMLL01000002.1 |
| *Klebsiella pneumoniae subsp. pneumoniae* WGLW3 AMLN01000007.1 |
| *Klebsiella pneumoniae subsp. pneumoniae* DSM 30104 |
| *Laribacter hongkongensis* HLHK9 |
| Marine gamma proteobacterium HTCC2143 |
| *Methyloversatilis universalis* FAM5 contig00046 |
| *Providencia stuartii* ATCC 25827 |
| *Providencia stuartii* MRSN 2154 |
| *Pseudomonas putida* F1 |
| *Pseudomonas putida* GB-1 |
| *Salmonella enterica subsp. enterica* serovar Newport |
| *Vibrio splendidus* 12B01 |
| *Vibrio splendidus* 12F01 |
| *Vibrio splendidus* 1F-157 |
| *Vibrio splendidus* FF-162 |
| *Vibrio splendidus* FF-500 |
| *Vibrio tasmaniensis* 5F-79 |
| *Yersinia pseudotuberculosis* IP 31758 |
| *Yersinia ruckeri* ATCC 29473 |
| *Yersinia intermedia* ATCC 29909 |

## 4.3 Discussion

### 4.3.1 Excision mechanics

Excision of the prophage element is perfectly correlated with high number of SNPs, and the resulting frameshift in *mutS* provides an obvious mechanism. The transition bias reported for *mutS* deficient strains has been discussed in the previous chapter. We do not detect an episome (circular or linear), found in other reported cases [26, 102]. We also never detect re-entry of the prophage to the genome in later rounds.

The fate of the prophage DNA after excision from the host genome is not yet known. The gene content of the *mutS* element does not contain a typical set of phage structural genes, though two hypothetical genes are indicated as phage tail proteins by an artificial neural network predictor of phage structural genes [103]. Details are in Table 4.3. The element may be packaged into virions by a helper phage elsewhere in the genome, as in phage P2/P4 [29, 109]. There are several other prophage in the 12B01 genome, some with homologs to genes also found in the contig 25 prophage. The mosaic structure of prophage make this not uncommon [22].

Table 4.3: Predicted phage tail structural genes from contig 25 prophage.

| MicrobesOnline gene ID | RefSeq | a.a. | model score |
|---|---|---|---|
| hypothetical protein 2680760 | ZP_00991392.1 GI:84390130 | 122 | 0.999908 |
| hypothetical protein 2680766 | ZP_00991398.1 GI:84390136 | 107 | 0.874339 |

### 4.3.2 Genomic context

The prophage was present in the genome of wild-type isolate of 12B01, and related *V. splendidus* genomes contain similar elements, but with varying gene content and length. Genomes of some close relatives lack this infection, so the trait is not monophyletic. The different sizes and gene contents imply that this is a degraded lysogen. Lysogens are reported to be subject to inactivation by gene deletions, due perhaps to relaxed selection [23]. Our experiments show that the excision is still possible, even if packaging may not be. Observation of particles in the media coupled with the disappearance of this prophage from the genome raises the possibility of packaging of virions with this prophage's DNA using genes elsewhere in the genome.

### 4.3.3 Larger phenomenon of prophage disruption of host genes

There are growing reports of similar gene interruptions by mobile elements. Typically the host elements have been ubiquitous genes with high sequence conservation, such as tRNA genes. Similarly, *mutS* is a universally conserved gene, so it is always an available target for homologous sequence integration. The 3' end of genes is more often an integration target, but there are some cases of 5' integration and gene maintenance by mobile elements, such as the SXT integrative-conjugative element in *V. cholerae prfC* [58] and a lambdoid prophage in *E. coli prfC* [6].

Prophage are not well annotated within genomes, and automated gene calling algorithms used to annotate sequence databases do not recognize the change in *mutS*, since the phage

element provides a start ATG in the correct coding frame and upstream of the insertion site. Were this a simple gene disruption, the absence of any potential coding sequence for *mutS* would be alarming and immediately noticeable, but the phage has disguised itself both to the cell and to bioinformatics.

### 4.3.4 Model of potential relevance of this mechanism to bacterial evolution

We have found a previously uncharacterized mechanism for altering mutation rate. This prophage excision is potentially a switch between discrete mutation rates mediated by *mutS* inactivation. We speculate that this phenomenon discovered during evolution in the lab could be affecting wild strains, since the excision takes place with observable frequency even at the normal salinity of wild Vibrios' marine habitat.

Though mutators are detectable in wild populations, how often they originate and reach levels in the population high enough to be significant in adaptive evolution are still open questions [44]. The excision process can happen in many cells simultaneously, so this means of becoming a mutator is not limited by clonal expansion. This phenomenon would provide one means for mutators to become a significant portion of a population within a generation.

In our experiments, hypermutating lines show an ever-increasing amount of mutations. The long-term fate of such strains is unknown, but this phenotype is thought to be eventually lethal. In the wild, given rampant recombination, loss of this element may not doom a strain. Adaptive mutations generated in a mutator can be preserved by recombination into nonmutator relatives, or conversely, hypermutators could recover normal repair function by recombination. The latter scenario is supported by some phylogenetic data showing the increased transfer of repair genes [38].

### 4.3.5 Further examination of *mutS* as integration site

This family of elements has a common integration site centered on the HTP(MM) motif coded for in *mutS*. The insertion takes advantage of this conserved part of the MutS protein. It appears that codon variation and the homology for recombination are associated, since by mimicking the host protein coding sequence, even using different synonymous codons from the host, the phage still has enough homology in its *attP* site to integrate.

Even though the prophage provides a new start codon and in-frame sequence prior to the HTP motif, we do not yet know the functional level of infected MutS, which has a prophage-host hybrid peptide sequence. We believe no one has yet studied *mutS* expression or protein function in infected versions (even inadvertently). MutS structure and, more likely, its regulation may be already affected by insertion, and the scarred excision represents complete inactivation. In this scenario, a paper that competed single-mutator (high mutation rate vs. wild type) and double-mutator (even higher) strains of *E. coli* might be relevant [50]. That study found the same pattern between single and double mutators as studies comparing wild-type and single mutators. Higher mutator alleles accumulated if their beneficial mutation supply rate was higher than that of the lower mutators. Smaller population sizes or hard selection favored the higher mutators.

The other common feature is the integrases, which are more closely related by homology to each other than to integrases found in phages or in prophages with different insertion points. This association allows identification of related elements, since the most closely related integrases are within prophages inserted into *mutS*.

The exact residues in the integrase that confer site-specificity are unknown. Because the integrases found at the *mutS* insertion site are more closely related to each other, deconvolving the phylogenetic signal from the site-specific signal is very difficult.

### 4.3.6 Broad presence of this element

Members of a family of related prophage elements are found in at least 50 genome-sequenced proteobacteria, including genera of important human pathogens as well as non-clinical environmental isolates from at least 3 continents and habitats including plant endophyte, ocean, subsurface groundwater, and soil. Infected strains also include human commensals and pathogens. Lysogenized genera include *Vibrio cholerae*, *Burkholderia* spp., *E. coli*, *Acinetobacter* spp., and *Klebsiella pneumoniae*, bacteria with great clinical importance. Horizontal gene transfer and mutator strains in these pathogens have been connected to the spread of antibiotic resistance, and this element may be connected to that rise, by facilitating the conversion of infected cells to mutators that can more rapidly generate resistance mutations that may be then transferred among non-mutator strains. Sequence identity of the integrase and insertion at the 5' end of *mutS* are the primary common features of this family.

In Vibrios carrying the prophage, the host's upstream and N-terminal regions of *mutS* remain identical to the pre-infection sequence inferred from close relatives, even though they are now separated from the gene by thousands of bases. We do not yet have the phylogenetic analysis to rigorously examine this point, but it may support our model of transience. After integration of the prophage, the now-orphaned host sequence for the N-terminal region should be under relaxed selection and accumulate mutations. There are at least two explanations for why it has not. Integration may be a recent event in all these strains, and not enough time has passed yet for mutation to affect this region. Another possibility is that there is *not* relaxed selection on the orphaned host start sequence. If excision events are common and the host start is reattached upon excision events, the host start would remain under selection to ensure proper DNA repair.

MutS was an early subject gene for phylogenomics due to its ubiquity and importance in DNA repair [43]. The presence of the *mutS*-element affects the phylogenetic information if the starting sequence is part of the sequence comparison. In the case of closely related Vibrio strains with low % divergence, the phylogenetic tree is greatly affected by the signal of insertion, with strains containing the element grouping together, in disagreement with a tree based on the rest of the protein or on alignable regions of the whole genome. The varying amino acid sequences might have been attributed to relaxed selection on the region.

## 4.4 Conclusions

As in these experiments, experimental evolution has rewarded prior investigations with the discovery of unexpected pathways and outcomes of adaptation. These are anecdotal cases with broadly relevant lessons, such as the importance of composition of growth media [27], unexpected morphotypes [95], and new physiologies [16].

Mobile elements and means of gene transfer, such as phage and prophage, drive rapid and dramatic evolution visible in both comparative sequencing approaches and single-strain evolution experiments. Here we have presented another example of the latter, with dramatic genetic results and potential for relevance to evolution in the wild as well.

## 4.5 Methods

### 4.5.1 PCR assays

Primer sequences generated with NCBI primer-blast and checked with blastn (E<1, word size 7) vs. 12B01.

Table 4.4: Primers used to amplify *attL*, *attR*, *attB*, *attP* sites in *V. splendidus* 12B01.

| Primer | Sequence | Region |
|---|---|---|
| 12B01_c25_*attL*_v1_FWD | GCGTCATCGTAGAAAAGCTCG | 44800-44820 |
| 12B01_c25_*attL*_v1_REV | ACAAAGACATAAAAGCAGCGCA | |
| 12B01_c25_*attR*_v1_FWD | AGTTTAGGTTGCTTTAAGTTGTTGT | |
| 12B01_c25_*attR*_v1_REV | GCACGAAACGCTTTCACGTA | 71924-71905 |
| 12B01_c25_*attL*_v2_FWD | CCATGCGGTAGAACAGCAAA | |
| 12B01_c25_*attL*_v2_REV | TGCAAAAAGAAACCACTAAAAACAA | |
| 12B01_c25_*attR*_v2_FWD | GTTTAGGTTGCTTTAAGTTGTTGTT | |
| 12B01_c25_*attR*_v2_REV | TCACGTATCCTAAGCCGCAA | |

### 4.5.2 Database searches for related prophage

NCBI BLAST [2] was used for finding related prophage. The integrase gene protein sequence (NCBI accession ZP_00991386.1) was the query for tblastn and blastp searches against bacterial genomes, including drafts. Hits were then examined for position adjacent to *mutS* either with genbank annotations or by blastx of neighboring nucleotide sequence to confirm coding for *mutS*. The reading frames of integrase and *mutS* always code in opposite directions. Integrases found adjacent to *mutS* in other bacteria were then also used as tblastn queries to expand the search.

To find the att sites of infected strains, the nucleotide sequence of the 5' intergenic region plus the start of *mutS* until the HTPMMQQ motif was used as a blastn query against the same genome. In some cases, tree examination of the gene order in an uninfected close relative indicated the gene that would normally be at the 5' end of *mutS*. In infected strains, this gene (now separated from *mutS*) indicates the resumption of host sequence and limits the boundaries of the search for att sites. MicrobesOnline was used as an invaluable tool for genome comparison and visualization [37]. The prophage genes are sometimes annotated as such, but are often dominated by annotation as hypothetical.

### 4.5.3 *V. splendidus* 12B01 contig 25 prophage gene contents

Annotations from MicrobesOnline [37]:

N.B. Though it is an intriguing annotation, the DNA repair protein within the prophage has an inaccurate annotation. It is a RadC homolog, but RadC's function is actually unknown, and it has been shown not to have the attributed DNA repair function [7].

```
Phage_integrase__2680754__1
hypothetical_protein__2680755__1
```

transcriptional_regulator__2680757__1
ribonuclease_H__2680758__1
hypothetical_protein__2680759__1
hypothetical_protein__2680760__1
hypothetical_protein__2680761__1
hypothetical_protein__2680762__1
hypothetical_protein__2680763__1
DNA_repair_protein__2680764__1
deoxyguanosinetriphosphate_triphosphohydrolase__2680765__1
hypothetical_protein__2680766__1
hypothetical_protein__2680767__1
hypothetical_protein__2680768__1
hypothetical_protein__2680769__1
hypothetical_protein__2680770__1
type_I_restriction-modification_system,_R_subunit__2680771__1
type_I_restriction-modification_system,_M_subunit__2680772__1
type_I_restriction_enzyme_specificity_protein__2680773__1
GTPase_subunit_of_restriction_endonuclease-like__2680774__1
hypothetical_protein__2680775__1
Uncharacterized_phage-associated_protein__2680776__1
transcriptional_regulator,_putative__2680777__1

# Chapter 5

# Future work

This thesis began with simple experiments that eventually revealed the dramatic effects of a previously unnoticed prophage element. This discovery shifted the focus of the work from understanding adaptation to an environmental stressor to understanding the effects of the prophage and its interaction with host adaptation. This new discovery merits further investigation.

Potential future work exists for all the preceding chapters, but it should first address the unprecedented mutators we have found through the experimental evolution process. Continuing from Chapter 4, there are still many questions about the distribution, effects and behavior of the prophage that led to these mutators. In following up on chapter 3, more analysis of mutational dynamics and spectra could reveal the basis of uneven mutation levels in the nonmutators. Also, the frozen set of 40 evolved strains are an interesting avenue to studying further questions of connections between phenotype and genotype, as well as niche dynamics and tradeoffs. Having an archive of organisms that can be regrown at the investigator's whim is a valuable facet of experimental bacterial evolution. The sections of this chapter run in roughly the reverse order of the prior chapters, since the mutator phenotype and prophage-based mechanism are the most pressing topics to pursue further.

## 5.1 Mutators and *mutS* prophage (work continued from Chapter 4)

### 5.1.1 Can naive strains be infected with the *mutS* prophage element?

Showing the transmissibility of this element after it has been a lysogen (as we found it in 12B01) and then excised would be key to understanding its dynamics in a population. We have tried infecting mutator cells that have completely lost the element by exposing them to supernatant from infected cells. These trials have failed, but that may be due to the excision scar in the mutators. The 2 base pair scar leaves a different sequence in the area of insertion from that in cells that have not been infected. If we could detect transfer of these genes to a naive cell by exposing it to supernatant from infected cells, it would obviate much of the need for DNA extraction from particles. This may well be possible, as there are close relatives lacking the prophage in the large strain collection the Polz lab has isolated. These close strains should be very similar in phenotype. Versions of these strains that express GFP or RFP have been constructed. These markers, coupled with PCR, would make possible a screen for a naive strain with the prophage integrated. Marked strains also allow precise

fitness comparisons using flow cytometry.

### 5.1.2 Do naive strains represent a different wild-type/negative control?

Measurement of mutation rate and spectrum in strains with a clean genomic background may more closely reflect typical *mutS* function. Since 12B01 has presumably been infected since its isolation, the wild-type *is* its genome containing the prophage. By sequence comparison, close relative of this strain appear to be naive–strains that have never had the lysogen, or perhaps have experienced a scarless excision. The latter possibility is especially intriguing, but the two alternatives are impossible to distinguish just by looking at the insertion site. Strains that have been infected may have other signals in mutational spectrum or SNPs affecting host sequence homology to the phage. These strains may be better representatives of the host's behavior as a baseline for our 12B01 wild-type and mutator strains. Only because of the expansion of whole genome sequencing, applied to this whole family of more than 80 Vibrios, do we have such close relatives differing in this key feature.

### 5.1.3 What is the fate of the excised prophage DNA? What are the particles observed by fluorescence and their contents?

The fate of the excised prophage DNA is so far unknown, but it is key to understanding the significance of this element for populations of Vibrio and other lysogenized bacteria. A circular episomal form of the element has not amplified with PCR. Such a form has been reported in some cases of switchable lysogenic conversion [34, 94, 102], but our prophage may be degraded after excision.

Another possibility is packaging by structural proteins encoded in another phage. This helper (or competitor) phage phenomenon is known to occur, being best-studied in the P2/P4 phage system [109]. There are other prophage in the 12B01 genome that share homologous genes with this element, making this a reasonable possibility.

By SYBR Gold staining and epifluorescence microscopy, we have observed punctate, phage-like particles in the media of wild-type 12B01. These particles are not present in supernatant from cultures of cells that have all previously lost the element.

Extraction of DNA from these particles has been unsuccessful. DNase treatment of the supernatant before extraction is necessary to prevent host contamination, but DNase is sensitive to the salt concentration in our growth medium. We attempted to lower salt concentration in the media as we concentrated the particles with a 10kD filter, but this process appeared to cause rupture of the particles based on gel images. We believe a smaller change in salt concentration and a more salt-tolerant DNase could solve this problem. Adjustment of conditions during concentration and DNAse treatment could yield DNA isolated from within these particles and uncontaminated by host genomic DNA. Because these do not seem to be lytic phage (we never see clearing of cultures), the yield of DNA from extrachromosomal elements is lower and sensitive to contamination.

PCR amplification in the extracted DNA by primers matching the element, but not by control primers for a host gene, would prove packaging of the element. If this is unsuccessful but DNA is detectable, full sequencing of the enclosed DNA could be undertaken, as has been done with Vibrio phage isolated from the wild.

If the excised DNA is degraded in the cell, then this mutational mechanism is one that dooms the cell and does not affect relatives via transfer of the infective element. However, the cell with the deletion could provide mutations to the broader gene pool by recombination,

where adaptive alleles would be favored by selection and could spread in non-mutators. The mutator cell could itself regain a functional *mutS* by recombination, though if transfer likelihoods are agnostic to gene locus, this would seem a less likely scenario. Depending on insertion sequence, a related phage integrating at the same site might reintroduce an in-frame start site. This phage-based rescue scenario would render the bacteria dependent on the phage for survival as well as modulation of mutation rate.

### 5.1.4 What are the baseline frequency and modulating factors of excision?

The PCR assay we have been using to detect the state and presence of the genome and element can be modified to a quantitative PCR assay to measure the fraction of genomes where the prophage has excised. This fraction, excised/total, corresponds to the ratio of attB/host amplicons. Knowing the input mass of DNA, we can use a strain completely lacking the prophage, such as an evolved mutator (2.1) or mutator isolated from wild type (c94) for an attB calibration curve, and a host gene for calibration of total genomes. Then we could ask what factors modulate excision, such as stress or growth phase. Salt was not required for excision and did not seem to raise the mutation rate *per se*, but perhaps high salt raises the proportion of cells that become mutators. If not, that is evidence that the enrichment of mutators we saw in our evolved lines (3 of 9 lines instead of 1 in 96 colonies) is due to selection favoring adaptive mutations being found more frequently by a mutator, as has been reported.

### 5.1.5 Bioinformatic study of the *mutS* prophage family found in many sequenced proteobacteria

Sequence comparison in this family of elements can help us infer the coevolution of phage integrases and regions of integration (att sites and nearby) in both phage and host regions. This bioinformatic approach will help establish a host range for this element, how it has spread, and the set of susceptible sequences. The phage relies on homology with a conserved region of a universal gene so that it may integrate, but perhaps the bacterial host also benefits from the modulation of mutation rate. This modulation of mutation rate, happening either in just some cells or reversibly, may produce adaptive mutations that can spread within a population without dooming it to extinction, a death by 1000 cuts (in this case, 1000s of mutations).

If this element is responsible for higher mutation in clinical strains, it would be a good target for screening methods or inactivation. Mutators are a common topic in the literature on pathogens, as mutators enable antibiotic resistance and adaptation to the host environment during disease progression. Reviewing that literature for strains containing the element would associate a causal mechanism with this clinically important phenotype.

With comparative sequencing, we should look for signatures of hypermutation (elevated transitions) in strains containing the element at *mutS*, along with closely related strains that may have had it or recombined genetic material with mutators. Excess abundance of transitions in adaptive mutations are an indication that these mutators could be generating adaptive alleles that transfer into the nonmutator population, as our conceptual model predicts.

### 5.1.6 What are the effects of phage interruption of *mutS*?

We know from sequence analysis that the prophage always provides a start codon within a short coding distance of the integration site within *mutS* coding sequence, but the expression and function of *mutS* in cells with the prophage integrated are unknown. In our study, we unintentionally began with a strain containing this prophage, and we observed its improper excision during experiments. Our data compares these two states: infected and excised. There is a third, naive, state with intact *mutS* and without a prophage. Several isolated Vibrio strains lack the prophage but are otherwise nearly identical in nucleotide sequence to 12B01. Comparisons of the naive state with the integrated state would reveal the effect of the upstream regulatory region and first 10-20 amino acids provided by the phage prior to the host *mutS* sequence. We hypothesize that infected *mutS* retains function, since the results of a truncation are so dramatically different, but it seems likely that regulation of its expression would change. The difference in *mutS* between naive and infected cells is important for interpreting the widespread presence of the prophage in other genera of bacteria, since we observe the infected state in their genomes, rather than the excised state with its more dramatic mutator phenotype.

## 5.2 Genomics of salt-tolerant mutants (work continued from Chapter 3)

### 5.2.1 Did we discover salt-tolerance mutations?

This experiment in selection and resequencing aimed to discover mutations that could be tested for their adaptive benefit in high salt, since which changes confer tolerance are unknown. The SNP analysis includes some interesting potential target genes, such as those related to glycine betaine, a known osmolyte. These loci are enriched but not statistically significant. The few transversion mutations are also potentially enriched in adaptive mutations, as transversions are less likely to have been generated by random hypermutation. While the overall count of mutations is daunting, the few rounds that have 1-5 mutations present a manageable set of variants to test. These rounds were expected to be the norm when we began the experiment, hoping that the adaptive mutations would be immediately discernible as salt tolerance increased.

In order to convincingly attribute benefits to putatively adaptive SNPs, such as those affecting ion transport or the glycine betaine pathway, SNPs seen in evolved lines could be introduced into the ancestral background strain to test for a salt advantage. Allele swaps to introduce the mutant version to wild-type 12B01 would be attempted. The salt tolerance of these single-site mutants would be compared to wild-type to look for benefits. There is a caveat to working with a non-canonical organism, in that we have limited tools for genetic manipulation of these strains. Gene replacements and deletions made by conjugation have been made in closely related strains, but not yet in 12B01.

Alternately, the same experiment could be repeated with a different strain that lacks this prophage element. If the prophage is causing the mutation rate increase, then a selection lacking it might return fewer, but more informative mutations. 12B01 was chosen for this work as one of few environmental Vibrios with a known genome sequence at the start of the experiment. Prior reference genomes are no longer a requirement, due to increases in sequencing capacity, so any culturable strain(s) could be evolved.

Finding beneficial mutations could have application for engineering more tolerant organisms, if target proteins and functional effects of mutant alleles could be identified. Leads generated here would constrain the variables present in an entire genome for a general trait. The more established process of directed gene evolution could then construct and test more variants of specific targets.

### 5.2.2 Testing mutator survival

The long-term fate of mutators is unknown [50]. Our strains are accumulating nonsynonymous mutations at an astounding rate. So far, they continue to grow. We hypothesize that this level of mutation is ultimately unsustainable without rescue of the mutator phenotype. If they accumulate mutations at a rate faster than can be repaired or selected against, then they should eventually be overwhelmed with deleterious mutations. The rate of added mutations has seemed linear and shows no sign of diminishing, but how long can it continue? We believe that the mutation rate is now fixed at a high level, even without salt stress.

A mutation accumulation experiment would address this question. Clones from each mutator line could be passaged in streaks on plates from single colonies, archived, and resequenced to measure changes in the rate of mutation accumulation and to measure how many further mutations they can possibly accumulate. Generations could be counted carefully to calculate a rate of mutation/generation by counting changes at four-fold degenerate, silent sites. If they remain viable, resequencing these survivors would reveal even more sites that are neutral or not especially deleterious. Though what fraction of mutations are neutral is still an open question, if most are, then we could see huge shifts in the genome before extinction [44].

We also speculate that in the wild, mutators could both be doomed and generate adaptive mutations that survive via horizontal gene transfer. In related wild populations, recombination is a major factor in distributing adaptive mutations [104], so alleles generated in mutators could spread to nonmutators, indirectly selecting for the maintenance of potential for some cells to become mutators. If mutators with this mechanism generated adaptive mutations for the rest of the population, we can test computationally for signal of the extreme transition bias in recently recombined DNA that has not undergone purifying selection.

### 5.2.3 Detecting tradeoffs

Tradeoffs were already of interest from an ecological perspective in the previous chapter, even before we had mutational data. Now, there are also questions of tradeoffs due to mutational load. We expect the mutator strains to have suffered severe tradeoffs in areas other than what we suspect to be the main priority of the selection, growth on salty rich media. Tests for auxotrophy along metabolic or environmental axes would determine the nature and extent of the tradeoffs these strains have experienced. Phenotypes of mutators, nonmutators, and wild-type would be compared. The nonmutators present an intermediate case of strains adapted to salt, but without as many mutations, so we should be able to detect if tradeoffs are more due to the mutational load or to adaptation along the particular axis of salinity.

There may also be tradeoffs just along one niche axis. Further growth experiments under a range of salinities would measure whether strains adapted to higher salt have expanded their bounds of tolerance, or simply shifted them, perhaps revealing biophysical constraints on adaptation.

Earlier work in the lab addressed two-dimensional niche space, so we already have a basis for comparison with wild strains [80]. Though temperature and salinity response are thought to be linked, there are several possible outcomes of strains selected on one of these factors. The mutations could improve tolerance of both stresses, shifting the niche shape outwards. Alternately, under strong selection for one stressor, mutations could shift tolerance of only that stressor. Temperature tolerance could be unaffected, or be reduced if there is an underlying tradeoff. The distribution of mutational effects among these categories is an open question, but these tests would add further empirical input to the theory of interacting stressors of the previous work.

### 5.2.4 Further study of the cause of variability in the number of SNPs/round in the nonmutators.

The semblance of transient hypermutation in the nonmutators is one of the most intriguing findings, and merits more investigation. The phenomenon visible in the spikes of mutation may reveal more about the effects or life cycle of the *mutS* prophage.

### 5.2.5 Closing the genome with *de novo* assembly of existing reads

Currently 12B01's genome is not closed, but is in 119 contigs in Genbank (AAMR01000001 - AAMR01000119). We should take advantage of having 1200X coverage of essentially one strain to combine contigs and attempt to close the reference genome. Even SNPs among the 9 mutator strains will be a minority of base calls because each line was sequenced to about the same coverage and the SNPs were not parallel across lines. A closed 12B01 genome would facilitate assembly of the many related Vibrio genomes in the Polz lab collection.

### 5.2.6 Small insertions and deletions

From the patterns of assembled reads in the evolved genomes, small indels do not appear to be a common mutational mechanism. However, I have not yet analyzed or assembled the reads with a program that specifically looks for these changes. If there are loss-of-function adaptations, then these mutations could easily cause that loss of gene function by shifting the gene's translational reading frame or disrupting a regulatory region.

## 5.3 Selection method, Salt tolerance phenotype (work continued from Chapter 2)

### 5.3.1 Comparing selective methods

Can using hard selection on plates lead to more rapid improvement focused on tolerance, not growth? We have shown strain improvement with our method, but have not compared it directly with other experimental evolution techniques.

Results of selection experiments also have applied purpose now that engineered bacteria are also being to chemical production, sensing, remediation, and other industrial needs. Both the resulting improved strains and knowledge of how best to improve bacteria will be helpful–indicating either particular genetic targets for engineering or methods for strain improvement [14, 112, 139]. A new version of a strain with improved tolerance to a controllable environmental factor would allow selection for that strain in an mixed culture industrial setting or allow operation in a regime less susceptible to contamination.

Work from Pilpel and colleagues found in yeast that different selection stress dynamics led to different evolutionary dynamics and a notable difference in mechanism (transient aneuploidy) [139]. Gradual rather than abrupt application of stress favored alternative solutions and possibly more overall improvement than jumps, so a future experiment could test that idea in this context or use a chemostat and compare the results with our gradient plate method. Their study used a microbial eukaryote, which is single-celled like our Vibrio, but capable of different genetic mechanisms, including maintained genome duplication. The general concept of different regimes of evolutionary rate and adaptive mechanism could still hold. The gradual selection approach could be adapted by stretching out the gradient, plating more cells in a smaller range to select for mutations of smaller effect one at a time, and comparing that adaptive potential to that of mutants isolated from large leaps on the plate. Do many small increments ultimately produce more improvement than fewer bigger ones?

A morbidostat experiment using increasing salt concentration as a growth-limiting factor would present an interesting contrast. The morbidostat would produce strains with the same growth rate as the ancestor, but at whatever higher salt concentration adaptation would allow. Mutators could still hitchhike with beneficial mutations, but they would not immediately sweep the population as they did in our colony bottlenecks. The rate of salinity increases required to keep growth rate constant might reflect the appearance of adaptive mutations in the culture, providing a direct connection between the dynamics of phenotypic improvement and genetic changes.

### 5.3.2 What are the dynamics and limits of salt adaptation?

Even further improvement could still be possible with our own method. We stopped selecting after 5 rounds, given that we had enough mutants for sequencing. So far, each new round has always produced new, fitter mutants growing beyond the confluent boundary. Future work on the dynamics and limit of adaptation, perhaps by extending this process with the same strains, would be very helpful.

Another informative experiment could look at the extension of the salt tolerance range in the opposite direction to compare the improvements. Salt minima were difficult to measure since the growth medium may have contributed sodium and chloride. Titration of salt to lower concentrations in a more defined growth medium would reveal the lower limits and potential tradeoffs of tolerance range. Selection for high salt tolerance may lead to adaptations that limit the cell's growth at lower concentrations. Mutations that permit survival at lower salinity would be interesting in their own right, since the requirement of marine bacteria for some salt (obligate halophily) is a defining feature of their physiology. Related species, including the epidemic pathogen *Vibrio cholerae*, can survive in fresh water. Finding the genetic basis for that ability might be uncovered by experiments to adapt marine Vibrios to fresh water, which might help us connect genetics with lifestyle.

Salinity was only manipulated with addition of sodium chloride, supplementing a rich growth medium that surely contains some level of salts already. We cannot not attempt to replicate the complexity of the natural environment and all its biotic and abiotic conditions, and here we simplified salinity to just these two ions, which are the predominant ones by mass in sea water. There is not one accepted standard for testing salinity tolerance. Published experiments have used solid or liquid media and a variety of culture volumes and growth measurements to evaluate tolerance. It is reasonable to expect differences between growth on solid and liquid media of the same salinity, even if agar is the only variable ingredient. For

purposes of this experiment, solid and liquid media conditions corresponded well enough that strains survived the transitions between solid and liquid and back to solid salt-supplemented media. Reconsidering the experiment might involve finer control over salinity and nutrients, and fewer transitions between solid and liquid stages. These changes might better isolate the selection pressures (solid vs. liquid, set threshold for time or growth, rate vs. yield).

Perhaps the variation seen in experimental evolution on suspected environmental traits of importance (nutrients, temperature, salinity, pH) will correspond to the observed natural diversity on an allele or locus level. If so, we could assign putative adaptive benefits to sequence variants present in the wild.

## 5.4 Conclusion

Experimental evolution enables unintended discoveries that are among its greatest rewards. Surprises and enigmatic results lead to revised hypotheses and results, revealing new phenomena behind bacterial genetics, physiology, and behavior. Further application of experimental evolution and resequencing to different environments and different species will continue to tell us the multitude of ways bacteria may naturally adapt, as well as how we can engineer improved strains and guide evolution towards desired traits.

# Bibliography


[1] D Agashe, N C Martinez-Gomez, D A Drummond, and C J Marx. Good Codons, Bad Transcript: Large Reductions in Gene Expression and Fitness Arising from Synonymous Mutations in a Key Enzyme. *Molecular Biology and Evolution*, 30(3):549–560, February 2013.

[2] S F Altschul, T L Madden, and A A Schäffer. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids ...*, 1997.

[3] James B Anderson, Jason Funt, Dawn Anne Thompson, Snehit Prabhu, Amanda Socha, Caroline Sirjusingh, Jeremy R Dettman, Lucas Parreiras, David S Guttman, Aviv Regev, and Linda M Kohn. Determinants of Divergent Adaptation and Dobzhansky-Muller Interaction in Experimental Yeast Populations. *Current Biology*, 20(15):1383–1388, August 2010.

[4] Susana P Armada, Rosa Farto, Maria J Perez, and Teresa P Nieto. Effect of temperature, salinity and nutrient content on the survival responses of Vibrio splendidus biotype I. *Microbiology*, 149(2):369–375, February 2003.

[5] Frances H Arnold, Patrick L Wintrode, Kentaro Miyazaki, and Anne Gershenson. How enzymes adapt: lessons from directed evolution. *Trends in Biochemical Sciences*, 26(2):100–106, February 2001.

[6] Masahiro Asakura, Atsushi Hinenoya, Mohammad S Alam, Kensuke Shima, Shamim Hasan Zahid, Lei Shi, Norihiko Sugimoto, A N Ghosh, T Ramamurthy, Shah M Faruque, G Balakrish Nair, and Shinji Yamasaki. An inducible lambdoid prophage encoding cytolethal distending toxin (Cdt-I) and a type III effector protein in enteropathogenic Escherichia coli. *Proceedings of the National Academy of Sciences*, 104(36):14483–14488, September 2007.

[7] Laetitia Attaiech, Chantal Granadel, Jean-Pierre Claverys, and Bernard Martin. RadC, a misleading name? *Journal of Bacteriology*, 190(16):5729–5732, August 2008.

[8] Tomoya Baba, Takeshi Ara, Miki Hasegawa, Yuki Takai, Yoshiko Okumura, Miki Baba, Kirill A Datsenko, Masaru Tomita, Barry L Wanner, and Hirotada Mori. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol*, 2(1):–, February 2006.

[9] Jeffrey E Barrick, Dong Su Yu, Sung Ho Yoon, Haeyoung Jeong, Tae Kwang Oh, Dominique Schneider, Richard E Lenski, and Jihyun F Kim. Genome evolution and adaptation in a long-term experiment with Escherichia coli. *Nature Genetics*, 461(7268):1243–1247, October 2009.

[10] Albert F Bennett and Bradley S Hughes. Microbial experimental evolution. *American journal of physiology. Regulatory, integrative and comparative physiology*, 297(1):R17–25, July 2009.

[11] Albert F Bennett, R E Lenski, and J E Mittler. Evolutionary Adaptation to Temperature. I. Fitness Responses of Escherichia coli to Changes in its Thermal Environment. *Evolution; international journal of organic evolution*, 46(1):16–30, 1992.

[12] Albert F Bennett and Richard E Lenski. Evolutionary Adaptation to Temperature II. Thermal Niches of Experimental Lines of Escherichia coli. *Evolution; international journal of organic evolution*, 47(1):1–12, February 1993.

[13] Ivana Bjedov, Olivier Tenaillon, Bénédicte Gérard, Valeria Souza, Erick Denamur, Miroslav Radman, Francois Taddei, and Ivan Matic. Stress-induced mutagenesis in bacteria. *Science (New York, N.Y.)*, 300(5624):1404–1409, May 2003.

[14] Ian K Blaby, Benjamin J Lyons, Ewa Wroclawska-Hughes, Grier C F Phillips, Tyler P Pyle, Stephen G Chamberlin, Steven A Benner, Thomas J Lyons, Valérie de Crécy-Lagard, and Eudes de Crécy. Experimental evolution of a facultative thermophile from a mesophilic ancestor. *Applied and Environmental Microbiology*, 78(1):144–155, January 2012.

[15] Daniel Blankenberg, Assaf Gordon, Gregory Von Kuster, Nathan Coraor, James Taylor, Anton Nekrutenko, and Galaxy Team. Manipulation of FASTQ data with Galaxy. *Bioinformatics (Oxford, England)*, 26(14):1783–1785, July 2010.

[16] Zachary D Blount, Jeffrey E Barrick, Carla J Davidson, and Richard E Lenski. Genomic analysis of a key innovation in an experimental Escherichia coli population. *Nature*, 489(7417):513–518, September 2012.

[17] L E Brand. The salinity tolerance of forty-six marine phytoplankton isolates. *Estuarine, Coastal and Shelf Science*, 18(5):543–556, 1984.

[18] Michael A Brockhurst, Nick Colegrave, and Daniel E Rozen. Next-generation sequencing as a tool to study microbial evolution. *Molecular ecology*, 20(5):972–980, September 2010.

[19] Vernon Bryson and W Szybalski. Microbial Selection. *Science (New York, N.Y.)*, 116(3003):45–51, July 1952.

[20] J J Bull, M R Badgett, H A Wichman, J P Huelsenbeck, D M Hillis, A Gulati, C Ho, and I J Molineux. Exceptional convergent evolution in a virus. *Genetics*, 147(4):1497–1507, December 1997.

[21] Carlos Canchaya, Frank Desiere, W Michael McShan, Joseph J Ferretti, Julian Parkhill, and Harald Brüssow. Genome analysis of an inducible prophage and prophage remnants integrated in the Streptococcus pyogenes strain SF370. *Virology*, 302(2):245–258, October 2002.

[22] Carlos Canchaya, Ghislain Fournous, and Harald Brüssow. The impact of prophages on bacterial chromosomes. *Molecular Microbiology*, 53(1):9–18, June 2004.

[23] Carlos Canchaya, C Proux, G Fournous, A Bruttin, and Harald Brüssow. Prophage Genomics. *Microbiology and Molecular Biology Reviews*, 67(2):238–276, June 2003.

[24] Peter A Carr and George M Church. Genome engineering. *Nature biotechnology*, 27(12):1151–1162, December 2009.

[25] Sherwood R Casjens. Prophages and bacterial genomics: what have we learned so far? *Molecular Microbiology*, 49(2):277–300, June 2003.

[26] Sherwood R Casjens, Eddie B Gilcrease, Wai Mun Huang, Kim L Bunny, Marisa L Pedulla, Michael E Ford, Jennifer M Houtz, Graham F Hatfull, and Roger W Hendrix. The pKO2 linear plasmid prophage of Klebsiella oxytoca. *Journal of Bacteriology*, 186(6):1818–1832, March 2004.

[27] Hsin-Hung Chou, Julia Berthet, and Christopher J Marx. Fast growth increases the selective advantage of a mutation arising recurrently during evolution under metal limitation. *PLoS Genetics*, 5(9):e1000652, September 2009.

[28] Hsin-Hung Chou and Christopher J Marx. Optimization of gene expression through divergent mutational paths. *Cell reports*, 1(2):133–140, February 2012.

[29] Gail E Christie and Terje Dokland. Pirates of the Caudovirales. *Virology*, 432(2):277–282, November 2012.

[30] Tom M Conrad, Nathan E Lewis, and Bernhard O Palsson. Microbial laboratory evolution in the era of genome-scale science. *Mol Syst Biol*, 7:509, 2011.

[31] Tim F Cooper, Daniel E Rozen, and Richard E Lenski. Parallel changes in gene expression after 20,000 generations of evolution in Escherichia coli. *Proceedings of the National Academy of Sciences*, 100(3):1072–1077, February 2003.

[32] Otto X Cordero, Hans Wildschutte, Benjamin Kirkup, Sarah Proehl, Lynn Ngo, Fatima Hussain, Frédérique Le Roux, Tracy Mincer, and Martin F Polz. Ecological populations of bacteria act as socially cohesive units of antibiotic production and resistance. *Science (New York, N.Y.)*, 337(6099):1228–1231, September 2012.

[33] Rafael Couñago, Stephen Chen, and Yousif Shamoo. In vivo molecular evolution reveals biophysical origins of organismal fitness. *Molecular cell*, 22(4):441–449, 2006.

[34] Aurélie Daccord, Daniela Ceccarelli, Sébastien Rodrigue, and Vincent Burrus. Comparative analysis of mobilizable genomic islands. *Journal of Bacteriology*, 195(3):606–614, February 2013.

[35] J Arjan G M de Visser. The fate of microbial mutators. *Microbiology (Reading, England)*, 148(Pt 5):1247–1252, May 2002.

[36] J Arjan G M de Visser and Daniel E Rozen. Clonal interference and the periodic selection of new beneficial mutations in Escherichia coli. *Genetics*, 172(4):2093–2100, April 2006.

[37] P S Dehal, M P Joachimiak, M N Price, J T Bates, J K Baumohl, D Chivian, G D Friedland, K H Huang, K Keller, P S Novichkov, I L Dubchak, Eric J Alm, and Adam P Arkin. MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Research*, 38(Database):D396–D400, December 2009.

[38] E Denamur, G Lecointre, P Darlu, Olivier Tenaillon, C Acquaviva, C Sayada, I Sunjevaric, R Rothstein, J Elion, Francois Taddei, M Radman, and I Matic. Evolutionary implications of the frequent horizontal transfer of mismatch repair genes. *Cell,*, 103(5):711–721, November 2000.

[39] Erick Denamur, Stéphane Bonacorsi, Antoine Giraud, Patrick Duriez, Farida Hilali, Christine Amorin, Edouard Bingen, Antoine Andremont, Bertrand Picard, Francois Taddei, and Ivan Matic. High frequency of mutator strains among human uropathogenic Escherichia coli isolates. *Journal of Bacteriology*, 184(2):605–609, January 2002.

[40] Jeremy R Dettman, Nicolas Rodrigue, Anita H Melnyk, Alex Wong, Susan F Bailey, and Rees Kassen. Evolutionary insight from whole-genome sequencing of experimentally evolved microbes. *Molecular ecology*, 21(9):2058–2077, May 2012.

[41] J W Drake. Contrasting Mutation Rates from Specific-Locus and Long-Term Mutation-Accumulation Procedures. *G3: Genes/Genomes/Genetics*, 2(4):483–485, April 2012.

[42] D E Dykhuizen and D L Hartl. Selection in chemostats. *Microbiological Reviews*, 47(2):150, 1983.

[43] J A Eisen. A phylogenomic study of the MutS family of proteins. *Nucleic Acids Research*, 26(18):4291–4300, 1998.

[44] Santiago F Elena and Richard E Lenski. Microbial genetics: Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Genet*, 4(6):457–469, June 2003.

[45] Stephen S Fong, Andrew R Joyce, and Bernhard O Palsson. Parallel adaptive evolution cultures of Escherichia coli lead to convergent growth phenotypes with different gene expression states. *Genome research*, 15(10):1365–1372, October 2005.

[46] M P Forsyth and D J Kushner. Nutrition and distribution of salt response in populations of moderately halophilic bacteria. *Canadian journal of microbiology*, 16(4):253–261, 1970.

[47] P L Foster, A J Hanson, H Lee, E M Popodi, and H Tang. On the Mutational Topology of the Bacterial Genome. *G3 Genes/Genomes/Genetics*, 3(3):399–407, March 2013.

[48] M P Francino, L Chao, M A Riley, and H Ochman. Asymmetries generated by transcription-coupled repair in enterobacterial genes. *Science (New York, N.Y.)*, 272(5258):107–109, April 1996.

[49] Pauline Funchain, Annie Yeung, Jean Stewart, Wendy M Clendenin, and J H Miller. Amplification of mutator cells in a population as a result of horizontal transfer. *Journal of Bacteriology*, 183(12):3737–3741, June 2001.

[50] Christopher F Gentile, Szi-Chieh Yu, Sebastian Akle Serrano, Philip J Gerrish, and Paul D Sniegowski. Competition between high- and higher-mutating strains of Escherichia coli. *Biology letters*, 7(3):422–424, June 2011.

[51] A Giraud, M Radman, I Matic, and Francois Taddei. The rise and fall of mutator bacteria. *Current Opinion in Microbiology*, 4(5):582–585, October 2001.

[52] Antoine Giraud, Ivan Matic, Miroslav Radman, Michel Fons, and Francois Taddei. Mutator bacteria as a risk factor in treatment of infectious diseases. *Antimicrobial agents and chemotherapy*, 46(3):863–865, March 2002.

[53] Ingegerd Gustafsson, Maria Sjölund, Erik Torell, Marie Johannesson, Lars Engstrand, Otto Cars, and Dan I Andersson. Bacteria with increased mutation frequency and antibiotic resistance are enriched in the commensal flora of patients with high antibiotic usage. *The Journal of antimicrobial chemotherapy*, 52(4):645–650, October 2003.

[54] Simon R Harris, Edward J Feil, Matthew T G Holden, Michael A Quail, Emma K Nickerson, Narisara Chantratita, Susana Gardete, Ana Tavares, Nick Day, Jodi A Lindsay, Jonathan D Edgeworth, Hermínia de Lencastre, Julian Parkhill, Sharon J Peacock, and Stephen D Bentley. Evolution of MRSA during hospital transmission and intercontinental spread. *Science (New York, N.Y.)*, 327(5964):469–474, January 2010.

[55] Terry C Hazen, Eric A Dubinsky, Todd Z DeSantis, Gary L Andersen, Yvette M Piceno, Navjeet Singh, Janet K Jansson, Alexander Probst, Sharon E Borglin, Julian L Fortney, William T Stringfellow, Markus Bill, Mark E Conrad, Lauren M Tom, Krystle L Chavarria, Thana R Alusi, Regina Lamendella, Dominique C Joyner, Chelsea Spier, Jacob Baelum, Manfred Auer, Marcin L Zemla, Romy Chakraborty, Eric L Sonnenthal, Patrik D'haeseleer, Hoi-Ying N Holman, Shariff Osman, Zhenmei Lu, Joy D Van Nostrand, Ye Deng, Jizhong Zhou, and Olivia U Mason. Deep-sea oil plume enriches indigenous oil-degrading bacteria. *Science (New York, N.Y.)*, 330(6001):204–208, October 2010.

[56] Z He, A Zhou, E Baidoo, Q He, M P Joachimiak, P Benke, R Phan, A Mukhopadhyay, C L Hemme, K Huang, Eric J Alm, M W Fields, J Wall, D Stahl, T C Hazen, J D Keasling, Adam P Arkin, and J Zhou. Global Transcriptional, Physiological, and Metabolite Analyses of the Responses of Desulfovibrio vulgaris Hildenborough to Salt Adaptation. *Applied and Environmental Microbiology*, 76(5):1574–1586, February 2010.

[57] R W Hendrix. Bacteriophage genomics. *Current Opinion in Microbiology*, 2003.

[58] B Hochhut and Matthew K Waldor. Site-specific integration of the conjugal Vibrio cholerae SXT element into prfC. *Molecular Microbiology*, 32(1):99–110, April 1999.

[59] Dana E Hunt, Lawrence A David, Dirk Gevers, Sarah P Preheim, Eric J Alm, and Martin F Polz. Speciation by Resource Partitioning Among Closely-Related Bacterioplankton. *Science (New York, N.Y.)*, 320(5879):1081–1085, May 2008.

[60] H K Huse, T Kwon, J E A Zlosnik, David P Speert, E M Marcotte, and M Whiteley. Parallel Evolution in Pseudomonas aeruginosa over 39,000 Generations In Vivo. *mBio*, 1(4):e00199–10–e00199–17, August 2010.

[61] F Jabr. Engineered bugs make oil with help from the sun. *New Scientist*, 2011.

[62] N Jezequel, M C Lagomarsino, F Heslot, and P Thomen. Long-Term Diversity and Genome Adaptation of Acinetobacter baylyi in a Minimal-Medium Chemostat. *Genome Biology and Evolution*, 5(1):87–97, January 2013.

[63] A Jolivet-Gougeon, B Kovacs, S Le Gall-David, H Le Bars, L Bousarghin, M Bonnaure-Mallet, B Lobel, F Guille, C J Soussy, and P Tenke. Bacterial hypermutation: clinical implications. *Journal of Medical Microbiology*, 60(5):563–573, April 2011.

[64] T Kaneko and R R Colwell. Ecology of Vibrio parahaemolyticus in Chesapeake Bay. *Journal of Bacteriology*, 113(1):24–32, January 1973.

[65] C W Kaspar and M L Tamplin. Effects of temperature and salinity on the survival of Vibrio vulnificus in seawater and shellfish. *Applied and Environmental Microbiology*, 59(8):2425–2429, 1993.

[66] Rees Kassen, Martin Llewellyn, and Paul B Rainey. Ecological constraints on diversification in a model adaptive radiation. *Nature*, 431(7011):984–988, October 2004.

[67] Tadeusz J Kawecki, Richard E Lenski, Dieter Ebert, Brian Hollis, Isabelle Olivieri, and Michael C Whitlock. Experimental evolution. *Trends in Ecology & Evolution*, 27(10):547–560, October 2012.

[68] J Kelly. The Organism Is the Product. *ACS Synthetic Biology*, 2012.

[69] Daniel P Keymer, Michael C Miller, Gary K Schoolnik, and Alexandria B Boehm. Genomic and phenotypic diversity of coastal Vibrio cholerae strains is linked to environmental factors. *Applied and Environmental Microbiology*, 73(11):3705–3714, June 2007.

[70] Frédérique Le Roux, Mohamed Zouine, Nesrine Chakroun, Johan Binesse, Denis Saulnier, Christiane Bouchier, Nora Zidane, Laurence Ma, Christophe Rusniok, Aurélie Lajus, Carmen Buchrieser, Claudine Médigue, Martin F Polz, and Didier Mazel. Genome sequence of Vibrio splendidus: an abundant planctonic marine species with a large genotypic diversity. *Environmental Microbiology*, 11(8):1959–1970, August 2009.

[71] J E LeClerc, B Li, W L Payne, and T A Cebula. High mutation frequencies among Escherichia coli and Salmonella pathogens. *Science (New York, N.Y.)*, 1996.

[72] Heewook Lee, Ellen Popodi, Haixu Tang, and Patricia L Foster. Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing. *PNAS*, September 2012.

[73] Ming-Chun Lee, Hsin-Hung Chou, and Christopher J Marx. Asymmetric, bimodal trade-offs during adaptation of Methylobacterium to distinct growth substrates. *Evolution; international journal of organic evolution*, 63(11):2816–2830, November 2009.

[74] Ming-Chun Lee and Christopher J Marx. Repeated, selection-driven genome reduction of accessory genes in experimental populations. *PLoS Genetics*, 8(5):e1002651, 2012.

[75] Richard E Lenski, Michael R Rose, Suzanne C Simpson, and Scott C Tadler. Long-term experimental evolution in Escherichia coli. I. Adaptation and divergence during 2,000 generations. *American Naturalist*, pages 1315–1341, 1991.

[76] R Leplae, G Lima-Mendez, and A Toussaint. ACLAME: A CLAssification of Mobile genetic Elements, update 2010. *Nucleic Acids Research*, 38(Database):D57–D61, December 2009.

[77] Shengkang Li, Xiang Xiao, Jinyuan Li, Jinxian Luo, and Fengping Wang. Identification of genes regulated by changing salinity in the deep-sea bacterium Shewanella sp. WP3 using RNA arbitrarily primed PCR. *Extremophiles*, 10(2):97–104, August 2005.

[78] E F Mao, L Lane, J Lee, and J H Miller. Proliferation of mutators in a cell population. *Journal of Bacteriology*, 1997.

[79] Christopher J Marx. Evolution as an experimental tool in microbiology: 'Bacterium, improve thyself!'. *Environmental Microbiology Reports*, 3:12–14, April 2011.

[80] Arne C Materna, Jonathan Friedman, Claudia Bauer, Christina David, Sara Chen, Ivy B Huang, April Gillens, Sean A Clarke, Martin F Polz, and Eric J Alm. Shape and evolution of the fundamental niche in marine Vibrio. *The ISME journal*, July 2012.

[81] I Matic, M Radman, Francois Taddei, B Picard, C Doit, E Bingen, E Denamur, and J Elion. Highly variable mutation rates in commensal and pathogenic Escherichia coli. *Science (New York, N.Y.)*, 277(5333):1833–1834, September 1997.

[82] A Mena, E E Smith, J L Burns, David P Speert, S M Moskowitz, J L Perez, and A Oliver. Genetic adaptation of Pseudomonas aeruginosa to the airways of cystic fibrosis patients is catalyzed by hypermutation. *Journal of Bacteriology*, 190(24):7910–7917, December 2008.

[83] J H Miller. Mutators in Escherichia coli. *Mutation research*, 409(3):99–106, December 1998.

[84] A Mukhopadhyay, Z He, Eric J Alm, Adam P Arkin, E E Baidoo, S C Borglin, W Chen, T C Hazen, Q He, H Y Holman, K Huang, R Huang, D C Joyner, N Katz, M Keller, P Oeller, A Redding, J Sun, J Wall, J Wei, Z Yang, H C Yen, J Zhou, and J D Keasling. Salt Stress in Desulfovibrio vulgaris Hildenborough: an Integrated Genomics Approach. *Journal of Bacteriology*, 188(11):4068–4078, May 2006.

[85] L M Naughton, S L Blumerman, M Carlberg, and E Fidelma Boyd. Osmoadaptation among Vibrio Species and Unique Genomic Features and Physiological Responses of Vibrio parahaemolyticus. *Applied and Environmental Microbiology*, 75(9):2802–2810, April 2009.

[86] Lucinda Notley-McRobb and Thomas Ferenci. Adaptive mgl-regulatory mutations and genetic diversity evolving in glucose-limited Escherichia coli populations. *Environmental Microbiology*, 1(1):33–43, 02 1999.

[87] Lucinda Notley-McRobb, Shona Seeto, and Thomas Ferenci. Enrichment and elimination of mutY mutators in Escherichia coli populations. *Genetics*, 162(3):1055–1062, November 2002.

[88] Aaron Novick and Leo Szilard. Experiments with the chemostat on spontaneous mutations of bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 36(12):708, 1950.

[89] A Oliver and A Mena. Bacterial hypermutation in cystic fibrosis, not only for antibiotic resistance. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 16(7):798–808, July 2010.

[90] E Z Panagou, P N Skandamis, and G-J E Nychas. Use of gradient plates to study combined effects of temperature, pH, and NaCl concentration on growth of Monascus ruber van Tieghem, an Ascomycetes fungus isolated from green table olives. *Applied and Environmental Microbiology*, 71(1):392–399, January 2005.

[91] Su-Chan Park and Joachim Krug. Clonal interference in large populations. *Proceedings of the National Academy of Sciences*, 104(46):18135–18140, November 2007.

[92] Julian Parkhill. Time to remove the model organism blinkers. *Trends in microbiology*, 16(11):510–511, November 2008.

[93] A L Prunier and R Leclercq. Role of mutS and mutL Genes in Hypermutability and Recombination in Staphylococcus aureus. *Journal of Bacteriology*, 187(10):3455–3464, May 2005.

[94] Lev Rabinovich, Nadejda Sigal, Ilya Borovok, Ran Nir-Paz, and Anat A Herskovits. Prophage excision activates Listeria competence genes that promote phagosomal escape and virulence. *Cell,*, 150(4):792–802, August 2012.

[95] P B Rainey and Michael Travisano. Adaptive radiation in a heterogeneous environment. *Nature*, 394(6688):69–72, July 1998.

[96] Y Ram and L Hadany. The Evolution of Stress-Induced Hypermutation in Asexual Populations. *Evolution; international journal of organic evolution*, 66(7):2315–2328, 2012.

[97] Mark A Randa, Martin F Polz, and Eelin Lim. Effects of Temperature and Salinity on Vibrio vulnificus Population Dynamics as Assessed by Quantitative PCR. *Applied and Environmental Microbiology*, 70(9):5469–5476, September 2004.

[98] Douglas B Rusch, Aaron L Halpern, Granger Sutton, Karla B Heidelberg, Shannon Williamson, Shibu Yooseph, Dongying Wu, Jonathan A Eisen, Jeff M Hoffman, Karin Remington, Karen Beeson, Bao Tran, Hamilton Smith, Holly Baden-Tillson, Clare Stewart, Joyce Thorpe, Jason Freeman, Cynthia Andrews-Pfannkoch, Joseph E Venter, Kelvin Li, Saul Kravitz, John F Heidelberg, Teresa R Utterback, Yu-Hui Rogers, Luisa I Falcón, Valeria Souza, Germán Bonilla-Rosso, Luis E Eguiarte, David M Karl, Shubha Sathyendranath, Trevor Platt, Eldredge Bermingham, Victor Gallardo, Giselle Tamayo-Castillo, Michael R Ferrari, Robert L Strausberg, Kenneth Nealson, Robert Friedman, Marvin Frazier, and J Craig Venter. The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biology*, 5(3):e77 OP, March 2007.

[99] Gerda Saxer, Michael Doebeli, and Michael Travisano. The Repeatability of Adaptive Radiation During Long-Term Experimental Evolution of Escherichia coli in a Multiple Nutrient Environment. *PloS one*, 5(12):e14184, December 2010.

[100] R M Schaaper and R L Dunn. Spectra of spontaneous mutations in Escherichia coli strains defective in mismatch correction: the nature of in vivo DNA replication errors.

*Proceedings of the National Academy of Sciences of the United States of America*, 84(17):6220–6224, September 1987.

[101] Julie Scott, Scott V Nguyen, Catherine J King, Christina Hendrickson, and W Michael McShan. Phage-Like Streptococcus pyogenes Chromosomal Islands (SpyCI) and Mutator Phenotypes: Control by Growth State and Rescue by a SpyCI-Encoded Promoter. *Frontiers in Microbiology*, 3, 2012.

[102] Julie Scott, Prestina Thompson-Mayberry, Stephanie Lahmamsi, Catherine J King, and W Michael McShan. Phage-associated mutator phenotype in group A streptococcus. *Journal of Bacteriology*, 190(19):6290–6301, October 2008.

[103] Victor Seguritan, Nelson Alves, Michael Arnoult, Amy Raymond, Don Lorimer, Alex B Burgin, Peter Salamon, and Anca M Segall. Artificial neural networks trained to detect viral and phage structural proteins. *PLoS Computational Biology*, 8(8):e1002657, 2012.

[104] B Jesse Shapiro, Jonathan Friedman, Otto X Cordero, Sarah P Preheim, Sonia C Timberlake, Gitta Szabo, Martin F Polz, and Eric J Alm. Population genomics of early events in the ecological differentiation of bacteria. *Science (New York, N.Y.)*, 336(6077):48–51, April 2012.

[105] Aaron C Shaver, Peter G Dombrowski, Joseph Y Sweeney, Tania Treis, Renata M Zappala, and Paul D Sniegowski. Fitness evolution and the rise of mutator alleles in experimental Escherichia coli populations. *Genetics*, 162(2):557–566, October 2002.

[106] Aaron C Shaver and Paul D Sniegowski. Spontaneously arising mutL mutators in evolving Escherichia coli populations are the result of changes in repeat length. *Journal of Bacteriology*, 185(20):6076–6082, October 2003.

[107] E C Siegel and Vernon Bryson. Mutator gene of Escherichia coli B. *Journal of Bacteriology*, 94(1):38–47, July 1967.

[108] F L Singleton, R Attwell, S Jangi, and R R Colwell. Effects of temperature and salinity on Vibrio cholerae growth. *Applied and Environmental Microbiology*, 44(5):1047–1058, November 1982.

[109] E W Six and CAC Klug. Bacteriophage P4: a satellite virus depending on a helper such as prophage P2. *Virology*, 1973.

[110] W Soto, J Gutierrez, M D Remmenga, and M K Nishiguchi. Salinity and Temperature Effects on Physiological Responses of Vibrio fischeri from Diverse Ecological Niches. *Microbial Ecology*, 57(1):140–150, June 2008.

[111] Christine C Spencer, Jabus Tyerman, Melanie Bertrand, and Michael Doebeli. Adaptation increases the likelihood of diversification in an experimental bacterial lineage. *Proceedings of the National Academy of Sciences*, 105(5):1585–1589, 2008.

[112] Gregory Stephanopoulos, Hal Alper, and Joel Moxley. Exploiting biological complexity for strain improvement through systems biology. *Nat Biotech*, 22(10):1261–1267, October 2004.

[113] A Stern and R Sorek. The phage-host arms race: Shaping the evolution of microbes. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 2011.

[114] Gitta Szabo, Sarah P Preheim, Kathryn M Kauffman, Lawrence A David, Jesse Shapiro, Eric J Alm, and Martin F Polz. Reproducibility of Vibrionaceae population structure in coastal bacterioplankton. *The ISME journal*, 7(3):509–519, March 2013.

[115] W Szybalski and Vernon Bryson. Genetic studies on microbial cross resistance to toxic agents. I. Cross resistance of Escherichia coli to fifteen antibiotics. *Journal of Bacteriology*, 64(4):489–499, October 1952.

[116] Mark M Tanaka, Carl T Bergstrom, and Bruce R Levin. The evolution of mutator genes in bacterial populations: the roles of environmental change and timing. *Genetics*, 164(3):843–854, July 2003.

[117] G M Tantillo, M Fontanarosa, A Di Pinto, and M Musti. Updated perspectives on emerging vibrios associated with human infections. *Letters in applied microbiology*, 39(2):117–126, 2004.

[118] Olivier Tenaillon, A Rodriguez-Verdugo, R L Gaut, P McDonald, Albert F Bennett, A D Long, and B S Gaut. The Molecular Diversity of Adaptive Convergence. *Science (New York, N.Y.)*, 335(6067):457–461, January 2012.

[119] Jr Theodore Garland and Michael R Rose. *Experimental Evolution.* Concepts, Methods, and Applications of Selection Experiments. Univ of California Press, December 2009.

[120] Linda V Thomas and J W Wimpenny. Investigation of the effect of combined variations in temperature, pH, and NaCl concentration on nisin inhibition of Listeria monocytogenes and Staphylococcus aureus. *Applied and Environmental Microbiology*, 62(6):2006–2012, 1996.

[121] Fabiano L Thompson, Tetsuya Iida, and Jean Swings. Biodiversity of Vibrios. *Microbiology and Molecular Biology Reviews*, 68(3):403–431, September 2004.

[122] Janelle R Thompson, Sarah Pacocha, Chanathip Pharino, Vanja Klepac-Ceraj, Dana E Hunt, Jennifer Benoit, Ramahi Sarma-Rupavtarm, Daniel L Distel, and Martin F Polz. Genotypic diversity within a natural coastal bacterioplankton population. *Science (New York, N.Y.)*, 307(5713):1311–1313, February 2005.

[123] Janelle R Thompson, Mark A Randa, Luisa A Marcelino, Aoy Tomita-Mitchell, Eelin Lim, and Martin F Polz. Diversity and dynamics of a north atlantic coastal Vibrio community. *Applied and Environmental Microbiology*, 70(7):4103–4110, July 2004.

[124] Michael Travisano and Richard E Lenski. Long-term experimental evolution in Escherichia coli. IV. Targets of selection and the specificity of adaptation. *Genetics*, 143(1):15–26, 1996.

[125] D S Treves, S Manning, and J Adams. Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of Escherichia coli. *Molecular Biology and Evolution*, 15:789–797, April 1998.

[126] Jabus Tyerman, N Havard, G Saxer, Michael Travisano, and M Doebeli. Unparallel diversification in bacterial microcosms. *Proceedings of the Royal Society B: Biological Sciences*, 272(1570):1393–1398, July 2005.

[127] Tim van Opijnen, Kip L Bodi, and Andrew Camilli. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature methods*, 6(10):767–772, September 2009.

[128] G J Velicer. Comprehensive mutation identification in an evolved bacterial cooperator and its cheating ancestor. *Proceedings of the National Academy of Sciences of the United States of America*, 103(21):8107–8112, May 2006.

[129] A Ventosa, J J Nieto, and A Oren. Biology of moderately halophilic aerobic bacteria. *Microbiology and Molecular Biology Reviews*, 62(2):504–544, 1998.

[130] A Ventosa, F Rodriguez-Valera, J S Poindexter, and W S Reznikoff. Selection for moderately halophilic bacteria by gradual salinity increases. *Canadian journal of microbiology*, 30(10):1279–1282, 1984.

[131] L Wang, B Spira, Z Zhou, and L Feng. Divergence involving global regulatory gene mutations in an Escherichia coli population evolving under phosphate limitation. *Genome biology and ...*, 2010.

[132] Xiaoxue Wang, Younghoon Kim, Qun Ma, Seok Hoon Hong, Karina Pokusaeva, Joseph M Sturino, and Thomas K Wood. Cryptic prophages help bacteria cope with adverse environments. *Nature Communications*, 1(9):147–9, December 2010.

[133] Daniel M Weinreich, Nigel F Delaney, Mark A Depristo, and Daniel L Hartl. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science (New York, N.Y.)*, 312(5770):111–114, April 2006.

[134] W B Whitaker, M A Parent, L M Naughton, G P Richards, S L Blumerman, and E Fidelma Boyd. Modulation of Responses of Vibrio parahaemolyticus O3:K6 to pH and Temperature Stresses by Growth at Different Salt Concentrations . *Applied and Environmental Microbiology*, 76(14):4720–4729, July 2010.

[135] H A Wichman, J Millstein, and J J Bull. Adaptive Molecular Evolution for 13,000 Phage Generations A Possible Arms Race. *Genetics*, 2005.

[136] Rachel A F Wozniak and Matthew K Waldor. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nature Reviews Microbiology*, 8(8):552–563, July 2010.

[137] A C Wright, R T Hill, J A Johnson, M C Roghman, R R Colwell, and J G Morris. Distribution of Vibrio vulnificus in the Chesapeake Bay. *Applied and Environmental Microbiology*, 62(2):717–724, 1996.

[138] K C Wrighton, B C Thomas, I Sharon, C S Miller, C J Castelle, N C VerBerkmoes, M J Wilkins, R L Hettich, M S Lipton, K H Williams, P E Long, and J F Banfield. Fermentation, Hydrogen, and Sulfur Metabolism in Multiple Uncultivated Bacterial Phyla. *Science (New York, N.Y.)*, 337(6102):1661–1665, September 2012.

[139] Avihu H Yona, Yair S Manor, Rebecca H Herbst, Gal H Romano, Amir Mitchell, Martin Kupiec, Yitzhak Pilpel, and Orna Dahan. Chromosomal duplication is a transient evolutionary solution to stress. *Proceedings of the National Academy of Sciences of the United States of America*, 109(51):21010–21015, December 2012.