The role of ecological specialization in diversification of microbial populations

by

Alana Schick

B.Sc., The University of British Columbia, 2007
M.Sc., The University of British Columbia, 2013

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Biology)

THE UNIVERSITY OF OTTAWA

(Ontario, Canada)

December 2019

Abstract

The emergence of diversity in populations is a complex process that remains challenging to predict in the natural world despite a solid theoretical understanding of the mechanisms that generate diversity. In this work, I explore several fundamental questions about the dynamics of diversification in microbial populations. I focus on the role of ecological specialization in this process by examining the variation that arises de novo in experimental populations adapting to novel, complex environments. First, using previously adapted populations *Pseudomonas fluorescens*, I investigate specialization by directly measuring fitness effects of adaptive mutations in multiple environments. I find that adaptive mutations commonly have deleterious effects in other environments, generating fitness trade-offs associated with ecological specialization. Then, using initially homogenous populations *Pseudomonas aeruginosa*, I explore the emergence of phenotypic variation in a set of environments of varying complexity asking whether environments with more ecological opportunity result in more diverse populations. Further, as *P. aeruginosa* is an important human pathogen, commonly infecting the lungs of patients with cystic fibrosis (CF), the environments tested also vary in similarity to the ecological conditions of the CF lung. I find that populations rapidly and repeatedly become phenotypically diverse, with many of the same phenotypes that are known to arise in CF patients. I show that adding ecological complexity increases diversity within populations and increases divergence among populations, demonstrating that ecological complexity provides a larger breadth of evolutionary trajectories and may play an important role in pathogen adaptation. Finally, I investigate the genetic variation that underlies these evolutionary trajectories using whole genome sequencing of evolved populations. I find large amounts of genetic variation present within populations and varying amounts of genetic parallelism among populations, depending on the environment. Taken together, this thesis advances our knowledge of the generation of diversity and how the ecological conditions of an environment can influence that process.

Abstrait en français

L'émergence de la diversité des populations est un processus complexe qui reste difficile à prédire dans le monde naturel malgré une solide compréhension théorique des mécanismes qui génèrent la diversité. Dans ce travail, j'explore plusieurs questions fondamentales sur la dynamique de la diversification des populations microbiennes. Je me concentre sur le rôle de la spécialisation écologique dans ce processus en examinant la variation qui apparaît *de novo* chez des populations expérimentales s'adaptant à des environnements nouveaux et complexes. Tout d'abord, en utilisant des populations de *Pseudomonas fluorescens* précédemment adaptées, j'étudie la spécialisation en mesurant directement les effets des mutations adaptatives sur la condition physique dans plusieurs environnements. Je trouve que les mutations adaptatives ont généralement des effets negatif dans d'autres environnements, générant des compromis de remise en forme liés à la spécialisation écologique. Puis, en utilisant des populations *Pseudomonas aeruginosa* initialement homogènes, j'explore l'émergence de variations phénotypiques dans un ensemble d'environnements de complexité variable en demandant si des environnements offrant davantage d'opportunités écologiques entraînent des populations plus diversifiées. De plus, comme *P. aeruginosa* est un agent pathogène humain important, infectant généralement les poumons des patients atteints de fibrose kystique (FK), les environnements testés présentent également des similitudes avec les conditions écologiques du poumon FK. Je trouve que les populations présentent une diversité phénotypique rapide et répétitive, avec bon nombre des mêmes phénotypes que ceux connus chez les patients atteints de FK. Je montre que l'ajout de la complexité écologique accroît la diversité au sein des populations et la divergence entre les populations, ce qui montre que la complexité écologique élargit l'éventail des trajectoires évolutives et peut jouer un rôle important dans l'adaptation des agents pathogènes. Enfin, j'étudie la variation génétique qui sous-tend ces trajectoires évolutives en utilisant le séquençage du génome entier de populations évoluées. Je trouve de grandes quantités de variations génétiques présentes au sein des populations et des quantités variables de parallélisme génétique parmi les populations, en fonction de l'environnement. Pris ensemble, cette thèse fait progresser notre connaissance de la génération de la diversité et de la façon dont les conditions écologiques d'un environnement peuvent influencer ce processus.

Lay Summary

How does a population of initially identical individuals become diverse? Ultimately, the source of variation in any population is genetic mutation, but if a mutation arises that increases fitness, what prevents that mutation from spreading over the entire population and resulting in a changed, but identical, population? In environments that vary in terms of selection pressures, there may be more than one way to increase fitness in that environment. Here, I investigate how the process of diversification is influenced by the ability of organisms to specialize to their particular environment. I find that an increase in fitness in one environment often causes a decrease in fitness in other environments, generating a trade-off associated with specialization. I then explore the effect of two factors on diversification: the range of resources, or niches, in the environment and the spatial structure of the environment. I find that both factors increase diversification in different ways, highlighting the need to focus on all ecological aspects of an environment when trying to predict evolutionary paths of diversification.

Preface

A version of Chapter 2 has been published as "Schick, A., Bailey, S., & Kassen, R. (2015). Evolution of fitness trade-offs in locally adapted populations of *Pseudomonas fluorescens*. American Naturalist, 186(S1), S48-S49". Myself and R. Kassen conceived the original project. I performed laboratory work using strains evolved by S. Bailey and with mutation information from sequence data analyzed by S. Bailey. I wrote the initial draft of the manuscript and R. Kassen and I contributed major revisions, with minor feedback from S. Bailey.

A version of Chapter 3 has been published as "Schick, A. & Kassen, R. (2018). Rapid diversification of *Pseudomonas aeruginosa* in cystic fibrosis lung-like conditions. Proceedings of the National Academy of Scienes, 115(42), 10714-10719". I conceived the project with advice with R. Kassen. I performed the laboratory work and performed the analyses with advice from R. Kassen. I prepared the initial draft, R. Kassen and I both contributed major revisions to the manuscript.

A version of Chapter 4 has been submitted for puplication as "Schick, A. & Kassen, R. (2019). Genomics of experimental diversification of *Pseudomonas aeruginosa* in cystic fibrosis lung-like conditions". I conceived the project with advice from R. Kassen. I performed the laboratory work and all bioinformatic and statistical analysis, with advice from R. Kassen. I prepared the initial draft of the manuscript, and R. Kassen contributed revisions to the manuscript.

Table of Contents

List of Tables

List of Figures

Acknowledgements

This work would not have been possible without the tremendous support and assistance I received from so many people. First and foremost, my advisor, Rees Kassen. Thank you for being a human first and a scientist second. You are the only PhD advisor I've ever had and hopefully the last, but the best I could imagine. Your balance of guidance and encouragement allowed me to find the confidence to succeed. It's been a pleasure to have you as an advisor, mentor, and advocate.

Thank you to my academic friend/mentors Milica Mandic and Aleeza Gerstein for your advice and support over the years, and for being badass role models and wonderful friends. Thank you to Alex Wong for your mentorship, feedback on my work, and help and advice for working with the SCFM (and for teaching a course at Carleton that I could skate-commute to). Thank you to the faculty of the University of Ottawa's Department of Biology, with special thanks to the members of my thesis advisory committee, Julie Morand-Ferron, Alex Poulain, and Shaun Aaron, for your feedback and discussions, especially during the planning portion of my research.

Thank you to Aaron Hinz for all your help in the wet-lab, you were an invaluable resource for setting up all of my experiments, a strong source of moral support in the lab, and a dependable attendee of post-work beers. Thank you to Susan Bailey for helping me get started in the lab (by giving me your evolved strains) and for being on the same conference schedule as me. Thank you to Jasmine Ono for your feedback on my proposal, help with thesis formatting, and visits to Ottawa. Thank you to Kassen lab members, past and present, for a fantastic lab environment, especially Jeremy Dettman, Anita Melnyk, Susanne Kraemer, Nicolas Rode, Gabriel Perron, Patrick Chen, Alanna Leale, Eleonore Lebeuf-Taylor, Nick McCloskey, Jonathan Dench, Felipe Dargent, and Sonal Shewaramani.

Thank you as well to The University of Ottawa and the Ontario Provincial government for providing funding for my research as well as funding to travel to international conferences.

This work was at least equally as challenging personally as it was scientifically. I would like to thank all the people who supported me throughout this process on a personal level, especially my dad, Peter Schick, and my amazing brothers Ryan Schick, Dan Schick, and Kevin Schick. Thank you for the comic relief, the casual (and not-casual) pints, for supporting my questionable life choices (i.e. lending me money) and for being generally amazing. Thank you to Alanna Goobie and Chantal Mansuy for your unwavering confidence in me and for giving me some of my favourite Ottawa memories. Thank you to Sam Yeaman for your many years of academic, emotional, and recreational support. Thank you to Lorraine Whale and Neil Mansuy, the two 'adults' I turned to for advice when I was deciding whether or not to do a PhD. Thank you for encouraging me to do it. Lastly, thank you to my skiing and frisbee family – I would not have survived without you, especially Liz Westburg, Grant Skully, Greg DeBoer, Jonathan Griffiths, and Tyler Hochhalter.

Dedication

This thesis is dedicated to my mom, a biologist, who died just days before the letter of acceptance into grad school arrived. And who I once told "I will never be a biologist, it's too boring". I was wrong Mom, and this is for you.

Chapter 1

Introduction

Evolution by natural selection occurs when individuals with certain phenotypes are more likely to reproduce. If natural selection operates to increase the frequency of the highest fitness individuals, we might expect this process to end with a single optimal type. This is clearly not the case in nature. Biological diversity is abundant across all branches of life, sometimes breathtakingly so. Indeed, even Charles Darwin used the qualifier "endless" to describe the "forms most beautiful and most wonderful" (DARWIN 1859). Where do these forms come from and what makes their formation "endless"? Decades of evolutionary theory and empirical examination have provided a solid framework for the mechanisms that generate diversity and for predicting conditions under which that diversity should be maintained. In the natural world, however, diversification, a subprocess of speciation, remains difficult to predict. One reason for this is that the conditions in any environment are never constant. As organisms interact with their environment by consuming resources, for example, changing conditions are inevitable. And changing conditions in turn impose changing selection pressures that result in changing interactions. The addition of complex ecological interactions between organisms and their environment creates a feedback loop where change is the only constant. In this way, evolution is circular, or endless.

One prominent feature of life in the natural world is that organisms are often well-suited to the conditions of the environment in which they are found. They are often specialized for a particular set of ecological challenges or opportunities. On a small enough scale, specializing to a specific environment could be viewed as a decrease in diversity. On a large scale that contains many sub-environments, however, specializing in each of those sub-environments could result in an increased amount of diversity. In this thesis, I investigate the role of ecological specialization in the process of diversification. In this introductory chapter, I first describe the experimental backbone of this research: 'evolve and resequence' experiments. Then, I provide an overview of the dynamic mechanisms that are known to generate diversity within populations, considering the underlying genetic changes that produce the phenotypes for selection to act on. From this, I will explore the complicated mapping of genotype to phenotype and discuss the role of this mapping in the repeatability of diversification. I will describe how the answers to these fundamental questions in evolutionary biology can be applied directly to the field of medicine, specifically to understanding adaptation in an important human pathogen. Lastly, I will describe what each chapter contributes to this thesis and how they connect to inform the dynamics of diversification in heterogeneous environments.

1.1 Evolve and resequence experiments

The research presented in this thesis utilizes an experimental technique referred to as 'evolve and resequence', or ER. As the term suggests, this technique is comprised of two parts: allowing an organism or population to evolve and sequencing the evolved organism or population to determine what changed. This approach combines experimental evolution and next-generation sequencing to 'replay the tape of life' (GOULD 1989) many times over and play 'spot the difference' with all the tapes. This technique is often utilized in microbial systems due to short generation times and large population sizes, but has been used in recent years for numerous studies in Drosophila (reviewed in SCHLÖTTERER *et al.* 2015). Evolution in natural systems is very complex, but these ER experiments can be used to control many of the often confounding variables in natural systems. By modifying a single component of an organism's environment, the effect of that component on selection can be isolated. This is the experimental backbone of my research; I modify environmental variables to investigate key questions about the role of those variables in adaptation and diversification in novel environments.

1.1.1 Experimental evolution

Experimental evolution puts the 'E' in 'ER'. In general, populations are allowed to evolve in a given environment or set of environments and observed. In microbial populations, this is accomplished typically by diluting a saturated culture into a fresh batch of growth media at regular intervals. Experimental evolution allows us to reduce the complexity of adaptation to a manageable number of variables. By controlling environmental, historical, and developmental variables, we can repeat the process of adaptation and diversification to isolate effects of single variables at a time. This technique is not new; the infamous long term evolution experiment (LTEE) in the lab of Richard Lenski has been ongoing since 1988. In this case, twelve populations of *Escherichia coli* have been propagated by a daily 100-fold dilution, corresponding to a rate of 6.64 generations per day, putting the populations at approximately 73,000 generations old in 2018. For reference, this corresponds to over 2 million years of human evolution. The development of this experimental technique gave researchers the unprecedented ability to observe evolution and adaptation in real time and catalog the results, opening up a whole new field of research questions (reviewed in KASSEN 2014; KAWECKI *et al.* 2012; LENSKI 2017; LONG *et al.* 2015).

Depending on the research question of interest, evolution experiments can begin with either genetically homogenous or polymorphic populations. Because the questions addressed in this thesis pertain to the accumulation of beneficial mutations and the generation of genetic diversity, the evolution experiments here all begin with genetically invariable populations. This allows us to attribute all genetic diversity to mutations that arose *de novo* during the course of the experiment. Combined with technology from the current era of next-generation sequencing technologies, discovering the identity and properties of those mutations from many replicate populations is possible. Our capabilities to sequence many whole populations dramatically increases the value of the simple evolution experiment. Evidence of this is clear when considering that the LTEE is still producing novel and high impact insights. For example, deep sequencing of the LTEE populations every 500 generations over 60,000 generations recently demonstrated the complexity of long-term

adaptation in a seemingly simple environment (GOOD *et al.* 2017). This molecular catalogue of evolution revealed three major insights: rates of molecular evolution remain high as adaptation proceeds, long-term coexistence is common, and targets of natural selection shift over time.

1.1.2 Model organism

Model organisms are invaluable resources for studying evolution, especially when combined with the technological advances in sequencing. There are many benefits of using a commonly studied model organism, largely from the rich body of previous research findings, often including a detailed understanding of many phenotypes, many well-characterized genes, and a well-curated openly-available reference genome. Bacteria in the genus Pseudomonas have been commonly used for experimental evolution studies for a number of reasons. Pseudomonads are easy to culture, reproduce asexually, have short generation times, and can be preserved in a frozen state. Over deep evolutionary time, they have become a particularly diverse genus of bacteria, with some species sharing only 50-60% of their genes (KOEHORST *et al.* 2016). This diversity has been largely mediated by horizontal gene transfer, allowing strains to explore novel nutritional niches and acquire new metabolic capabilities (MATHEE *et al.* 2008). As a result, Pseudomonads are ubiquitous in natural environments ranging from water to soil and are able to adapt readily to novel environments due to their intrinsic metabolic flexibility (YANG *et al.* 2011). Compared to other bacterial species, the genomes of Pseudomonads contain a higher number of transcriptional regulators and two-component regulatory systems, providing them with higher metabolic versatility and physiologic adaptability (KLOCKGETHER *et al.* 2011).

In the experiments presented here, I use two species of Pseudomonads: *P. fluorescens* and *P. aeruginosa*. Both species are facultative anaerobes, preferring to grow aerobically when oxygen is present but able to grow anaerobically, using nitrate or nitrite instead of oxygen as a final electron acceptor in cellular respiration. *Pseudomonas aeruginosa* is of particular interest due to its role as a human pathogen, causing disease in immunocompromised individuals, including infection of burn wounds, urinary tract infections, and infections in the lungs of people with cystic fibrosis (CF). The importance of this microbe and prevalence of use in genetic research has led to a well-annotated genome reference database, containing 335 complete genomes (WINSOR *et al.* 2015), providing even further utility as a model organism.

1.2 Mechanisms that generate diversity

In recent years, the combination of experimental evolution and next generation sequencing has dramatically illuminated the prevalence of within-species diversity in microbial populations. How this diversity is generated and maintained remains an unanswered problem in biology. My research addresses this fundamental question of how diversity arises by simply allowing populations initially lacking diversity to evolve under a range of conditions. Historically, evolution in microbial populations with large population sizes was thought to proceed via a mechanism called 'periodic selection', a mode in which a single adaptive mutation arises and sweeps throughout the population to reach fixation before the next mutation arises (ATWOOD *et al.* 1951). Under this regime, populations would be mostly genetically uniform, aside from the time during which the current beneficial mutation is rising to fixation. However, in the experiments presented in this the-

sis, bacterial populations harbour larger amounts of diversity, which arises over a relatively short amount of time. Moreover, this phenomenon is becoming an increasingly common observation in microbial selection experiments in both bacteria (DE VISSER and ROZEN 2006; DESAI *et al.* 2007; ELENA and LENSKI 2003; FRIESEN *et al.* 2004; RAINEY and TRAVISANO 1998) and yeast (KAO and SHERLOCK 2008; LANG *et al.* 2013). In this section, I first discuss the evolutionary dynamics that are specific to microbial populations and then summarize the extensive body of theoretical models for predicting how this diversity arises and is maintained.

1.2.1 Dynamics of evolution in microbial populations

Adaptive evolution in populations typically depends on some combination of the evolutionary forces of selection, genetic drift, mutation, and recombination. Unlike populations of higher taxa in the wild, the dynamics of asexually reproducing organisms in very large populations in simple lab environments are much less complex. One simplification is that there is no genetic recombination as a result of meiotic recombination in clonal microbes. A second simplification is that the presence of mutations due to the stochasticity of genetic drift can largely be ignored. To see why this must be the case, consider that in selection experiments with daily 1:100 dilution (as discussed in the previous section), population sizes (N) fluctuate between 10^6 and 10^8 individuals. The probability of a fixation for any neutral mutation is equal to $1/N_e$, where N_e is the effective population size, approximately equal to the harmonic mean in this case, corresponding to a probability of $1/10^7$. What about mutations at intermediate frequencies? Under neutral evolution, governed only by genetic drift, the number of generations it would take for a mutation to reach a given frequency can be calculated by:

$$g = -log_2 f/\mu, \tag{1.1}$$

where g is the number of generations, f is the frequency of a mutation, and μ is the mutation rate (COOPER 2018). For a mutation rate of 1×10^{-10} mutations per nucleotide per generation, it would take ~33 billion generations for a mutation to reach a frequency of 0.1. Even in populations with mutation rates several orders of magnitude higher, the number of generations it would take for a neutral mutation to reach a detectable frequency remains implausibly high. Therefore, mutations cannot be present due to the action of mutation and drift alone. They must either be under selection or connected by genetic linkage to mutations under selection (hitchhiking). Further, coefficients of selection need not be strong for the dynamics to be driven by selection. When $Ne * s >> 1$, selection is the dominant force, and considering an N_e on the order of 10^7, even very weak coefficients of selection leads to situations where $Ne * s >> 1$. After ignoring the effects of genetic drift, we focus on genetic mutations and the complex relationship between those mutations and their effects on fitness.

At the root of all evolutionary change lies genetic mutation; adaptation can occur if and only if a change at the DNA sequence level leads to a change in phenotype that leads to a change in fitness. Because selection operates at the level of fitness, the role a given mutation has in adaptation depends on the effect that mutation

has on fitness. The connection between mutation and fitness is very complex for myriad reasons. Here, I focus on two factors that determine the fitness effect of a given mutation. First, the link between mutation and fitness is mediated by phenotype; selection acts on phenotypes and therefore depends on the mapping between genotype and phenotype. This mapping is complicated by pleiotropy, occurring when one genetic mutation affects more than one phenotype, and by epistasis, occurring when the phenotypic effects of a mutation depend on the genetic background on which they arise. Second, the link between mutation and fitness is environment-dependent. A mutation may confer a fitness benefit in the environment it arises, but under fluctuating conditions or other environments, there could be a variety of fitness effects. Together, these two complexities lead to difficulty predicting evolutionary trajectories. In Chapter 2, I investigate properties of several beneficial mutations to explore the role that pleiotropy and epistasis play in adaptation. I show that the fitness effects are indeed environment-dependent, and demonstrate how these environment-specific fitness effects could contribute to diversification.

1.2.2 Rate of beneficial mutation and clonal interference

One possible source of the abundance of diversity observed in evolving microbial populations is clonal interference. This occurs when multiple beneficial mutations arise simultaneously and compete for fixation, resulting in a pattern of overlapping segregating mutations. The prevalence of clonal interference in a population depends primarily on the supply rate of beneficial mutations (HERMISSON and PENNINGS 2005) and is therefore especially common in populations of microbes and viruses, where large population sizes lead to high supply rate of beneficial mutations (LANG *et al.* 2013; MESSER and PETROV 2013). The supply rate of beneficial mutations is also dependent on the selective conditions; a completely different set of mutations may be adaptive in one environment versus another (nicely demonstrated in PAYEN *et al.* 2016.

The effects of clonal interference on evolutionary adaptation have been explored extensively, both theoretically (FOGLE *et al.* 2008; GERRISH and LENSKI 1998; GOOD *et al.* 2012) and experimentally (DE VISSER and ROZEN 2006; LANG *et al.* 2013; WOODS *et al.* 2011). To summarize this body of work, there are three main effects of clonal interference on the dynamics of an adapting population. First, a supply rate of beneficial mutations that is larger than the rate of fixation increases the amount of genetic diversity present in a population. Second, the presence of multiple beneficial mutations reduces the efficiency of selection, increasing the fixation time of a beneficial mutation. This slows the rate of adaptive evolution, because multiple beneficial mutations are not able to be combined into a single lineage (DE VISSER *et al.* 1999). Third, the competition between beneficial mutations reduces the role of small-effect mutations; even relatively large-effect mutations can be outcompeted by mutations with a slightly larger selective advantage. One notable implication of the reduced efficiency of selection is the effect this has in spatially structured environments. If the population has some degree of substructure present, the varying amounts of clonal interference in each subpopulation increases overall diversity further. In the realm of speciation, this mechanism of diversification is analogous to mutation-order speciation, a result of different mutations arising and fixing in separate groups adapting to similar selection pressures (SCHLUTER 2009). Without the reproductive isolation that results from genetic incompatibilities during sexual reproduction, however, diversity that arises as a product of clonal interference is transient. It fluctuates as new beneficial mutations arise and remains only until one lin-

eage reaches fixation. This is in contrast to stable diversity that can arise as a product of divergent ecological selection (analogous to ecological speciation; SCHLUTER 2009) or ecological interactions between individuals within a population. The distinction is important when comparing drivers of diversification; diversity from clonal interference is neutral and present as a byproduct of adaptation while diversity from divergent ecological selection is adaptive, increasing the mean fitness of the population as a whole.

1.2.3 Ecologically-driven diversification

While clonal interference acts to increase diversity transiently, the interaction between evolutionary and ecological dynamics in a population can lead to more sustained and stable diversity. In environments that contain some degree of spatial variation in conditions, it is relatively straightforward to see how diversity would arise. Selection toward different ecological niches is generated by heterogeneity in the environment. As individuals within the population become more specialized to different habitats, diversity grows. There are many classic examples of this mechanism in natural populations, perhaps the most famous being the diversity of Darwin's finches in the Galapagos Islands. In this case, finches on different islands experienced different vegetation due to variation in climate and altitude, and adaptation to this island-specific vegetation resulted in diverse morphological traits such as beak size and beak shape (GRANT and GRANT 1989). This mechanism is contingent on at least some degree of subpopulation structure that limits mixing between divergent niches. In the case of the finches and other examples in higher taxa, the patterns of environmental heterogeneity and limited mixing are often clear. In microbial populations, however, these ecological variables are much harder to discern and subpopulation structure is not obvious. Many ER experiments are carried out in liquid broth, covered but not airtight. If grown in a static (ie. not shaken) culture, one source of environmental heterogeneity is the availability of oxygen. For example, RAINEY and TRAVISANO 1998 showed that the oxygen gradient present in static broth cultures of *Pseudomonas fluorescens* led to rapid and highly repeatable diversification into niche-specialized ecotypes. Here, the researchers were able to determine the selective pressure underlying the heterogeneity, but ecology on a microbial scale is challenging. Further, interactions between microbes and their environment, by metabolizing resources present, modify the resources available, thereby generating the heterogeneity in selection themselves. Some species of bacteria also interact by producing toxins, which can to lead to diversification and coexistence of types with varying levels of resistance, as observed in *Escherichia coli* by KERR *et al.* 2002. Like diversity that arises as a result of clonal interference, diversity that arises as a result of environmental heterogeneity depends on the supply rate of beneficial mutations and the spatial structure of the population. In contrast, diversity in response to environmental heterogeneity also depends on the strength of selection across space, occurring at a rate proportional to size of the difference in selection coefficients (reviewed in FELSENSTEIN 1976; HEDRICK *et al.* 1976.

In the examples and theory outlined above, diversification is contingent on environmental heterogeneity. However, diversification is often observed in microbial ER studies that are performed in well-mixed environments, often designed to reduce environmental heterogeneity (ELENA and LENSKI 1997; FRENKEL *et al.* 2015; FRIESEN *et al.* 2004; MADDAMSETTI *et al.* 2015). In these cases of experiments in homogeneous environments, diversification is driven by negative frequency-dependent selection. Resource competition

and subsequent specialization leads to this type of disruptive selection, resulting in a selective benefit for rare types in a population. This is often due to variation in metabolic pathways; organisms can obtain energy by many different routes. In *E. coli*, glucose is the preferred carbon source, and acetate is the byproduct of glycolysis that can be used as a secondary carbon source when glucose concentrations become low (SAXER *et al.* 2010). When grown in an environment containing both glucose and acetate, populations repeatedly and rapidly diversify, specializing on either resource (FRIESEN *et al.* 2004). Theoretically, this mechanism is predicted to generate diversity in a wide range of ecological conditions (DOEBELI and DIECKMANN 2000), including environments containing a single limiting resource (GUDELJ *et al.* 2007). Diversity in response to negative frequency-dependent selection in homogeneous environments depends on the same factors as diversity in heterogenous environments: supply rate of beneficial mutations and strength of selection, minus the spatial aspect. In homogeneous environments though, the strength of selection is not determined by a selection gradient, but instead by the geometry of the underlying network of ecological interactions, similar to a metabolic network. Therefore, diversity by this mechanism is expected to increase in environments with a larger range of resources, due to the increased number of trait dimensions increasing the likelihood of disruptive selection (DÉBARRE *et al.* 2014)).

In summary, a plethora of ecological interactions promote diversification in microbial populations. Disentangling the relative effects of the mechanisms above is crucial to understanding and predicting evolutionary trajectories. In Chapters 3 and 4, I investigate these mechanisms of diversification by comparing adaptation and diversification in environments that vary along two ecological axes: spatial structure and resource complexity. I contrast phenotypic and genetic diversity as a result of increasing both of these factors to investigate drivers of diversification within populations.

1.3 Repeatability of evolution

How repeatable is adaptation? Answering this question informs how well we are able to predict the outcomes of evolution, a tool that is becoming increasingly useful every day as climate change presents environmental changes to a large number of species. In natural populations, insight into repeatability comes from observations of independent parallel adaptation, such as the repeated evolution of similar morphotypes of Anolis lizards on different islands (LOSOS 1992), the loss of body armour in stickleback (SCHLUTER *et al.* 2004), and the evolution of wing pigmentation spots in Drosophila (PRUD'HOMME *et al.* 2006). In constrast to these examples, replicate evolving populations in ER experiments can be subjected to identical conditions, a feat that is difficult and arguably impossible to achieve in natural populations with many sources of environmental variation. ER experiments therefore provide a strong experimental framework to measure repeatability. Many such studies have revealed varying amounts of repeatability, reviewed thoroughly in LOBKOVSKY and KOONIN 2012, who conclude that though evolutionary trajectories are more predictable than expected, the problem remains far from a general solution.

Repeatability depends on the balance between the stochastic and deterministic processes of evolution. For reasons described in Section 1.2.1, genetic drift, one source of stochasticity, can be ignored in most microbial ER experiments, with the exception of mutation accumulation (MA) experiments, which are designed

to drastically reduce the effects of selection to observe the properties of spontaneously occurring mutations. A second source of stochasticity is mutation. At the molecular level, genetic mutations occur as a result of an error during DNA replication, an inherently random process. This is not a completely random process, however; across the genome of *P. aeruginosa*, DETTMAN *et al.* 2016 found that the distribution of spontaneous mutations was non-random. How much does this heterogeneity in mutation rate reduce stochasticity and therefore increase the repeatability of evolution? The answer likely depends on population size, suggests BAILEY *et al.* 2017, explaining that larger populations are much less likely to be limited by mutation supply, reducing the effect of mutational heterogeneity on parallel evolution. In addition, the degree of heterogeneity in mutation rate observed by DETTMAN *et al.* 2016 in *P. aeruginosa* is small, less than an order of magnitude. Taken together, this evidence allows us to assume that observations of parallel evolution in the experiments presented here are primarily the result of selection.

Considering only the action of selection to produce repeated evolutionary outcomes, repeatability then depends on the relationship between genetic mutation and fitness. As discussed in Section 1.2.1, this relationship is comprised of two components, the genotype to phenotype map and the environment in which adaptation occurs. If many genotypes lead to the same adaptive phenotype, adaptation is less constrained at the genetic level. This breadth of genetic routes to a given phenotype reduces evolutionary repeatability because there is a larger set of solutions to the problem (modelled in CHEVIN *et al.* 2010. In contrast, if only a handful of mutations are beneficial, adaptation will be more constrained and therefore highly repeatable. Further, the number of beneficial mutations depends on the environment such that a larger number of traits under selection leads to a larger number of potentially beneficial mutations and therefore less repeatability. Typically, evolutionary repeatability is measured at the genetic level, by quantifying the number of times genes are reused in replicate populations adapting to the same environment. Without considering repeatability at the trait level, a pattern of low genetic repeatability could signify either of the factors discussed here: many potentially adaptive phenotypes or many genetic routes to the one or few adaptive phenotypes. In Chapter 4, I investigate the effects of environment on genetic repeatability by comparing the breadth of mutations that arise in replicate populations. From this, we are able to infer how the experimental conditions may have constrained evolutionary trajectories.

1.4 Clinical applications of fundamental research

Microbes are globally ubiquitous and completely intertwined, ecologically and evolutionary, with human health and disease. A wide range of both commensal and harmful species and often a complex community of both can colonize various body sites, including the skin (COGEN *et al.* 2008), gut (GILL *et al.* 2006), urinary tract (FOXMAN 2010), and lungs (BECK *et al.* 2012). Concepts from theoretical and experimental evolutionary biology are critical to understanding how these populations adapt to the complex environments of the human body. Further, as bacterial pathogens evolve resistance to antibiotics and we enter the "post-antibiotic era" (ALANIS 2005), the integration of evolutionary thinking and medicine will become even more critical.

The microbe introduced in Section 1.1.2 is one such important human pathogen. *Pseudomonas aerugi-*

nosa is an opportunistic pathogen, infecting the lungs of patients with cystic fibrosis (CF) and other immunocompromised individuals. It is the most prevalent pathogen the CF lung and responsible for the symptoms that lead to patient morbidity and mortality (FOLKESSON *et al.* 2012; HARRISON 2007). CF is the most common fatal disease affecting Canadian children and young adults and though improvements in treatment have lengthened the typical lifespan of those with the disease, there still remains no cure (STEPHENSON *et al.* 2015). Infections of this species are particularly difficult to eradicate likely due to its metabolically versatile genome and ability to adapt (YANG *et al.* 2011). In Chapter 3, I allow populations of *P. aeruginosa* to evolve under a range of environments, some of which are designed to replicate the CF lung environment. Though this experiment is motivated by the fundamental question of the emergence of biological diversity, the choice of a medically relevant system makes this interdisciplinary research uniquely positioned to inform the process of adaptation in a deadly human pathogen. In this section, I first provide a brief overview of the field of microbial evolutionary medicine and then describe the role of diversity in pathogen adaptation, both in *P. aeruginosa* and in general. In the last section, I attempt to connect the research questions addressed in this thesis to their clinical counterpart.

1.4.1 Microbial evolutionary medicine

The concept of applying the principles of evolution by natural selection to understanding aspects of human health is not new. In the early 1990s, WILLIAMS and NESSE 1991 described how applying the adaptationist program to human disease could provide new insight into the causes of various medical disorders, originally dubbing the field Darwinian medicine. One key shift in mindset was that evolutionary principles apply to pathogens causing disease as well as human bodies trying to resist disease (NESSE 2001). Framing the causes of disease into questions from the point of view of the pathogen stands to offer a better understanding of how a pathogen will respond to treatment and a therefore greater chance of disease control. For example, this shift has illuminated the identity of mutations that confer resistance to antibiotics (FARHAT *et al.* 2013), the role of interactions between species in pathogen adaptation (DE VOS *et al.* 2017), and is beginning to illuminate the complex dynamics of the human gut microbiome (GROUSSIN *et al.* 2017).

What is relatively new, however, is the application of ER experiments to medically-relevant systems. The benefits of applying ER experiments to medicine are two-fold. First, the setting of the laboratory controls environmental conditions, allowing for true replication of a population adapting to a given environment. Second, by seeding an evolution experiment with cells from a single bacterial colony, the initial genetic diversity can be controlled, attributing any diversity that arises to *de novo* mutations, as opposed to other sources in a clinical setting such as other patients. These two benefits, combined with technological advances in large scale genomics, gives researchers unprecedented power to understand and potentially alleviate human diseases caused by microbes, old and new.

1.4.2 Importance of diversity in pathogen adaptation

Until recently, diversity in populations of infectious pathogens was rarely discussed. In the context of *P. aeruginosa* infecting the CF lung, descriptions of the extent of the diversity present were first published only 6 years ago (ASHISH *et al.* 2013; LIEBERMAN *et al.* 2014). Why this diversity existed and more importantly,

what effect this diversity had on clinical outcomes was a complete unknown. At the time, the common assumption was that a single dominant lineage of *P. aeruginosa* becomes established in the lung and over time evolves in a relatively predictable way towards a phenotype categorized as chronic based on characteristics such as reduced motility, increased resistance to antibiotics, altered quorum sensing, and reduced virulence factor production. In the last couple of years, however, studies including the analysis of more than a single isolate have caused a dramatic shift in focus. These studies discovered a wide and dynamics range of phenotypes present in a single population, often diverged from a common ancestor (CLARK *et al.* 2015; MARKUSSEN *et al.* 2014). Further, the lack of a correlation between the prevalence of any phenotypic trait and the symptoms of CF (WORKENTINE *et al.* 2013) made two things clear. First, that sampling a single isolate from any given patient was not meaningful and second, that the composition and complexity of the population is what determines pathogenicity, rather than a specific phenotypic or genetic variant present. Relatively little is known about what governs the composition and complexity of a population and for this reason understanding the factors that drive diversification of this pathogen could have significant implications for the treatment of CF.

What is the relationship between diversity and adaptability? Theoretically, the potential of a given population to adapt is directly proportional to genetic variation present, so a more diverse population should be more able to adapt (FISHER 1930). At present, I know of no explicit test of the effect of intra-species microbial diversity on the ability to adapt or colonize a novel environment, though there is some recent evidence that more diverse populations of *P. aeruginosa* are more resilient to antibiotic treatment (reviewed in CLARK *et al.* 2018). On the multi-species, community level, however, there is some evidence that diversity is a key factor. A diverse human gut microbiome has been shown to be associated with improved health and resistance to disease (MOSCA *et al.* 2016). A specific example of this effect is that colonization of *Clostridium difficile* of the human gut is more likely to occur when diversity is low (SMITS *et al.* 2016). At the genetic level, more diverse populations are more adaptable, as is likely in the case of resistance to antibiotics. Diversity at the phenotypic level could play an important role too. If a less diverse population has more open niches, more opportunity for colonization of problematic species could lead to disease. In the case of *C. difficile* infections, this could be the underlying mechanism. Either way, diversity plays an important role and elucidating this role is on the frontier of microbial evolutionary medicine.

1.4.3 Direct applications of this research

Though rooted in fundamental questions in evolutionary biology, this body of research, is highly translational, both broadly and to specific issues in clinical medicine. Chapter 2 addresses environment-specific fitness effects of beneficial mutations. Though far from direct translation, this could be applied to fitness effects of AMR (antimicrobial resistance) mutations when conditions change. Chapters 3 and 4 apply to clinical medicine more specifically by addressing phenotypic and genetic adaptation of a human pathogen to a human lung-like environment.

Chapter 2 investigates the genetics of local adaptation by measuring fitness effects of adaptive mutations in novel environments. When a population of microbes adapts to a particular environment, it becomes specialized to use the resources present in that environment by accumulating mutations that increase fitness,

typically by increasing growth efficiency. When conditions change, however, these mutations may no longer be beneficial. And when conditions change, the effect of those mutations on fitness under the new environment may determine whether the population will be able to grow and survive. If a beneficial mutation incurs a fitness cost in a new environment, that mutation carries a fitness trade-off; it is beneficial in one environment and deleterious in another. A general question when treating bacterial infections with antibiotics is whether or not the mutations that lead to AMR carry these fitness trade-offs. Elucidating the extent of environment-dependent fitness effects can inform how the fitness effects of AMR mutations change when environmental conditions change, such as stopping a given antibiotic treatment or starting a new one.

Chapters 3 and 4 are more directly applied to clinical medicine, by addressing adaptation and the emergence of diversity in populations of *Pseudomonas aeruginosa* in the CF lung, with Chapter 3 focused on phenotypic changes and Chapter 4 on genetic changes. Prolonged infections in CF patients are accompanied by a suite of phenotypic characteristics and marked diversification into coexisting types, but little is known about what drives these changes in vivo. The suite of traits that commonly arise are thought of as markers of chronic infection that are a result of *P. aeruginosa* adaptation to the unique conditions of the CF lung environment. A direct comparison of the traits that arise in simple lab environments to those that often arise in the lung can highlight the selection pressures driving the emergence of those traits. For the diversification into coexisting types, it is commonly thought that diversification is driven by the environmental heterogeneity generated by a highly compartmentalized structure (MARKUSSEN *et al.* 2014; WILLNER *et al.* 2012). Much less attention has been paid to the effect of the nutritional content of the human lung. By investigating the emergence of diversity in environments that mimic different aspects of the ecological conditions of the lung, we can parse the ultimate source of diversity, potentially leading to more sustainable methods for controlling population diversification.

In addition to the source of diversity, the highly replicated nature of evolve and resequence experiments allows us to answer another big question in clinical medicine: how personalized should treatment be? With advancing genomic technologies, precision medicine is a rapidly growing method for treating disease. However, this method only has value if evolutionary trajectories are somewhat predictable and experimental evolution can inform this. Populations from different patients or from different compartments of a single patient could vary in terms of ecological conditions and selective pressures, whereas controlled experimental environments allow us to ask how much of the divergence we observe is independent of differences between hosts, or patients. The degree to which adaptation and diversification is parallel in these populations provides a measure of predictability. For example, convergent evolutionary paths implies constraint, which may be important as highly constrained traits may be more readily targeted by antibiotics or other treatment. In the context of the experiments presented here, comparing parallelism between populations within treatments allows us to compare the evolutionary constraints imposed by the various environments. Ultimately, a better understanding of the breadth of evolutionary trajectories may lead to better treatment choices for clinicians.

1.5 Summary

Overall, the work presented in this thesis advances our understanding of the role of diversification and ecological specialization in adaptation to novel environments. Using the techniques of evolve and resequence experiments, each chapter examines variation at a different biological level: at the level of fitness, phenotype, and genotype. In Chapter 2, I investigate variation in fitness effects across environments by directly measuring fitness effects of adaptive mutations in multiple envioronments. I find that adaptive mutations commonly have deleterious effects in other environments, generating fitness trade-offs associated with ecological specialization. In Chapter 3, I investigate the emergence of phenotypic variation in a range of environments, asking whether environments with more ecological opportunity result in more diverse populations. I show that adding ecological complexity increases diversity within populations and increases divergence among populations, suggesting that ecological complexity provides a larger breadth of evolutionary trajectories. In Chapter 4, I investigate the genetic variation that underlies these evolutionary trajectories. Whole genome sequencing of evolved populations revealed large amounts of polymorphic loci and varying amounts of genetic parallelism, depending on the environment. Taken together, this thesis advances our knowledge of the generation of diversity and how the ecological conditions of an environment can influence that process.

Chapter 2

Environment-dependent fitness effects of adaptive mutations

A version of this chapter has been published as "Schick, A., Bailey, S., & Kassen, R. 2015. Evolution of fitness trade-offs in locally adapted populations of *Pseudomonas fluorescens*. American Naturalist." The original project was conceived by myself and R. Kassen. I preformed the laboratory work using strains evolved by S. Bailey and with mutation information from sequence data analyzed by S. Bailey. I wrote the initial draft of the manuscript and R. Kassen and I contributed major revisions, with minor feedback from S. Bailey.

2.1 Summary

The degree to which a population is adapted to its environment, a phenomenon called local adaptation, seems to be common in natural systems, but the genetic causes of its evolution remain poorly understood. Here we characterize the genetic causes of trade-offs generating local adaptation in populations of *Pseudomonas fluorescens* that had previously been evolved for specialization on three different carbon resources. We measured the fitness effects of mutations that arose during selection in that environment and in alternative environments to quantify the degree of specialization. We find that all mutations are beneficial in the environment of selection and that those arising later during adaptation are associated with increasingly antagonistic effects in alternative environments compared with those arising earlier, consistent with a multioptima version of Fisher's geometric model of adaptation. We also find that fitness of pairs of beneficial mutations are consistently less than additive in selection environments, producing a pattern of diminishing returns, but are more variable in alternative environments, being either positive or negative. Finally, we find that mutations in genes associated with loss of motility are beneficial across all environments, whereas mutations involving other functions, such as gene regulation, had more variable effects, being more environment specific. Taken together, these results provide a detailed account of the genetics of specialization and suggest that the evolution of trade-offs associated with local adaptation may often result from the antagonistic effects of beneficial mutations substituted later in adaptation.

2.2 Introduction

Our explanations for the maintenance of biodiversity, whether at the scale of genetic variation within populations or among species in communities, rely on the simple and not unreasonable condition that fitness

trade-offs across environments prevent any single genotype (or species) from being fittest across all conditions of growth. The available evidence largely supports this view. Building on the foundational work on specialization and cross-environment fitness trade-offs (FRY 1996; FUTUYMA and MORENO 1988), recent reviews of local adaptation find that between 60-80 percent of studies surveyed find some evidence for local adaptation (FRASER *et al.* 2011; GREISCHAR and KOSKELLA 2007; HEREFORD 2009; HOEKSEMA and FORDE 2008; LEIMU and FISCHER 2008; SANFORD and KELLY 2011) in the sense that a native population outperforms a foreign population in the native's home environment. Specialization, which represents the magnitude of the fitness difference across habitats or niches, is also common, though often less pronounced than our intuition might suggest. HEREFORD 2009, for example, estimates the correlation between home and away fitness across a wide range of study systems to be just -0.14, hardly evidence of extreme specialization to abiotic conditions of growth. Finally, data recently collated from microbial selection experiments that have studied the emergence of specialization resulting from divergent selection leave little doubt that trade-offs evolve readily and quickly: 20/21 studies examined, across a range of species including bacteria, algae and fungi, showed evidence of trade-offs that evolved on time scales ranging from a few hundred to a few thousand generations (KASSEN 2014).

Despite the prevalence of trade-offs, we still lack, with a few notable exceptions (ANDERSON *et al.* 2013; BANTINAKI *et al.* 2007; RODRIGUEZ-VERDUGO *et al.* 2014), a detailed understanding of the genetics governing their evolution. There are two reasons for this. The first is that access to the identity of mutations causing adaptation in general, and local adaptation in particular, has only become possible in the past few years as the costs associated with whole-genome sequencing have fallen. The second is that compelling predictions from detailed multi-locus models for the genetics of adaptation have focused almost entirely on adaptation to a single phenotypic optimum (MARTIN and LENORMAND 2006; ORR 2005).

To remedy this situation, we have studied the genomics of local adaptation in evolving populations of the Gram-negative bacterium, *Pseudomonas fluorescens*, during adaptation to distinct carbon resources. We outline a conceptual framework for thinking about the evolution of fitness trade-offs across environments leading to local adaptation that builds on Fisher's geometric model of adaptation. We then test two compelling predictions emerging from this model and provide empirical insight into how local adaptation and trade-offs evolve by unpacking the genetics of local adaptation in a microbial selection experiment.

The basic principles governing the evolution of ecological specialization and the fitness trade-offs that accompany it are straightforward. Strong divergent selection, often generated by an environment that contains distinct niches or habitats that vary in the conditions of growth, causes the evolution of different optimal phenotypes in each niche. These phenotypes are expected to be ecologically specialized to the prevailing conditions of growth, in the sense that they have highest fitness in their own niche and lower fitness elsewhere, because adaptation proceeds by substituting mutations that are beneficial in those conditions, irrespective of their fitness in any other environment. Most of the time, a gene that increases fitness in a focal environment will not increase fitness by as much, and may even be deleterious, in an alternative environment because they have environment-specific effects (ANDERSON *et al.* 2013; FALCONER 1960; KASSEN 2014). The pleiotropic effects of beneficial mutations in different environments can thus vary from being positive to negative (or, equivalently, antagonistic), and understanding what governs the sign and magnitude of those

pleiotropic effects is key to understanding the evolution of specialization.

Fisher's geometric model of adaptation (FISHER 1930) provides a useful starting place. In its simplest form, adaptation is modeled as a sequence of mutational moves or steps in phenotype space that brings that population closer to the optimal combination of traits for a single environment. Two key predictions emerge from this model: (1) mutations with large beneficial effects tend to be substituted first, followed by mutations of smaller effect as the population approaches the optimum; (2) because of the bias towards substituting large effect mutations first, adaptive walks - the sequential substitution of beneficial mutations by selection - tend to be quite short, often no more than a handful of mutations before fitness plateaus.

To gain some insight into the pleiotropic effects of beneficial mutations in alternative environments in this model, imagine there is a second phenotypic optimum located some distance away from the first (modeled in MARTIN and LENORMAND 2015). This might be the case, for example, when a population gains access to a novel environment through dispersal. Assume that selection happens towards one optimum only (i.e. divergent selection is strong relative to gene flow). It is now easy to see that a mutation beneficial in this environment has a pleiotropic effect in the other environment that depends on starting position of the wild-type genotype. When the starting genotype is located far away from both fitness peaks, then a mutation beneficial in one environment is also likely to improve fitness in the other environment and we say that the mutation exhibits synergistic or positive pleiotropy. However if the starting genotype is between the fitness peaks, then a mutation beneficial in one environment will be deleterious in the other, and the mutation is antagonistically pleiotropic. The pleiotropic effects of beneficial mutations are thus likely to change as a result of adaptation, becoming more negative as adaptation proceeds.

As a result, the model predicts that the degree of ecological specialization, and therefore the strength of fitness trade-offs across environments, should increase as adaptation proceeds. The evolution of ecological specialization may be further supported or exaggerated through two additional mechanisms. First, mutations that are neutral in the environment of selection but deleterious in alternative conditions can be fixed either by drift or through hitchhiking alongside beneficial mutations in asexual populations (KAWECKI 1994; KAWECKI *et al.* 1997). Second, additional mutations that get substituted through drift, hitchhiking, or selection can further contribute to the evolution of trade-offs if their combined effects are more deleterious in an alternate environment than expected from an additive model. This mechanism is reminiscent of Dobhzansky-Mueller interactions, where beneficial mutations substituted independently in different populations are, in combination, deleterious. In both cases the combined effects of multiple mutations remain untested by selection, and so are not constrained to be beneficial.

To test these predictions, we have investigated the genomics of specialization in a series of populations of the Gram-negative bacterium *Pseudomonas fluorescens* that had previously evolved for 1000 generations on three distinct carbon sources. The dynamics of fitness and specialization in these populations have been reported previously (BAILEY and KASSEN 2012) and we recently published a detailed study of the genetics of adaptation in one line that involved the substitution of strongly beneficial synonymous mutations (BAILEY *et al.* 2014). Importantly for our present purpose, all lineages evolved fitness trade-offs across the environments studied. Here we provide a more complete account of the genetics underlying these fitness trade-offs, by genotyping isolates from a range of populations and timepoints and testing competitive fitness of specific

mutations. Our results show that the distribution of pleiotropic fitness effects in alternative environments becomes more antagonistic as adaptation proceeds, in line with the predictions of Fisher's two-optima model. We find no evidence that mutation accumulation contributes to trade-offs in this experiment but instead that trade-offs are the result of pleiotropic fitness effects in alternative environments. Furthermore, we find that non-additive fitness effects in alternative environments can modulate the degree of specialization and that mutations in motility related genes tend to have much less specialized effects than those in regulatory genes.

2.3 Materials and Methods

2.3.1 Bacterial strains

The bacterial strains used in this study were isolated from a previous selection experiment designed to explore the impact of different forms of environmental variation on the dynamics of fitness (for details see BAILEY and KASSEN 2012). Briefly, three replicate populations in each of five environments (3 single-resource environments, and 2 multi-resource environments) were founded from a single clonal isolate of *Pseudomonas fluorescens* strain SBW25:lacZ. The single-resource environments consisted of a minimal salt medium supplemented with glucose, mannose, or xylose as the sole energy source. One multi-resource environment contained all three sugars mixed together (MIX) while the other consisted of a spatially structured environment composed of each of the three carbon sources contained in a different well and randomly dispersing a sample of the population among wells every 24 hours (SPAT). We picked 12 colonies at random at 100 generation intervals from each of eight populations in this original experiment. More specifically, we studied three populations selected in mannose, two selected in xylose, one selected in glucose, and one each from the MIX and SPAT treatments. Our decision to focus on these populations specifically rather than all 15 evolved populations stems from our interest in the evolution of specialization, which is most likely to occur in single-resource environments, and the fact that the theory being tested requires more than one mutation to have been substituted during the experiment. One glucose line and one xylose line were omitted because sequencing revealed only a single mutation present at the endpoint. A second glucose line was omitted because it has been investigated in detail in a previous paper (BAILEY *et al.* 2014).

2.3.2 Genomic evolution over 1000 generations

We reconstructed the dynamics of substitution in each population by interrogating each of 12 isolates at every time point in a given evolved lineage for mutations identified through whole genome sequencing of the endpoint populations. We designed 20 pairs of primers for genotyping the isolates for this study (see appendix Table A.1 for primer sequences). All polymerase chain reactions (PCR) consisted of 1 μL forward primer (10 μM), 1 μL reverse primer (10 μM), 1 μL bacterial culture, 12.5 μL Promega PCR Master Mix, and 10.5 μL nuclease free water. For all reactions, the basic thermocycler protocol was as follows: denaturation at 98°C for 10 minutes, followed by 30 cycles of 98°C for 30 seconds, 56°C for 30 seconds, and 72°C for 90 seconds. Annealing temperatures were varied to improve amplification (optimal temperature ranges for each primer pair are reported in Table A.1). PCR products were assayed on 1% agarose gels and sent

away for commercial sequencing. Unpurified PCR products were sequenced by the McGill University and Genome Quebec Innovation in Montreal, Canada, which uses Applied Biosystem's 3730xl DNA Analyzer technology for Sanger sequencing. On top of the 120 PCR products (12 isolates $\times$ 10 time points) for each of the 23 mutations, we examined an additional 15 isolates from each of generations 600 - 1000 in one of the glucose-selected lines, because we failed to recover one genotype identified by whole genome sequencing in our initial survey of 12 isolates.

2.3.3 Measuring fitness

We estimated fitness of each evolved strain or genotype using competitive fitness assays. A focal strain and a competitor strain with the opposite lacZ marker were grown as pure cultures overnight from frozen stocks in 2 mL of LB media at 30 degrees C, 250 rpm. Optical density at 600nm (OD) of each culture was measured and volumes adjusted to ensure an approximately equal number of cells. 20 μl of the mixture was inoculated into 2 mL of fresh medium and allowed to grow for 24 hours. All competitions were performed in triplicate in each of the three single resource environments (minimal salts supplemented with either glucose, mannose, or xylose) yielding a total of 9 competitions per genotype. Initial and final frequencies of the focal and competitor strain were obtained by counting at least 60 colonies on agar plates supplemented with X-gal (40 μg/mL). Fitness was determined as follows:

$$ln(f_{final}) = ln(f_{initial}) + s \times t \tag{2.1}$$

where f_{final} and f_{inital} are ratios of the frequency of the focal and competitor colonies after and before competition, s is the selection coefficient, and t is the number of generations that occur during the competition period (DYKHUIZEN 1990). Rearranging for s and substituting $w = 1 + s$ gives the following equation for fitness:

$$w = 1 + ln(f_{final}/f_{initial})/t \tag{2.2}$$

All competitions were carried out for 24 hours and fitness was calculated by dividing the natural log of the ratios by 6.64, the estimated number of doublings associated with a 1:100 dilution factor.

We performed two kinds of competitive fitness assay that differed in the identity of the competitor strain. In one, the common competitor was the reciprocally marked SBW25 ancestor. The second assay involved competing the focal strain against its reciprocally marked immediate ancestor by removing the *lacZ* gene from one of the strains. Details on how the marker was removed are provided below. Note that to ensure there was no effect of marker removal on fitness, we constructed marked and unmarked versions of both genotypes used in these assays.

2.3.4 Allelic replacement

Since all populations were founded using an ancestor marked with a *lacZ* marker (from ZHANG and RAINEY 2007), we used allelic replacement to remove the *lacZ* marker in order to measure the relative fitness of two evolved genotypes. This was done with all 23 unique evolved genotypes. The wild-type (*lacZ-*) fragment was amplified and ligated into a plasmid vector (*pUIC3-sacB*), which was then transformed by electroporation into *E. coli* DH5λpir. After screening the transformants for the presence of the insert, this strain was used as the donor in triparental matings using each of the 23 evolved strains of *P. fluorescens* as the recipients. Following the mating, clones that had incorporated the plasmid by a single crossover event were selected for using the tetracycline and ampicillin resistance markers carried by the *pUIC3-sacB* plasmid. Excision of the plasmid by a second crossover event was selected for using a sucrose sensitivity marker (*sacB*) on the plasmid. Finally, the absence of the *lacZ* marker (and therefore presence of the wild-type fragment) was confirmed by plating on agar supplemented with X-gal and screening for white colonies. Targeted sequencing was used to confirm the marker had been removed from the *lacZ* locus.

2.3.5 Distribution of fitness effects

We estimated the fitness of individual mutations substituted during the experiment from the competitive fitness assays described above. In those cases where the mutant was the first to be substituted, fitness was determined in competition against the SBW25 ancestor. Where the mutation arose in a genetic background that contained a previously substituted mutation, we used the estimate of fitness from the marker-removal experiments described above. Fitness was assayed in the selection and non-selection environments, the latter being expressed as the mean fitness across each of the two non-selection conditions.

2.3.6 Effects of multiple mutations on specialization

To obtain a clearer picture of how the substitution of multiple mutations impacts specialization, and so the degree of local adaptation, we investigated the interaction between beneficial mutations at two loci. We did this by comparing the individual effects of each mutation on competitive fitness to the effect of both mutations in combination. More specifically, we compared the expected (additive) fitness for each mutation estimated from competitions against their immediate ancestor against the fitness of the double mutant in competition against SBW25. The non-additive effect on fitness for a pair of mutations A and B (e_{AB}) was calculated as follows:

$$e_{AB} = w_{ABvsSBW25} - (w_{AvsSBW25} + w_{ABvsA}), \tag{2.3}$$

in each environment. A non-additive fitness effect can be caused by epistasis between the two loci, but could also be caused by frequency-dependent fitness effects or intransitive fitness effects. Since the fitness of each genotype is inferred directly through competition, we cannot determine the mechanism of non-additivity.

2.3.7 Extent of specialization amongst different functional classes of mutation

Different functional classes of mutation might be expected to have different effects on fitness across multiple environments. To explore this possibility in more detail, we categorized all the mutations in our study into three broad functional groups (previously identified by BAILEY *et al.* 2015, who provide methodological details) and used our estimates of fitness across the three carbon resource environments to calculate the extent of specialization associated with each mutation. Mutations were categorized as motility, regulatory, or other (specific functions listed in Table A.2), with those of unknown function being classified as other and the variance in competitive fitness across environments was used to quantify specialization.

2.3.8 Data Accessibility

To facilitate data reuse, all raw data have been deposited in the Dryad Digital Repository doi: http://dx.doi.org/10.5061/dryad.51mg4125.

2.4 Results

2.4.1 Local adaptation

The mean fitness of all isolates across all three single-resource environments at the end of the experiment is shown in Figure 2.1. Note that the data presented in Figure 2.1 were previously reported, in a slightly different form, in BAILEY and KASSEN 2012. Selection was effective at generating local adaptation in each single-resource environment, as evidenced by notable gains in fitness in each selection environment and less substantial gains, or sometimes even losses, of fitness in the respective alternative environments. This result is confirmed by the significant interaction between selection environment and assay environment in analysis of variance (two-factor ANOVA, $F = 9.28$, $df = 8$, $p < 0.0001$). The trade-offs associated with local adaptation to xylose were underlain by a cost of adaptation to glucose and mannose. By contrast, trade-offs in mannose were associated with increases in fitness in both glucose and xylose, and trade-offs in glucose were associated with increased fitness in mannose, but no effect on fitness in xylose.

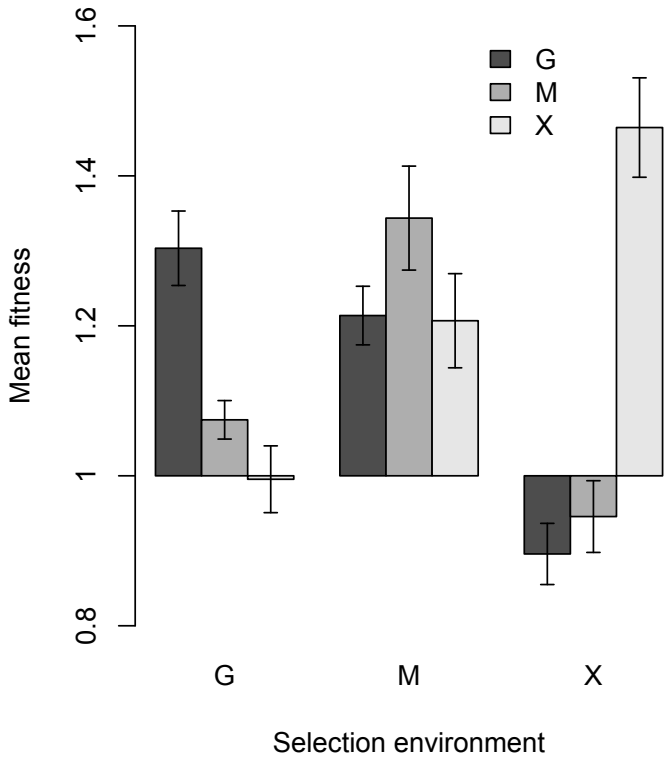

Figure 2.1: Mean relative fitness at 1000 generations. Fitness values shown are the mean fitness (+/- SE) of six clones from each line, measured by competition against the ancestral strain (SBW25) in glucose (G), mannose (M), and xylose (X). Each group represents the environment of selection, and bars within groups represent the environment of the fitness assay.

2.4.2 Genome sequencing

Detailed results on the identity and patterns of phenotypic and genomic parallel evolution observed in this experiment are being reported elsewhere (BAILEY *et al.* 2015). Briefly, genome sequencing of the end-point populations revealed a total of 52 genetic changes including 24 SNPs, 8 insertions and 20 deletions. All but two mutations were non-synonymous. Half of the genetic changes were fixed, the other half being polymorphic. Among all mutations identified, xylose populations contained most (3.7 +/- 0.2) followed by mannose (2.7 +/- 0.8) and glucose (2.3 +/- 0.8).

2.4.3 Dynamics of genome evolution

Figure 2.2 and Figure 2.3 display illustrative examples, both from independently evolved lines in mannose, of two representative patterns of substitution during the experiment. The first pattern, shown in Figure 2.2, is consistent with the classic model of periodic selection put forward by ATWOOD *et al.* 1951. The sequential substitution and replacement of one genotype for another is associated with a steady fitness increase in the selection environment and more modest fitness gains the alternative environments. The second pattern,

Figure 2.3, shows a more complicated pattern indicative of clonal interference, with multiple beneficial mutations arising and competing for fixation. Nevertheless the dynamics of fitness in the selection environment are not too dissimilar from what was seen previously in the example of periodic selection, with fitness increasing steadily over the course of the experiment and reaching a plateau by the end. Again, the pleiotropic effects of the mutations being substituted appear to be on the whole synergistic, with fitness increased to a lesser extent in the alternative environments compared to the focal environment by the end of the experiment. Visual inspection of the dynamics of genomic evolution in the remaining lineages reveals there is no tendency for periodic selection or clonal interference to be associated with a particular selection regime.

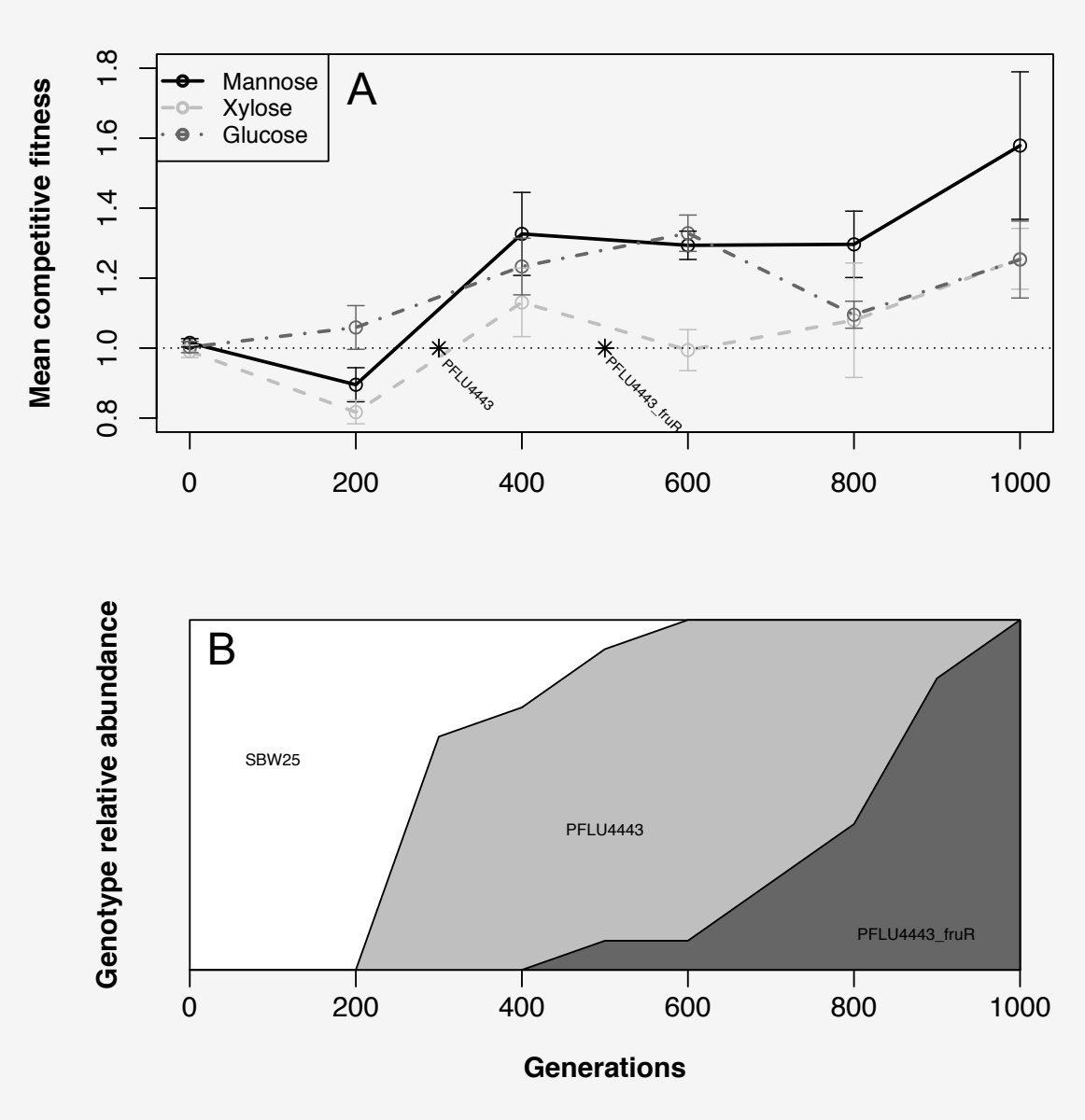

Figure 2.2: Dynamics of adaptive mutations in mannose-selected line m1. A. Mean competitive fitness (+/- SE) at 200 generation intervals, assayed in mannose (black, solid), xylose (light gray, dashed) and glucose (dark gray, dotted). Stars indicate the first timepoint at which mutations were detected, labelled with the gene those mutations were found in. B. Frequency of adaptive mutations, measured by presence/absence in 12 clones at 100 generation intervals.

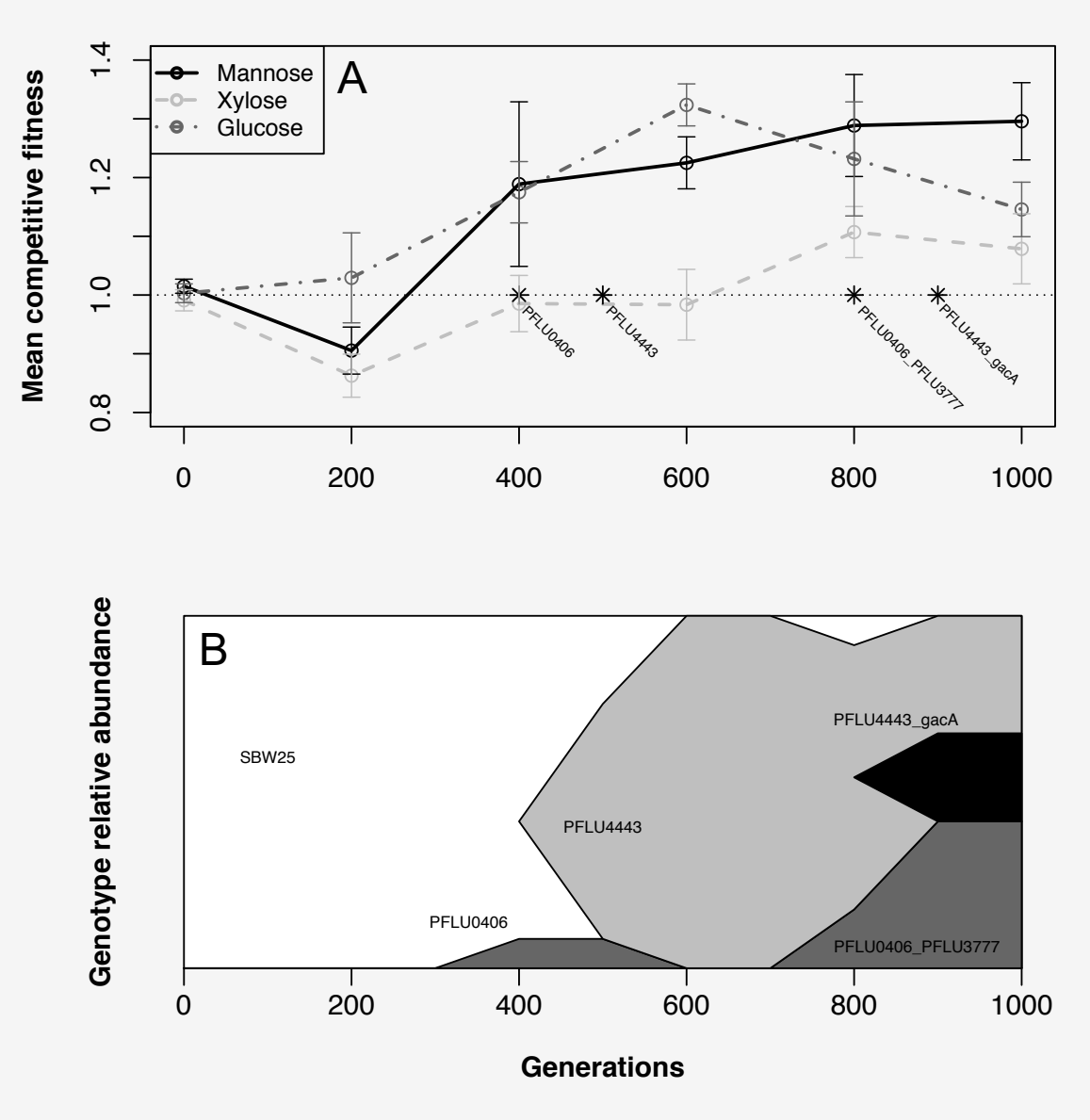

Figure 2.3: Dynamics of adaptive mutations in mannose-selected line m3. A. Mean competitive fitness (+/- SE) at 200 generation intervals, assayed in mannose (black, solid), xylose (light gray, dashed) and glucose (dark gray, dotted). Starts indicate the first timepoint at which mutations were detected, labelled with the gene those mutations were found in. B. Frequency of adaptive mutations, measured by presence/absence in 12 clones at 100 generation intervals.

2.4.4 Distribution of fitness effects among beneficial mutations

Figure 2.4 shows the distribution of fitness effects among beneficial mutations for the first and second mutations substituted in the environment of selection (filled circles) and in both alternative environments (expressed as the mean fitness effect across both alternative environments; open circles). The mean fitness

effect of both the first and second step in the focal environment is always positive, as expected. Consistent with theory (MARTIN and LENORMAND 2008; ORR 2006) and previous work (BARROSO-BATISTA *et al.* 2014; MILLER *et al.* 2011; SCHOUSTRA *et al.* 2009), the first step in the focal environment is larger than the second (mean first step = 0.213, mean second step = 0.063, paired t-test, t = 4.1089, df = 10, $p = 0.0034$). The pleiotropic fitness effects of the first mutations were also larger, on average, than those of the second (i.e. fitness in the alternative environments; mean first step = 0.089, mean second step = 0.006, paired t-test, t = 3.6773, df = 10, $p = 0.0062$), consistent with the expectation from Fisher's model with multiple optima that the pleiotropic fitness effects of mutations become more negative as adaptation proceeds (MARTIN and LENORMAND 2015).

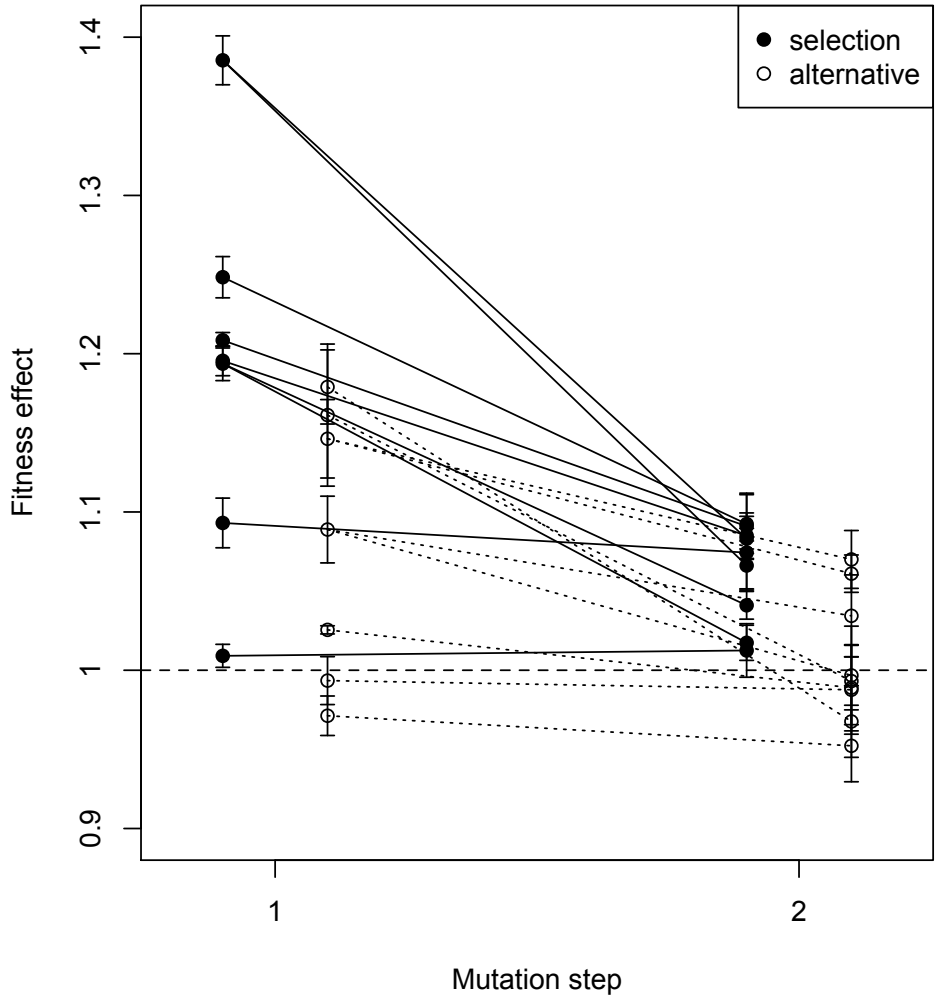

Figure 2.4: Distribution of fitness effects of beneficial mutations. Fitness effects of the first mutations are relative to the ancestral strain (SBW25) and fitness effects of the second step mutations are relative to the single mutant they evolved from. Mutations that arose in the same lineage are connected. Fitness effects above the dashed line are beneficial and those below are deleterious. Error bars denote standard error.

2.4.5 Multiple mutations and specialization

The impact of multiple mutations on the degree of specialization is illustrated for lineages in Figure 2.5 and summarized for all lineages in Figure 2.6. One way to interpret these results is to compare the fitness effects of the double mutant to the additive expectation based on the combined effects of each individual mutation. Within the first example (Figure 2.5A; line x2), the fitness of the double mutant is opposite in sign to the additive expectation in mannose, less than the additive expectation in xylose, and more negative than the additive expectation in glucose. A similar range in the non-additive fitness effects can be seen in the second example (Figure 2.5B; line m2). Despite the variation in the effects that can occur, in all cases that we studied, the difference between the fitness of the double mutant and the additive expectation was negative within the selection environment (Figure 2.6). This was not the case for the alternative environment, where responses were more varied and this difference was less negative on average (mean in selection environment = -0.124, mean in alternative environment = -0.048, paired t-test; t = -2.5861, df = 13.8, p = 0.02172).

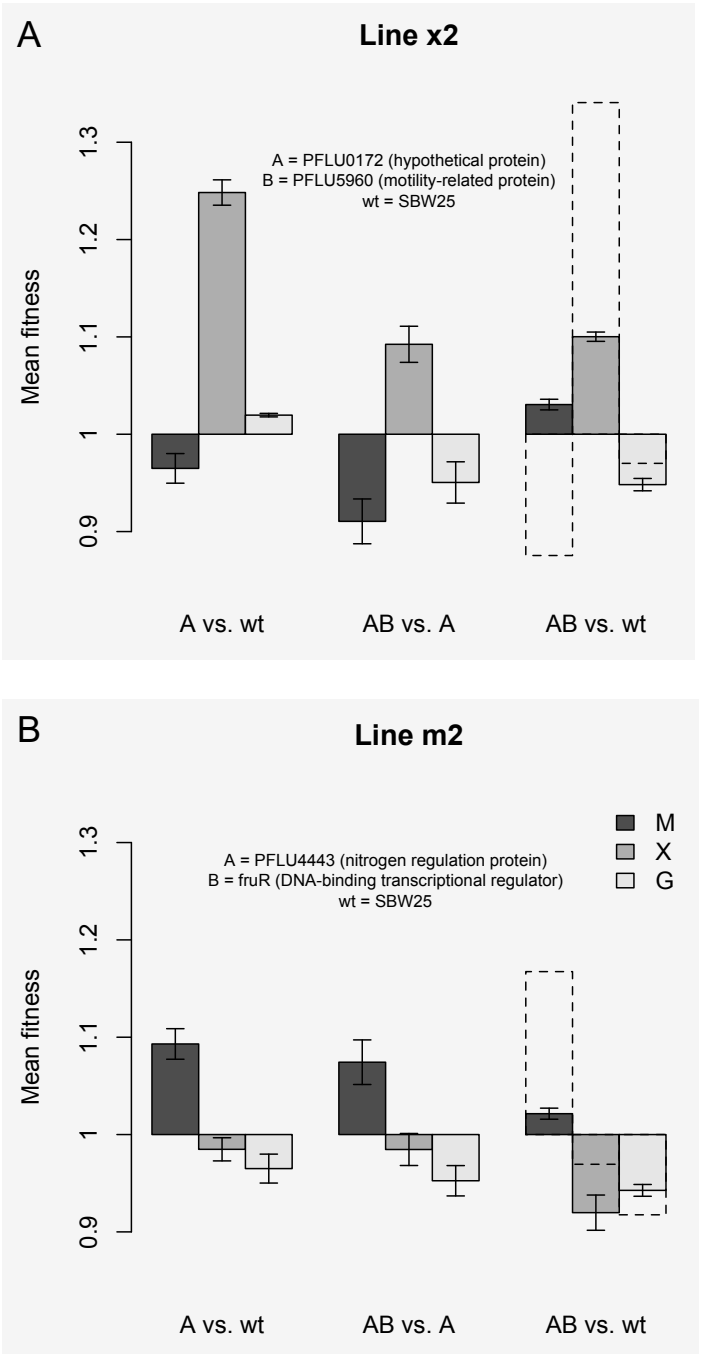

Figure 2.5: Effect of multiple mutations on specialization. Mean fitness of beneficial mutations in a xylose-selected population (line x2, panel A) and a mannose-selected population (line m2, panel B), assayed by competition in mannose (M), xylose (X) and glucose (G). Dashed lines show expected fitness effect if the fitness effects of the two mutations are additive.

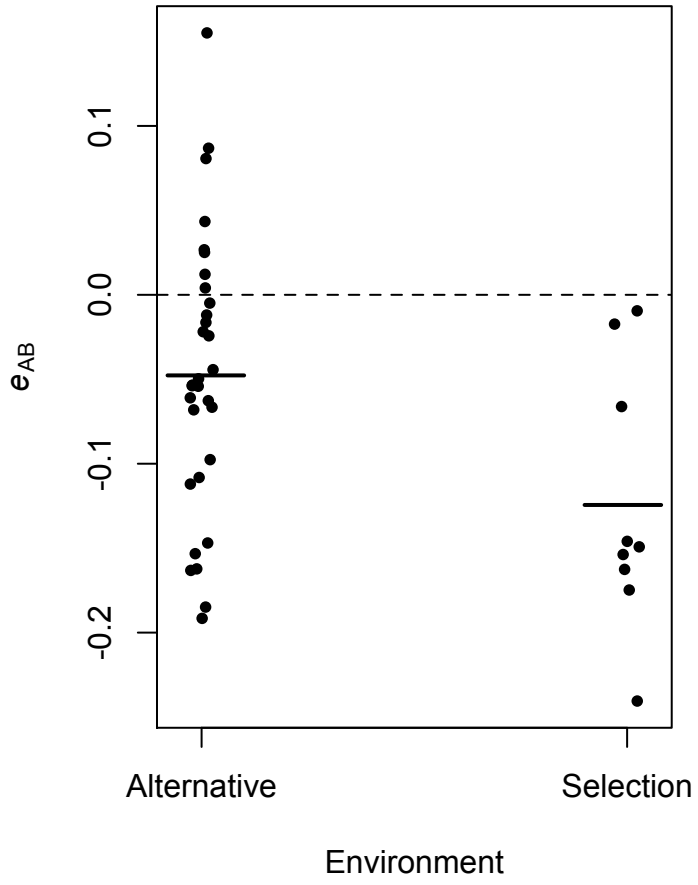

Figure 2.6: Non-additive fitness effects. The difference between the observed and expected fitness effects of two mutations, given the expectation of additive fitness. Horizontal bars denote the mean.

2.4.6 Specialization among functional classes of mutation

Based on functional annotations identified in our previous work (BAILEY *et al.* 2015), we identified three broad classes of mutations in our strain collection: those occurring in genes associated with motility, regulatory functions, and an other category of as yet uncharacterized genes. Mutations in genes impacting motility, which were most often expected to be loss-of-function mutations, had a lower variance in fitness across environments than mutations affecting regulatory processes or whose function could not be predicted (one-way ANOVA, $F_{2,20} = 14.605$, $p < 0.001$; Figure 2.7).

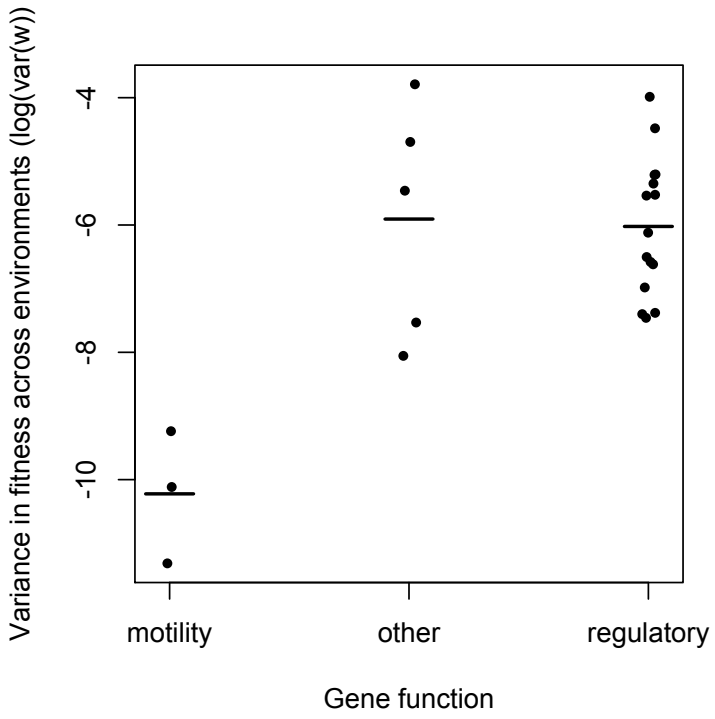

Figure 2.7: Variance in mean fitness effects across environments. Mutations are categorized by the function of the gene they are located in, with genes of unknown function grouped in other. Variance is calculated for each beneficial mutation using the mean competitive fitness assayed in all three single-resource environments (M, G and X).

2.5 Discussion

Local adaptation is thought to evolve due to strong divergent selection that leads to niche specialization underlain by trade-offs in fitness across environments. We have used a combination of experimental evolution and whole-genome sequencing to provide insight into the genetic mechanisms causing trade-offs. To examine the genomics of adaptive evolution in more detail, we have directly estimated the fitness effects of putatively adaptive mutations in both the environment of selection and in alternative environments, and examined the effects of multiple mutations on fitness and specialization. Our leading result is that specialization in this system evolves because beneficial mutations in the environment of selection have increasingly antagonistic effects in alternative environments (Figure 2.4), consistent with a simple extension of Fisher's geometric model of adaptation to multiple optima (MARTIN and LENORMAND 2015). While non-additive interactions among multiple mutations can sometimes exaggerate the degree of specialization, they often seem to lessen it. Finally, we find no evidence that conditionally neutral mutations contribute to specialization in this system. Below we discuss the implication of these results in more detail.

2.5.1 The dynamics of genome evolution

The dynamics of genomic evolution in our experiment included both classic periodic selection, the sequential substitution of beneficial mutations, and clonal interference, where multiple beneficial mutations compete for fixation. Whether a lineage evolves via periodic selection or clonal interference should depend on the mutation supply rate, which is set by the combination of the genome-wide mutation rate to beneficial mutations, U_b, and population size, N. When $N \times U_b << 1$, periodic selection is expected to dominate whereas clonal interference dynamics should prevail when the reverse is true. However this simple explanation does not seem to explain our results. Even when the mutation supply rate was identical, as it must have been for the lineages shown in Figure 2.2 and Figure 2.3, both of which were evolved in mannose and were founded from the same ancestral strain, replicate lineages showed patterns of adaptation consistent with both kinds of dynamics. The cause of this variation is not clear. It is possible that these populations are evolving on or near the point at which $N \times U_b$ ~1 such that in some instances the dynamics of genomic evolution resemble periodic selection and in others it is closer to clonal interference. This interpretation gains some, admittedly cautious, support from what we know about the mutation supply rate in this system. KASSEN and BATAILLON 2006 previously estimated the beneficial mutation rate for a single target of selection to be ~$8 \times 10^{-}11$ mutations per cell division and BAILEY and KASSEN 2012 estimated the effective population size in the current experiment to be ~3×10^7 CFU ml^{-1}. *P. fluorescens* has ~6700 genes, not all of which will be under selection in this environment. If we conservatively (and arbitrarily) assume that between 100-1000 sites could be under selection, then this corresponds to a mutation supply rate of approximately 0.24-2.4 mutations per generation across the entire genome, values not too far off of the expected value of $N \times U_b$ ~1. While our assumption of the number of sites under selection can be criticized as an overestimate, seeing as though we found ~50 sites mutated for the entire experiment, we note that there are probably many more mutations our sequencing protocol will have missed because they arose earlier in the experiment but were not present, or were very rare, in the endpoint populations. It therefore does not seem unreasonable to suggest that the differences in genomic dynamics we observe stem from the fact that these populations are evolving in a region of parameter space near the boundary of where we expect to differentiate between periodic selection and clonal interference.

2.5.2 The genetic causes of specialization

Our results provide compelling evidence that the main genetic cause of specialization in this system stems from the changing nature of the pleiotropic fitness effects of beneficial mutations. The evidence for this interpretation stems from four observations. First, we could find no indication that conditionally neutral mutations contributed to specialization. All the mutations we uncovered in our experiment were beneficial when competed against their immediate genetic ancestor and, moreover, carried detectable fitness consequences in other environments. Second, we found that the mean pleiotropic fitness effect in the alternative environments of each step in adaptation was less than that in the environment of selection. This result is consistent with the idea that most beneficial mutations have environment-specific fitness effects and that the gain in fitness associated with adaptation to a focal environment is greater than that to an alternative environment, a necessary condition for the evolution of a trade-off (KASSEN 2014). Third, the mean pleiotropic fitness

effects in the alternate environments changed in a manner suggesting that pleiotropy evolves to become more antagonistic as adaptation proceeds. More specifically, the first mutation substituted had a mean pleiotropic fitness effect significantly greater than one, while the second mutational step did not, fitness being on average indistinguishable from one and including many individual mutations with antagonistically pleiotropic effects. The direction of these changes is consistent with that predicted from a multiple-optima extension of Fisher's geometric model: pleiotropic fitness effects evolve from being positive when the starting population is far from the different optima to being less positive, and even occasionally antagonistic, as the population adapts (MARTIN and LENORMAND 2015). Notably, JASMIN and ZEYL 2013 saw a similar result in evolving yeast populations. Fourth, while we found strong evidence of less than additive fitness effects in the selection environment, we also found inconsistent effects among multiple mutations on the degree of specialization, sometimes enhancing specialization and sometimes reducing it. We suspect that this pattern is likely caused by epistasis between the beneficial mutations, where the fitness effect of multiple mutations is less (or more) than the sum of each mutation individually. Consistent with this view, our observation that non-additive fitness effects among mutations are consistently negative is in line with previous observations of negative epistasis for fitness during adaptation to a single environment in other systems (CHOU *et al.* 2011). In alternative environments we have no reason to expect a particular pattern of non-additive fitness effects, where mutations have not been tested by selection, leading to the large range of non-additive fitness effects we observed in those environments.

2.5.3 Genes responsible for specialization

Our ability to make strong inferences about the functional properties of genes associated with specialization is limited. The molecular genetics of *P. fluorescens*, or pseudomonads in general, is less well understood than in other microbial model systems such as Escherichia coli or yeast. Nevertheless, we are able to distinguish two broad functional classes in of mutation in our study, those that result in a putative loss-of-function in genes impacting motility and those affecting gene regulation.

The selection of mutants that lack motility is not surprising in light of how we do our experiments: our founder strain, which is flagellated and motile, is propagated in shaken liquid cultures where motility is unnecessary and likely costly. We might expect, then, that non-motile mutants that avoid paying the cost of manufacturing and maintaining a competent motility system would be selected. Moreover, these mutants were among the most broadly adapted to have evolved, having the lowest environmental variance of fitness among any of the mutations identified. The breadth of adaptation associated with these mutations is not surprising, given that the loss of motility should be adaptive under all conditions studied in our experiment. The loss of motility therefore constitutes an example of domestication resulting from adaptation to the general conditions of growth associated with laboratory culture.

The selection of mutations impacting gene regulation is also not entirely unexpected. DETTMAN *et al.* 2012, for example, pointed out that it is not uncommon in microbial selection experiments to see mutations arising that restore the ability of a cell to grow and reproduce in conditions that would otherwise be too stressful for them to do so. It is tempting to interpret the functional effect of all or some of these mutations in this light, however a more rigorous test of this hypothesis is beyond the scope of this paper and will have to

await a better characterization of how these genes work. Notably, mutations in this category have a range of effects on the environmental variance of fitness for reasons that are similarly hard to parse in the absence of additional knowledge of mechanisms. A fruitful avenue for future research will be to connect the mutational effects on fitness, the degree of specialization, and the molecular biology of gene function.

2.5.4 The multiple-optima extension of Fisher's model

The multiple-optima extension of Fisher's model (MARTIN and LENORMAND 2015) is an attractive interpretation of the pleiotropic fitness effects of beneficial mutations because it seems to account for a number of results commonly observed in experimental evolution. The fitness effects of single beneficial mutations are often not strongly antagonistic when measured under a wide range of environments, for example (BATAILLON *et al.* 2011; OSTROWSKI *et al.* 2005). Many microbial selection experiments also find that adaptation to one set of conditions does not result in a loss of fitness, relative to the ancestor from which the evolved lines are derived, in other conditions (reviewed in Kassen 2014). To the extent that these results are generally true of what happens in nature when a population gains access to a novel environment to which it is not well adapted, we would expect that adaptation to that environment will not, initially, at least, involve strongly antagonistically pleiotropic mutations.

2.5.5 Caveat to competitive fitness measures

It is important to note that these estimates of fitness associated with beneficial mutations were not measured in an environment exactly identical to the one in which those mutations arose and that may have a non-trivial effect on fitness. For example, in line X, we know that mutation A arose in the background of mutation B, therefore we can measure the fitness of the double mutant relative to the single mutation by directly competing the two strains in equal proportions. However, because it is unlikely that there are other polymorphisms present in the population at the time the second mutations arises, this fitness measured from this type of competition experimental may not accurately reflect the fitness increase associated with the natural occurrence of the mutation. It would be interesting to measure how sensitive the fitness effects of these mutations are to the competition environment. For example, one could either vary the initial frequencies of both competitors (to get an indication of frequency dependence) or add a third competitor to the mix. We predict the results would be highly mutation-dependent. Some mutations, especially those associated with functions like metabolic regulation, will have environment-dependent fitness effects, and other mutations, especially those with motility related functions will have fitness effects that are less environment-dependent.

2.5.6 Conclusions

Despite the increasing speed and decreasing cost of genome sequencing, the insight gained with regards to the genetics of local adaptation are limited without further examination of the specific loci involved. To understand the genetic trade-offs involved in adaptation, it is important to go beyond identifying mutations by characterizing the properties of those mutations through direct measurements of fitness. We have shown here that trade-offs arise because beneficial mutations arising later in an adaptive walk become increasingly

deleterious in alternative environments, as opposed to being conditionally neutral. It will be interesting to see if similar results are found for different ranges of environments or in applied contexts, such as adaptation to the human gut, lung, or to novel antibiotics.

Chapter 3

Ecological factors that drive phenotypic diversification

A version of this chapter has been published as "Schick, A. and Kassen, R. 2018. Rapid diversification of *Pseudomonas aeruginoas* in cystic fibrosis lung-like conditions. PNAS." I conceived the project with advice from R. Kassen. I preformed the laboratory work and performed the analysis. I prepared the initial draft of the manuscript and R. Kassen and I both contributed major revisions.

3.1 Summary

Chronic infection of the cystic fibrosis (CF) airway by the opportunistic pathogen *Pseudomonas aeruginosa* is the leading cause of morbidity and mortality for adult CF patients. Prolonged infections are accompanied by adaptation of *P. aeruginosa* to the unique conditions of the CF lung environment, as well as marked diversification of the pathogen into phenotypically and genetically distinct strains that can coexist for years within a patient. Little is known, however, about the causes of this diversification and its impact on patient health. Here, we show experimentally that, consistent with ecological theory of diversification, the nutritional conditions of the CF airway can cause rapid and extensive diversification of *P. aeruginosa*. Mucin, the substance responsible for the increased viscosity associated with the thick mucus layer in the CF airway, had little impact on within-population diversification but did promote divergence among populations. Furthermore, *in vitro* evolution recapitulated traits thought to be hallmarks of chronic infection, including reduced motility and increased biofilm formation, and the range of phenotypes observed in a collection of clinical isolates. Our results suggest that nutritional complexity and reduced dispersal can drive evolutionary diversification of *P. aeruginosa* independent of other features of the CF lung such as an active immune system or the presence of competing microbial species. We suggest that diversification, by generating extensive phenotypic and genetic variation on which selection can act, may be a key first step in the development of chronic infections.

3.2 Introduction

Pseudomonas aeruginosa is a globally ubiquitous, metabolically flexible Gram-negative opportunistic pathogen. While it is an important nosocomial pathogen causing a range of acute infections, *P. aeruginosa* is also commonly recovered from the airways of adult cystic fibrosis (CF) patients where it causes chronic endobronchial infections in the majority of adult patients and is the leading cause of morbidity and mortality in this population (HOIBY and PRESSLER 2006; RAJAN and SAIMAN 2002; SCHAEDEL *et al.* 2002). The majority of

chronic CF lung infections are thought to be the result of colonization by *P. aeruginosa* from environmental sources, although highly transmissible epidemic strains are responsible for 25% of infections in Canadian patients (AARON *et al.* 2010). Once established, chronic infections can remain persistently associated with a host for decades, being virtually impossible to eradicate with standard antibiotic therapy (GIBSON *et al.* 2003).

The transition from environmental strain to one that causes chronic infection is characterized by a few repeatable phenotypic changes underlain by a much larger suite of genetic changes. Free-living environmental strains are typically motile, virulent, and nonmucoid whereas strains isolated from chronically infected CF patients are often non-motile, avirulent, mucoid, and highly antibiotic resistant (CIOFU *et al.* 2010; MOWAT *et al.* 2011; POOLE 2005; SMITH *et al.* 2006; SOUSA and PEREIRA 2014). These phenotypes represent parallel adaptations to a range of CF lung-specific stressors including increased viscosity, osmotic stress, low oxygen, high concentrations of antibiotics and immune system attack. This parallelism notwithstanding, the other striking feature common to *P. aeruginosa* isolates from CF patients is their diversity. Phenotypes resembling both environmental and chronically-infected strains can persist in the same patient for long periods of time and isolates from within (ASHISH *et al.* 2013; CLARK *et al.* 2015; DARCH *et al.* 2015; FOTHERGILL *et al.* 2010; JORTH *et al.* 2015; MOWAT *et al.* 2011; WORKENTINE *et al.* 2013) and among (DETTMAN *et al.* 2013; WILLIAMS *et al.* 2015) patients can be highly diverse, both phenotypically and genetically. This diversity makes it difficult to identify reliable markers for the onset of chronic infection (WINSTANLEY *et al.* 2016) and predict clinical outcomes on the basis of single colony isolates alone. Indeed, the lack of correlation between the prevalence of a particular phenotype and clinical outcomes (WORKENTINE *et al.* 2013) hints that the diversity of the population, rather than the abundance of any particular phenotype, could be what makes CF lung infections by *P. aeruginosa* so recalcitrant to treatment (DARCH *et al.* 2015).

Diversity in the CF lung typically evolves rapidly and *de novo* following colonization (DARCH *et al.* 2015; FOLKESSON *et al.* 2012; JORTH *et al.* 2015; WORKENTINE *et al.* 2013), with genetically and phenotypically diverse clones often persisting for years within the same host (CIOFU *et al.* 2012; FELIZIANI *et al.* 2014; SOUSA and PEREIRA 2014). This dynamic has many of the hallmarks of an adaptive radiation, the rapid diversification of a lineage into a range of niche specialist types. Both theory (LEWONTIN *et al.* 1974; NEVO 1978; WHITTAKER and LEVIN 1975) and experiment (reviewed in FUTUYMA and MORENO 1988; KASSEN 2009, 2014; SCHLUTER 2000) suggest that adaptive diversification is most likely to occur when divergent selection, imposed by ecological opportunities in the form of vacant niche space or underutilized resources, is strong relative to the rate of dispersal. The extent of phenotypic divergence may be further exaggerated, or its rate accelerated, through ecological interactions such as resource competition or predation (DOEBELI and DIECKMANN 2000; FELSENSTEIN 1981; FRIESEN *et al.* 2004; ROZEN and LENSKI 2000).

We suspect that the complex ecological conditions of the CF lung likely promote diversification in colonizing *P. aeruginosa*. The CF airway contains a rich spectrum of resources and nutrients that could provide ample ecological opportunity to drive specialization (MARVIG *et al.* 2014; PALMER *et al.* 2007; WORKENTINE *et al.* 2013) and the thick mucus layer, a by-product of the impaired ability to transport chloride ions resulting from mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene, serves to reduce dispersal among subpopulations. Dispersal will be further reduced due to the highly compart-

mentalized nature of the lung itself, with distinct patches or habitats such as right and left lobes, upper and lower respiratory tract, and the alveoli and bronchioles on different branches of the respiratory tree. The resource profile and spatial structure of the CF lung thus provide ideal conditions for strong divergent selection to drive diversification. Additional aspects of life in the CF lung such as the emergence of hypermutator lineages (FELIZIANI *et al.* 2014; OLIVER 2000), high levels of antibiotic use (WRIGHT *et al.* 2013), and interactions between *P. aeruginosa* and co-infecting species (KORGAONKAR *et al.* 2013; SIBLEY *et al.* 2008), bacteriophages (BROCKHURST *et al.* 2005; JAMES *et al.* 2015), and the host immune system (JENSEN *et al.* 2010; MCKEON *et al.* 2010) may further contribute to the genetic and phenotypic variability seen among isolates from chronically infected patients.

Here, we evaluate the contributions of resource complexity and reduced dispersal to *P. aeruginosa* diversification by tracking the extent of phenotypic diversification within and among independently evolved populations descended from *P. aeruginosa* strain Pa14 after ~220 generations of selection in environments that varied in how closely they resemble the resource profile and viscosity of the CF lung. The most CF-like environment consisted of the nutritionally complex synthetic cystic fibrosis medium (SCFM), a defined medium based on the free amino acid, anion, cation, and carbon source profiles from CF sputum (PALMER *et al.* 2007), supplemented with mucin, the major protein component of mucus, to mimic the thick mucus layer in the lumen of CF patients. Several studies have shown that the presence of mucin increases the viscosity of the media and reduces the motility of *P. aeruginosa* (FUNG *et al.* 2010; LANDRY *et al.* 2006; WANG *et al.* 1996) and is therefore very likely to reduce dispersal. The least CF-like environment was composed of a minimal medium supplemented with glucose as the sole carbon source in the absence of mucin. The experimental conditions consisted of factorial combinations of two levels of nutrient complexity (minimal medium versus SCFM) and viscosity (with or without mucin). Phenotypic divergence among independently evolved replicate populations provided an estimate of the contribution of spatial compartmentalization to diversification. We scored both colony morphology and ten traits thought to be associated with patho-adaptation tied to chronic infection of the CF lung (growth rate in LB, pyocyanin and pyoverdine production, biofilm formation, swim and twitch motility, and resistance to four classes of antibiotics: ciprofloxacin, ceftazidime, colistin, and tobramycin) for 12 randomly sampled isolates from each of 12 independently evolved populations within each treatment (for a total of 576 isolates) as well as a collection of 24 *P. aeruginosa* isolates from the lungs of CF patients from across Ontario (DETTMAN *et al.* 2013) to gauge the extent to which our experimental conditions recapitulate the markers of phenotypic divergence associated with clinical strains.

3.3 Materials and Methods

3.3.1 Bacterial strains and growth conditions

The clinical isolate of *Pseudomonas aeruginosa*, Pa14, originally isolated from a human burn wound, was used as the founder strain for the selection experiment. Cultures were grown overnight in Luria Bertani broth (LB) and plated on LB agar to isolate individual colonies. Replicate populations were founded from distinct, single colony isolates from these plates and grown at 37°C in unshaken 24-well plates. All other clinical

isolates used in this study were collected during a study of adult patients attending cystic fibrosis clinics in six cities across Ontario, Canada, between September 2005 and September 2008. See AARON *et al.* 2010; DETTMAN *et al.* 2013 for details.

3.3.2 Selection experiment

We propagated 30 replicate populations in each of four selection environments (120 populations in total) by daily serial transfer for approximately 220 generations. Every 24 hours, 15 μL of overnight culture was transferred into 1.5 mL of fresh liquid media. We conducted a factorial experiment involving two levels of nutrient complexity (MIN: M9 minimal salts medium supplemented with 0.7% glucose; SCFM: synthetic cystic fibrosis medium, a defined medium resembling the nutritional conditions of the CF lung; see PALMER *et al.* 2007) crossed with two levels of viscosity: supplemented or not with 50 mg/mL of porcine mucin following WONG *et al.* 2012.

3.3.3 Morphological and phenotypic diversity of endpoint populations

Aliquots from frozen archives were first revived in their respective selection environment for 24 hours at 37°C and then plated on LB plates to assess colony morphology. A single observer visually identified the number of distinct colony morphs from at least 100 colonies from each population based on eight characteristics (pigmentation, opacity, iridescence, surface texture, margin, halo, autolysis, and small colony variant). Additional phenotypes, described in detail below, were measured for each of 12 random colonies from 12 random populations of each treatment for a total of 576 isolates. All phenotype assays were performed in triplicate unless specified otherwise.

3.3.4 Motility assays

Swimming and twitching motility were measured as described in CLARK *et al.* 2015. Briefly, swim motility was assayed by stab-inoculating a sample into 0.3% LB agar plates, incubated for 24 hours at 37 degrees C, and their zone of growth measured. For twitch motility, colonies were stab inoculated into 1.5% LB agar and incubated for 72 hours (37°C for 48 hours then room temperature for 24 hours). The agar was then removed from the petri dish, and the zone of growth was measured by staining the dish with 1% crystal violet.

3.3.5 Pyocyanin and pyoverdine production

Cultures were first grown in LB for 24 hours and then pigment production quantified by measuring the optical density (OD) of the supernatant at 405 nm and 695 nm for pyoverdine and pyocyanin, respectively, and standardized by cell density estimated from OD at 600 nm before centrifugation and extraction of the supernatant.

3.3.6 Biofilm formation

Biofilm formation was assayed as described by O'TOOLE 2011. Cultures were diluted 1:100 in fresh LB, seeded into 4 replicate wells of a 96-well non-tissue culture treated microtitre plate and incubated for 24 hours at 37°C under static conditions. Plates were then rinsed and stained with 125 uL of a 0.1% solution of crystal violet in water. After incubating at room temperature for 15 minutes, the plates were rinsed several times with water and allowed to dry for 48 hours. The stain was then dissolved in 125 μL of 30% acetic acid in water and incubated at room temperature for 15 minutes. Biofilm formation was quantified by measuring the optical density of this final solution at 550nm.

3.3.7 Antimicrobial resistance

We used a two-fold dilution series for each of four antibiotics (ciprofloxacin, ceftazidime, tobramycin, and colistin) in LB media to estimate minimum inhibitory concentration (MIC; the lowest concentration of drug that prevents growth) for each evolved isolate and the founding strain. Growth was assessed at 24 hours and 48 hours by measuring optical density at 600 nm and the MIC determined as the drug concentration where OD was less than 0.100. In cases where growth was observed at all concentrations of antibiotic, the MIC was reported, conservatively, as the next highest concentration in the dilution series.

3.3.8 Statistical analysis

All statistical analyses were performed in R, version 3.3.0 [ref: R]. All phenotypic traits were treated as continuous variables, with antibiotic resistance traits analyzed on a logarithmic scale. To determine the effect of media, mucin, and the interaction on each phenotype individually, we used the R package nlme to construct linear mixed effects models with population modeled as a random effect and media and mucin as fixed effects. We performed post hoc pairwise t-tests, Holm-adjusted for multiple comparisons, to evaluate the statistical significance among treatments. Trait values of evolved populations were compared to those of the unevolved founder using one-sample t-tests within each treatment and corrected for multiple comparisons (40 in total) using a Bonferroni corrected p-value of $\alpha = 0.0025$. Diversity within a population was determined by calculating the Euclidean distance on standardized z-scores for each trait between all possible pairs of isolates. Diversity between populations was determined by calculating the Euclidean distance between the means of all populations within a treatment group for all possible pairs of populations. A randomization test was performed to determine significance by resampling treatment labels and calculating a null distribution of F-statistics. The R package ggplot2 was used to generate plots of Euclidean distance. To determine the distance between the clinical strains and the evolved isolates, we calculated the Mahalanobis distance between each clinical strain and the distribution of treatment group. Significance was tested using the Hotelling T2 test. Correlations between phenotypes were determined by calculating Spearman correlation coefficients and visualized using the R package gplots.

Table 3.1: Summary of phenotypic evolution. Arrows represent significant deviations from the ancestral value in each treatment group, Bonferroni corrected for multiple comparisons. Antibiotic resistance was measured by minimum inhibitory concentration (MIC), the lowest concentration of drug at which growth was inhibited. See Table B.1 for test statistics and p-values.

Trait	SCFM	MIN	SCFM+mucin	MIN+mucin
Growth density	-	-	-	-
Pyoverdine production	↓	↓	↓	↓
Pyocyanin production	-	-	↑	-
Biofilm formation	↓	↓	↑	↓
Swim motility	↓	↓	↓	↓
Twitch motility	↓	↓	↓	↓
MIC ciprofloxacin	↑	-	↑	↑
MIC ceftazidime	-	-	↑	↑
MIC colistin	-	-	-	↓
MIC tobramycin	-	-	-	-

3.4 Results and Discussion

3.4.1 Markers of chronic infection evolve in the most lung-like environments

Adaptation to the CF lung is often accompanied by a suite of phenotypic changes that include loss of virulence factors, loss of motility, and increased biofilm formation. We determined the extent to which similar phenotypic changes evolved in our in vitro experiment by scoring a range of traits associated with CF lung adaptation and using a linear mixed effects model with media and mucin as fixed effects and population as a random effect to examine the impact of different treatment combinations on trait evolution. Our results are summarized for all traits in Table 3.1 (and Table B.1) and for four traits commonly associated with CF lung adaptation in Figure 3.1.

We observed marked trait evolution, including a reduction in pyoverdine production (the main siderophore and measure of iron-scavenging ability; Figure 3.1A), swim motility (Figure 3.1C), and twitch motilty (Figure 3.1D), and an increase in biofilm production (Figure 3.1B), in the most CF lung-like conditions, all changes that mirror those seen during adaptation to the CF lung. These phenotypic changes persist when colonies are subcultured, implying that they have a genetic basis. Given that our experiment was conducted in vitro without any competing microflora or an active immune system, these results suggest that selection driven by the combined effects of resource complexity and viscosity is sufficient to explain changes in these putatively pathoadaptive traits characteristic of *P. aeruginosa* infections of the CF lung.

This interpretation might be questioned on the grounds that many of these trait changes, especially those involving a decrease in trait value relative to the common ancestor (pyoverdine production, swim motility, and twitch motility), reflect general adaptation to in vitro culture conditions rather than specific adaptation to CF-like environments. This explanation cannot, however, explain the treatment-specific changes in these and other traits (Tables B.1 and B.2). Pyoverdine production, for example, declined less in the presence of

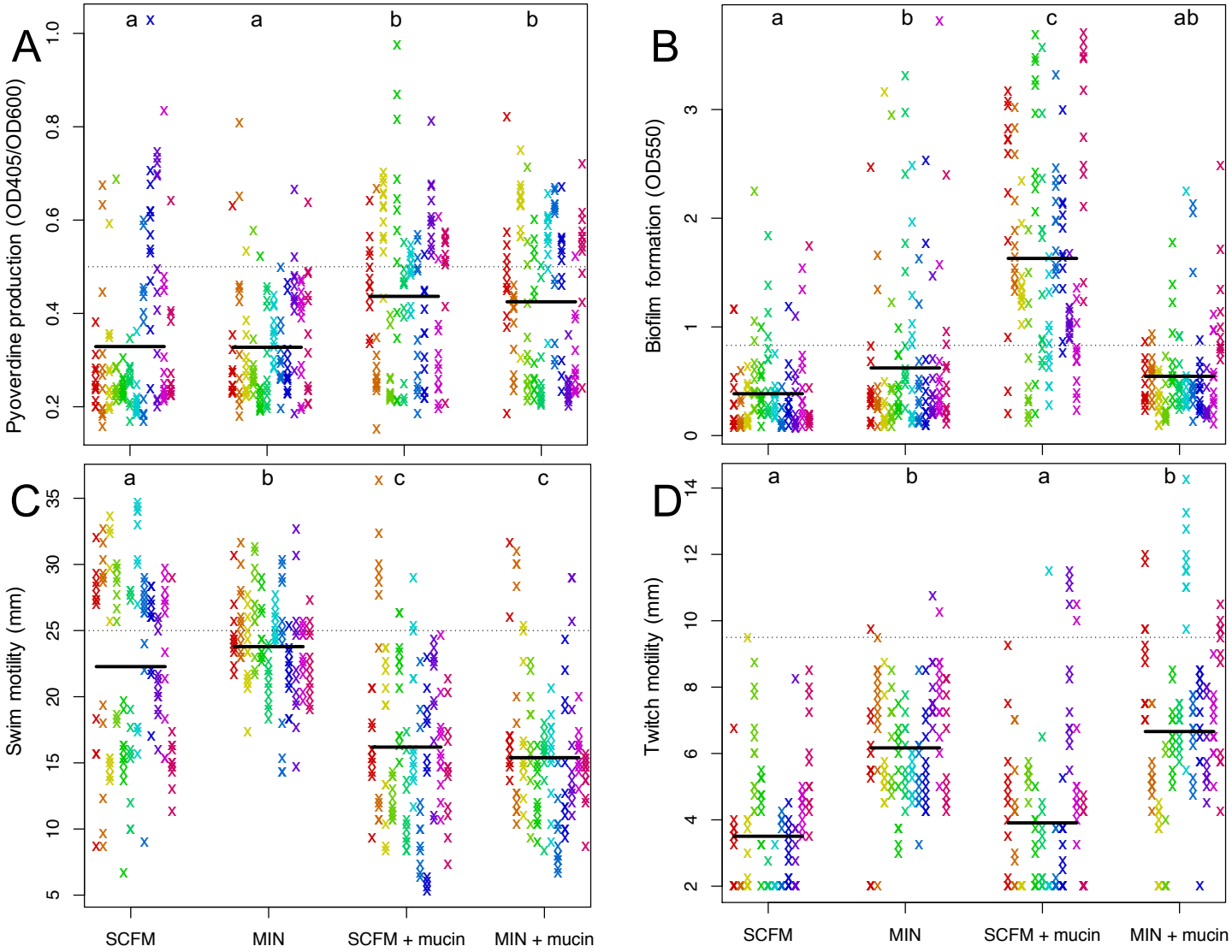

Figure 3.1: Phenotypic adaptation of evolved isolates for pyoverdine production (A), biofilm formation (B), swim motility (C), and twitch motility (D). Within each treatment, each column/colour represents a replicate evolved population (n = 12 populations for each treatment). Solid lines represent treatment means, and dashed lines represent the value of that phenotype in the ancestral strain (Pa14). Lowercase letters across the top of each panel denote significantly different groups.

mucin than in its absence, while swim motility showed the reverse trend. Changes in twitch motility, on the other hand, were independent of mucin and decreased more in the nutrient-rich conditions of SCFM than in minimal medium. Pyocyanin and biofilm production both increased in the most CF lung-like conditions (SCFM+mucin), with either no change (pyocyanin) or decreases (biofilm) in the other treatments. The increase in pyocyanin (Figure B.1), a virulence factor, was unexpected in light of the commonly observed loss of virulence factors among chronic CF isolates. Nonetheless, it has been noted that pyocyanin production is increased in isolates from early stage infections and decreases as disease severity progresses (FUNG *et al.* 2010), suggesting that reduced pyocyanin production in isolates from chronic infections evolves as a by-product of prolonged infection or in response to an aspect of ecology that we did not investigate, with immune system attack or competition from other species being the most obvious candidates. Taken together, such treatment-specific responses to selection suggest that our results represent trait changes specifically associated with CF lung-like conditions. We note that while this may not suggest that the laboratory environment recapitulates the conditions of the human lung, it does demonstrate that the selection pressures driving these parallel phenotypic changes are recapitulated in the laboratory.

Intriguingly, our results also suggest that antibiotic resistance can evolve in these populations as a pleiotropic effect of adaptation, a phenomenon observed in several bacterial species (HERSHBERG 2017). While a lack of evolutionary change in resistance to antibiotics such as colistin and tobramycin was unsurprising, as no drugs were used in the selection experiment, it was surprising to see a marked increase in ciprofloxacin resistance across all conditions except the simplest environment [M9 minimal salts medium supplemented with 0.7% glucose (MIN)] and an increase in ceftazidime resistance in environments containing mucin (Table 3.1). Although resistance is commonly observed following antibiotic treatment in clinical settings, our results suggest that both ciprofloxacin and ceftazidime resistance can evolve as a pleiotropic by-product of adaptation to CF lung-like conditions. Spontaneous resistance to antibiotics has been shown to arise via mutations to the MexGHI-OpmD efflux pump that are also associated with biofilm production (SAKHTAH *et al.* 2016). This mechanism seems unlikely to explain the increases in resistance in our experiment because we did not observe the expected positive relationship between biofilm formation and resistance for either drug class, with the correlation being significantly negative for ciprofloxacin and indistinguishable from zero for ceftazidime (Figure B.3). Interestingly, a recent study found that resistance to ceftazidime arose in the absence of this antibiotic as a result of an increase in beta-lactamase production (DAVIES *et al.* 2017), a mechanism that warrants further exploration.

We did not observe any mucoid individuals in our evolved populations, despite this being a common phenotype of lung- evolved isolates that can often be associated with aggregation and biofilm formation. If mucoidy represents an adaptation to prolonged infection, it is possible that we would have observed it in a longer experiment. An alternative explanation is that mucoidy evolves in response to selection for some other factor not captured in our experiment. The mucoid colony morphology is often linked to the overproduction of alginate (MARVIG *et al.* 2015), a compound that can also effectively scavenge reactive oxygen species (HASSETT *et al.* 2009; SIMPSON *et al.* 1989). If mucoidy is a pleiotropic result of adaptation for tolerance to oxidative burst from macrophages (DETTMAN *et al.* 2013) or neutrophils (MAKAM *et al.* 2009), for example, we would not expect to see it evolving in our experiment.

3.4.2 Within-population phenotypic diversity is driven by resource complexity

We observed extensive phenotypic diversity in all evolved populations. To capture the general trends associated with multivariate phenotypic diversification within and among populations, we used all 10 traits described above to calculate the Euclidean distance between pairs of isolates within a population (3.1) and between the means of pairs of populations within each treatment (3.2) as follows:

For isolates,

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{3.1}$$

For populations,

$$d(a,b) = \sqrt{\sum_{i=1}^{n}(\bar{x}_{a,i} - \bar{y}_{b,i})^2} \tag{3.2}$$

Here, x_i, y_i are standardized z-scores for trait i of individuals x, y, respectively, and $x_{a,i}$, $y_{b,i}$ are standardized z-scores for trait i of an individual in populations a,b, respectively. Note that calculating among-population distance by using the population means controls for any effect of variation within populations on the measure of among-population distance.

Our results, shown in Figure 3.2, are striking. Populations evolved in the complex nutritional environment of SCFM supported higher amounts of within-population diversity than those evolved in the simpler nutritional environment of minimal medium with glucose as the sole carbon and energy source (Figure 3.2A; two-factor ANOVA, $F_{1,44}$ = 10.21, p = 0.003), independent of the presence of mucin ($F_{1,44}$ = 0.65, p = 0.423). Moreover, we found no significant interaction between resource complexity and mucin ($F_{1,44}$ = 0.79, p = 0.786). Our results contradict the prevailing view that the within-host diversification by *P. aeruginosa* accompanying chronic infection is driven primarily by restricted dispersal due to the thick mucus layer in the lung lumen. In evolutionary terms, the nutritional complexity of the lung likely represents abundant ecological opportunity for colonizing *P. aeruginosa* that generates strong divergent selection, leading to phenotypic divergence. Competition for resources among incipient niche specialists in the evolving population could further accelerate phenotypic diversification and contribute to coexistence, perhaps through negative frequency-dependent selection (reviewed in KASSEN 2014; RAINEY and COOPER 2004), leading to long-term persistence of phenotypically divergent *P. aeruginosa* variants within a single host.

3.4.3 Nutritional complexity and spatial structure cause populations to diverge

P. aeruginosa isolates from distinct, spatially separated compartments within the CF airway are often genetically and phenotypically distinct, a result that has been taken to imply niche-specific adaptive diversification following colonization (MARKUSSEN *et al.* 2014). An alternative interpretation is that spatial separation of subpopulations in distinct lung compartments leads to genetic and phenotypic divergence, independent of niche-specific adaptation due to stochastic effects in the timing and order of mutations contributing to adap-

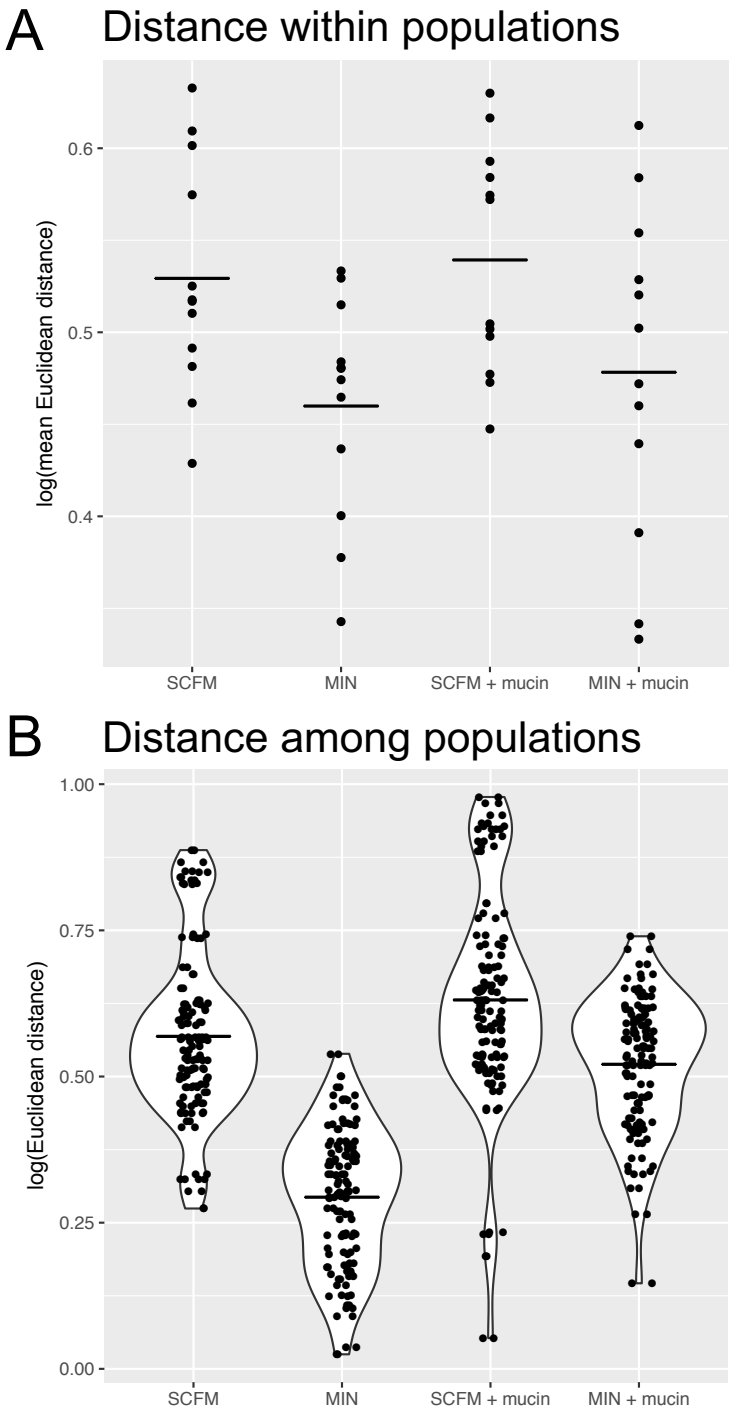

Figure 3.2: Euclidean distance within (A) and among (B) populations. (A) Each point represents a population with the distance within that population determined by the average of all pairwise comparisons of individuals within that population. (B) Each point represents the distance between the mean values of two populations. Solid lines represent treatment means. All distance calculations are Euclidean distance based on all 10 z-score-transformed phenotypic traits.

tation. Our results allow us to quantify the extent of phenotypic divergence among replicate evolved populations within each treatment in our experiment as a proxy for the extent of among-compartment divergence. Both nutritional complexity and mucin had statistically significant effects on among-population divergence (randomization test where treatment labels were resampled and a null distribution of F statistics calculated; p-values were <0.0001, <0.0001, and 0.0048 for medium, mucin, and the interaction, respectively, from a two-factor ANOVA), with the SCFM+mucin treatment displaying the most extensive between-population diversity, followed by the SCFM and MIN+mucin treatments and the MIN treatment (Figure 3.2B). Since replicate populations within a treatment evolved, by design, under the same selective pressure, these results lend support to the idea that mutation-order effects in the most CF-like conditions can support substantial among-population divergence of multivariate phenotypes.

3.4.4 Comparison of evolved populations and clinical isolates from CF patients across Ontario

To what extent do the conditions in our experiment recapitulate the range of variation associated with *P. aeruginosa* isolates from CF patients? To answer this question, we measured the same 10 phenotypic traits in each of 24 *P. aeruginosa* strains isolated from the lungs of CF patients from across the Canadian province of Ontario, a collection that included both nonepidemic and epidemic strains (AARON *et al.* 2010). We then calculated the Mahalanobis distance between each clinical strain and the multivariate distribution of all evolved isolates from each treatment group, separately, to obtain a measure of phenotypic similarity between the clinical strains and the isolates evolved under different conditions. MIN-evolved isolates were the most different (largest Mahalanobis distance) from the clinical strains (Figure 3.3A; ANOVA, $F_{3,92}$ = 17.9, $p < 0.001$). The clinical strains were most similar to those evolved in SCFM, but not statistically more similar than those evolved in SCFM+mucin or MIN+ mucin, evaluated by post-hoc pairwise t-tests, Holm-adjusted for multiple comparisons. Interestingly, strains classified as epidemic were consistently less similar to laboratory-evolved isolates than those classified as nonepidemic (two-factor ANOVA, effect of epidemic/nonepidemic, $F_{1,88} = 6.21$, $p = 0.015$; Figure 3.3A). This difference is largely attributable to levels of antibiotic resistance, as excluding antibiotic resistance traits eliminates the difference in Mahalanobis distance between epidemic and nonepidemic strains from our in vitro isolates (Figure 3.3B); two-factor ANOVA, effect of epidemic/nonepidemic, $F_{1,88} = 1.69$, $_{p} = 0.197$). This result further supports those noted by DETTMAN *et al.* 2013 that the key distinguishing feature of epidemic strains is that they are more resistant to antibiotics than nonepidemic strains. Importantly, epidemic and nonepidemic strains are indistinguishable in the other phenotypic dimensions measured here, making it difficult to distinguish them on the basis of nonantibiotic phenotypes alone.

Interestingly, the Mahalanobis distance between the clinical isolates and those from even the most CF-like conditions in our experiment was always significantly greater than zero, whether or not we included antibiotic resistance traits in the analysis (Hotelling's T2 test, $p < 10^{-7}$ for all comparisons). There are two, not mutually exclusive, interpretations of this result. The first is that a positive Mahalanobis distance reflects the absence in our experiment of additional selective forces commonly experienced by *P. aeruginosa* in the lungs of CF patients such as an active immune system or competition from a diverse microbial community.

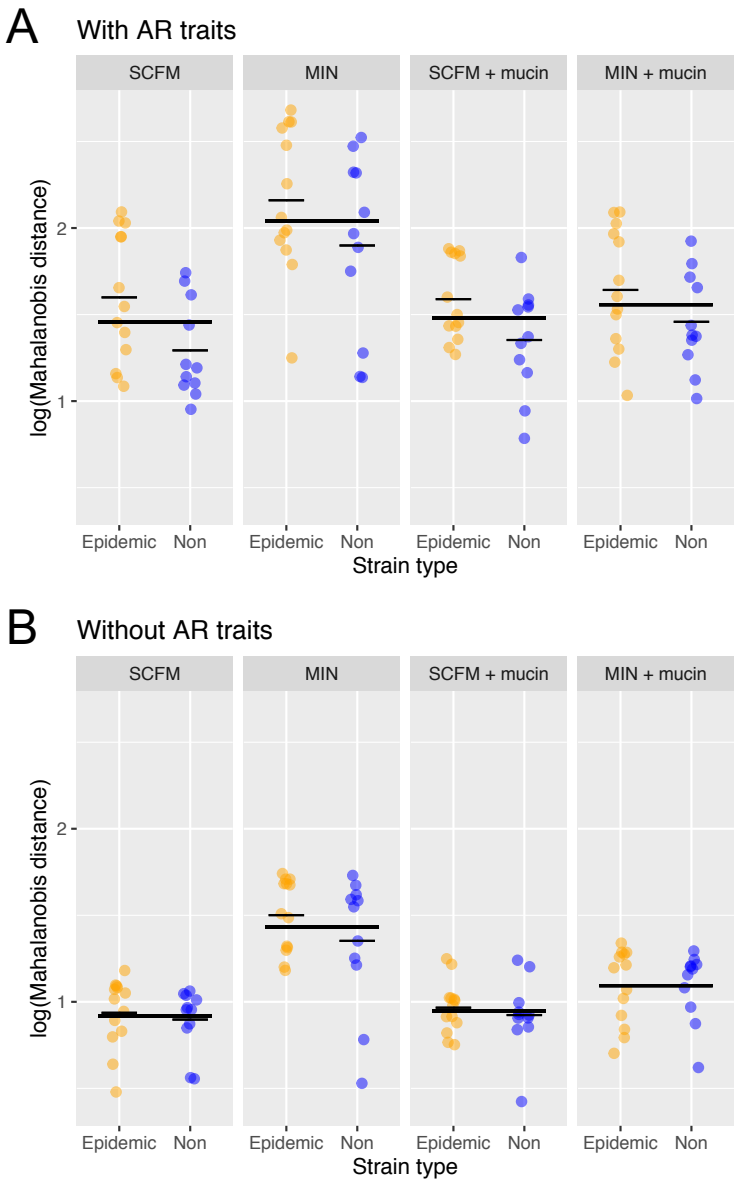

Figure 3.3: Comparison of clinical isolates to laboratory-evolved isolates. Each point represents the multivariate (Mahalanobis) distance between a clinical strain and the distribution of all individual isolates from replicate populations in each treatment. Mahalanobis distances are calculated for all traits (A) and excluding antibiotic resistance (AR) traits (B). Blue circles are nonepidemic clinical isolates, and orange circles are epidemic isolates. Solid lines represent treatment means for nonepidemic and epidemic separately and together.

The second stems from the observation that most chronic CF infections arise from *de novo* colonization by genetically unique strains of *P. aeruginosa*. If so, then some portion of the phenotypic divergence among isolates from different patients could be associated with phylogenetic diversity among colonizing strains. The relative contribution of phylogenetic vs. selective forces in driving CF lung-associated adaptation and divergence is unclear and constitutes a subject for future investigation.

3.4.5 Variation in colony morphology is also linked to nutritional complexity

Colony morphologies are often used as indicators for progression to chronic infection, so we assayed colonies from each evolved population for eight morphological characteristics (pigmentation, opacity, iridescence, surface texture, margin, halo, autolysis, and small colony variant). We identified a total of 18 distinct morphotypes across all 120 populations and an average of ~2 morphotypes per population (range from 1 to 4; Figure B.4). Notably, and consistent with the results presented above, populations evolved in SCFM contained more distinct colony morphs on average than those evolved in MIN (ANOVA, medium, $F_{1,116} = 7.36$, $p = 0.008$), while the presence of mucin had no significant effect on the number of colony morphs (ANOVA, presence of mucin, $F_{1,116} = 0.64$, $p = 0.427$). Moreover, there was little correspondence between colony morphology and the changes in the suite of putatively patho-adaptive traits we measured, as revealed by inspection of Spearman rank correlations for all pairs of traits (Figure B.3). These results lend further support to the growing consensus that colony morphotypes are unreliable markers of adaptation and the onset of chronic infection.

3.4.6 Clinical significance and implications

Our results provide direct experimental evidence that phenotypic diversification of *P. aeruginosa* in the airways of CF patients can be driven by divergent selection imposed by the ecological opportunity associated with the nutritionally complex lung environment. Divergence among subpopulations can be further exaggerated by reduced dispersal resulting from the thick mucus layer associated with the CFTR defect and colonization of different airway compartments. Together, these results suggest that the combination of nutritional complexity and reduced dispersal alone is sufficient to drive rapid and repeatable diversification, independently of other sources of selection associated with adaptation such as immune evasion, resource competition from other microbial species, or redox stress.

This interpretation must be qualified, of course, by the fact that our experiments were done in vitro under conditions that were a far cry from the more complex and dynamic environment of the CF lung itself. The advantage of our approach is that it affords us the opportunity to construct focused, highly replicated tests of the role of nutrient complexity and mucin in driving diversification on scales that would not otherwise be possible. The disadvantage, of course, is that our environments can only represent a crude simulacrum of the conditions actually experienced over the course of an infection. That said, it is worth noting that many of the phenotypes thought to be hallmarks of chronic infection, especially loss of motility and biofilm formation, were also observed in the more CF-like conditions in our experiment, suggesting that we have done a reasonable job of recapitulating some of the selective conditions driving adaptation to the CF lung. These changes are likely due to loss-of-function mutations, a result commonly observed in the initial stages of adaptation to

novel environments in microbial selection experiments (KASSEN 2014). While the detailed genetic changes underlying these phenotypes will have to await whole-genome sequencing (currently underway), the striking phenotypic parallelism between adaptation in vitro and in vivo suggests that the initial stages of colonization of the CF lung can be understood as a specific instance of the more general phenomenon of adaptation and diversification to a novel, nutrient-rich environment.

There are two important implications of our results for clinical practice. First, our results provide direct evidence that CF lung-like conditions promote rapid phenotypic and colony morph diversification, a result that is in line with both longitudinal (CLARK *et al.* 2015; MARKUSSEN *et al.* 2014), and cross-sectional (FOTHERGILL *et al.* 2010; WORKENTINE *et al.* 2013) studies at the level of both phenotype and genotype. Such rapid diversification, together with the observation that a number of hallmark phenotypes evolved rapidly in vitro, imply that colony morphology or other phenotypic biomarkers are not reliable diagnostic traits of the transition to chronic infection. Second, rapid and extensive diversification both within and among independently evolved populations suggests that the transition to chronic infection can occur by many different phenotypic and genetic routes. The presence of genetically distinct subpopulations within the lung or among different patients complicates treatment because it means that no single therapy targeting *P. aeruginosa* is likely to be effective at clearing or managing infection for all patients. Rather, a more tailored, patient-specific approach that focuses therapy on the phenotypic and genomic properties of the strain infecting a given host may prove more useful.

The long-term fate of diversity and its consequences for patient health remain unclear. Longitudinal studies suggest that *P. aeruginosa* diversification can occur rapidly following colonization and can persist for decades within a single host (MARKUSSEN *et al.* 2014), likely due to specialization of subpopulations to different conditions of growth in distinct lung compartments. The comparatively short duration of our experiment does not allow us to distinguish whether the diversity we observed was transient, being the product of high mutation supply rates generating competition among independently arising genotypes, or stable, being supported by negative frequency-dependent selection linked to resource specialization. Nevertheless, our results do lend support to the idea that the nutritional conditions of the CF lung provide a substrate that spurs strain diversification and supports higher levels of genetic variation than would be otherwise available in a more nutritionally homogenous environment. It may be that this diversity provides the raw material for colonization of different airway compartments, leading to compartment-based specialization. Under this view, the transition to chronic infection is intimately tied to, and is in fact the result of, diversification.

3.5 Conclusion

In summary, this study shows that within population diversification occurs rapidly in lab-evolved populations of *Pseudomonas aeruginosa*, even in environments that are simple compared to the human lung. We show that within population diversity is driven largely by the nutritional complexity of the medium, suggesting that competition for resources may be an important factor driving adaptation and diversification. We also show that among population diversity is driven by both the nutritional complexity and the spatial structure of the medium, demonstrating the importance of both of these aspects of ecology for understanding the evolution

of this opportunistic pathogen. Further, our observation that common markers of chronic infection arise in lab-evolved demonstrates the utility of the technique of experimental evolution for investigating variables that affect how *P. aeruginosa* adapts to new environments.

Chapter 4

Genetic changes underlying diversifying populations

A version of this chapter has been submitted for publication as "Schick, A. and Kassen, R. 2019. Genomics of experimental diversification of *Pseudomonas aeruginosa* in cystic fibrosis lung-like conditions". I conceived the project with advice from R. Kassen. I performed the laboratory work and all bioinformatic and statistical analysis, with advice from R. Kassen. I prepared the initial draft of the manuscript, and R. Kassen contributed revisions to the manuscript.

4.1 Summary

Pseudomonas aeruginosa is among the most problematic opportunistic pathogens for adults with cystic fibrosis (CF), causing repeated and resilient infections in the lung and surrounding airways. Mounting evidence suggests that long-term infections are associated with diversification into specialized types but the underlying cause of that diversification and the effect it has on the persistence of infections remains poorly understood. Here, we use evolve and resequence experiments to investigate the genetic changes accompanying rapid, *de novo* phenotypic diversification in lab environments designed to mimic two aspects of the complex ecology of the human lung: spatial structure and complex nutritional content. After ~220 generations of evolution, we find extensive genetic variation present in all environments, including those that most closely resemble the CF lung. We use data on the abundance and frequency of nonsynonymous and synonymous mutations to estimate the ratio of mutations that are selectively neutral (hitchhikers) to those that are under selection (drivers). A significantly lower proportion of driver mutations in the presence of mucin, the protein component of mucus that creates spatial structure by increasing viscosity and reducing dispersal, suggests that spatial structure generates subpopulations with reduced effective population size, decreasing the supply of beneficial mutations and causing more divergent evolutionary trajectories. In addition, we find mutations in a handful of genes typically associated with chronic infection in the CF lung, including one gene associated with antibiotic resistance. This demonstrates that many of the genetic changes considered to be hallmarks of adaptation to the CF lung are the result of colonization and adaptation to a novel environment and do not arise as a result of antimicrobial treatment, immune system suppression, or competition from other microbial species.

4.2 Introduction

Of the many complications associated with the genetic disorder cystic fibrosis (CF), arguably the most difficult to manage and treat is chronic infection of the CF airways by the opportunistic pathogen *Pseudomonas aeruginosa*. Chronic infection occurs in 60-70% of adult CF patients in Canada and is associated with increased morbidity and mortality, irrespective of lung function (HOIBY and PRESSLER 2006; RAJAN and SAIMAN 2002; SCHAEDEL *et al.* 2002). The majority of infections are thought to result from colonization by environmental strains that adapt to the stressful conditions of the CF lung, leading to characteristic phenotypic and genetic changes including the loss of motility and virulence, a tendency to form mucoid colonies, and the acquisition of high levels of antibiotic resistance (MOWAT *et al.* 2011; POOLE 2005; SMITH *et al.* 2006). The transition to chronic infection is also marked – and, indeed, may be caused – by rapid diversification of *P. aeruginosa* into phenotypically and genetically distinct clones, some of which coexist for years within the same host (ASHISH *et al.* 2013; FOWERAKER *et al.* 2005; MARKUSSEN *et al.* 2014; WORKENTINE *et al.* 2013; WRIGHT *et al.* 2013). The causes of diversification, and the contribution of this diversity to the long-term persistence of *P. aeruginosa* infections, remain poorly understood.

Two features of the ecology of microbial life in the CF lung have been shown to promote phenotypic diversification that recapitulates the phenotypic variation seen among clinical isolates from CF patients (SCHICK and KASSEN 2018). The first is the nutritional complexity of the airways and the second is the reduced dispersal among subpopulations caused by both the thick mucous layer covering the lung epithelia and the spatially compartmentalized nature of the lung. Our previous work showed that nutritional complexity was the main driver of within-population diversification, a result consistent with the view that ecological opportunity, the range of underutilized resources, generates strong divergent selection that can drive diversification (FUTUYMA and MORENO 1988; KASSEN 2009; SCHLUTER 2000). Substantial phenotypic divergence among populations was observed as well, suggesting that reduced dispersal associated with spatial segregation allows distinct populations to explore divergent evolutionary routes to adaptation (MARKUSSEN *et al.* 2014; SCHICK and KASSEN 2018). The rapid and repeated diversification of *Pseudomonas aeruginosa* demonstrated both empirically and experimentally suggest that this diversification, by generating extensive phenotypic and, presumably, genetic variation on which selection can act, may be a key first step in the development of chronic infections.

Two properties governing the genetics of diversification remain unclear. The first is the extent to which phenotypic variation is matched by comparable levels of genetic variation. Theory suggests that there should be a close correspondence between the two, for three reasons. First, mutation and genetic drift, the fluctuations in allele frequencies associated with finite population size, are stochastic processes that cause replicate populations descended from a common ancestor to diverge through time even in a common environment. Second, strong divergent selection generated by ecological opportunity is expected to lead to the evolution of genetically distinct niche specialists SCHICK *et al.* 2015 that can co-exist for prolonged periods of time or even indefinitely (BEHRINGER *et al.* 2018; LEALE and KASSEN 2018). Third, the reduced dispersal associated with spatial segregation allows independently arising beneficial mutations in distinct subpopulations to persist longer than in a well-mixed system because competition among distinct beneficial mutations for fixation arising from clonal interference is weaker. Genetic divergence should thus be closely linked to phe-

notypic disparity, being most pronounced in ecologically complex environments characterized by abundant ecological opportunity and spatial structure, precisely the conditions thought to characterize the CF airway.

The second is the repeatability of the genetic changes associated with divergence. The probability of parallel evolution, the repeated evolution of the same genetic changes in independently evolved populations under directional selection, is largely a function of population size (BAILEY *et al.* 2017): in large populations rare, large-effect beneficial mutations are both more abundant and likely to outcompete independently arising beneficial mutations of smaller effect. For a given population size, however, the probability of parallelism under divergent selection should be lower than under directional selection due to the evolution of genetically distinct niche specialists (KASSEN 2014). Spatial structure should also decrease the probability of parallelism relative to a well-mixed system because it reduces effective population size by creating sub-populations that evolve more or less independently of each other. Within patients, then, parallelism is expected to be low.

To test these predictions, we sequenced end-point populations from our previous experiment where *P. aeruginosa* strain Pa14 was allowed to evolve and diversify for ~220 generations in conditions designed to mimic the environmental conditions encountered during colonization of the CF lung (SCHICK and KASSEN 2018). Briefly, the environments varied in how closely they resemble the resource profile and viscosity of the CF lung, the most CF lung-like environment being composed of a synthetic CF medium (SCFM) based on the chemical contents of samples of CF sputum and mucin, the major protein component of mucus (see SCHICK and KASSEN 2018 for details). The least CF lung-like environment was composed of a minimal medium with glucose as the sole carbon source (MIN) and no mucin. The two intermediate environments were SCFM without the addition of mucin and MIN with mucin, completing the factorial design. We sequenced 81 evolved populations from this experiment to a mean of 120-fold coverage across the genome to uncover all genetic changes that arose and spread to a detectable frequency. These in vitro experiments by no means capture all the dimensions of life in the CF lung, as they lack key features such as a diverse microbial community that accompanies *P. aeruginosa* infections and an active immune system. Nevertheless, our results provide insight into the spectrum and quantity of genetic changes associated with diversification associated with nutrient complexity and spatial structure, two features that are thought to play important roles in governing the dynamics of diversity during the early stages of colonization.

4.3 Materials and Methods

4.3.1 Bacterial Strains

Bacterial strains and populations used for this study were from the end-point of a selection experiment described in SCHICK and KASSEN 2018. Briefly, a total of 120 populations derived from a common ancestor, *Pseudomonas aeruginosa* strain 14 (Pa14), were propagated in daily batch culture for ~220 generations in one of four environments: SCFM (synthetic cystic fibrosis medium), MIN (M9 minimal salts + glucose), SCFM + mucin, and MIN + mucin. The factorial design of the experiment allows us to examine the main effects of nutritional complexity (SCFM versus MIN), spatial structure (mucin versus no mucin), and their interaction on phenotypic and genetic diversification. We randomly chose 24 populations from each of

the SCFM and SCFM + mucin treatments, and 23 populations from each of the MIN and MIN + mucin treatments, for sequencing. Two reference strains, both ancestral to the selection experiment (Pa14 and a lacZ-marked Pa14, were also sequenced to facilitate genome assembly and to identify genetic variants evolved over the course of the experiment.

4.3.2 Whole-genome sequencing

Populations were revived overnight from frozen in liquid Luria Bertani (LB) broth at 37°C. LB instead of the selection media was used to revive populations for two reasons: to be consistent across treatment groups and because the mucin interfered with the ability to extract DNA. Genomic DNA was extracted for whole-genome sequencing from samples using the MO BIO Ultraclean 96 Microbial DNA kit (now sold as QIAGEN DNeasy UltraClean 96 Microbial kit), following the manufacturer's recommended protocol. Library preparation and sequencing was performed by Genome Quebec at McGill University on the Illumina HiSeq 4000 platform, using paired-end sequencing of 2x100 base-pair reads.

4.3.3 Sequence processing and generating variant table

Whole genome sequencing yielded a total of 75 Gb of raw data, with a median coverage of 120-fold. Analyses were performed in-house using a custom pipeline. Briefly, sequencing reads were quality trimmed using Trimmomatic version 0.36 (BOLGER *et al.* 2014), removing the leading and trailing bases below a quality score of 5 as well as scanning the read with a 5-base sliding window, cutting when the average quality per base drops below 20. Reads shorter than 20 base pairs were also discarded. Reads were then aligned to *Pseudomonas aeruginosa* reference genome UCBPP-PA14 109 (from WINSOR *et al.* 2016) using the bwa-mem algorithm of BWA version 0.7.12 and Samtools version 1.3.1 (LI *et al.* 2009). Picard Tools version 2.9.2 was then used to mark PCR duplicates and add read group information. Variants were called using two independently developed algorithms. First, Breseq version 0.30.0 (DEATHERAGE and BARRICK 2014) was used, a tool specifically designed for detecting mutational variants in microbial genomes. For the evolved populations, a mixed-alleles model was used by setting the mode to predict polymorphism (option -p) in order to detect sites with multiple alleles in populations and the frequencies of those alleles. For the two ancestral strains, the mode was set to consensus (the default option) because these two populations were homogenous. To confirm no major biases in variant calling, the HaplotypeCaller function of GATK version 3.6 (MCKENNA *et al.* 2010) was used as well. To identify only mutations that arose over the course of the selection experiment, the set of variants found in our two ancestral strains (Pa14 and Pa14-lacZ) were compared to those found in our evolved populations. Variants in common across ancestral strains and all evolved populations were discarded. Of the 94 populations sequenced, we found 4 populations to have evidence of cross-contamination and 9 populations to have increased mutation rates, evidenced by both a larger number of variants present (> 30) and variants in one of the following genes: *mutS* (2), *lexA*, *dnaA*, *dnaX*, *recQ*, *uvrD*, *ruvB*, and PA14_25780, all of which have been linked to increased mutation rates. Populations with these putative hypermutators present were excluded from subsequent analysis.

4.3.4 Estimating rates of genomic hitchhiking

To estimate the proportion of mutations that are selectively neutral but arose in a genome containing a beneficial mutation (a phenomenon referred to as hitchhiking), we compared the observed number of nonsynonymous mutations to the expected number of nonsynonymous mutations under neutrality, inferring that the excess mutations over the expected number are likely non-neutral, or 'driver' mutations. The expected number of nonsynonymous mutations under neutrality is based on the observed number of synonymous mutations, genetic changes assumed to have no effect on fitness. Under neutral evolution, nonsynonymous and synonymous mutations arise at the same rate, accumulating to numbers proportional to the number of sites of each type. Using this estimate of the number of hitchhiker mutations (expected nonsynonymous mutations under neutrality), we determine the estimated number of driver mutations by subtracting the expected number of nonsynonymous mutations from the observed number of nonsynonymous mutations. We apply this calculation to all populations grouped together as well as populations within a given treatment. For all populations grouped together, we estimate that 28.4% of nonsynonymous SNPs are drivers (see Table 4.1 for mutation numbers). Further, after estimating the relative rates of hitchhiker and driver mutations, we compare the distribution of mutation frequencies of synonymous and nonsynonymous mutations to find that nearly all synonymous mutations are present at frequencies less than 0.16. Hitchhiker mutations are likely to be present at frequencies equal to or less than their respective driver mutations, suggesting that low frequency mutations are most likely to be hitchhikers. Therefore, we exclude low-frequency mutations to match the estimated proportion of hitchhiker mutation as calculated above. This corresponds to a frequency of 0.16, shown in Figure 4.3.

4.3.5 Genetic diversity

We estimated within-population genetic diversity by calculating heterozygosity, defined here as the mean heterozygosity ($2pq$, where p is the estimated allele frequency and $q = 1 - p$) of all polymorphic loci in a population, excluding synonymous and low-frequency variants (as defined above). Treatment group ranks were preserved when low-frequency variants were included. Significant differences between treatment groups were tested for using a single-factor ANOVA. To quantify genetic divergence among evolved populations, we performed a principal coordinates analysis (PCoA) using the R package vegan (version 2.5.2) on a Euclidean distance matrix. We estimated genetic divergence within a treatment by calculating mean distance to the spatial median using the 'betadisper' function. This function determines if treatment groups differ in dispersion (variance), a multivariate analogue of the Levene test for homogeneity of variances. Following this, we performed a Tukey HSD test to determine which treatment groups differed significantly in mean dispersion. Dispersion was also used a measure of parallelism within treatment, discussed in the section below.

4.3.6 Parallelism and repeatability

We quantified population levels of parallelism using three different metrics: dispersion, Jaccard, and C-scores. All three metrics were calculated using the same set of data: frequencies of mutations in all genes,

after excluding synonymous and low-frequency variants. For dispersion, we calculated the mean distance between a population and the treatment centroid, following a principal coordinates analysis (PCoA) on a Euclidean distance matrix. The measurement corresponds to treatment level genetic variance with larger mean dispersion signifying more divergent populations and therefore less parallelism. We also include standard error of dispersion. To determine significance, we performed an ANOVA with Euclidean distance as the response variable and treatment as the explanatory. For the Jaccard measure, we used the Jaccard index to calculate the dissimilarity between all pairs of populations within a treatment. We then report the mean and standard error of all pairwise comparisons, with larger values signifying larger dissimilarities and therefore less parallelism. C-score is a metric for repeatability that uses the hypergeometric distribution to calculate the deviation between the observed amount of parallelism and the expectation under random gene-use (YEAMAN *et al.* 2018). The magnitude of the C-score represents the magnitude of the deviation with larger C-scores signify higher repeatability and therefore more parallelism. To determine significance for both Jaccard and C-score metrics, we performed an exact test (permutation test) by randomizing treatment labels (number of permutations = 10000) and calculating a null distribution of F-values.

Parallelism at the gene level was defined as the proportion of populations with mutations in that gene, both globally for all treatments and within specific treatments. To test for significance, we calculated the probability of our observed results if gene use was random, using the binomial distribution with the number of populations as the number of trials, number of times a gene was mutated as the number of successes, and proportion of total populations across all treatments with a mutation in that gene as the probability of success. From this, if the probability of an observation was less than 0.05, we considered that gene to be treatment specific.

4.4 Results and Discussion

Comparing whole-genome sequence data from the 81 evolved populations against the ancestral Pa14 genome revealed a total of 656 unique genetic variants, corresponding to an average of 8.10 variants per population. The identity and genomic location of mutations recovered in our analysis are summarized in Figure 4.1 and Figure 4.2. The majority of variants were present at low frequencies (mean frequency = 0.207 and median frequency = 0.077), with only 35 variants fixed (*i.e.*, variant frequency = 1). Variants were distributed across the genome (Figure 4.1), with peaks corresponding to genomic positions containing genes that are often mutated in *P. aeruginosa* isolates from CF patients (*mexT*, *lasR*; discussed in more detail below). Nonsynonymous SNPs (single-nucleotide polymorphisms) were the most common variant class in the experiment, being nearly 3 times more common than the next most abundant class, small indels (Figure 4.2A), and there was little variation among environments in the distribution of variant classes (Figure 4.2B). This last result suggests that, at this level of categorization at least, there is little genomic signature to distinguish among any of the environmental conditions in our experiment. This interpretation does not hold for the total and average number of variants segregating or fixed among environments, which varied substantially across treatments due to fewer variants being recovered in MIN relative to the other three environments (Figure 4.2B; Figure C.1; ANOVA, $F_{3,77} = 6.24$, $p < 0.001$).

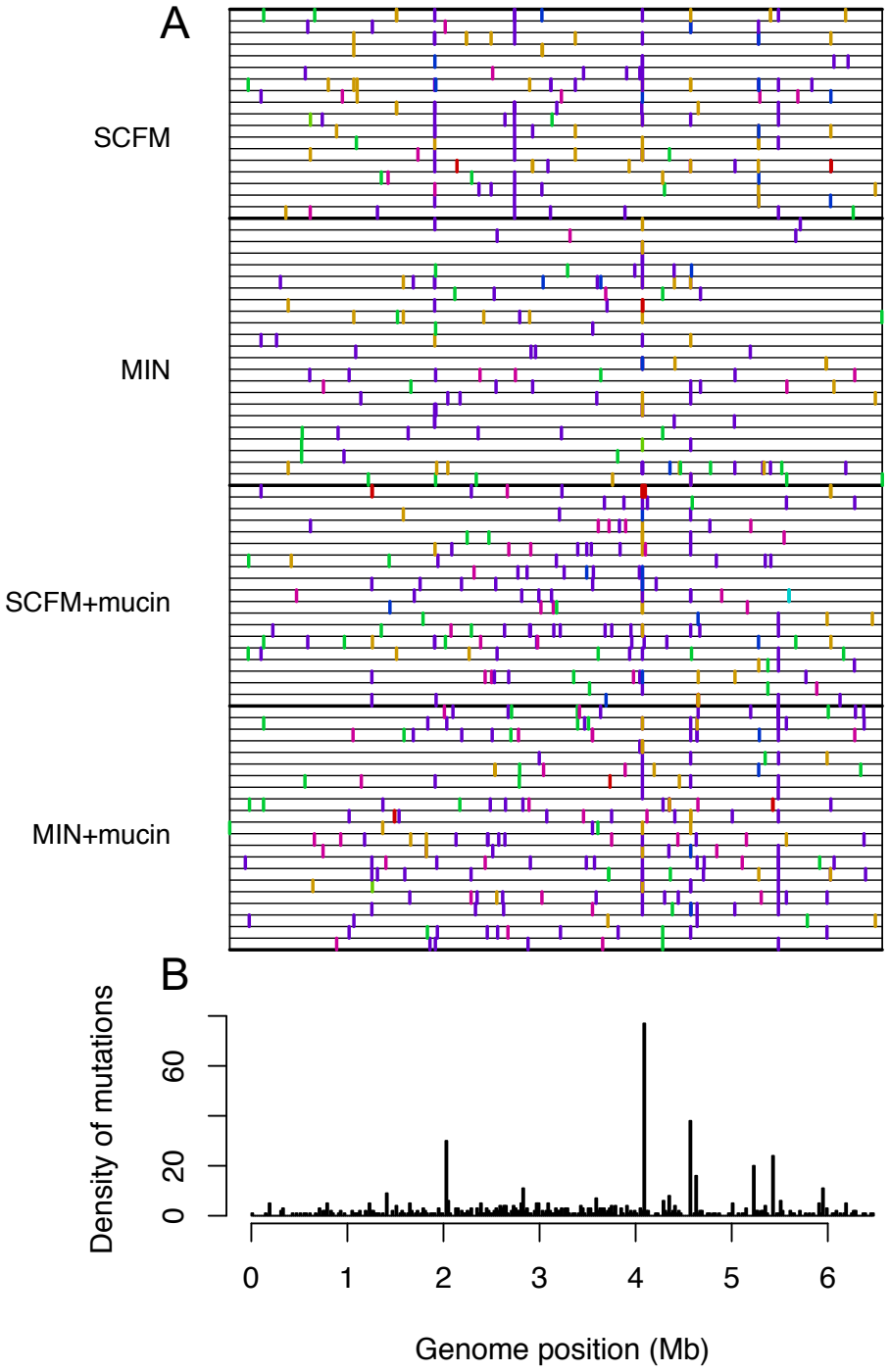

Figure 4.1: Location of all mutations discovered in evolved populations. A. Position in the genome of mutations per population; each row is a population and the genome is represented horizontally. Mutational types are coloured as in Figure 2. B. The density of mutations along the genome in 10 kb windows.

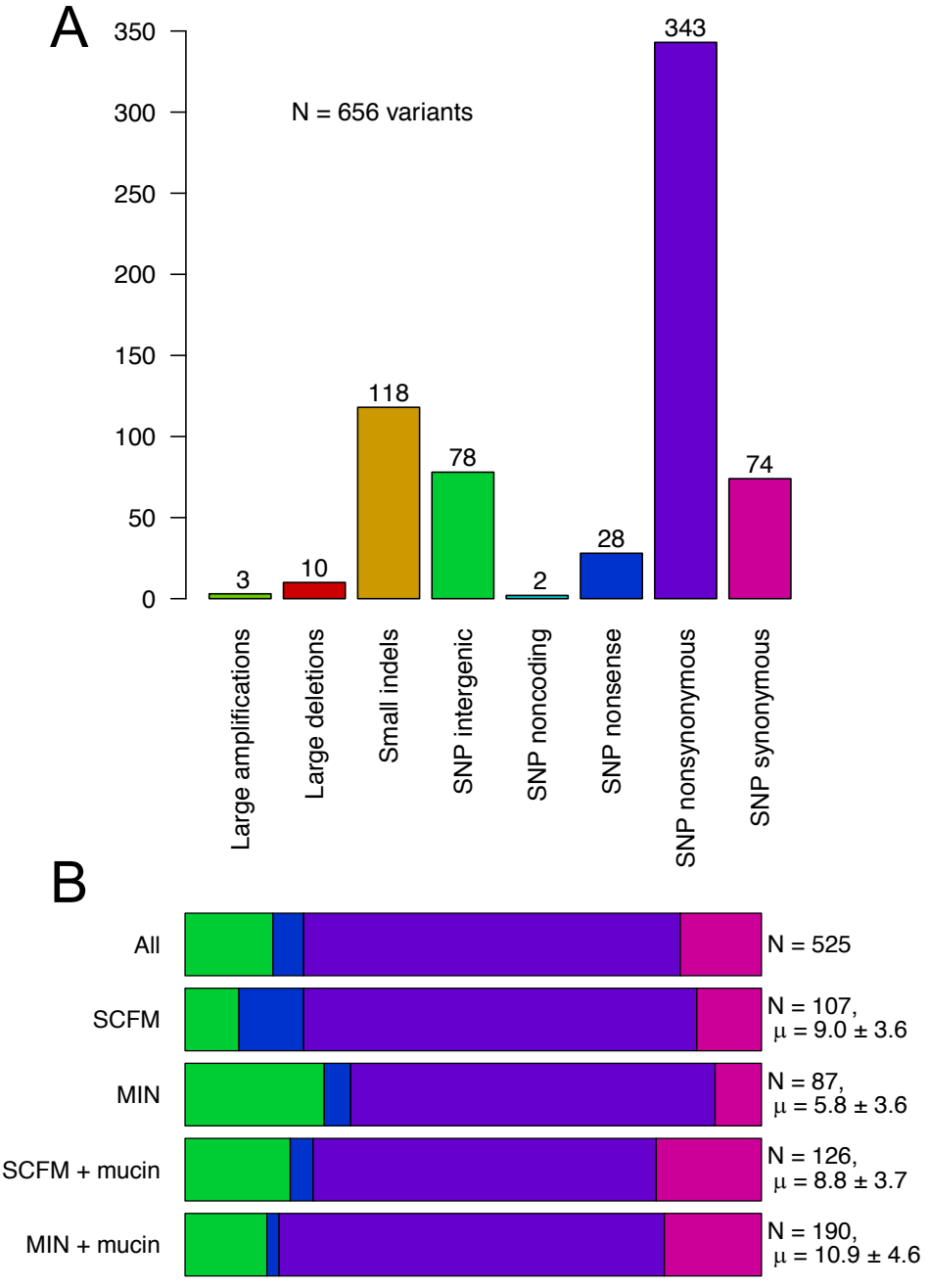

Figure 4.2: Types of mutations. A. A. The distribution of mutations by mutational type. B. The distribution of the different types of only point mutations (SNPs) by selection environment. N is the total number of variants for all populations within each group, with mean and standard deviation of the number of variants per population.

4.4.1 Positive selection and the prevalence of genomic hitchhiking

To what extent do the patterns of genomic variation we observe in our experiment reflect the action of selection, as opposed to other mechanisms influencing the distribution of genetic variation like the stochastic effects of drift? The large population sizes (approximately 10^8 cfu/ml) and relatively short duration of our experiment mean that genetic drift is very likely too weak for neutral mutations to reach observable frequencies on their own. However, neutral or even mildly deleterious variants can reach appreciable frequencies if they occur in the same genetic background as one or more beneficial mutations, a process termed 'hitchhiking'. The genomic variation we observe in our populations is therefore likely a mixture of beneficial 'driver' mutations and hitchhiking mutations; ideally, we would like to know which is which. This task is relatively straightforward when mutations are few in number, as their fitness effects can be evaluated directly through competition experiments (for example BUSKIRK *et al.* 2017; KHAN *et al.* 2011; SCHICK *et al.* 2015). For studies such as this one, where there are many evolved populations and large amounts of genomic variation within each, direct measurements of fitness on individual variants are impractical and rates of hitchhiking must be inferred using statistical methods.

As a first step, we calculate the ratio of nonsynonymous to synonymous mutations under the assumption that the majority of synonymous mutations are neutral with respect to fitness (although reports to the contrary suggest synonymous mutations can occasionally be adaptive; see AGASHE *et al.* 2016; BAILEY *et al.* 2014; KRISTOFICH *et al.* 2018; LEBEUF-TAYLOR *et al.* 2019. Of the 445 SNPs we observed in protein coding regions, 371 and 74 were nonsynonymous and synonymous, respectively, which translates into a frequency of 16.6% synonynmous SNPs. By surveying all protein-coding regions of the reference genome, YANG *et al.* 2011 reported that 25.1% of single basepair changes would be synonymous, resulting in no change in the amino acid being coded. Given this expected rate of synonymous changes, nonsynonymous SNPs are vastly over-represented relative to synonymous SNPs in our experiment, suggesting that positive selection is driving a large number of these polymorphisms to high frequency. Further evidence supporting this inference comes from examining the SNPs present in the five most frequently mutated genes in our experiment (*lasR*, *morA*, *mvfR*, *orfH*, and PA14_32420): all five contain only nonsynonymous changes, suggesting strong positive selection on protein altering mutations in these genes.

Next, we estimate the proportion of driver to hitchhiker mutations by comparing the observed number of nonsynonymous mutations to that expected from the observed number of synonymous mutations, again under the assumption that the latter are neutral. The rationale here is that if nonsynonymous hitchhiker mutations are also neutral, this class of mutation should be at least as frequent as that of synonymous sites in our experiment. The number of putative driver mutations – excluding those beneficial mutations arising late in the experiment that have not had sufficient time to become common – can therefore be calculated by subtracting the number of expected nonsynonymous SNPs under neutrality (hitchhikers) from the number of nonsynonymous SNPs observed. Our results, which are summarized in Table 1, show that 28.4% of all observed SNPs are likely drivers when all populations are considered together, a value that is not too far off the value of 20% estimated for evolving populations of yeast (BUSKIRK *et al.* 2017). Disaggregating this result by environment reveals a striking result: environments lacking mucin harbour a larger proportion of putative drivers than those with mucin (Table 4.1; $\chi^2 = 7.638$, $p = 0.008$). As population sizes and

Table 4.1: Estimated proportion of SNPs that are driver mutations. The estimated number of hitchhiker SNPs is calculated as the observed number of synonymous SNPs times three (following the expected ratio under neutral evolution of 25.1:74.9 of synonymous to nonsynonymous sites). The estimated number of driver mutations is calculated by subtracting the expected number of nonsynonymous hitchhiker SNPs from the observed number of nonsynonymous SNPs. Proportion is determined by dividing number of nonsynonymous driver mutations by the total number of observed SNPs (including SNPs in noncoding regions).

Treatment	Observed nonsynonyous SNPs	Observed synonymous SNPs	Estimated number of hitchhiker SNPs	Estimated number of driver SNPs	Estimated proportion of driver SNPs (%)
All	371	74	222	149	28.4
SCFM	85	12	36	49	45.8
MIN	59	7	21	38	43.7
SCFM + mucin	82	22	66	16	12.7
MIN + mucin	131	32	96	35	28.4

mutation rates are similar across environments, these results are unlikely to reflect lower mutation supply rates in the presence of mucin. If anything, the opposite is the case, as the total number of nonsynonymous and synonymous mutations tends to be higher in the presence of mucin than in its absence. The simplest explanation for this result is that the presence of mucin, by increasing viscosity and reducing dispersal, creates subpopulations with lower effective population sizes (N_e) relative to a well-mixed system and so reduces the supply rate of beneficial mutations (because only those with selection coefficients greater than $1/N_e$ stand an appreciable chance of fixing).

4.4.2 Frequency spectra of nonsynonymous and synonymous SNPs

Examining the frequency spectra of nonsynonymous and synonymous SNPs from all populations reveals the distributions to be very different. Most synonymous SNPs were rare, 97% having a frequency of 0.16 or less, whereas nonsynonymous SNPs were present at a range of frequencies from rare to fixed (Figure 4.3). The paucity of high frequency synonymous SNPs suggests that the majority of this class of mutation do not contribute to adaptation (with the possible exception of two high-frequency synonymous SNPs), a result that is not surprising in light of the widely accepted view that the majority of synonymous mutations are neutral with respect to fitness. The majority of synonymous SNPs are thus likely hitchhikers whose frequencies, by definition, must be equal to or less than those of their respective driver mutations. If the same is true for rare nonsynonymous SNPs, then a conservative estimate (because it ignores potentially rare beneficial nonsynonymous mutations) of the frequency of driver SNPs is 31.2%. Reassuringly, this estimate is not far off that derived above based on inferences based on the ratio of nonsynonymous to synonymous mutations, independent of mutant frequency.

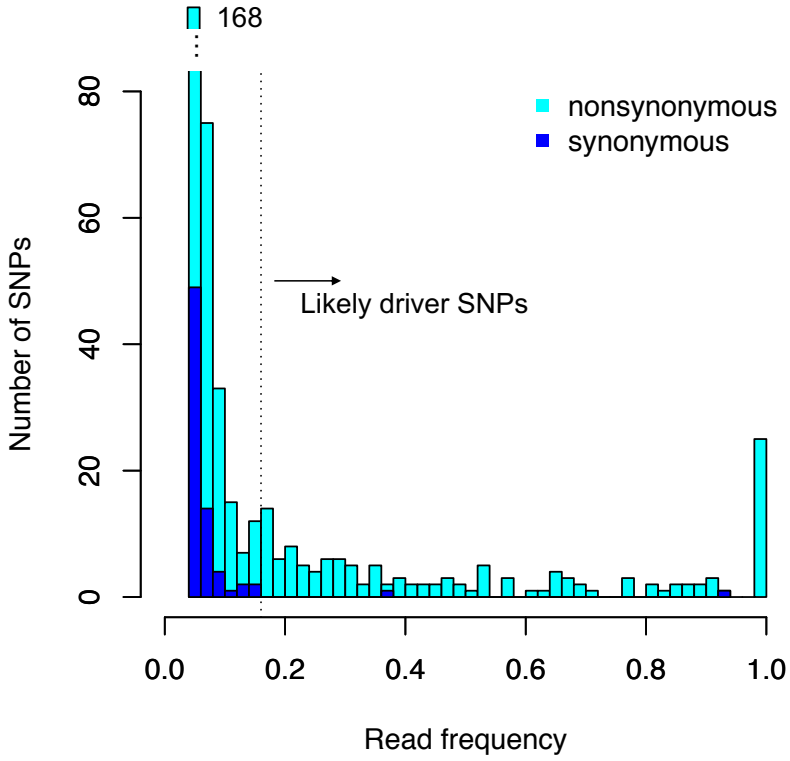

Figure 4.3: Frequency spectra for nonsynonymous and synonymous SNPs. Distribution of read frequencies for all SNPs found across all evolved populations, nonsynonymous (light blue) and synonymous (dark blue). The majority of nonsynonymous SNPs present at frequencies greater than 0.16 (the range of frequencies of most synonymous SNPs) are likely to be driver mutations.

4.4.3 The correlation between genetic diversity and phenotypic disparity

To what extent is the phenotypic disparity we observed in our previous work (SCHICK and KASSEN 2018) underlain by comparable levels of genetic diversity, as predicted by theory? To answer this question, we first calculated mean heterozygosity within each population after excluding synonymous and low-frequency (<0.16) nonsynonymous SNPs (as defined above) and then regressed these values against the extent of phenotypic disparity, calculated as the multivariate Euclidean distance among pairs of isolates from the same population (details provided in SCHICK and KASSEN 2018. Average heterozygosity was significantly greater than zero for all treatments (t-test, $p < 0.0001$ for all treatments; Figure 4.4A), a result that is likely a consequence of high mutation supply rates that generate both clonal interference and introduce hitchhiking mutations alongside genetic variants associated with adaptive divergence. As expected from theory, we see a statistically significant positive correlation between heterozygosity and phenotypic disparity among populations within environments (Pearson coefficient = 0.393, $p = 0.006$; Figure 4.4B). While SCFM, the condition that resembles the nutritionally complex conditions of the CF lung, did have highest average heterozygosity, as expected, we could not detect an effect of environment on the slope nor the y-intercept of the regression of heterozygosity on phenotypic diversity (ANCOVA, $p = 0.131$ and $p = 0.630$ for treatment and interaction term for treatment and heterozygosity). These results suggest that rapid and extensive phenotypic diversification associated with nutritional complexity is driven by a few mutations of large effect that can be easily missed by purely genomic approaches when mutation supply rates are high.

4.4.4 Genome scale patterns of parallelism and constraint

Adaptation to the CF lung typically involves a handful of phenotypic changes involving loss of motility, reduced virulence, increased antibiotic resistance, and mucoidy (MOWAT *et al.* 2011; SMITH *et al.* 2006). Such repeated, or parallel, evolution of the same phenotypes in independent lineages is often taken to be a marker of strong selection. Whether these parallel phenotypic changes are underlain by parallel genetic changes remains unknown. The answer is important because high levels of parallelism that are unique to the CF lung could be used as a genetic marker of the onset of chronic infection. Our experiment allows us to answer this question directly and test hypotheses about how parallelism is impacted by the environmental complexity associated with the CF lung.

To quantify gene-level parallelism among mutations likely to be under selection, we first restrict our analysis to putative driver mutations by excluding synonymous and low-frequency SNPs that we expect to be hitchhiking, as discussed above. While this approach likely inflates our estimates of parallelism, it should do so in a way that is common across treatments and so should not introduce any systematic bias. Our results are summarized visually in Figure 4.5 and presented more quantitatively in Table 4.2. As there is no broadly accepted metric for quantifying gene-level parallelism, we used three distinct measures: variance in dispersion of Euclidean distances between populations, Jaccard index, and observed repeatability relative to expectation under randomness using the hypergeometric distribution (C-score; YEAMAN *et al.* 2018. Details on how each metric is calculated are provided in the Methods. For all metrics reported in Table 4.2, divergence was lowest (and thus parallelism was highest) in the MIN environment and next lowest in SCFM, consistent with the hypothesis that divergent selection reduces the probability of parallelism relative

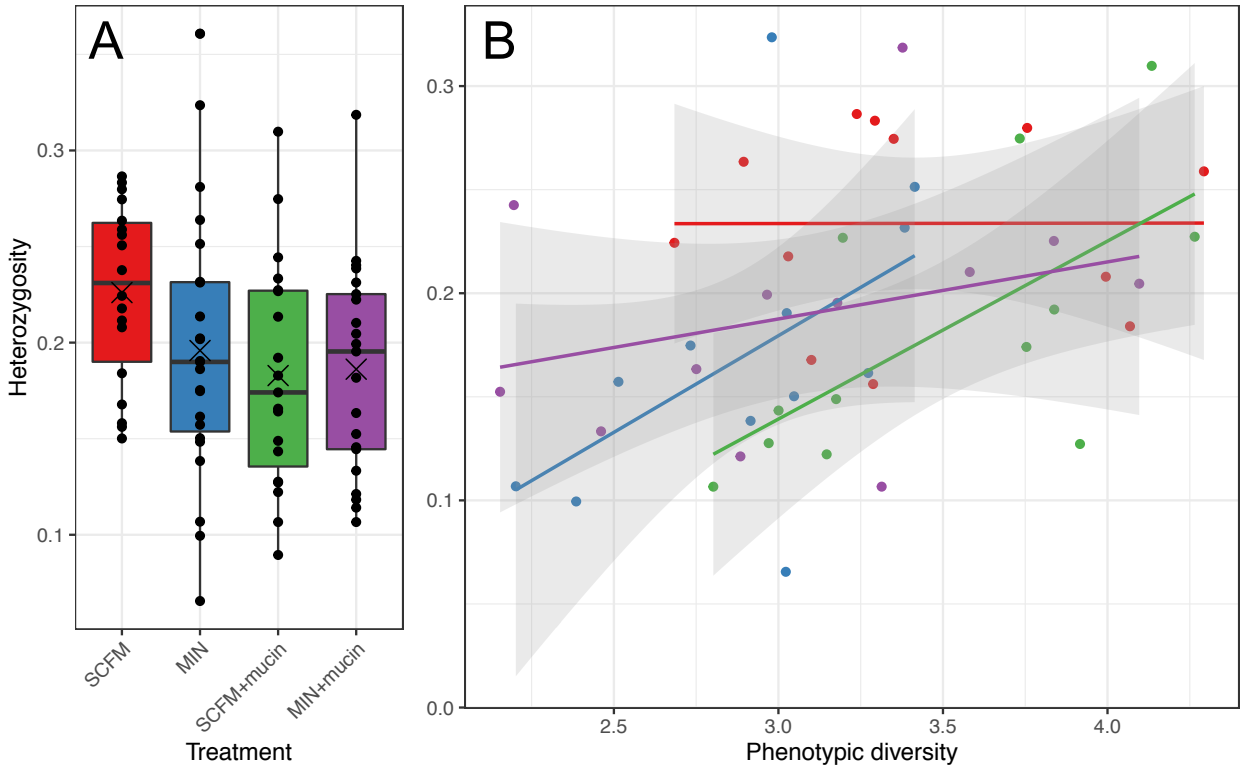

Figure 4.4: Genetic diversity within populations. A. Population mean heterozygosity, calculated using frequencies of variants at all loci with a variant present in a given population, excluding synonymous and low-frequency variants. Horizontal bars represent treatment median and crosses represent treatment mean. B. Correlation between heterozygosity and phenotypic diversity (measured by mean euclidean distance between pairs of populations for a suite of phenotypic characteristics, see SCHICK and KASSEN 2018 for details).

Table 4.2: Parallelism and constraint. Population level patterns of parallelism within each treatment. Dispersion is the mean distance from the centroid of a PCoA (principal coordinates analysis) based on euclidean distance; larger values signify more divergent populations and therefore less parallelism. Jaccard distance is dissimilarity between sets; a Jaccard distance of 1 for any two populations signifies they are completely dissimilar and have no gene variants in common, therefore larger values signify more divergent populations and less parallelism. C-scores measure observed repeatability relative to expectations under randomness; smaller values signify less constraint and therefore less parallelism. For all metrics, SE is standard error and treatments are ranked from most divergent (1) to least divergent (4), equivalent to least parallel (1) to most parallel (4). All three metrics were calculated excluding synonymous and low frequency SNPs. Significance is determined by an ANOVA for Dispersion and by an exact test (permutation test with 10,000 permutations) for Jaccard Index and C-score.

Measure	**Dispersion**			**Jaccard Index**			**C-score**		
	Mean	+/- SE	Rank	Mean	+/- SE	Rank	Value	+/- SE	Rank
SCFM	0.710	0.054	3	0.833	0.013	3	4.70	0.351	3
MIN	0.457	0.055	4	0.810	0.022	4	4.98	0.464	4
SCFM + mucin	0.760	0.058	2	0.856	0.020	1	4.04	0.482	1
MIN + mucin	0.855	0.063	1	0.851	0.012	2	4.17	0.285	2
Significance	$F_{3,77} = 4.78$; $p = 0.004$			Exact test; $p = 0.254$			Exact test; $p = 0.261$		

to directional selection. Divergence was highest (parallelism lowest) in environments containing mucin, although the rank-order depends on which metric is used, consistent with the idea that spatial structure and reduced dispersal lead to more divergence among subpopulations within a lineage.

4.4.5 Gene-level parallelism and environmental specificity

The spectrum of parallel genetic changes evolved in each environment, shown in Figure 4.6, reveals two striking features. The first is that most genes show modest levels of parallelism, with repeated changes occurring in just a handful of lines. Notable exceptions are mutations in *lasR* and *mvfR*, both transcriptional regulators involved in quorum sensing, which were shared by the majority of lines across all treatments. The second is that, with a handful of exceptions (*morA* in MIN+mucin or *pilA* and PA14_32420 in SCFM), there is no strong signal of environment-specific parallelism. To quantify this effect, we ask whether there is a significantly higher probability of gene-level parallelism, measured as the proportion of populations with non-synonymous mutations in a given gene, within than across treatments. We included only genes that were mutated in more than one population (N = 55 genes) and defined significance using a binomial distribution to calculate the probability of seeing the observed proportion of populations if gene use was random across treatments. A probability less than 0.05 suggests parallelism in that gene is treatment specific. We found 13 genes to be significantly treatment specific in at least one selection environment. The SCFM environment showed the largest number of environmentally specific genes (n= 8) though mutations in three genes (*pilA*, *pilB*, and *pilC*) likely result in a similar phenotypic effect, namely, a reduction in twitch motility (BURROWS 2012). Both the SCFM+mucin and MIN environments were found to have only a single gene mutated in more

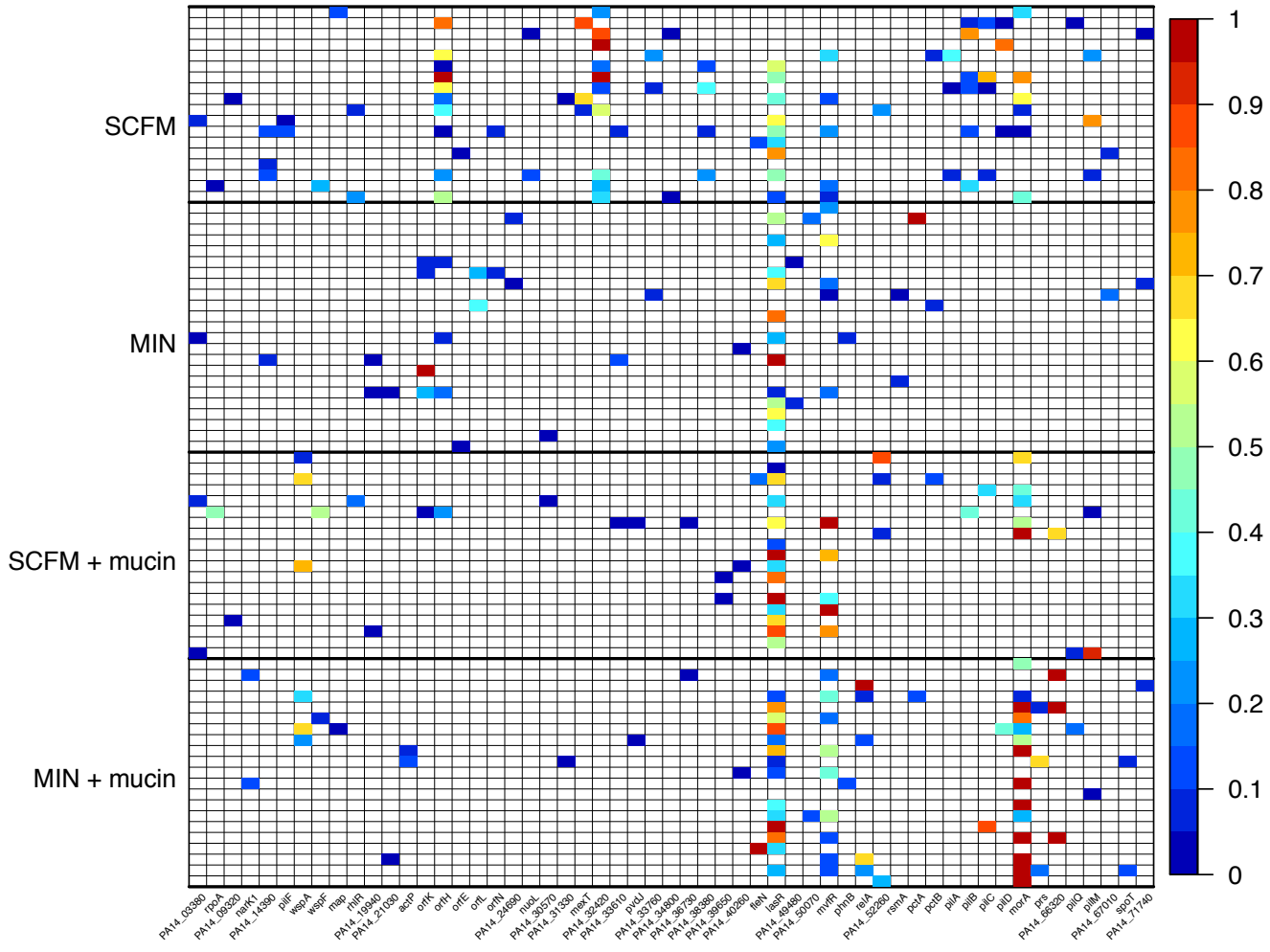

Figure 4.5: Variant frequencies for multiple use genes. Frequency of variants for genes that have variants in more than one population, after excluding synonymous SNPs (N = 55 genes). Each row represents a population, with the presence of a mutation in each gene indicated by a coloured box (high frequency mutations in red and low frequency mutations in blue). Genes are arranged by genomic position.

populations than expected (Pa14_39650 and *orfK*, respectively). Grouping genes by functional classification revealed no patterns, unsurprising given the limited number of treatment specific genes (n =13) and the large number of functional classes. Together, these results suggest that, with the exception of a handful of genes, parallel genetic evolution may be an unreliable marker of CF-specific adaptation.

4.4.6 Genetic targets of selection

Our previous work showed that many of the characteristic phenotypic changes that are the hallmark of the onset of chronic infection also evolve under conditions that mimic the nutritional complexity and spatial structure of the CF lung *in vitro*, in the absence of an active immune system and diverse microflora (SCHICK and KASSEN 2018). Is the same true at the genetic level? To some extent, yes. Many of the genes mutated in our experiment are also observed in genomic analyses of isolates from the CF lung, including *lasR*, *mexT*, *morA*, *wspA/F*, and the *pilA-D/Q* genes (DIAZ CABALLERO *et al.* 2015; FRESCHI *et al.* 2018; MARVIG *et al.* 2015). Few, however, show strong associations with specific components of the CF lung, with the possible exception of *mexT*, that arose multiple times in SCFM, and *morA*, which evolved repeatedly in all environments except MIN. We also failed to observe some characteristic mutations altogether, especially those linked to alginate production such as *mucA*, *gacA/S*, or *algG/U*, however this is not surprising since we did not observe mucoid colonies in our evolved populations. Taken together our results suggest that many of the genetic changes thought to be characteristic of chronic infections are not specific to the nutritional complexity and viscosity of the CF lung but are, rather, favoured across a range of environmental conditions. Rather, this result is consistent with simple models of adaptation that predict the first mutations fixed during an adaptive walk should not have strongly antagonistic effects across different environments (MARTIN and LENORMAND 2006; SCHICK *et al.* 2015), implying that the CF airway represents a novel environment to colonizing *P. aeruginosa*.

Two additional features of our results deserve mention. The first is the high repeatability of *lasR* mutations, with 74 unique variants being identified in 48 populations in our experiment. Putatively loss-of-function mutations in *lasR* are commonly observed in CF isolates, their selective advantage thought to be the result of reduced expression of acute virulence factors or growth advantages linked to amino acid metabolism (D'ARGENIO *et al.* 2007; HOFFMAN *et al.* 2009; LAFAYETTE *et al.* 2015). While we do not know for certain the effect of our mutations have on the gene product of *lasR* itself, given that all but one (a large amplification that occurred in the MIN environment) were either nonsynonymous (n = 43), nonsense (n = 5), or small indels (n = 25), it seems likely that our collection includes many loss-of-function mutants as well. That *lasR* mutations evolved repeatedly in both SCFM and minimal media suggests that growth advantages associated with amino acid metabolism are not responsible for their prevalence *in vivo*. The complexity of quorum-sensing regulation in *P. aeruginosa* can allow mutations in *lasR* to have little or no effect on other quorum-sensing gene products (FELTNER *et al.* 2016), suggesting that the fitness advantage of a *lasR* mutants may derive simply from reducing the metabolic cost of expression. This is a hypothesis that awaits further testing.

The second is the prevalence of putatively antibiotic-resistant mutations in our experiment despite the fact that antibiotics were not present at any time in our experiment. Specifically, we observed the evolution

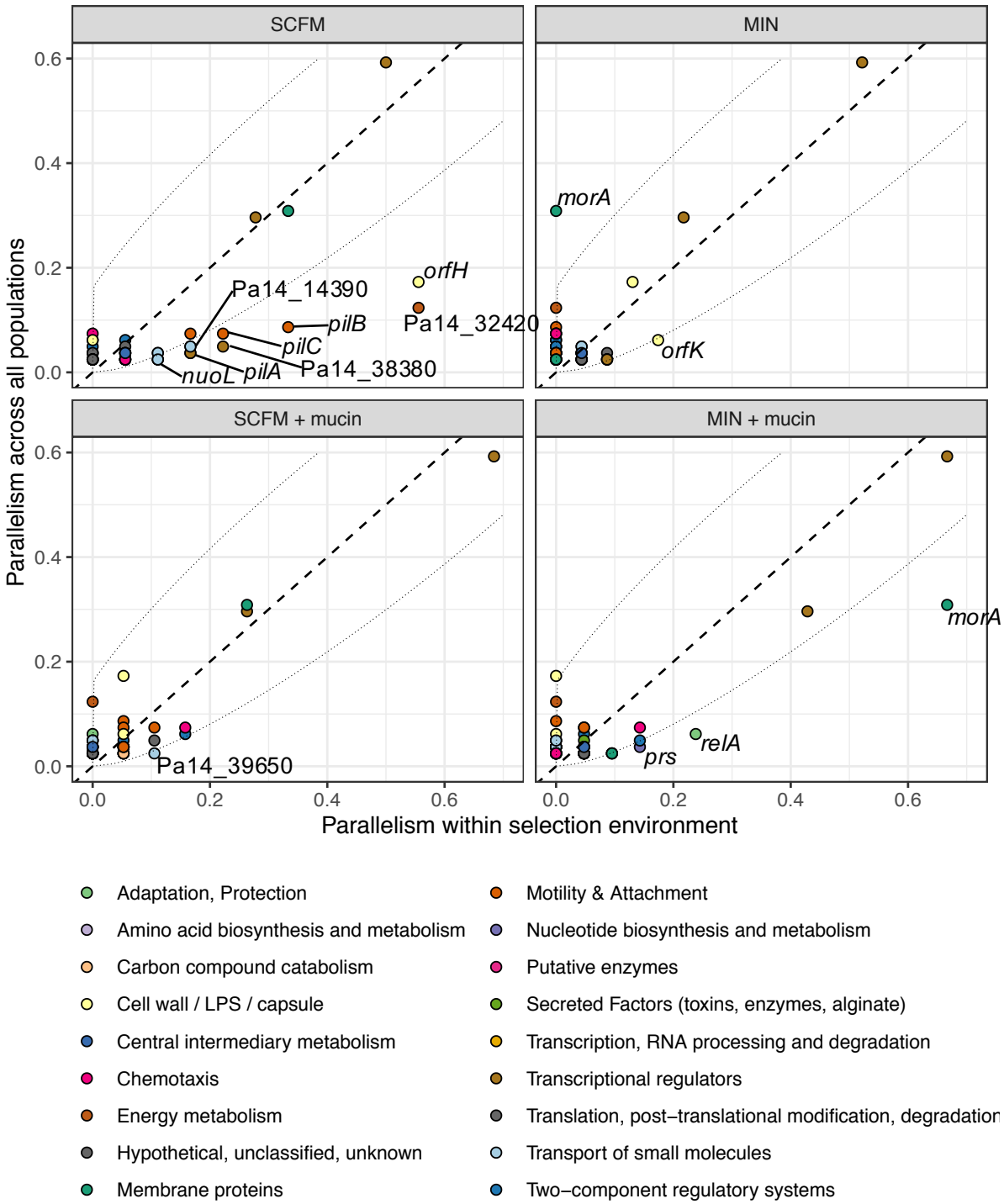

Figure 4.6: Gene-level parallelism. Parallelism within selection environment versus parallelism across all populations, for all genes with variants in more than one population, after excluding synonymous SNPs. Parallelism for a gene is defined as the proportion of populations with mutations in that gene. Dashed lines represent the line where parallelism within selection environment is equal to parallelism across all populations; points below this line are genes that are more commonly mutated in that environment and points above this line are genes less commonly mutated. Dotted lines represent statistical significance, based on the binomial distribution and an alpha = 0.05. Genes that fall outside the dotted lines are labelled.

of resistance to the fluoroquinolone ciprofloxacin, commonly used to manage *P. aeruginosa* infections in CF patients, in three evolved populations from SCFM (SCHICK and KASSEN 2018). Our genomic analysis points to mutations in *mexT*, a transcriptional regulator of the MexEF-OprN efflux pump, as the likely cause (BREIDENSTEIN *et al.* 2011). Two of the three distinct nonsynonymous SNPs were present at high frequencies of 0.89 and 0.71, suggesting strong selection at this locus, with the third present at a lower frequency of 0.09. The mechanism by which mutations in this gene confer an antibiotic-independent increase in fitness is unknown, but it could be a response to disulfide stress (FARGIER *et al.* 2012) or, as in the case of *lasR*, simply a reduction in the metabolic cost of maintaining an active regulatory response. Regardless of the specifics of the mechanism, our results suggest that mutations in *mexT* may evolve for reasons other than antibiotic selection. Others have observed the evolution of resistance in the absence of drug selection as well, and the genetic causes are often due to mutations in housekeeping genes like *rpoB* and *gyrA* (reviewed in HERSHBERG 2017. To the best of our knowledge, ours is the first study to show that *mexT* mutations can be selected in the absence of drug as well.

4.5 Conclusion and clinical significance

Our work provides a snapshot of genomic changes associated with rapid adaptation and diversification in populations of *P. aeruginosa* evolving under conditions that mimic, to varying degrees, the nutritional complexity and viscosity of the CF lung environment. Our leading result is that genetic variation is abundant across all conditions including those resembling the CF lung and, unlike what was observed for phenotypic disparity (SCHICK and KASSEN 2018), is not substantially higher in CF-like conditions. This result is attributable to high mutation supply rates resulting from large population sizes (~10^8 CFU/ml) and consistent with that seen in other microbial evolution experiments (GOOD *et al.* 2017; SCHICK *et al.* 2015). Because population densities in our experiment are, by design, similar to those in chronically-infected patients, this result suggests that *P. aeruginosa* populations in chronic infections can be highly diverse due to mutation alone. Genetic diversity may be higher still if distinct strains co-infect a single patient (DIAZ CABALLERO *et al.* 2015) and recombination can generate additional variants (DARCH *et al.* 2015).

We also saw low levels of genomic repeatability across all environments, a result in line with what is often seen in evolve-and-resequence experiments in bacteria and yeast (BAILEY *et al.* 2017). Importantly, however, evolution in CF-like conditions causes parallelism to be substantially lower than in the MIN environment, the least CF-like treatment in our experiment. This result is attributable both to divergent selection generated by nutritional complexity and spatial structure imposed by mucin. Divergent selection promotes the evolution of divergent niche specialists, reducing the likelihood of parallelism relative to directional selection towards a single fitness optimum. Mucin, for its part, reduces dispersal and creates spatially structured subpopulations with smaller effective population sizes than would be found under well-mixed conditions, making it less likely that the same beneficial mutations can be found and fixed by selection repeatedly in independently-evolved populations. Evidently, adaptation to the CF airway can follow many genetic routes which can make it difficult to identify reliable genomic signals marking the transition from transient to chronic infection.

Despite the low levels of parallelism, on average, in our experiment, we did recover mutations in a num-

ber of genes thought to be important during adaptation to the CF airway. Many of these genes are likely to impact regulatory functions associated with quorum sensing or motility (*lasR*, *mvfR*, *wspF*) or motility itself (*wspA*, *pilA-D*). Importantly few, if any, of these mutations were specific to CF-like conditions and some, like *lasR*, even evolved repeatably across every environment in our experiment. These results suggest that many of the genetic changes observed in our experiment represent general adaptations to novel conditions rather than specific adaptations to nutrient complexity or mucin in the CF lung. Indeed, changes to patterns of gene regulation, often mediated by loss-of-function mutations in non-essential genes, are commonly observed during the early stages of adaptation to novel, stressful environments in many selection experiments (DETTMAN *et al.* 2012; KASSEN 2014). Our results suggest the same might be true for the suite of genetic changes commonly recovered among *P. aeruginosa* isolates from CF patients. If so, then adaptation to the CF lung may be best understood as a particular instance of the more general phenomenon of adaptation to a novel, stressful environment.

The strength of these inferences must, of course, be tempered by the fact that our experiment was done in the lab under conditions that are substantially different from those encountered in the CF lung. Most obviously, the populations we have studied here have evolved in the absence of an active immune system, competing microflora, or the regular administration of antibiotics. It is not immediately obvious how the quantity of genetic variation and degree of parallelism would respond to these additional sources of selection. On the one hand we might expect lower levels of genetic variation and higher parallelism if these stressors represent additional filters that favour only those genotypes that can withstand the multiple sources of selection in the CF lung. Alternatively, if no single genotype is superior across all niche dimension, then these additional stressors could serve to actively preserve genetic diversity in the CF airway. Nevertheless, it is clear that the large population sizes typical of *P. aeruginosa* infections can supply high levels of genetic variation through mutation alone, a necessary condition for adaptive evolution through selection. This interpretation lends support to the idea that the dynamics of genetic variation among *P. aeruginosa* in the CF lung are caused mainly by mutation-driven selection.

Chapter 5

Discussion

In this thesis, I have addressed several fundamental questions about the dynamic process of diversification in microbial populations. By employing the powerful tools of evolve and resequence experiments, I have been able to achieve high levels of replication and observe evolutionary outcomes unfold in real time. Combined with the appropriate statistical analyses, these tools can be used to address a very wide range of questions regarding evolutionary dynamics. Here, I have focused on the diversity that arises de novo in evolving populations. To do this, I used two closely related species of bacteria: *Pseudomonas fluorescens* and *Pseudomonas aeruginosa*. Microbial model systems are an invaluable resource for inference based on evolve and resequence experiments due to their ease of manipulation in the laboratory and to the wealth of genomic resources for reference. Further to this, using an important human pathogen, *P. aeruginosa*, allowed me to address questions about diversification in the context of human health and medicine. This work has addressed diversification in heterogeneous environments, focusing on the role of ecological conditions and specialization in those environments. The chapters of my thesis have quantified the fitness effects of adaptive mutations in various environments (Chapter 2), investigated the ecological factors driving phenotypic diversification (Chapter 3), and explored the characteristics and repeatability of the genetic changes underlying that diversity (Chapter 4). In this last chapter, I summarize the main findings of each chapter, expand on the interpretations and limitations of these findings and discuss the overall contribution of this research to the field of evolutionary biology.

5.1 Thesis summary

5.1.1 Chapter 2: Environment-dependent fitness effects of adaptive mutations

In Chapter 2, I explored how the fitness effects of adaptive mutations depend on the environment. I quantified the degree to which adaptive mutations were specialized by measuring fitness effects in multiple environments. To do this, I used populations of *Pseudomonas fluorescens* that had previously been evolved for specialization on one of three different carbon sources: glucose, mannose, or xylose. The degree of specialization of a glucose-adapted mutation, for example, was then determined by the fitness effect of that mutation in mannose and xylose environments. I found that antagonistic effects were common; mutations with positive fitness effects in their environment of selection were often deleterious in the other environments. Further, when I compared the specialization of mutations that arose later during adaptation to those that arose earlier, I found that the deleterious effects increased over the course of adaptation. For the case of multiple adaptive mutations arising in the same genetic background, I found that the combined fitness effects were often less than additive, consistent with a commonly observed pattern of diminishing-returns epistasis. Lastly, I found

that the degree to which mutations were specialized was dependent on the predicted function of the gene in which the mutation was found.

By measuring the fitness effects of a mutation in multiple environments, we can determine the pleiotropic effects of that mutation. Adaptive mutations are beneficial in the environment in which they arose (by definition), irrespective of their fitness effect in any other environment. Mutations that have the same effect in other environments are generalist mutations, while mutations that have neutral or deleterious effects in other environments are specialists. General theoretical predictions for how specialization changes over the course of adaptation come from a multi-optima extension of Fisher's geometrical model of adaptation (FISHER 1930; MARTIN and LENORMAND 2015). In this extension, we consider an adaptive landscape with multiple peaks that are relatively close and a population that is initially far from both. Using this simplified metaphor, we can see that early mutations would bring a population closer to both optima but that subsequent steps towards one peak would result in movement away from the other. In other words, mutations would be predicted to become more specialized over the course of adaptation. Was this the case in my experiments? Yes, I did find that second-step mutations were much more likely to have deleterious effects in other environments than first-step mutations. In the context of laboratory evolution experiments, this is likely due to the similarity in environments tested and the emergence of 'domestication' mutations that are adaptive to the general conditions of growth, whose effects are likely to be similar under different carbon sources but may vary in vastly different environments.

How do the patterns of specialization I observed compare to other studies? There remains an absence of studies that examine fitness trade-offs over multiple steps of adaptation, but many studies have found that antagonistic pleiotropic effects are common in microbial populations during adaptation to different temperatures (CASPETA and NIELSEN 2015; DEATHERAGE *et al.* 2017), between classes of antibiotics (PÁL *et al.* 2015; STEINBERG and OSTERMEIER 2016), and between hosts for RNA viruses (BEDHOMME *et al.* 2015). Despite the numerous examples however, there is some disagreement; some studies show that antagonistic pleiotropy is rare and suggest that it's prevalence may be over-estimated as a result of observation bias and the focus on alleles of large effect (DILLON *et al.* 2016; SANE *et al.* 2018). Making any generalizations about the prevalence of antagonistic pleiotropy is difficult, largely due to a trade-off in depth and scale of investigation. For example, the number of mutations I examined was small (n = 23), limited by the need to construct strains using allelic replacement in order to measure direct fitness effects. Further, nearly all of the studies mentioned here involve laboratory-evolved populations and may not extend to natural populations that are already somewhat adapted to their environment. For this reason, these findings may be more representative of a population colonizing a novel environment, explored in much further detail in Chapters 3 and 4.

5.1.2 Chapter 3: Ecological factors that drive phenotypic diversification

In Chapter 3, I investigated the effect of two ecological factors, spatial structure and nutrient complexity, on diversification in populations of *Pseudomonas aeruginosa*. By allowing initially homogenous populations to evolve under different conditions, I was able to simulate the early stages of colonization and adaptation to a novel environment. This work was done in a laboratory environment with conditions designed to mimic two

ecological aspects of the CF lung: the spatial structure generated by both different compartments and thick mucous within those compartments, and the complex variety of nutrients available for metabolism. I found that populations rapidly became phenotypically diverse, in line with studies that find massive amounts of diversity within the CF lung. By using experimental evolution, however, I was able to replicate and analyze this diversification on a larger scale than has been achieved using strains from human subjects. Previously, most explanations of within-human diversification of *P. aeruginosa* have focused on environmental heterogeneity in selection as a result of the spatially structured environment of the lung. Much less attention has been paid to the complex resources present in the lung and the effect they may have on adapting pathogen populations. Surprisingly, I found that diversification was driven more by nutritional complexity than by spatial structure, suggesting that this aspect of ecology plays a key role in the CF lung.

The experimental conditions I used for this chapter were comprised of the same spectrum of resources found in the human lung. However, these environments did not include many aspects of the ecology of the lung that are almost certainly involved in driving adaptation and diversification, including the presence of antibiotics, immune system action, oxidative stress, oxygen availability, and competition with other microbial species. Despite the vast oversimplification of the lung environment, I found that many of the phenotypic changes that arose in my lab-evolved populations also commonly arise *in vitro*. This exciting observation suggests that these phenotypic changes are a result of *P. aeruginosa* colonizing a new nutrient-rich environment, and not necessarily specifically adaptive to the conditions of the human lung. An interesting extension of this chapter would be to include other ecological factors such as a microbial competitor or a fluctuating antibiotic treatment. However, due to the often factorial design and desire for large sample sizes, adding variables quickly becomes less practical, a common limiting factor in designing ER experiments. Further, there is often more data than can be analyzed. For example, I included both a static and shaken treatment in this experiment, doubling the size of the 2x2 factorial design, but was unable to analyze the static treatment in any depth beyond assaying colony morphology (summarized in Figure B.7).

One major limitation of inference from the results of this chapter is the short time period that populations were allowed to evolve, approximately 220 generations. It is difficult to measure growth rates *in vivo*, but one study estimates the doubling time of *P. aeruginosa* to be about 100 - 200 minutes, corresponding to about 14 generations per day (YANG *et al.* 2008). Based on this estimate, the duration of my selection experiment corresponds to about 16 days of adaptation in the CF lung. Therefore, the dynamics of adaptation and diversification I have presented here would only be applicable to the very initial stages of colonization. Though there is ample evidence of long-term persistence of polymorphic populations of *P. aeruginosa* in the human lung, the fate of the diversity in my experiments is unclear. Would it continue to increase over time, reach a stable equilibrium, or even decrease? In a recent experiment that tracked diversity in evolving populations of yeast using ~500,000 unique barcodes, for example, diversity was found to peak at around 200 generations and then crash due to a rare combination of mutations that were super fit (BLUNDELL *et al.* 2019). Going forward, it would be interesting to increase the length of this experiment as well as characterize the functionality of the diversity observed in order to determine its evolutionary fate.

5.1.3 Chapter 4: Genetic changes underlying diversifying populations

In Chapter 4, I explored the genetic changes associated with the patterns of diversification observed in Chapter 3. To do this, I sequenced whole-genomes of several populations at the endpoint of the selection experiment. By sequencing at the population level, I was able to obtain estimates of the frequency of each mutation present. I found extensive amounts of genetic variation, with the majority of mutations present at intermediate frequencies, likely due to a relatively high beneficial mutation supply rate. Using a combination of the ratio of nonsynonymous to synonymous mutations and the frequencies of those mutations, I estimated the ratio of adaptive mutations (drivers) to neutral mutations (hitchhikers). Comparing these ratios across treatments, I found that populations evolved under increased spatial structure provided by the presence of mucin had a significantly lower proportion of driver mutations. The increased viscosity of the medium reduces dispersal and therefore reduces the effective population size, likely resulting in a decreased supply of beneficial mutations. This result points to an effect of spatial structure that is likely an important factor in the CF lung. Though a much more dramatic case of spatial structure, it has been hypothesized that the compartmentalized nature of the lungs and surrounding airways results in heterogeneity in exposure to antibiotics that leads to 'reservoirs' of susceptible types (BREIDENSTEIN *et al.* 2011; CLARK *et al.* 2015).

When examining levels of genetic diversity present in evolved populations, I found no significant effect of treatment on genetic diversity either within or among populations, though I did observe a strong correlation between phenotypic and genetic diversity of a population. The lack of treatment effect is in contrast to the phenotypic data, in which I found a significant effect of nutritional content on diversity within populations and an effect of both nutritional content and spatial structure on diversity among populations. The comparison of phenotypic and genetic diversity is summarized in Figure 5.1. Because much of the genetic variation I observe is neutral, I hypothesize that the genetic changes underlying the significant phenotypic variation are only subset of the mutations I found, whose signal is being drowned out by hitchhiker mutations. Characterization of the effects of these genetic variants is needed to confidently connect genetic and phenotypic patterns of diversification. Furthermore, similar to the phenotypic results of Chapter 3, the fate of the genetic variation observed cannot be determined without this functional characterization.

Using a variety of metrics for quantifying parallelism, I found that populations evolved in spatially structured environments were consistently less parallel and therefore more divergent. Initially, I expected that both spatial structure and nutritional complexity would lead to increased divergence between populations. However, this result is consistent with the hypothesis that reduced dispersal among subpopulations allows for more divergent outcomes. In repeatability research, parallelism is often quantified by comparing gene re-use between two organisms or populations. If mutations arise independently in the same gene, parallel evolution has occurred. This data is typically binary, considering only whether or not a gene was mutated and is focused on repeatability of adaptation in general, ignoring the repeatability of diversification specifically. When quantifying repeatability of diversification, the relevant data is instead the frequencies of the polymorphisms present in populations. If repeated parallel adaptation is considered evidence of strong selection, I suggest that repeated diversification be considered evidence of selection for diversity. However, there is a distinct lack of statistical methods for quantifying the degree of parallelism in polymorphism. With my data, I attempted to compare parallelism at the gene level to frequency (Figure C.5), but did not have enough data

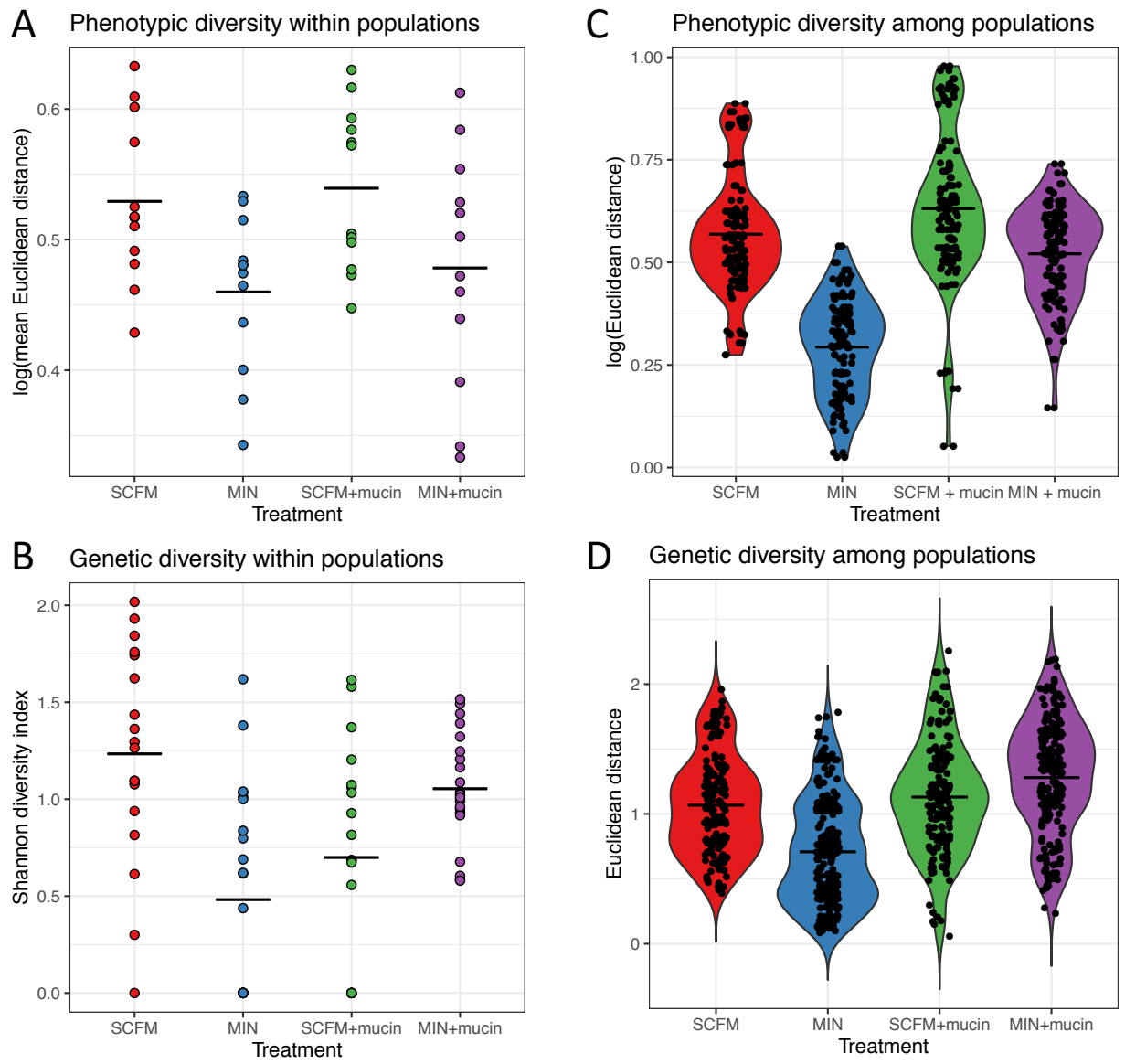

Figure 5.1: Phenotypic diversity and genetic diversity both within and among populations. A. Phenotypic diversity within populations is determined by the mean Euclidean distance between the phenotypes measured all pairs of individuals within a population. B. Genetic diversity within populations is calculated for each population by the Shannon diversity index, based on genetic variants at the gene-level. C. Phenotypic diversity among populations is determined by measuring pairwise Euclidean distance between the phenotypes measured for all populations within a treatment group. D. Genetic diversity among populations is determined by the Euclidean distance between all pairs of populations within a treatment, based on a matrix of genetic variants a the gene level. Horizontal bars represent treatment means. Note that panels A and C are included in Chapter 3, included here for comparison.

points to make any statistics-based inferences. As data from larger-scale ER experiments mounts, it will be interesting to quantify repeatability of diversification.

5.2 Drivers of diversification in microbial populations

There are several mechanisms that generate diversity in evolving populations, discussed in Chapter 1. Originally, one of the major motivating research questions of this work was to determine the relative contributions of these mechanisms. More specifically, I sought to disentangle diversity present as a result of selection in heterogeneous environments from diversity present as a result of ecological interactions within a population (such as competition for resources). In other words, to separate the effects of an extrinsic selection pressure generated by environmental conditions from the effects of an intrinsic selection pressure generated by dynamics within a population. Using the adaptive landscape metaphor, this can be described as divergent selection towards two fitness peaks versus disruptive selection away from a fitness minimum. To compare the effects of 'divergent' and 'disruptive' selection, I designed an experiment that compares the effects of spatial structure and nutritional complexity on diversification. The logic was that diversity that arose as a result of spatial structure was likely in response to divergent selection while diversity that arose as a result of nutritional complexity was more likely a response to disruptive, competition selection. What I quickly discovered, however, was that these two 'types' of selection are the same, depending on the scale at which you choose to define them. Competition for resources can result in a heterogeneous environment, indistinguishable from the heterogeneity present from all other ecological factors. For this reason, the factors generating diversity investigated in this work are simply the spatial structure and the nutritional complexity of the environment instead of the initially more arbitrary factors of 'extrinsic' and 'intrinsic' selection.

5.2.1 Differentiating clonal interference and adaptive diversity

Another major motivating research question was to determine the role of ecological specialization in diversification. Diversity as a result of specialization is adaptive, stably maintained diversity, in contrast to neutral, transient diversity that can arise as a byproduct of adaptation, referred to as clonal interference. When quantifying diversity, it is important to note that both niche-based ecological diversity and clonal interference would lead to the same pattern of multiple lineages competing for fixation. The difference is important; one is a byproduct of adaptation and one is adaptation. In Chapter 4, I observed a large amount of genetic variation in evolved populations, with the majority of mutations present at intermediate to low frequencies. In the absence of time series data, how can we determine whether mutations are being stably maintained by selection or are increasing in frequency and have not yet fixed? One method, described in the summary of Chapter 4, could be to investigate the relationship between parallelism and frequency. The logic here is that variants that are both highly parallel and present at intermediate frequencies are more likely to be under balancing selection than those at high frequencies with no evidence of parallelism. Unfortunately, I did not have enough data to test this hypothesis, though there are a number of other studies in which variants are maintained for long periods of time at low frequencies, often through the action of negative frequency-dependent selection (BEHRINGER *et al.* 2018; GOOD *et al.* 2017). A second method to distinguish the cause of mul-

tiple lineages would be to use invasion experiments to determine if specific mutations are under balancing selection, though these experiments are often limited to two competing lineages, often missing much of the population composition in which that polymorphism arose.

5.2.2 Evidence of resource specialization

In Chapter 2, I found strong evidence of resource specialization, demonstrated by antagonistic pleiotropic fitness effects of adaptive mutations. Perhaps unsurprisingly, mutations had positive fitness effects when measured in the environment in which they arose and negative fitness effects in environments containing alternative carbon sources. In this case, specialization contributes to diversification because it generates fitness trade-offs when conditions change. In Chapters 3 and 4, it is highly likely that resource specialization is contributing to diversification, but I do not have direct evidence of this. In an attempt to quantify the degree to which populations are niche-specialized, I sought to quantify the diversity of within-population auxotrophy, the dependence of isolates on specific amino acids. I hypothesized that populations with more phenotypic diversity, based on the traits I examined, would also exhibit more diversity in amino-acid use. However, when conducting a pilot experiment to test for auxotrophy, I found that all isolates were prototrophic, able to grow in minimal media without any supplemental amino acids. In other words, I found no variation in amino-acid use among evolved isolates. However, this experiment was conducted over a 24 hour period of growth, with isolates being deemed as auxotrophic or prototrophic based on whether or not any amount of growth occurred. Without measurements of growth rates in different environments, I am not able to quantify the degree of ecological specialization in the evolved populations, an extension of this work that would be fascinating to carry out.

5.3 The Emerging field of microbial evolutionary medicine

A major focus of preventing pathogen infection has been on the efficacy of various antimicrobial treatments. Though this work has led to vast improvements of the life-expectancy of people with cystic fibrosis, in the long-term, treating and controlling *Pseudomonas aeruginosa* infections with antibiotics is not sustainable. Classified by the World Health Organization as one of three top priority 'critical' pathogens, this superbug is often resistant to multiple antibiotics and if susceptible, able to become resistant quickly via a large collection of genes in which mutations confer resistance (BREIDENSTEIN *et al.* 2011). With limited prospects of novel antibiotics being developed (BROWN and WRIGHT 2016), there is a need to shift focus. Though no small task, controlling the spread of infection might be accomplished instead by controlling or manipulating the ecological conditions of the environment. The complex selective forces acting on a pathogen are often overlooked, aside from antimicrobial resistance, but could be a key part of the equation of disease prevention.

5.3.1 Role of microbial diversity in health

Generally speaking, within the field of evolutionary biology, biodiversity is good. A large portion of the global scientific community is concerned with maintaining biodiversity. In this context, biodiversity generally refers to a wide array of species. For human microbial pathogens however, diversity may not always be

a good thing. In the human gut, it is thought that more diverse communities are healthier. This comes from data that shows an association between a diseased state and a less diverse microbiome dominated by few species, as in the case of *C. difficile* infections, for example (SMITS *et al.* 2016). Though a clear statistical association exists in many cases, a causal relationship between gut microbiome diversity and health is difficult to demonstrate. Interestingly, the human vaginal microbiome seems to have the opposite relationship with diversity; a healthy vaginal community of microbes is dominated by a single species of the genus *Lactobacillus* and a community associated with bacterial vaginosis is much more diverse (RAVEL *et al.* 2013). In the human lung, similar to the gut, diseased states are associated with an over-representation of a particular species, *P. aeruginosa* in the case of CF, for example. However, within the species, more diverse populations are associated with more advanced states of infection, but again, there is no evidence of a causal relationship.

With this unknown relationship in mind, one much needed extension of this work is to investigate the relationship between diversity and virulence. Some novel methods are being developed to track virulence of *P. aeruginosa* over time, such as infecting mung bean seedlings with different strains or different compositions of strains (GARGE *et al.* 2018). Using this method, root development and mortality rates are being used as a proxy for virulence, allowing for direct tests of the effects of populations with varying amounts of inter- and intra-species diversity. With these methods, it will be interesting to see a comparison between clonal and diverse populations of *P. aeruginosa*, as well as between diverse populations evolved for varying numbers of generations.

This thesis investigates the dynamics of populations containing only a single species. The complexity of evolution in single-species populations demonstrates how challenging a task it is to understand the eco-evolutionary dynamics in communities with several hundred or thousand species. Thus far, there have been attempts from both directions. Medium-throughput studies, such as the work I present here, provide detailed insight into dynamics of a part of the community in isolation. Other similarly scaled studies are working to elucidate interactions between two species of the community (KORGAONKAR *et al.* 2013; SIBLEY *et al.* 2008). On the other end of the scale, high-throughput studies such as 16S sequencing of complete microbiome communities provide a characterization of the community composition, but are limited in making inferences about the eco-evolutionary dynamics. Both of these approaches are necessary and will, hopefully, become integrated into a more complete picture over time.

5.4 Conclusions

The work presented here has implications for understanding the emergence of diversity in microbial populations during adaptation to complex, novel environments. In general, I found that context matters. Various ecological aspects of the environment can promote diversification, and this can process can be mediated by specialization towards those aspects. In the context of understanding adaptation of *Pseudomonas aeruginosa* to the CF lung, my work provides two main insights. First, I show that the genetic and phenotypic changes often attributed to adaptation to the CF lung environment are simply byproducts of adaptation to a novel, resource-rich environment. For example, I observed increased virulence production in the absence of immune system pressure and increased antibiotic resistance in the absence of antibiotics. Second, by demon-

strating the effect of complex ecological conditions on diversification, this work highlights the importance of taking an eco-evolutionary approach to investigating pathogen adaptation. Colonizing a new environment or acquiring antimicrobial resistance, for example, is a process that involves myriad interactions within a constantly evolving community.

Much like the feedback loop that exists between ecological and evolutionary processes, research in this field is also a feedback loop between theory and experiment. Though the theoretical mechanisms underlying evolutionary dynamics may be well understood individually, the combination of them, as is almost always the case in natural systems, produces outcomes that are far from predictable. As theoretical models improve and incorporate more complex dynamics, experimental tests of their predictions are needed to improve those models and allow them to incorporate even more complex dynamics. Though small in scope, this body of work provides an experimental piece of the puzzle that is ultimately to understand and predict the emergence of diversity.

Bibliography

AARON, S. D., K. L. VANDEMHEEN, K. RAMOTAR, T. GIESBRECHT-LEWIS, E. TULLIS, *et al.*, 2010 Infection with transmissible strains of Pseudomonas aeruginosa and clinical outcomes in adults with cystic fibrosis. Jama **304**: 2145–2153.

AGASHE, D., M. SANE, K. PHALNIKAR, G. D. DIWAN, A. HABIBULLAH, *et al.*, 2016 Large-Effect Beneficial Synonymous Mutations Mediate Rapid and Parallel Adaptation in a Bacterium. Molecular Biology and Evolution **33**: 1542–1553.

ALANIS, A. J., 2005 Resistance to antibiotics: are we in the post-antibiotic era? Archives of medical research **36**: 697–705.

ANDERSON, J. T., C.-R. LEE, C. A. RUSHWORTH, R. I. COLAUTTI, and T. MITCHELL-OLDS, 2013 Genetic trade-offs and conditional neutrality contribute to local adaptation: GENETIC BASIS OF LOCAL ADAPTATION. Molecular Ecology **22**: 699–708.

ASHISH, A., S. PATERSON, E. MOWAT, J. L. FOTHERGILL, M. J. WALSHAW, *et al.*, 2013 Extensive diversification is a common feature of Pseudomonas aeruginosa populations during respiratory infections in cystic fibrosis. Journal of Cystic Fibrosis **12**: 790–793.

ATWOOD, K., L. SCHNEIDER, and F. RYAN, 1951 Periodic selection in escherichia coli. Proceedings of the National Academy of Sciences **37**: 146–155.

BAILEY, S. F., F. BLANQUART, T. BATAILLON, and R. KASSEN, 2017 What drives parallel evolution? how population size and mutational variation contribute to repeated evolution. BioEssays **39**: 1–9.

BAILEY, S. F., A. HINZ, and R. KASSEN, 2014 Adaptive synonymous mutations in an experimentally evolved Pseudomonas fluorescens population. Nature Communications **5**.

BAILEY, S. F., and R. KASSEN, 2012 Spatial Structure of Ecological Opportunity Drives Adaptation in a Bacterium. The American Naturalist **180**: 270–283.

BAILEY, S. F., N. RODRIGUE, and R. KASSEN, 2015 The Effect of Selection Environment on the Probability of Parallel Evolution. Molecular Biology and Evolution **32**: 1436–1448.

BANTINAKI, E., R. KASSEN, C. G. KNIGHT, Z. ROBINSON, A. J. SPIERS, *et al.*, 2007 Adaptive Divergence in Experimental Populations of *Pseudomonas fluorescens* . III. Mutational Origins of Wrinkly Spreader Diversity. Genetics **176**: 441–453.

BARROSO-BATISTA, J., A. SOUSA, M. LOURENÇO, M.-L. BERGMAN, D. SOBRAL, *et al.*, 2014 The First Steps of Adaptation of Escherichia coli to the Gut Are Dominated by Soft Sweeps. PLoS Genetics **10**: e1004182.

BATAILLON, T., T. ZHANG, and R. KASSEN, 2011 Cost of Adaptation and Fitness Effects of Beneficial Mutations in *Pseudomonas fluorescens*. Genetics **189**: 939–949.

BECK, J. M., V. B. YOUNG, and G. B. HUFFNAGLE, 2012 The microbiome of the lung. Translational Research **160**: 258–266.

BEDHOMME, S., J. HILLUNG, and S. F. ELENA, 2015 Emerging viruses: why they are not jacks of all trades? Current Opinion in Virology **10**: 1–6.

BEHRINGER, M. G., B. I. CHOI, S. F. MILLER, T. G. DOAK, J. A. KARTY, *et al.*, 2018 *Escherichia coli* cultures maintain stable subpopulation structure during long-term evolution. Proceedings of the National Academy of Sciences **115**: E4642–E4650.

BLUNDELL, J. R., K. SCHWARTZ, D. FRANCOIS, D. S. FISHER, G. SHERLOCK, *et al.*, 2019 The dynamics of adaptive genetic diversity during the early stages of clonal evolution. Nature ecology & evolution **3**: 293.

BOLGER, A. M., M. LOHSE, and B. USADEL, 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30**: 2114–2120.

BREIDENSTEIN, E. B., C. DE LA FUENTE-NÚÑEZ, and R. E. HANCOCK, 2011 Pseudomonas aeruginosa: all roads lead to resistance. Trends in Microbiology **19**: 419–426.

BROCKHURST, M. A., A. BUCKLING, and P. B. RAINEY, 2005 The effect of a bacteriophage on diversification of the opportunistic bacterial pathogen, Pseudomonas aeruginosa. Proceedings of the Royal Society B: Biological Sciences **272**: 1385–1391.

BROWN, E. D., and G. D. WRIGHT, 2016 Antibacterial drug discovery in the resistance era. Nature **529**: 336.

BURROWS, L. L., 2012 *Pseudomonas aeruginosa* Twitching Motility: Type IV Pili in Action. Annual Review of Microbiology **66**: 493–520.

BUSKIRK, S. W., R. E. PEACE, and G. I. LANG, 2017 Hitchhiking and epistasis give rise to cohort dynamics in adapting populations. Proceedings of the National Academy of Sciences **114**: 8330–8335.

CASPETA, L., and J. NIELSEN, 2015 Thermotolerant yeast strains adapted by laboratory evolution show trade-off at ancestral temperatures and preadaptation to other stresses. MBio **6**: e00431–15.

CHEVIN, L.-M., G. MARTIN, and T. LENORMAND, 2010 Fisher's model and the genomics of adaptation: restricted pleiotropy, heterogenous mutation, and parallel evolution. Evolution: International Journal of Organic Evolution **64**: 3213–3231.

CHOU, H.-H., H.-C. CHIU, N. F. DELANEY, D. SEGRE, and C. J. MARX, 2011 Diminishing Returns Epistasis Among Beneficial Mutations Decelerates Adaptation. Science **332**: 1190–1192.

CIOFU, O., L. F. MANDSBERG, T. BJARNSHOLT, T. WASSERMANN, and N. HOIBY, 2010 Genetic adaptation of Pseudomonas aeruginosa during chronic lung infection of patients with cystic fibrosis: strong and weak mutators with heterogeneous genetic backgrounds emerge in mucA and/or lasR mutants. Microbiology **156**: 1108–1119.

CIOFU, O., L. F. MANDSBERG, H. WANG, and N. HØIBY, 2012 Phenotypes selected during chronic lung infection in cystic fibrosis patients: implications for the treatment of *Pseudomonas aeruginosa* biofilm infections. FEMS Immunology & Medical Microbiology **65**: 215–225.

CLARK, S. T., J. DIAZ CABALLERO, M. CHEANG, B. COBURN, P. W. WANG, *et al.*, 2015 Phenotypic diversity within a Pseudomonas aeruginosa population infecting an adult with cystic fibrosis. Scientific Reports **5**.

CLARK, S. T., D. S. GUTTMAN, and D. M. HWANG, 2018 Diversification of pseudomonas aeruginosa within the cystic fibrosis lung and its effects on antibiotic resistance. FEMS microbiology letters **365**: fny026.

COGEN, A., V. NIZET, and R. GALLO, 2008 Skin microbiota: a source of disease or defence? British Journal of Dermatology **158**: 442–455.

COOPER, V. S., 2018 Experimental evolution as a high-throughput screen for genetic adaptations. Msphere **3**: e00121–18.

DARCH, S. E., A. MCNALLY, F. HARRISON, J. CORANDER, H. L. BARR, *et al.*, 2015 Recombination is a key driver of genomic and phenotypic diversity in a Pseudomonas aeruginosa population during cystic fibrosis infection. Scientific Reports **5**.

D'ARGENIO, D. A., M. WU, L. R. HOFFMAN, H. D. KULASEKARA, E. DÉZIEL, *et al.*, 2007 Growth phenotypes of pseudomonas aeruginosa lasr mutants adapted to the airways of cystic fibrosis patients. Molecular microbiology **64**: 512–533.

DARWIN, C., 1859 *On the origin of species, 1859*. Routledge.

DAVIES, E. V., C. E. JAMES, M. A. BROCKHURST, and C. WINSTANLEY, 2017 Evolutionary diversification of Pseudomonas aeruginosa in an artificial sputum model. BMC Microbiology **17**.

DE VISSER, J. A. G., and D. E. ROZEN, 2006 Clonal interference and the periodic selection of new beneficial mutations in escherichia coli. Genetics **172**: 2093–2100.

DE VISSER, M., C. W. ZEYL, P. J. GERRISH, J. L. BLANCHARD, R. E. LENSKI, *et al.*, 1999 Diminishing returns from mutation supply rate in asexual populations. Science **283**: 404–406.

DE VOS, M. G., M. ZAGORSKI, A. MCNALLY, and T. BOLLENBACH, 2017 Interaction networks, ecological stability, and collective antibiotic tolerance in polymicrobial infections. Proceedings of the National Academy of Sciences **114**: 10666–10671.

DEATHERAGE, D. E., and J. E. BARRICK, 2014 Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. In *Engineering and analyzing multicellular systems*. Springer, 165–188.

DEATHERAGE, D. E., J. L. KEPNER, A. F. BENNETT, R. E. LENSKI, and J. E. BARRICK, 2017 Specificity of genome evolution in experimental populations of escherichia coli evolved at different temperatures. Proceedings of the National Academy of Sciences **114**: E1904–E1912.

DÉBARRE, F., S. NUISMER, and M. DOEBELI, 2014 Multidimensional (co) evolutionary stability. The American Naturalist **184**: 158–171.

DESAI, M. M., D. S. FISHER, and A. W. MURRAY, 2007 The speed of evolution and maintenance of variation in asexual populations. Current biology **17**: 385–394.

DETTMAN, J. R., N. RODRIGUE, S. D. AARON, and R. KASSEN, 2013 Evolutionary genomics of epidemic and nonepidemic strains of Pseudomonas aeruginosa. Proceedings of the National Academy of Sciences **110**: 21065–21070.

DETTMAN, J. R., N. RODRIGUE, A. H. MELNYK, A. WONG, S. F. BAILEY, *et al.*, 2012 Evolutionary insight from whole-genome sequencing of experimentally evolved microbes: EXPERIMENTAL EVOLUTION AND GENOME SEQUENCING. Molecular Ecology **21**: 2058–2077.

DETTMAN, J. R., J. L. SZTEPANACZ, and R. KASSEN, 2016 The properties of spontaneous mutations in the opportunistic pathogen pseudomonas aeruginosa. BMC genomics **17**: 27.

DIAZ CABALLERO, J., S. T. CLARK, B. COBURN, Y. ZHANG, P. W. WANG, *et al.*, 2015 Selective Sweeps and Parallel Pathoadaptation Drive *Pseudomonas aeruginosa* Evolution in the Cystic Fibrosis Lung. mBio **6**: e00981–15.

DILLON, M. M., N. P. ROUILLARD, B. VAN DAM, R. GALLET, and V. S. COOPER, 2016 Diverse phenotypic and genetic responses to short-term selection in evolving escherichia coli populations. Evolution **70**: 586–599.

DOEBELI, M., and U. DIECKMANN, 2000 Evolutionary branching and sympatric speciation caused by different types of ecological interactions. The american naturalist **156**: S77–S101.

DYKHUIZEN, D. E., 1990 Experimental Studies of Natural Selection in Bacteria. Annual Review of Ecology and Systematics **2**: 373–398.

ELENA, S., and R. LENSKI, 2003 Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nature Rev. Genet **4**: 457.

ELENA, S. F., and R. E. LENSKI, 1997 Long-term experimental evolution in escherichia coli. vii. mechanisms maintaining genetic variability within populations. Evolution **51**: 1058–1067.

FALCONER, D. S., 1960 *Introduction to quantitative genetics*. Oliver And Boyd; Edinburgh; London.

FARGIER, E., M. MAC AOGAIN, M. J. MOOIJ, D. F. WOODS, J. P. MORRISSEY, *et al.*, 2012 MexT Functions as a Redox-Responsive Regulator Modulating Disulfide Stress Resistance in Pseudomonas aeruginosa. Journal of Bacteriology **194**: 3502–3511.

FARHAT, M. R., B. J. SHAPIRO, K. J. KIESER, R. SULTANA, K. R. JACOBSON, *et al.*, 2013 Genomic analysis identifies targets of convergent positive selection in drug-resistant mycobacterium tuberculosis. Nature genetics **45**: 1183.

FELIZIANI, S., R. L. MARVIG, A. M. LUJÁN, A. J. MOYANO, J. A. DI RIENZO, *et al.*, 2014 Coexistence and Within-Host Evolution of Diversified Lineages of Hypermutable Pseudomonas aeruginosa in Long-term Cystic Fibrosis Infections. PLoS Genetics **10**: e1004651.

FELSENSTEIN, J., 1976 The theoretical population genetics of variable selection and migration. Annual review of genetics **10**: 253–280.

FELSENSTEIN, J., 1981 Skepticism Towards Santa Rosalia, or Why are There so Few Kinds of Animals? Evolution **35**: 124.

FELTNER, J. B., D. J. WOLTER, C. E. POPE, M.-C. GROLEAU, N. E. SMALLEY, *et al.*, 2016 LasR Variant Cystic Fibrosis Isolates Reveal an Adaptable Quorum-Sensing Hierarchy in *Pseudomonas aeruginosa*. mBio **7**.

FISHER, R. A., 1930 The genetical theory of natural selection. Clarendon, Oxford .

FOGLE, C. A., J. L. NAGLE, and M. M. DESAI, 2008 Clonal interference, multiple mutations and adaptation in large asexual populations. Genetics **180**: 2163–2173.

FOLKESSON, A., L. JELSBAK, L. YANG, H. K. JOHANSEN, O. CIOFU, *et al.*, 2012 Adaptation of Pseudomonas aeruginosa to the cystic fibrosis airway: an evolutionary perspective. Nature Reviews Microbiology **10**: 841–851.

FOTHERGILL, J. L., E. MOWAT, M. J. LEDSON, M. J. WALSHAW, and C. WINSTANLEY, 2010 Fluctuations in phenotypes and genotypes within populations of Pseudomonas aeruginosa in the cystic fibrosis lung during pulmonary exacerbations. Journal of Medical Microbiology **59**: 472–481.

FOWERAKER, J., C. LAUGHTON, D. BROWN, and D. BILTON, 2005 Phenotypic variability of pseudomonas aeruginosa in sputa from patients with acute infective exacerbation of cystic fibrosis and its impact on the validity of antimicrobial susceptibility testing. Journal of Antimicrobial Chemotherapy **55**: 921–927.

FOXMAN, B., 2010 The epidemiology of urinary tract infection. Nature Reviews Urology **7**: 653.

FRASER, D. J., L. K. WEIR, L. BERNATCHEZ, M. M. HANSEN, and E. B. TAYLOR, 2011 Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. Heredity **106**: 404–420.

FRENKEL, E. M., M. J. MCDONALD, J. D. VAN DYKEN, K. KOSHELEVA, G. I. LANG, *et al.*, 2015 Crowded growth leads to the spontaneous evolution of semistable coexistence in laboratory yeast populations. Proceedings of the National Academy of Sciences **112**: 11306–11311.

FRESCHI, L., C. BERTELLI, J. JEUKENS, M. P. MOORE, I. KUKAVICA-IBRULJ, *et al.*, 2018 Genomic characterisation of an international Pseudomonas aeruginosa reference panel indicates that the two major groups draw upon distinct mobile gene pools. FEMS Microbiology Letters **365**.

FRIESEN, M. L., G. SAXER, M. TRAVISANO, and M. DOEBELI, 2004 Experimental evidence for sympatric ecological diversification due to frequency-dependent competition in Escherichia coli. Evolution **58**: 245–260.

FRY, J. D., 1996 The Evolution of Host Specialization: Are Trade-Offs Overrated? The American Naturalist **148**: S84–S107.

FUNG, C., S. NAUGHTON, L. TURNBULL, P. TINGPEJ, B. ROSE, *et al.*, 2010 Gene expression of Pseudomonas aeruginosa in a mucin-containing synthetic growth medium mimicking cystic fibrosis lung sputum. Journal of Medical Microbiology **59**: 1089–1100.

FUTUYMA, D. J., and G. MORENO, 1988 The evolution of ecological specialization. Annual Review of Ecology and Systematics **19**: 207–233.

GARGE, S., S. AZIMI, and S. P. DIGGLE, 2018 A simple mung bean infection model for studying the virulence of pseudomonas aeruginosa. BioRxiv : 229757.

GERRISH, P. J., and R. E. LENSKI, 1998 The fate of competing beneficial mutations in an asexual population. Genetica **102**: 127.

GIBSON, R. L., J. L. BURNS, and B. W. RAMSEY, 2003 Pathophysiology and Management of Pulmonary Infections in Cystic Fibrosis. American Journal of Respiratory and Critical Care Medicine **168**: 918–951.

GILL, S. R., M. POP, R. T. DEBOY, P. B. ECKBURG, P. J. TURNBAUGH, *et al.*, 2006 Metagenomic analysis of the human distal gut microbiome. science **312**: 1355–1359.

GOOD, B. H., M. J. MCDONALD, J. E. BARRICK, R. E. LENSKI, and M. M. DESAI, 2017 The dynamics of molecular evolution over 60,000 generations. Nature **551**: 45–50.

GOOD, B. H., I. M. ROUZINE, D. J. BALICK, O. HALLATSCHEK, and M. M. DESAI, 2012 Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. Proceedings of the National Academy of Sciences **109**: 4950–4955.

GOULD, S. J., 1989 *Wonderful life, the Burgess shale and the nature of history*. W Norton.

GRANT, B., and P. GRANT, 1989 Natural selection in a population of darwin's finches. The American Naturalist **133**: 377–393.

GREISCHAR, M. A., and B. KOSKELLA, 2007 A synthesis of experimental work on parasite local adaptation. Ecology Letters **10**: 418–434.

GROUSSIN, M., F. MAZEL, J. G. SANDERS, C. S. SMILLIE, S. LAVERGNE, *et al.*, 2017 Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. Nature Communications **8**: 14319.

GUDELJ, I., R. E. BEARDMORE, S. ARKIN, and R. C. MACLEAN, 2007 Constraints on microbial metabolism drive evolutionary diversification in homogeneous environments. Journal of evolutionary biology **20**: 1882–1889.

HARRISON, F., 2007 Microbial ecology of the cystic fibrosis lung. Microbiology **153**: 917–923.

HASSETT, D. J., M. D. SUTTON, M. J. SCHURR, A. B. HERR, C. C. CALDWELL, *et al.*, 2009 Pseudomonas aeruginosa hypoxic or anaerobic biofilm infections within cystic fibrosis airways. Trends in Microbiology **17**: 130–138.

HEDRICK, P. W., M. E. GINEVAN, and E. P. EWING, 1976 Genetic-Polymorphism in Heterogeneous Environments. Annual Review of Ecology and Systematics **7**: 1–32.

HEREFORD, J., 2009 A Quantitative Survey of Local Adaptation and Fitness Trade Offs. The American Naturalist **173**: 579–588.

HERMISSON, J., and P. S. PENNINGS, 2005 Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics **169**: 2335–2352.

HERSHBERG, R., 2017 Antibiotic-Independent Adaptive Effects of Antibiotic Resistance Mutations. Trends in Genetics **33**: 521–528.

HOEKSEMA, J., and S. FORDE, 2008 A Meta-Analysis of Factors Affecting Local Adaptation between Interacting Species. The American Naturalist **171**: 275–290.

HOFFMAN, L. R., H. D. KULASEKARA, J. EMERSON, L. S. HOUSTON, J. L. BURNS, *et al.*, 2009 Pseudomonas aeruginosa lasR mutants are associated with cystic fibrosis lung disease progression. Journal of Cystic Fibrosis **8**: 66–70.

HOIBY, N., and T. PRESSLER, 2006 Emerging pathogens in cystic fibrosis. EUROPEAN RESPIRATORY MONOGRAPH **35**: 66.

JAMES, C. E., E. V. DAVIES, J. L. FOTHERGILL, M. J. WALSHAW, C. M. BEALE, *et al.*, 2015 Lytic activity by temperate phages of Pseudomonas aeruginosa in long-term cystic fibrosis chronic lung infections. The ISME Journal **9**: 1391–1398.

JASMIN, J.-N., and C. ZEYL, 2013 Evolution of pleiotropic costs in experimental populations. Journal of Evolutionary Biology **26**: 1363–1369.

JENSEN, P. O., M. GIVSKOV, T. BJARNSHOLT, and C. MOSER, 2010 The immune system vs. *Pseudomonas aeruginosa* biofilms. FEMS Immunology & Medical Microbiology **59**: 292–305.

JORTH, P., B. J. STAUDINGER, X. WU, K. B. HISERT, H. HAYDEN, *et al.*, 2015 Regional isolation drives bacterial diversification within cystic fibrosis lungs. Cell host & microbe **18**: 307–319.

KAO, K. C., and G. SHERLOCK, 2008 Molecular characterization of clonal interference during adaptive evolution in asexual populations of saccharomyces cerevisiae. Nature genetics **40**: 1499.

KASSEN, R., 2009 Toward a General Theory of Adaptive Radiation: Insights from Microbial Experimental Evolution. Annals of the New York Academy of Sciences **1168**: 3–22.

KASSEN, R., 2014 *Experimental evolution and the nature of biodiversity*. Roberts.

KASSEN, R., and T. BATAILLON, 2006 Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. Nature Genetics **38**: 484–488.

KAWECKI, T. J., 1994 Accumulation of Deleterious Mutations and the Evolutionary Cost of Being a Generalist. The American Naturalist **144**: 833–838.

KAWECKI, T. J., N. H. BARTON, and J. D. FRY, 1997 Mutational collapse of fitness in marginal habitats and the evolution of ecological specialisation. Journal of Evolutionary Biology **10**: 407–429.

KAWECKI, T. J., R. E. LENSKI, D. EBERT, B. HOLLIS, I. OLIVIERI, *et al.*, 2012 Experimental evolution. Trends in Ecology & Evolution **27**: 547–560.

KERR, B., M. A. RILEY, M. W. FELDMAN, and B. J. BOHANNAN, 2002 Local dispersal promotes biodiversity in a real-life game of rock–paper–scissors. Nature **418**: 171.

KHAN, A. I., D. M. DINH, D. SCHNEIDER, R. E. LENSKI, and T. F. COOPER, 2011 Negative Epistasis Between Beneficial Mutations in an Evolving Bacterial Population. Science **332**: 1193–1196.

KLOCKGETHER, J., N. CRAMER, L. WIEHLMANN, C. F. DAVENPORT, and B. TÜMMLER, 2011 Pseudomonas aeruginosa Genomic Structure and Diversity. Frontiers in Microbiology **2**.

KOEHORST, J. J., J. C. J. VAN DAM, R. G. A. VAN HECK, E. SACCENTI, V. A. P. M. DOS SANTOS, *et al.*, 2016 Comparison of 432 Pseudomonas strains through integration of genomic, functional, metabolic and expression data. Scientific Reports **6**.

KORGAONKAR, A., U. TRIVEDI, K. P. RUMBAUGH, and M. WHITELEY, 2013 Community surveillance enhances Pseudomonas aeruginosa virulence during polymicrobial infection. Proceedings of the National Academy of Sciences **110**: 1059–1064.

KRISTOFICH, J., A. B. MORGENTHALER, W. R. KINNEY, C. C. EBMEIER, D. J. SNYDER, *et al.*, 2018 Synonymous mutations make dramatic contributions to fitness when growth is limited by a weak-link enzyme. PLoS genetics **14**: e1007615.

LAFAYETTE, S. L., D. HOULE, T. BEAUDOIN, G. WOJEWODKA, D. RADZIOCH, *et al.*, 2015 Cystic fibrosis-adapted Pseudomonas aeruginosa quorum sensing lasR mutants cause hyperinflammatory responses. Science Advances **1**: e1500199–e1500199.

LANDRY, R. M., D. AN, J. T. HUPP, P. K. SINGH, and M. R. PARSEK, 2006 Mucin-Pseudomonas aeruginosa interactions promote biofilm formation and antibiotic resistance. Molecular Microbiology **59**: 142–151.

LANG, G. I., D. P. RICE, M. J. HICKMAN, E. SODERGREN, G. M. WEINSTOCK, *et al.*, 2013 Lang2013Nature. Nature **500**: 571–574.

LEALE, A. M., and R. KASSEN, 2018 The emergence, maintenance, and demise of diversity in a spatially variable antibiotic regime. Evolution Letters **2**: 134–143.

LEBEUF-TAYLOR, E., N. MCCLOSKEY, S. F. BAILEY, A. HINZ, and R. KASSEN, 2019 The distribution of fitness effects among synonymous mutations in a gene under directional selection. eLife **8**: e45952.

LEIMU, R., and M. FISCHER, 2008 A Meta-Analysis of Local Adaptation in Plants. PLoS ONE **3**: e4010.

LENSKI, R. E., 2017 Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. The ISME Journal **11**: 2181–2194.

LEWONTIN, R. C., *et al.*, 1974 *The genetic basis of evolutionary change*, volume 560. Columbia University Press New York.

LI, H., B. HANDSAKER, A. WYSOKER, T. FENNELL, J. RUAN, *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**: 2078–2079.

LIEBERMAN, T. D., K. B. FLETT, I. YELIN, T. R. MARTIN, A. J. MCADAM, *et al.*, 2014 Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. Nature genetics **46**: 82.

LOBKOVSKY, A. E., and E. V. KOONIN, 2012 Replaying the tape of life: quantification of the predictability of evolution. Frontiers in genetics **3**: 246.

LONG, A., G. LITI, A. LUPTAK, and O. TENAILLON, 2015 Elucidating the molecular architecture of adaptation via evolve and resequence experiments. Nature Reviews Genetics **16**: 567–582.

LOSOS, J. B., 1992 The evolution of convergent structure in caribbean anolis communities. Systematic biology **41**: 403–420.

MADDAMSETTI, R., R. E. LENSKI, and J. E. BARRICK, 2015 Adaptation, clonal interference, and frequency-dependent interactions in a long-term evolution experiment with escherichia coli. Genetics **200**: 619–631.

MAKAM, M., D. DIAZ, J. LAVAL, Y. GERNEZ, C. K. CONRAD, *et al.*, 2009 Activation of critical, host-induced, metabolic and stress pathways marks neutrophil entry into cystic fibrosis lungs. Proceedings of the National Academy of Sciences **106**: 5779–5783.

MARKUSSEN, T., R. L. MARVIG, M. GOMEZ-LOZANO, K. AANAES, A. E. BURLEIGH, *et al.*, 2014 Environmental Heterogeneity Drives Within-Host Diversification and Evolution of Pseudomonas aeruginosa. mBio **5**: e01592–14–e01592–14.

MARTIN, G., and T. LENORMAND, 2006 A General Multivariate Extension of Fisher's Geometrical Model and the Distribution of Mutation Fitness Effects across Species. Evolution **60**: 893–907.

MARTIN, G., and T. LENORMAND, 2008 The Distribution of Beneficial and Fixed Mutation Fitness Effects Close to an Optimum. Genetics **179**: 907–916.

MARTIN, G., and T. LENORMAND, 2015 The fitness effect of mutations across environments: Fisher's geometrical model with multiple optima: FISHER'S MODEL ACROSS ENVIRONMENTS. Evolution **69**: 1433–1447.

MARVIG, R. L., S. DAMKIAER, S. M. H. KHADEMI, T. M. MARKUSSEN, S. MOLIN, *et al.*, 2014 Within-Host Evolution of Pseudomonas aeruginosa Reveals Adaptation toward Iron Acquisition from Hemoglobin. mBio **5**: e00966–14–e00966–14.

MARVIG, R. L., L. M. SOMMER, S. MOLIN, and H. K. JOHANSEN, 2015 Convergent evolution and adaptation of Pseudomonas aeruginosa within patients with cystic fibrosis. Nature Genetics **47**: 57–64.

MATHEE, K., G. NARASIMHAN, C. VALDES, X. QIU, J. M. MATEWISH, *et al.*, 2008 Dynamics of Pseudomonas aeruginosa genome evolution. Proceedings of the National Academy of Sciences **105**: 3100–3105.

MCKENNA, A., M. HANNA, E. BANKS, A. SIVACHENKO, K. CIBULSKIS, *et al.*, 2010 The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research **20**: 1297–1303.

MCKEON, D. J., K. A. CADWALLADER, S. IDRIS, A. S. COWBURN, M. C. PASTEUR, *et al.*, 2010 Cystic fibrosis neutrophils have normal intrinsic reactive oxygen species generation. European Respiratory Journal **35**: 1264–1272.

MESSER, P. W., and D. A. PETROV, 2013 Population genomics of rapid adaptation by soft selective sweeps. Trends in Ecology & Evolution **28**: 659–669.

MILLER, C. R., P. JOYCE, and H. A. WICHMAN, 2011 Mutational Effects and Population Dynamics During Viral Adaptation Challenge Current Models. Genetics **187**: 185–202.

MOSCA, A., M. LECLERC, and J. P. HUGOT, 2016 Gut microbiota diversity and human diseases: should we reintroduce key predators in our ecosystem? Frontiers in microbiology **7**: 455.

MOWAT, E., S. PATERSON, J. L. FOTHERGILL, E. A. WRIGHT, M. J. LEDSON, *et al.*, 2011 *Pseudomonas aeruginosa* Population Diversity and Turnover in Cystic Fibrosis Chronic Infections. American Journal of Respiratory and Critical Care Medicine **183**: 1674–1679.

NESSE, R. M., 2001 How is darwinian medicine useful? Western Journal of Medicine **174**: 358.

NEVO, E., 1978 Genetic variation in natural populations: patterns and theory. Theoretical population biology **13**: 121–177.

OLIVER, A., 2000 High Frequency of Hypermutable Pseudomonas aeruginosa in Cystic Fibrosis Lung Infection. Science **288**: 1251–1253.

ORR, H. A., 2005 The genetic theory of adaptation: a brief history. Nature Reviews Genetics **6**: 119–127.

ORR, H. A., 2006 The distribution of fitness effects among beneficial mutations in Fisher's geometric model of adaptation. Journal of Theoretical Biology **238**: 279–285.

OSTROWSKI, E. A., D. E. ROZEN, and R. E. LENSKI, 2005 Pleiotropic Effects of Beneficial Mutations in Escherichia coli. Evolution **59**: 2343–2352.

O'TOOLE, G. A., 2011 Microtiter Dish Biofilm Formation Assay. Journal of Visualized Experiments .

PÁL, C., B. PAPP, and V. LÁZÁR, 2015 Collateral sensitivity of antibiotic-resistant microbes. Trends in microbiology **23**: 401–407.

PALMER, K. L., L. M. AYE, and M. WHITELEY, 2007 Nutritional Cues Control Pseudomonas aeruginosa Multicellular Behavior in Cystic Fibrosis Sputum. Journal of Bacteriology **189**: 8079–8087.

PAYEN, C., A. B. SUNSHINE, G. T. ONG, J. L. POGACHAR, W. ZHAO, *et al.*, 2016 High-throughput identification of adaptive mutations in experimentally evolved yeast populations. PLoS genetics **12**: e1006339.

POOLE, K., 2005 Efflux-mediated antimicrobial resistance. Journal of Antimicrobial Chemotherapy **56**: 20–51.

PRUD'HOMME, B., N. GOMPEL, A. ROKAS, V. A. KASSNER, T. M. WILLIAMS, *et al.*, 2006 Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. Nature **440**: 1050.

RAINEY, P. B., and T. F. COOPER, 2004 Evolution of bacterial diversity and the origins of modularity. Research in Microbiology **155**: 370–375.

RAINEY, P. B., and M. TRAVISANO, 1998 Adaptive radiation in a heterogeneous environment. Nature **394**: 69.

RAJAN, S., and L. SAIMAN, 2002 Pulmonary infections in patients with cystic fibrosis. Seminars in respiratory infections **1**: 47–56.

RAVEL, J., R. M. BROTMAN, P. GAJER, B. MA, M. NANDY, *et al.*, 2013 Daily temporal dynamics of vaginal microbiota before, during and after episodes of bacterial vaginosis. Microbiome **1**: 29.

RODRIGUEZ-VERDUGO, A., D. CARRILLO-CISNEROS, A. GONZALEZ-GONZALEZ, B. S. GAUT, and A. F. BENNETT, 2014 Different tradeoffs result from alternate genetic adaptations to a common environment. Proceedings of the National Academy of Sciences **111**: 12121–12126.

ROZEN, D. E., and R. E. LENSKI, 2000 Long-term experimental evolution in Escherichia coli. VIII. Dynamics of a balanced polymorphism. The American Naturalist **155**: 24–35.

SAKHTAH, H., L. KOYAMA, Y. ZHANG, D. K. MORALES, B. L. FIELDS, *et al.*, 2016 The *Pseudomonas aeruginosa* efflux pump MexGHI-OpmD transports a natural phenazine that controls gene expression and biofilm development. Proceedings of the National Academy of Sciences **113**: E3538–E3547.

SANE, M., J. J. MIRANDA, and D. AGASHE, 2018 Antagonistic pleiotropy for carbon use is rare in new mutations. Evolution **72**: 2202–2213.

SANFORD, E., and M. W. KELLY, 2011 Local Adaptation in Marine Invertebrates. Annual Review of Marine Science **3**: 509–535.

SAXER, G., M. DOEBELI, and M. TRAVISANO, 2010 The repeatability of adaptive radiation during long-term experimental evolution of escherichia coli in a multiple nutrient environment. PLoS One **5**: e14184.

SCHAEDEL, C., I. DE MONESTROL, L. HJELTE, M. JOHANNESSON, R. KORNFÄLT, *et al.*, 2002 Predictors of deterioration of lung function in cystic fibrosis*: CFTR Genotype in Lung Function in CF. Pediatric Pulmonology **33**: 483–491.

SCHICK, A., S. F. BAILEY, and R. KASSEN, 2015 Evolution of Fitness Trade-Offs in Locally Adapted Populations of *Pseudomonas fluorescens*. The American Naturalist **186**: S48–S59.

SCHICK, A., and R. KASSEN, 2018 Rapid diversification of *Pseudomonas aeruginosa* in cystic fibrosis lung-like conditions. Proceedings of the National Academy of Sciences **115**: 10714–10719.

SCHLÖTTERER, C., R. KOFLER, E. VERSACE, R. TOBLER, and S. U. FRANSSEN, 2015 Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. Heredity **114**: 431–440.

SCHLUTER, D., 2000 *The ecology of adaptive radiation*. OUP Oxford.

SCHLUTER, D., 2009 Evidence for ecological speciation and its alternative. Science **323**: 737–741.

SCHLUTER, D., E. A. CLIFFORD, M. NEMETHY, and J. S. MCKINNON, 2004 Parallel evolution and inheritance of quantitative traits. The American Naturalist **163**: 809–822.

SCHOUSTRA, S. E., T. BATAILLON, D. R. GIFFORD, and R. KASSEN, 2009 The Properties of Adaptive Walks in Evolving Populations of Fungus. PLoS Biology **7**: e1000250.

SIBLEY, C. D., K. DUAN, C. FISCHER, M. D. PARKINS, D. G. STOREY, *et al.*, 2008 Discerning the Complexity of Community Interactions Using a Drosophila Model of Polymicrobial Infections. PLoS Pathogens **4**: e1000184.

SIMPSON, J. A., S. E. SMITH, and R. T. DEAN, 1989 Scavenging by alginate of free radicals released by macrophages. Free Radical Biology and Medicine **6**: 347–353.

SMITH, E. E., D. G. BUCKLEY, Z. WU, C. SAENPHIMMACHAK, L. R. HOFFMAN, *et al.*, 2006 Genetic adaptation by Pseudomonas aeruginosa to the airways of cystic fibrosis patients. Proceedings of the National Academy of Sciences **103**: 8487–8492.

SMITS, W. K., D. LYRAS, D. B. LACY, M. H. WILCOX, and E. J. KUIJPER, 2016 Clostridium difficile infection. Nature Reviews Disease Primers **2**: 16020.

SOUSA, A., and M. PEREIRA, 2014 Pseudomonas aeruginosa Diversification during Infection Development in Cystic Fibrosis Lungs—A Review. Pathogens **3**: 680–703.

STEINBERG, B., and M. OSTERMEIER, 2016 Shifting fitness and epistatic landscapes reflect trade-offs along an evolutionary pathway. Journal of molecular biology **428**: 2730–2743.

STEPHENSON, A. L., M. TOM, Y. BERTHIAUME, L. G. SINGER, S. D. AARON, *et al.*, 2015 A contemporary survival analysis of individuals with cystic fibrosis: a cohort study. European Respiratory Journal **45**: 670–679.

WANG, J., S. LORY, R. RAMPHAL, and S. JIN, 1996 Isolation and characterization of Pseudomonas aeruginosa genes inducible by respiratory mucus derived from cystic fibrosis patients. Molecular Microbiology **22**: 1005–1012.

WHITTAKER, R. H., and S. A. LEVIN, 1975 *Niche: theory and application*, volume 3. Halsted Pr.

WILLIAMS, D., B. EVANS, S. HALDENBY, M. J. WALSHAW, M. A. BROCKHURST, *et al.*, 2015 Divergent, Coexisting *Pseudomonas aeruginosa* Lineages in Chronic Cystic Fibrosis Lung Infections. American Journal of Respiratory and Critical Care Medicine **191**: 775–785.

WILLIAMS, G. C., and R. M. NESSE, 1991 The dawn of darwinian medicine. The Quarterly review of biology **66**: 1–22.

WILLNER, D., M. R. HAYNES, M. FURLAN, R. SCHMIEDER, Y. W. LIM, *et al.*, 2012 Spatial distribution of microbial communities in the cystic fibrosis lung. The ISME journal **6**: 471.

WINSOR, G. L., E. J. GRIFFITHS, R. LO, B. K. DHILLON, J. A. SHAY, *et al.*, 2015 Enhanced annotations and features for comparing thousands of pseudomonas genomes in the pseudomonas genome database. Nucleic acids research **44**: D646–D653.

WINSOR, G. L., E. J. GRIFFITHS, R. LO, B. K. DHILLON, J. A. SHAY, *et al.*, 2016 Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the Pseudomonas genome database. Nucleic Acids Research **44**: D646–D653.

WINSTANLEY, C., S. O'BRIEN, and M. A. BROCKHURST, 2016 Pseudomonas aeruginosa Evolutionary Adaptation and Diversification in Cystic Fibrosis Chronic Lung Infections. Trends in Microbiology **24**: 327–337.

WONG, A., N. RODRIGUE, and R. KASSEN, 2012 Genomics of Adaptation during Experimental Evolution of the Opportunistic Pathogen Pseudomonas aeruginosa. PLoS Genetics **8**: e1002928.

WOODS, R. J., J. E. BARRICK, T. F. COOPER, U. SHRESTHA, M. R. KAUTH, *et al.*, 2011 Second-order selection for evolvability in a large escherichia coli population. Science **331**: 1433–1436.

WORKENTINE, M. L., C. D. SIBLEY, B. GLEZERSON, S. PURIGHALLA, J. C. NORGAARD-GRON, *et al.*, 2013 Phenotypic Heterogeneity of Pseudomonas aeruginosa Populations in a Cystic Fibrosis Patient. PLoS ONE **8**: e60225.

WRIGHT, E. A., J. L. FOTHERGILL, S. PATERSON, M. A. BROCKHURST, and C. WINSTANLEY, 2013 Sub-inhibitory concentrations of some antibiotics can drive diversification of Pseudomonas aeruginosa populations in artificial sputum medium. BMC microbiology **13**: 170.

YANG, L., J. A. HAAGENSEN, L. JELSBAK, H. K. JOHANSEN, C. STERNBERG, *et al.*, 2008 In situ growth rates and biofilm development of pseudomonas aeruginosa populations in chronic lung infections. Journal of bacteriology **190**: 2767–2776.

YANG, L., L. JELSBAK, R. L. MARVIG, S. DAMKIAER, C. T. WORKMAN, *et al.*, 2011 Evolutionary dynamics of bacteria in a human host environment. Proceedings of the National Academy of Sciences **108**: 7481–7486.

YEAMAN, S., A. C. GERSTEIN, K. A. HODGINS, and M. C. WHITLOCK, 2018 Quantifying how constraints limit the diversity of viable routes to adaptation. PLoS genetics **14**: e1007717.

ZHANG, X.-X., and P. B. RAINEY, 2007 Construction and validation of a neutrally-marked strain of Pseudomonas fluorescens SBW25. Journal of Microbiological Methods **71**: 78–81.

Appendix

Appendix A

Appendix for Chapter 2: Environment-dependent fitness effects of adaptive mutations

A.1 Supporting Tables

Table A.1: Sequences of primers used for PCR and corresponding melting temperatures.

Gene	Forward primer sequence Reverse primer sequence	Melting Temp (0C)
PFLU0172	GTAAAACGACGGCCAGTTGTTGCTAGTATGGCCTCGG GGAAACAGCTATGACCATGAATCGAAGACAACAACCGCG	54.1 - 58.9
PFLU0406	GTAAAACGACGGCCAGTGCTGGTCGTAGGTGATGAGG GGAAACAGCTATGACCATGAAAGCGGACACATTCAATCC	52.0 - 58.9
fruR	GTAAAACGACGGCCAGTGTAGCTGGGGTTTTCCAGGT GGAAACAGCTATGACCATGCTGGAATCGGTTACGATTGG	54.1 - 56.6
PFLU1183	GTAAAACGACGGCCAGTAAATGGTACTGGGCGCCTTT GGAAACAGCTATGACCATGCCTCCTCAATCCCATGCCAG	54.1 - 56.6
PFLU1301	GTAAAACGACGGCCAGTTCGATCGCCTTCCGTTATGG GGAAACAGCTATGACCATGGACGGCGAATCTCAAAGTGC	52.0 - 60.0
gacA	GTAAAACGACGGCCAGTGAATCCTGACCTGTTGCCCA GGAAACAGCTATGACCATGGGTGGGCTGGAAAGACTTGA	52.6 - 59.4
PFLU2366	GTAAAACGACGGCCAGTCGGTAGGCCTGTTCAATGGT GGAAACAGCTATGACCATGCCTGAAACCGCCACAAAACC	52.0 - 52.6
PFLU2428	GTAAAACGACGGCCAGTGCGGCGTGAGGATATTCAGG GGAAACAGCTATGACCATGTGACCCTCAACCAGCAAGTG	52.0 - 60.0
PFLU3777	GTAAAACGACGGCCAGTGACCCTGTTGGAAGACATGG GGAAACAGCTATGACCATGGGTGCACACGTTCGATCA	52.0 - 59.4
fliP	GTAAAACGACGGCCAGTAGATCATGAAGCCGATTTGG GGAAACAGCTATGACCATGGGGCAAGGATCAGAAGGATA	52.6 - 58.9
fliF	GTAAAACGACGGCCAGTCACCCAACACTTGCTTGACG GGAAACAGCTATGACCATGCAATCCACCAGCGAACAACG	52.0 - 59.4
PFLU4443	GTAAAACGACGGCCAGTAAAACAGGCGTCAATCATCC GGAAACAGCTATGACCATGCGAAGAAATTCCGCTACGTC	52.0 - 60.0
PFLU4443	GTAAAACGACGGCCAGTGTCGACCAGTTGCTCGTCTT GGAAACAGCTATGACCATGCCCTGTTCCTTGACGAAATC	52.6 - 58.9
PFLU4443	GTAAAACGACGGCCAGTGTGATCGCCCCGGTAAAG GGAAACAGCTATGACCATGCCCCCAGCTACAGCAAAT	52.0 - 54.1
PFLU4443	GTAAAACGACGGCCAGTCGTACATCTCGCGATAGACCT GGAAACAGCTATGACCATGTGCTGTTTTCAGATTTCAGAC	52.0 - 58.9
flgL	GTAAAACGACGGCCAGTGACGATAGGGTCTTCGGTGA GGAAACAGCTATGACCATGCAACAACATGCTCTCGCAGT	54.1 - 56.6
flgH	GTAAAACGACGGCCAGTCCGTAGCCGATCAACTGGTT GGAAACAGCTATGACCATGCTGACCTCGTTGTTTGGTGC	52.0 - 60.0
PFLU4845	GTAAAACGACGGCCAGTGAAGAACAGCTGCGAACCAC GGAAACAGCTATGACCATGGGTGCGTTCTTTGATGTCGT	52.0 - 60.0
PFLU5822	GTAAAACGACGGCCAGTGTTGGCCGGTGATTACGTTG GGAAACAGCTATGACCATGAGTTATCCACGTTGCAGCCA	52.0 - 54.1
PFLU5960	GTAAAACGACGGCCAGTAGCGTGTTTTCCAGCAAACC GGAAACAGCTATGACCATGACGACAGTGTGACGTACCAC	54.1 - 56.6

Table A.2: Reference information for mutations investigated, including source population, gene name, genomic position and gene function.

Strain	Line	Gene	Location	Gene function
a2t3-3	mgx2	fruR (1)	904637	DNA-binding transcriptional regulator FruR
a2t6-11	mgx2	fruR (1)	904637	DNA-binding transcriptional regulator FruR
		PLFU1183	1315026	putative phage-like protein
a2t10-1	mgx2	fruR (1)	904637	DNA-binding transcriptional regulator FruR
		PFLU1183	1315026	putative phage-like protein
		fliP	4891052	flagellar biosynthesis protein FliP
b3t2-3	m-g-x3	PFLU4443 (1)	490173	putative two-component system response regulator Nitrogen regulation protein NR(I)
b3t5-4	m-g-x3	PFLU4443 (1)	490173	putative two-component system response regulator Nitrogen regulation protein NR(I)
		PFLU2366 (1)	2577913	putative GntR family regulatory protein
b3t8-1	m-g-x3	PFLU4443 (1)	490173	putative two-component system response regulator Nitrogen regulation protein NR(I)
		PFLU2366 (1)	2577913	putative GntR family regulatory protein
		fruR (2)	904688	DNA-binding transcriptional regulator FruR
g1t8-14	g1	PFLU4845	5324097	ABC transport system, membrane permease
g1t10-1	g1	PFLU4845	5324097	ABC transport system, membrane permease
		PFLU4443 (2)	4908340	putative two-component system response regulator Nitrogen regulation protein NR(I)
g1t9-3	g1	PFLU4845	5324097	ABC transport system, membrane permease
		PFLU4443 (3)	4907597	putative two-component system response regulator Nitrogen regulation protein NR(I)
m1t3-3	m1	PFLU4443 (4)	4907167	putative two-component system response regulator Nitrogen regulation protein NR(I)
m1t5-8	m1	PFLU4443 (4)	4907167	putative two-component system response regulator Nitrogen regulation protein NR(I)
		fruR (3)	904686	DNA-binding transcriptional regulator FruR
m2t3-4	m2	PFLU4443 (5)	4907167	putative two-component system response regulator Nitrogen regulation protein NR(I)
m2t5-2	m2	PFLU4443 (5)	4907167	putative two-component system response regulator Nitrogen regulation protein NR(I)
		fruR (4)	904686	DNA-binding transcriptional regulator FruR
m3t4-6	m3	PFLU0406	450662	putative penicillin-binding protein 1A
m3t8-4	m3	PFLU0406	450662	putative penicillin-binding protein 1A
		PFLU3777	4174835	hybrid sensory histidine kinase in two-component regulatory system with UvrY
m3t5-2	m3	PFLU4443 (6)	4908167	putative two-component system response regulator Nitrogen regulation protein NR(I)
m3t9-2	m3	PFLU4443 (6)	4908167	putative two-component system response regulator Nitrogen regulation protein NR(I)
		gacA	2372830	two-component system response regulator

x1t1-4	x1	PFLU2366 (3)	2578022	putative GntR family regulatory protein
x1t9-11	x1	PFLU2366 (3)	2578022	putative GntR family regulatory protein
		flgI	4920009	flagellar basal body P-ring protein
x1t9-4	x1	PFLU2366 (3)	2578022	putative GntR family regulatory protein
		flgH	4920009	flagellar basal body L-ring protein
x2t3-4	x2	PFLU0172	194140	hypothetical protein
x2t9-2	x2	PFLU0172	194140	hypothetical protein
		PFLU5960	3524328	putative EAL/GGDEF domain-containing protein

A.2 Supporting Figures

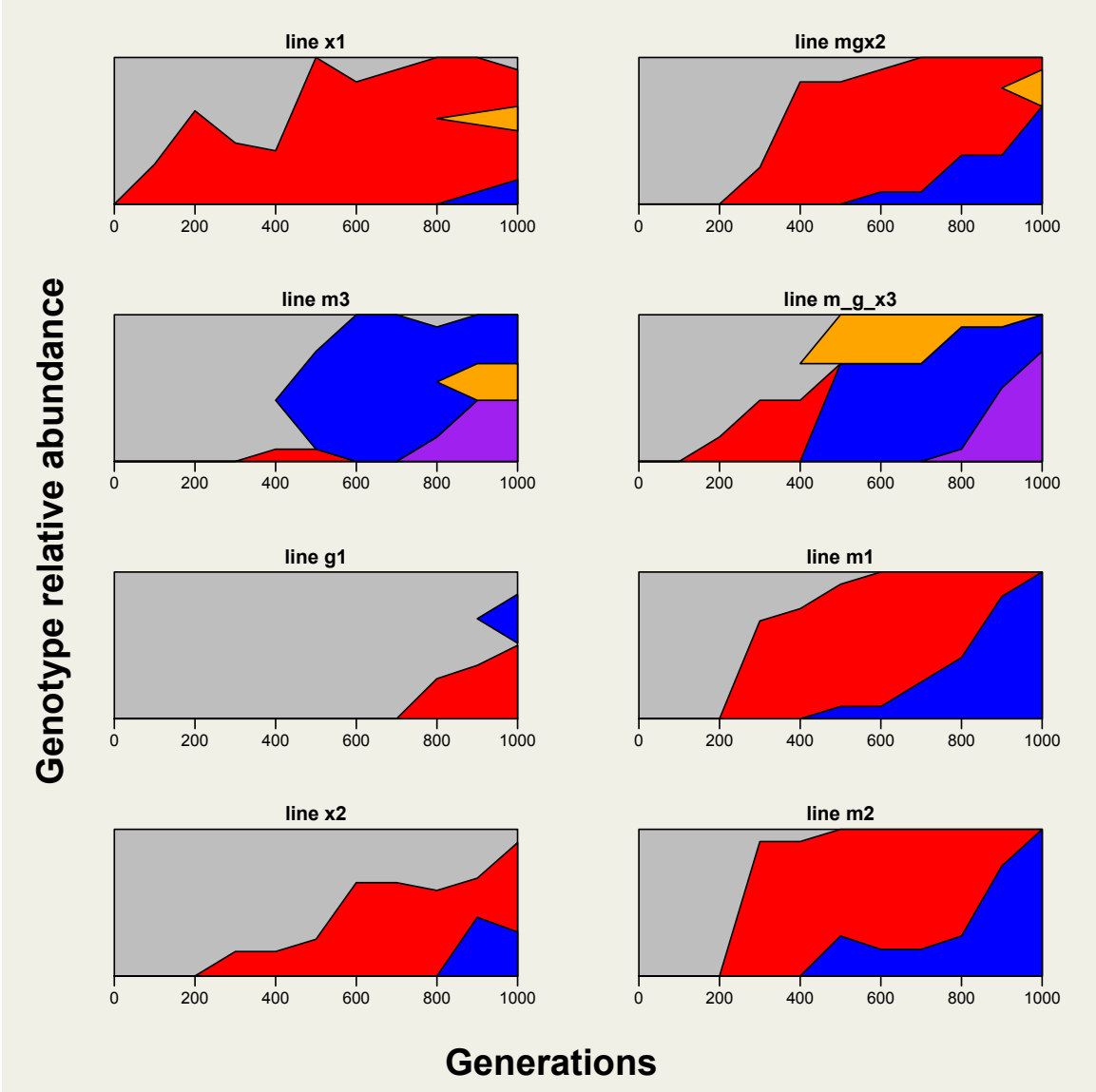

Figure A.1: Mueller plots of all lines. Genotype relative abundance is measured by presence or absence of focal mutation in 12 clones at 100 generation intervals, using PCR amplification of the region and sequencing. In each line, each colour represents a unique mutation. Grey represents the ancestor SBW25. See appendix Table A.2 for location and identity of mutations.

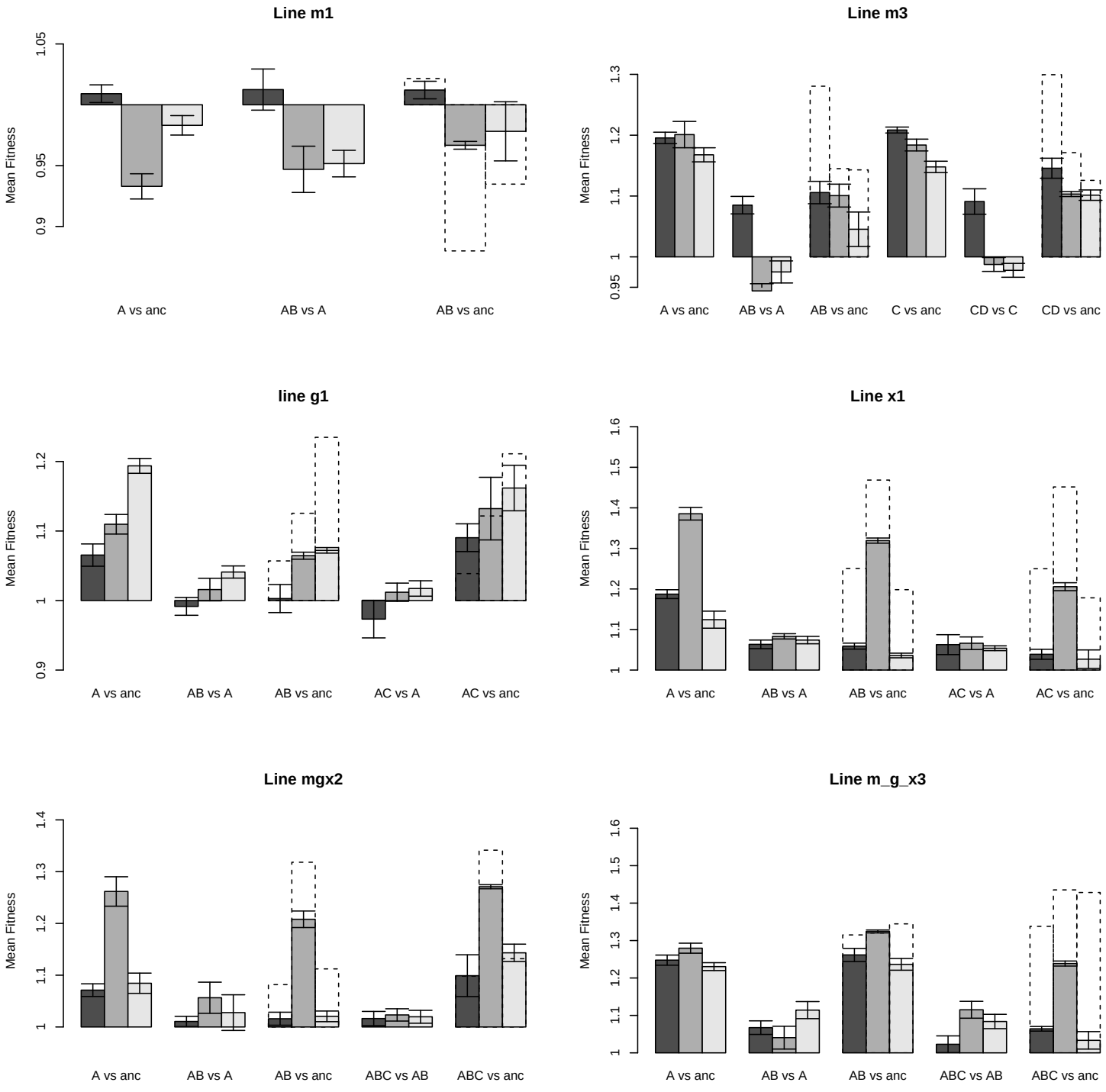

Figure A.2: Fitness effects of individual mutations in all environments. Dashed lines show expected fitness effect if the fitness effects are additive.

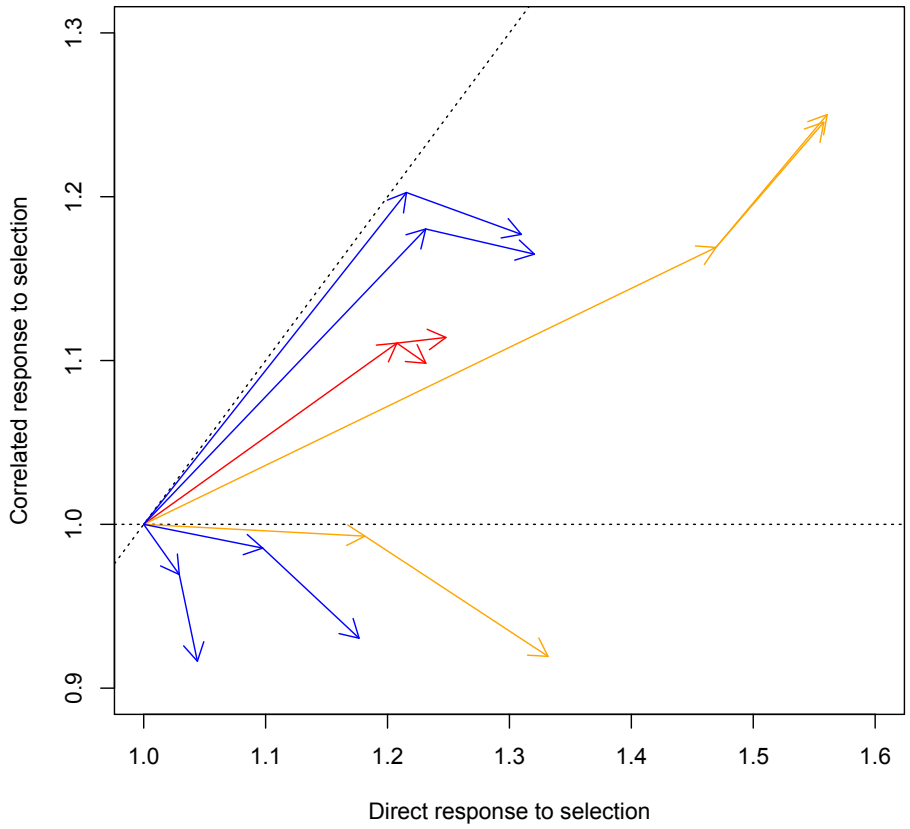

Figure A.3: Direct versus correlated responses to selection. Each individual arrow represents a single mutation with subsequent mutations in the same line connected head to tail. Direct response to selection is competitive fitness in the environment of selection for a given mutation. All mutations from single-resource environments are shown: Mannose (blue), glucose (red), and xylose (orange). Correlated response to selection is mean competitive fitness in alternative environments. Arrows above the horizontal dashed line are beneficial in alternative environments and arrows below this line are deleterious. The diagonal dashed line represents the line of equal direct and correlated response, or equal fitness in both selection and alternative environment, meaning no specialization.

Appendix B

Appendix for Chapter 3: Ecological factors that drive diversification

B.1 Supplementary Methods

B.1.1 Simulations

As a proxy for diversity among strains isolated from different patients, we also compared the clinical isolates to sets of evolved isolates by randomly sampling single isolates from different populations. The results of these simulations were congruent with those comparing the Mahalanobis distance between the clinical isolates distributions of evolved isolates, showing that sets sampled from MIN populations were the most different from the clinical strains (Fig B.6B). By comparing the phenotypic diversity present in our set of clinical isolates to the diversity that arose in our experimental conditions, we see that the diversity among clinical isolates falls between that of sets sampled from SCFM+mucin and SCFM (Fig B.6A).

B.2 Supporting Tables

Table B.1: Treatment effects on change in individual phenotypic traits. Values are from one-sample t-tests within each treatment, using the trait values f the ancestral strain (Pa14) as the null hypothesis and corrected for multiple comparisons (40 in total) using a Bonferroni corrected *p*-value of $\alpha = 0.0025$. Significant results are shown in bold.

	SCFM		MIN		SCFM + mucin		MIN + mucin	
Trait	t	*p*-value	t	*p*-value	t	*p*-value	t	*p*-value
Growth density	1.50	0.1358	-3.08	0.0025	-0.59	0.5561	1.85	0.0659
Pyoverdine	-12.0	**<0.0001**	-17.4	**<0.0001**	-4.57	**<0.0001**	-5.44	**<0.0001**
Pyocyanin	-1.06	0.2893	0.85	0.3962	4.22	**<0.0001**	1.06	0.2913
Biofilm	-13.71	**<0.0001**	-3.29	**0.0013**	10.16	**<0.0001**	-7.97	**<0.0001**
Swim motility	-4.65	**<0.0001**	-4.05	**<0.0001**	-17.3	**<0.0001**	-22.7	**<0.0001**
Twitch motility	-39.9	**<0.0001**	-23.7	**<0.0001**	-28.7	**<0.0001**	-13.1	**<0.0001**
MIC cirpofloxacin	20.2	**<0.0001**	3.70	0.0031	7.79	**<0.0001**	8.37	**<0.0001**
MIC ceftazidime	-2.12	0.0361	1.58	0.1161	4.77	**<0.0001**	3.95	**0.0001**
MIC colistin	0.75	0.4542	-0.03	0.9735	1.47	0.144	-5.09	**<0.0001**
MIC tobramycin	-0.08	0.9381	0.89	0.3749	-1.40	0.1632	0.74	0.461

Table B.2: Effects of nutrition and spatial structure on individual phenotypic traits. Values are from linear mixed effects models with media (MIN or SCFM) and mucin (presence or absence) as fixed effects and population as a random effect. P-values < 0.05 are in bold.

	media		mucin		interaction	
Trait	$F_{1,44}$ ratio	*p*-value	$F_{1,44}$ ratio	*p*-value	$F_{1,44}$ ratio	*p*-value
Growth density	0.56	0.457	0.90	0.347	6.25	**0.016**
Pyoverdine	0.05	0.832	11.56	**0.001**	0.03	0.875
Pyocyanin	0.80	0.375	3.47	0.069	2.88	0.097
Biofilm	13.75	**<0.001**	25.93	**<0.001**	25.29	**<0.001**
Swim motility	0.17	0.679	66.18	**<0.001**	1.73	0.196
Twitch motility	31.16	**<0.001**	0.85	0.362	0.01	0.917
MIC ciprofloxacin	31.07	**<0.001**	4.19	**0.047**	25.29	**<0.001**
MIC ceftazidime	0.36	0.553	7.36	**0.010**	1.54	0.222
MIC colistin	1.10	0.300	0.10	0.756	0.58	0.452
MIC tobramycin	1.73	0.195	0.40	0.531	0.27	0.608

B.3 Supporting Figures

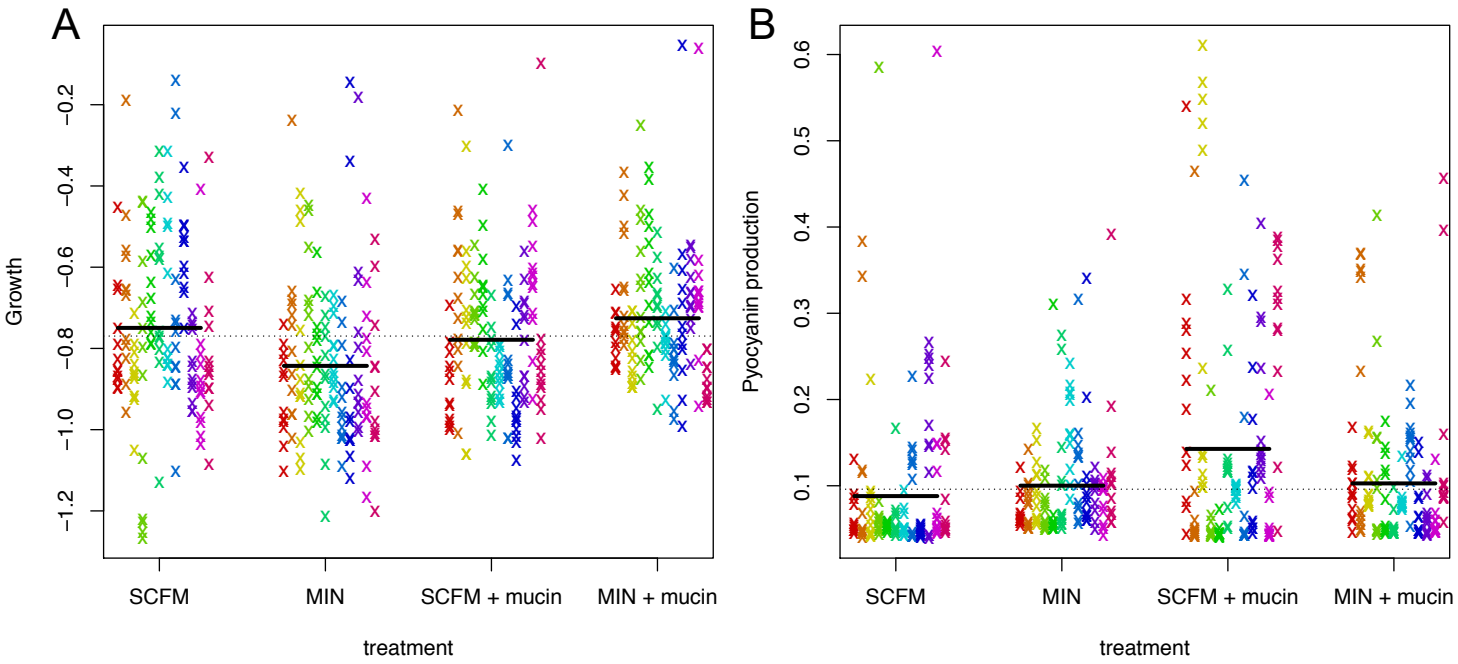

Figure B.1: Phenotypic adaptation in evolved isolates for growth (A) and pyocyanin production (B). Within each treatment, each column/colour represents a population. Solid lines represent treatment means and dashed lines represent the value of that phenotype in the ancestral strain (Pa14).

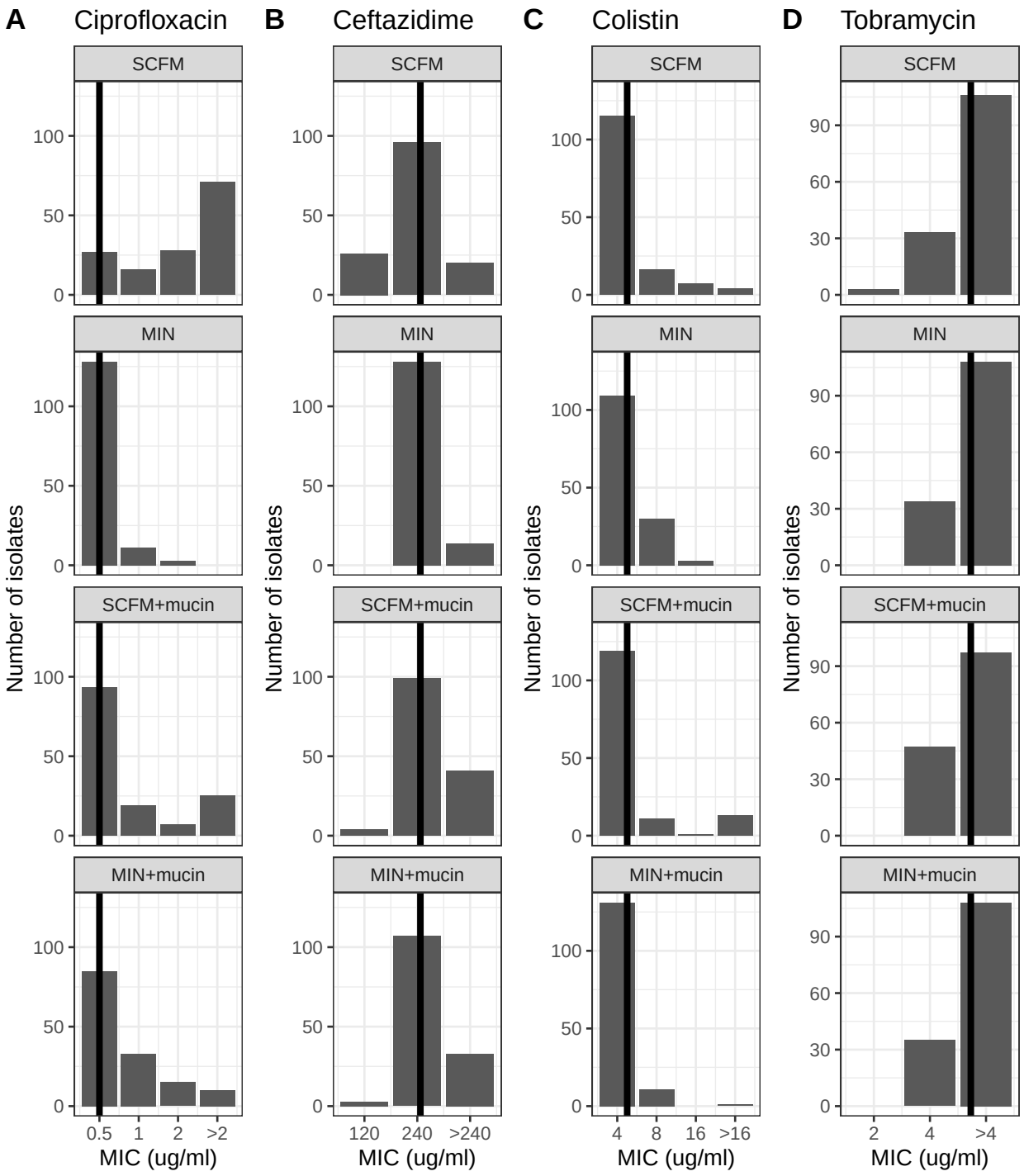

Figure B.2: Antibiotic resistance of all evolved isolates. Minimum inhibitory concentration (MIC) for four antibiotics: Ciprofloxacin (A), Ceftazidime (B), Colistin (C), and Tobramycin (D). Vertical lines represent ancestral MIC value for each antibiotic.

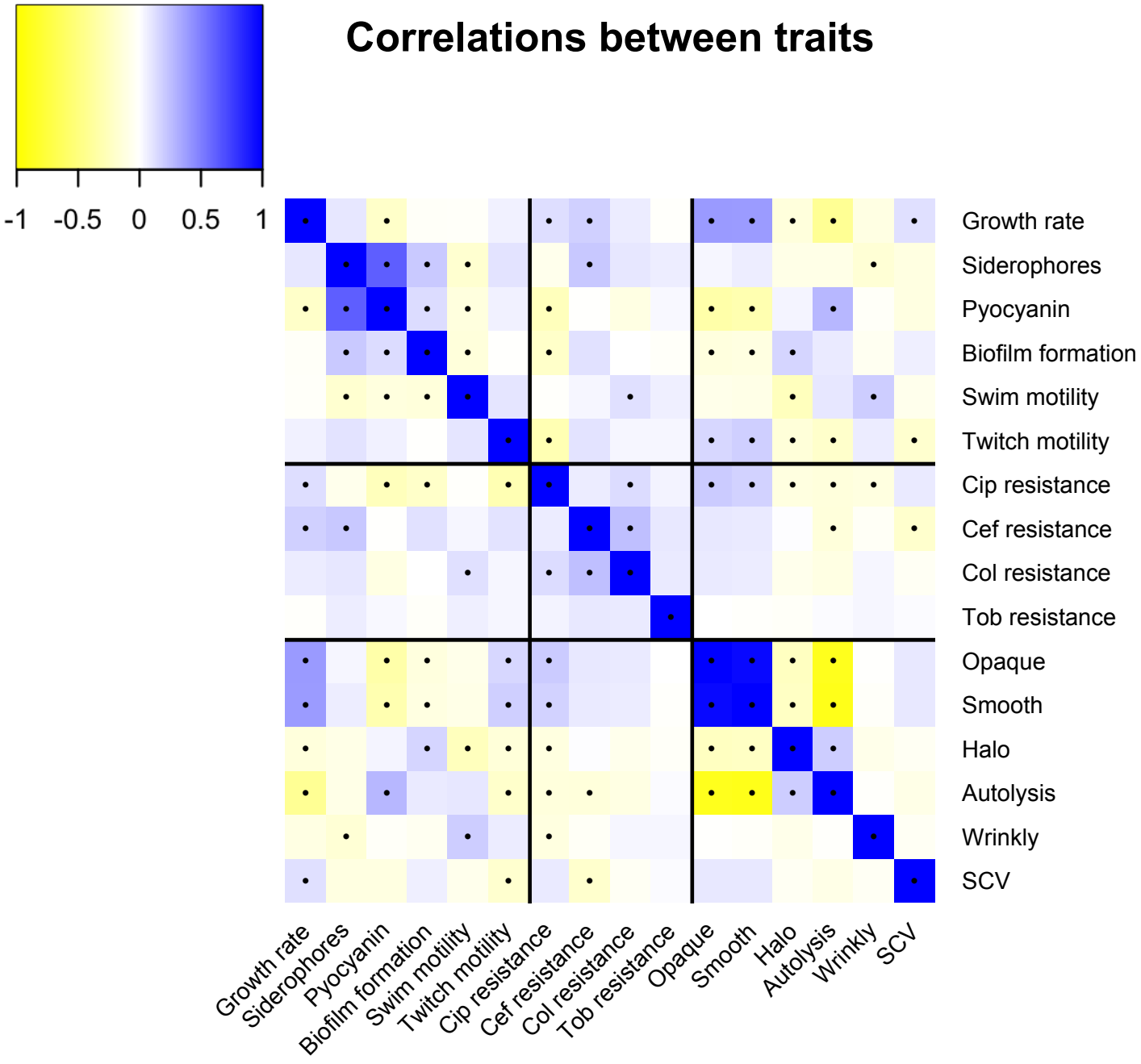

Figure B.3: Spearman rank correlations between all traits for all evolved isolates. Black lines separate groups of traits, divided into phenotypic traits, antibiotic-resistance, and colony morphology traits. Yellow indicates strong positive correlation between any pair of traits and blue indicates strong negative correlation. Black dots within boxes represent a statistically significant correlation ($p < 0.05$).

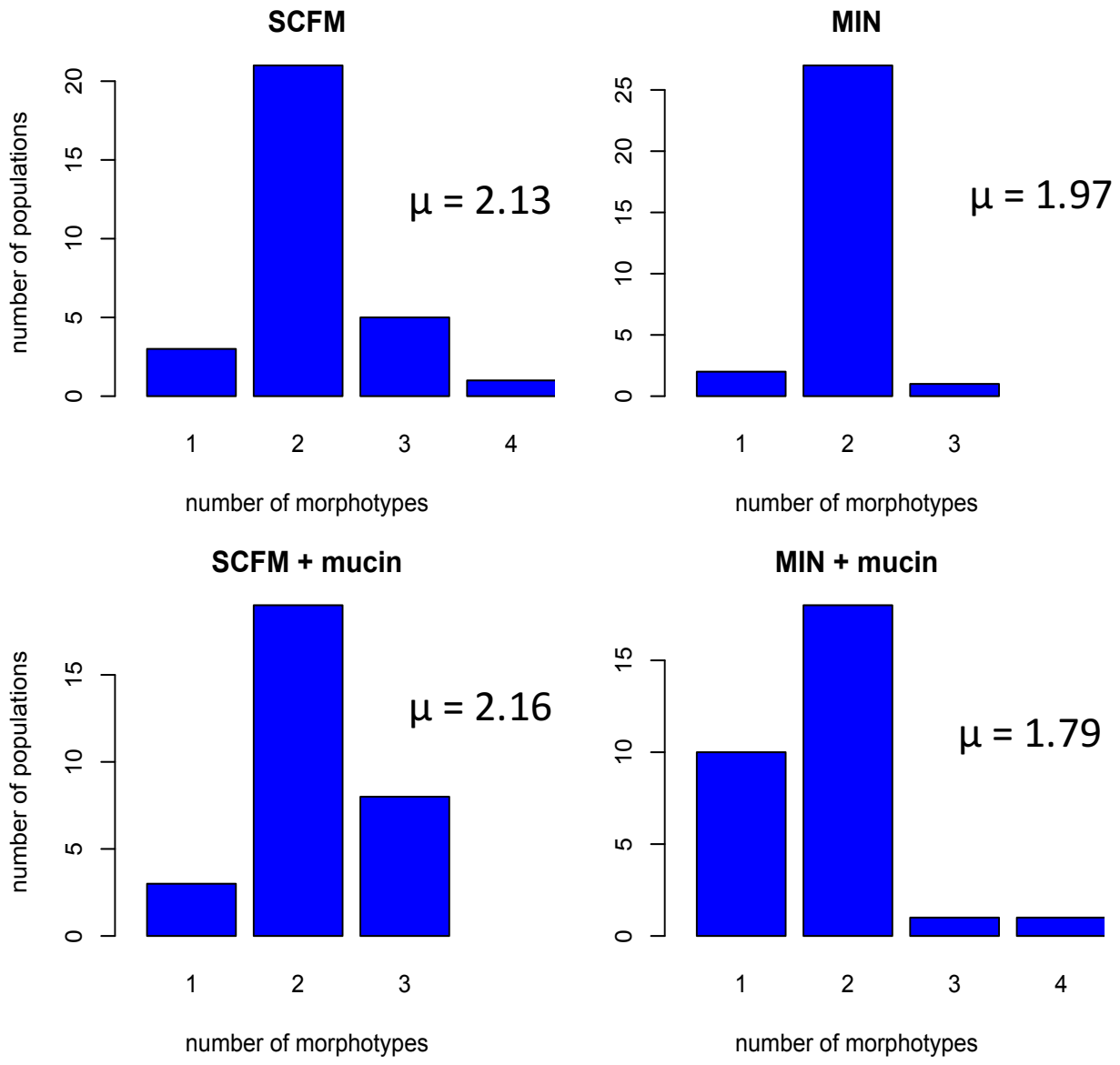

Figure B.4: Number of distinct morphologies present in all evolved populations. Colonies were categorized by a single observer using 8 binary characteristics: pigmentation, opacity, iridescence, surface texture, margin, halo, autolysis, and small colony variant.

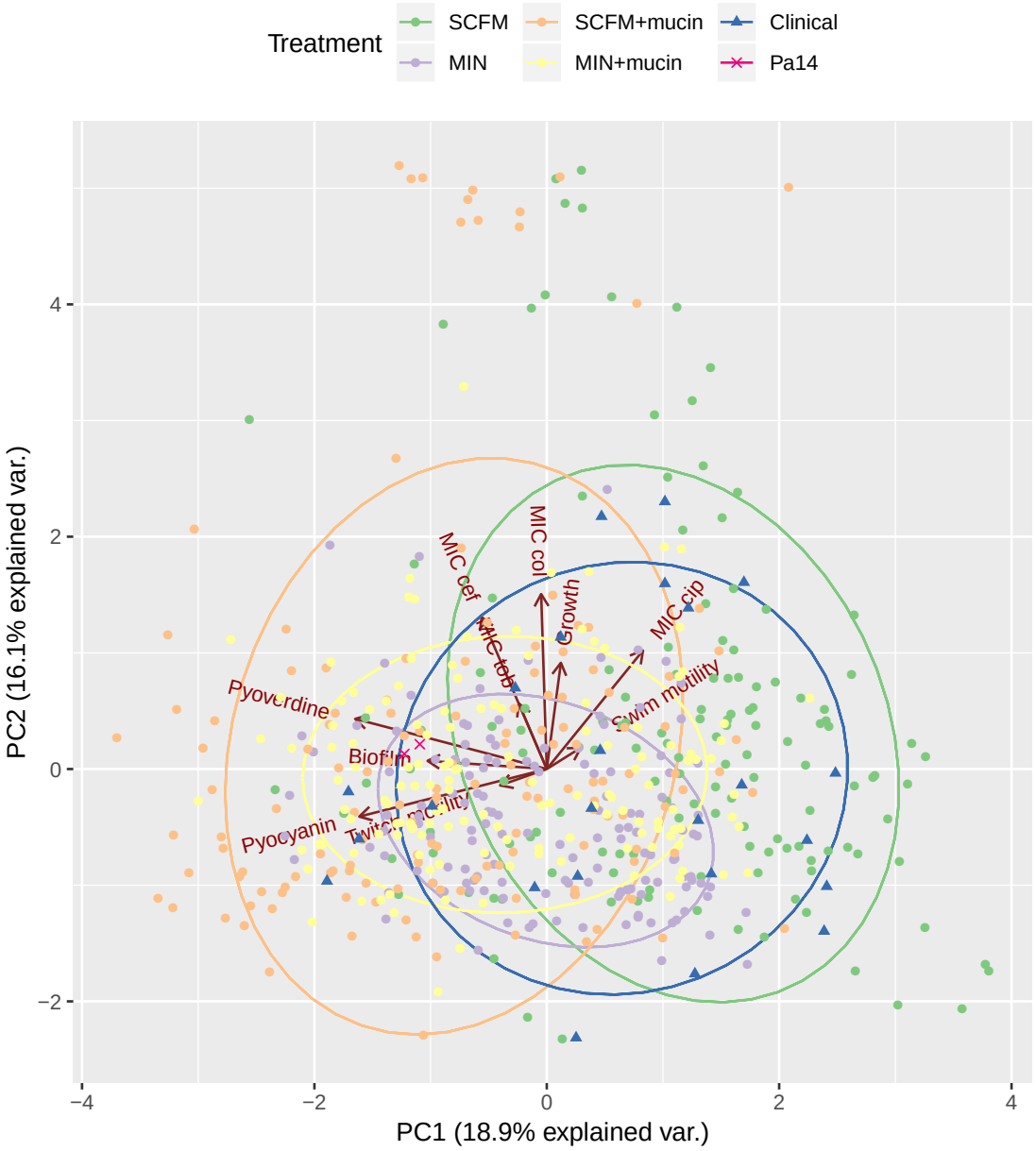

Figure B.5: Principal component analysis of all evolved isolates and clinical isolates. Each point represents a single isolate and each treatment is grouped.

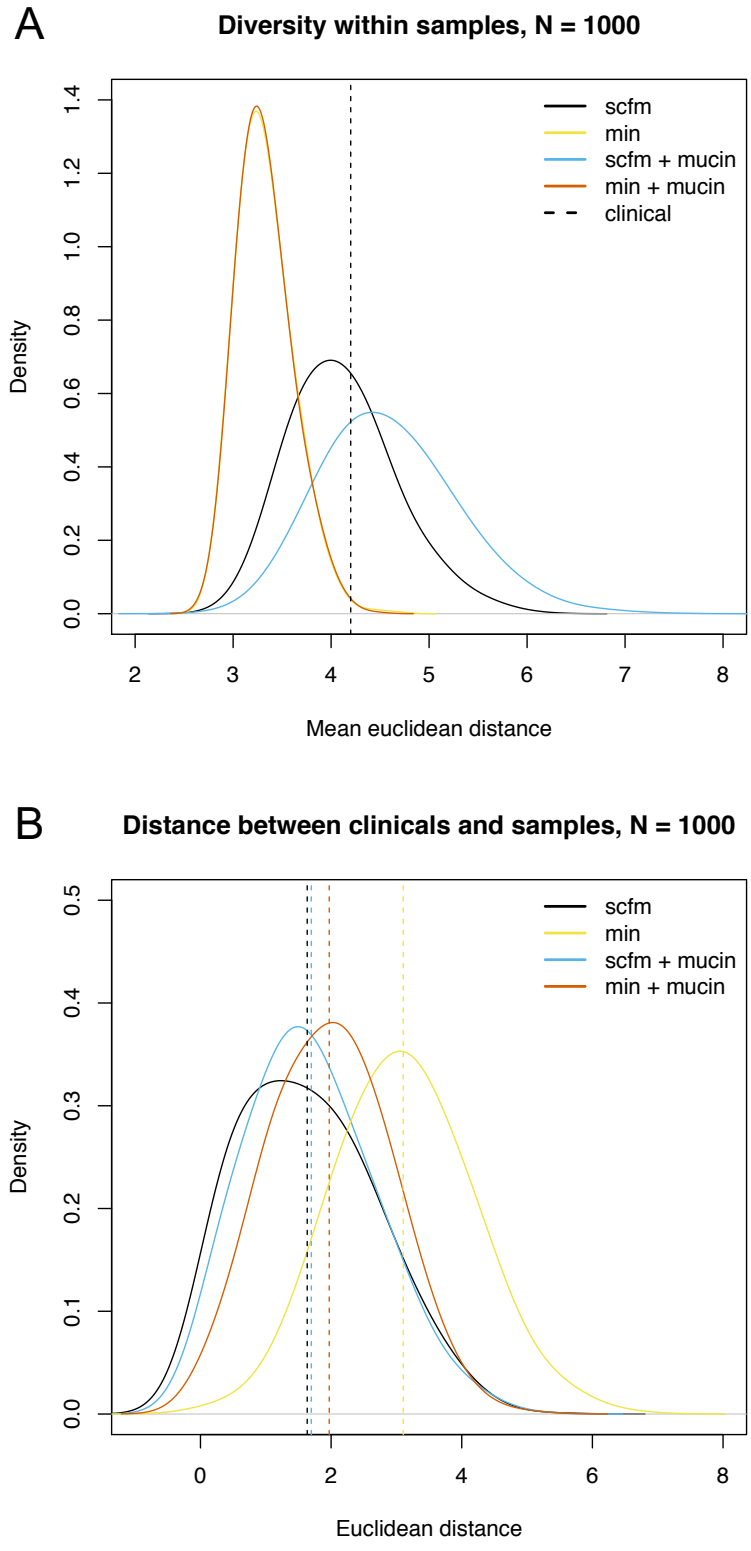

Figure B.6: Simulations comparing clinical isolates to random samples of evolved isolates. (Continued on following page.)

Figure B.6: (Previous page.) Simulations comparing clinical isolates to random samples of evolved isolates. A. Diversity, measured by mean pairwise Euclidean distance, within samples of 24 isolates from within a treatment, sampled $N = 1000$ times. Dashed line indicates this value for a set of 24 clinical isolates. B. Distance between groups of 12 isolates sampled from different populations, within a treatment, and 12 clinical isolates sampled from the clinical set, all of which are from different patients. Sampled $N = 1000$ times, without replacement. Dashed lines indicate mean distance for each treatment group.

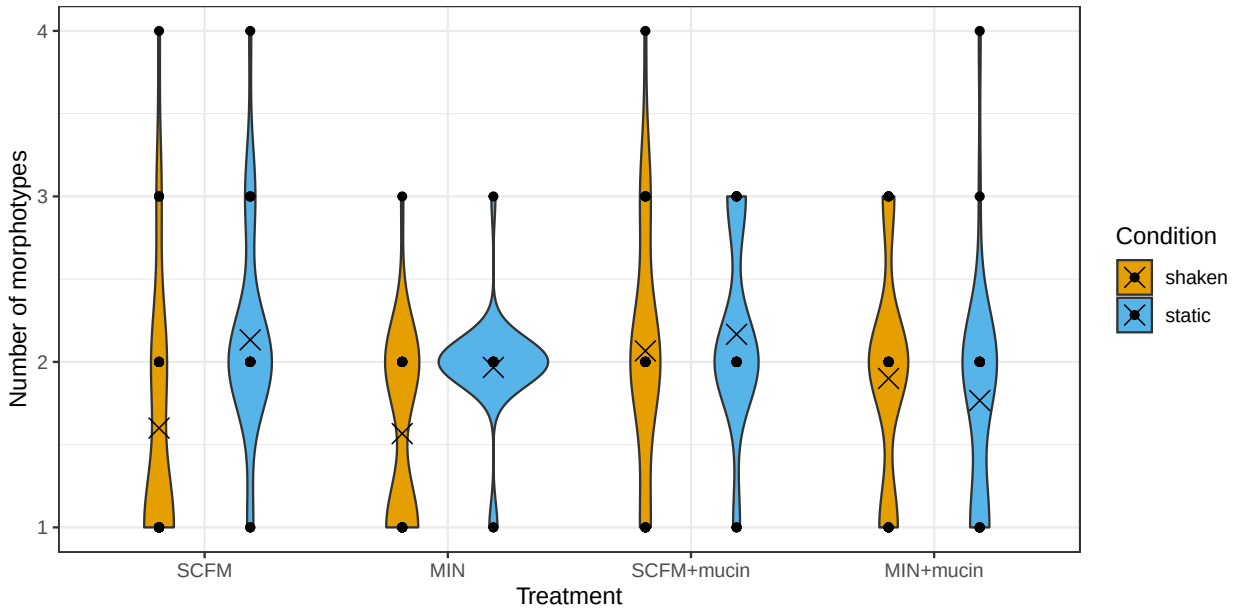

Figure B.7: Number of distinct morphologies present in all shaken and static populations. Colonies were categorized by a single observer using 8 binary characteristics: pigmentation, opacity, iridescence, surface texture, margin, halo, autolysis, and small colony variant.

Appendix C

Appendix for Chapter 4: Genetic changes underlying diversifying populations

C.1 Supporting Tables

Table C.1: Number of observed synonymous and nonsynonymous SNPs in protein coding regions of the genome. The ratio of the rate of nonsynonymous mutations (dN) to the rate of synonymous mutations (dS) was calculated using observed mutation rates, using a rate of 25.1% of all SNPs in protein-coding regions result in synonymous changes (Yang et al 2011). The expected number of nonsynonymous SNPs is the expected number under neutral evolution and probabilities are calculated from a Poisson distribution.

Treatment	dN/dS	Observed synonymous SNPs	Expected nonsynonymous SNPs	Observed nonsynonymous SNPs	Probability of neutrality
All	1.26	74	222	373	<0.0001
SCFM	1.77	12	36	85	<0.0001
MIN	2.11	7	21	59	<0.0001
SCFM + mucin	0.89	23	69	82	0.014
MIN + mucin	1.02	32	96	131	0.001

Table C.2: Mutations identified in evolved populations.

	Type	Position	Mut	Mutation category	Codon	Number	Pos	Ref	Gene name	Population	Treatment	Frequency
1	SUB	555745	27	small_indel	NA			NA	glcB	X1	scfm	0.11
2	SNP	797101	T	snp_synonymous	CTT	203	3	CTG	PA14_09300	X1	scfm	0.05
3	SNP	1462345	A	snp_nonsynonymous	TTC	168	2	TGC	map	X1	scfm	0.16
4	SNP	2820140	A	snp_nonsynonymous	GAT	56	2	GCT	PA14_32420	X1	scfm	0.27
5	SNP	3175128	A	snp_nonsynonymous	AAA	176	1	CAA	PA14_35700	X1	scfm	0.07
6	SNP	3912830	T	snp_nonsynonymous	ATG	167	1	CTG	sucC	X1	scfm	0.05
7	SNP	5432320	A	snp_nonsynonymous	AGC	1162	1	CGC	morA	X1	scfm	0.36
8	SNP	6172528	A	snp_intergenic	NA			NA	NA	X1	scfm	0.06
9	SNP	1058468	T	snp_synonymous	CGT	89	3	CGG	PA14_12260	X10	min_mucin	0.05
10	SNP	1982182	G	snp_nonsynonymous	CGC	92	2	CAC	PA14_22830	X10	min_mucin	0.05
11	SNP	2035917	T	snp_nonsynonymous	TCA	577	1	GCA	PA14_23420	X10	min_mucin	0.06
12	SNP	2953312	T	snp_nonsynonymous	AGC	108	1	CGC	PA14_33560	X10	min_mucin	0.05
13	SNP	3691947	A	snp_synonymous	GCA	72	3	GCC	ppiB	X10	min_mucin	0.05
14	SNP	4286137	A	snp_intergenic	NA			NA	NA	X10	min_mucin	0.19
15	SNP	5432189	A	snp_nonsynonymous	CAG	1118	2	CTG	morA	X10	min_mucin	0.50
16	SNP	1182107	C	snp_nonsynonymous	CCC	431	1	TCC	narK1	X11	min_mucin	0.13
17	SNP	1956730	T	snp_intergenic	NA			NA	NA	X11	min_mucin	0.07
18	SNP	2055691	T	snp_nonsynonymous	TGG	149	1	GGG	kdgA	X11	min_mucin	0.05
19	SNP	2549872	C	snp_nonsynonymous	ACC	1224	2	ATC	PA14_29400	X11	min_mucin	0.05
20	SNP	2651955	T	snp_nonsynonymous	CTG	273	2	CCG	PA14_30600	X11	min_mucin	0.05
21	SNP	2754373	A	snp_synonymous	GGA	271	3	GGC	PA14_31690	X11	min_mucin	0.06
22	SNP	3274352	T	snp_nonsynonymous	CAG	311	2	CCG	PA14_36730	X11	min_mucin	0.07
23	SNP	3845147	T	snp_nonsynonymous	GTG	277	2	GGG	PA14_43160	X11	min_mucin	0.05
24	SNP	4286130	C	snp_intergenic	NA			NA	NA	X11	min_mucin	0.65
25	SNP	4563013	C	snp_nonsynonymous	GAC	59	3	GAA	mvfR	X11	min_mucin	0.19
26	SNP	5912501	T	snp_nonsynonymous	CTG	32	2	CGG	PA14_66320	X11	min_mucin	1.00
27	SNP	2029395	C	snp_nonsynonymous	GCC	32	2	GTC	orfH	X14	scfm	0.87
28	SNP	2818975	T	snp_nonsynonymous	CAC	260	2	CGC	mexT	X14	scfm	0.89
29	DEL	5234203	22	small_indel	NA			NA	PA14_58740	X14	scfm	0.06
30	SNP	5235347	T	snp_nonsynonymous	TCG	357	1	CCG	pilB	X14	scfm	0.08
31	SNP	5236097	T	snp_nonsense	TAA	6	1	CAA	pilC	X14	scfm	0.15
32	INS	5237772	12	small_indel	NA			NA	pilD	X14	scfm	0.07
33	SNP	5948306	A	snp_nonsense	TAG	315	1	CAG	pilQ	X14	scfm	0.07
34	SNP	2030961	T	snp_synonymous	GGT	105	3	GGA	orfE	X15	scfm	0.06
35	SNP	2467367	A	snp_nonsynonymous	AAC	50	2	ACC	thrS	X15	scfm	0.05
36	SNP	2587875	G	snp_nonsynonymous	TCC	393	2	TTC	nuoL	X15	scfm	0.05

Continued on next page

Table C.2 – continued from previous page

	Type	Position	Mut	Mutation category	Codon	Number	Pos	Ref	Gene name	Population	Treatment	Frequency
37	SNP	2820100	G	snp_nonsynonymous	GAG	43	1	AAG	PA14_32420	X15	scfm	0.91
38	SNP	3090251	A	snp_nonsynonymous	AAC	27	2	ACC	PA14_34800	X15	scfm	0.06
39	SNP	4304120	T	snp_intergenic	NA			NA	NA	X15	scfm	0.06
40	DEL	5234961	1	small_indel	NA			NA	pilB	X15	scfm	0.78
41	SUB	6391571	24	small_indel	NA			NA	PA14_71740	X15	scfm	0.06
42	SNP	1372561	T	snp_intergenic	NA			NA	NA	X16	min	0.08
43	SNP	2038034	T	snp_intergenic	NA			NA	NA	X16	min	0.15
44	SNP	2440884	T	snp_intergenic	NA			NA	NA	X16	min	0.07
45	DEL	3789885	24	small_indel	NA			NA	pscP	X16	min	1.00
46	SNP	4563014	T	snp_nonsynonymous	TCC	60	1	CCC	mvfR	X16	min	0.06
47	DEL	4563280	1	small_indel	NA			NA	mvfR	X16	min	0.19
48	SNP	4565228	A	snp_nonsynonymous	TAT	245	1	GAT	phnA	X16	min	0.05
49	SNP	5514089	T	snp_intergenic	NA			NA	NA	X16	min	0.08
50	SNP	6461159	T	snp_intergenic	NA			NA	NA	X16	min	0.08
51	SUB	578898	27	small_indel	NA			NA	NA	X17	min	0.08
52	SUB	2049004	23	small_indel	NA			NA	PA14_23520	X17	min	0.09
53	SUB	2158814	25	small_indel	NA			NA	PA14_24690	X17	min	0.08
54	SNP	4085367	A	snp_nonsynonymous	GTC	231	2	GCC	lasR	X17	min	0.09
55	SNP	4085419	C	snp_nonsynonymous	GAT	214	1	AAT	lasR	X17	min	0.10
56	SNP	4085743	G	snp_nonsynonymous	CCG	106	1	TCG	lasR	X17	min	0.06
57	SNP	4085824	G	snp_nonsynonymous	CGT	79	1	TGT	lasR	X17	min	0.30
58	SNP	4358295	A	snp_nonsense	TAA	252	1	GAA	PA14_49070	X17	min	0.11
59	SUB	4451598	23	small_indel	NA			NA	PA14_50070	X17	min	0.20
60	SNP	4462490	T	snp_intergenic	NA			NA	NA	X17	min	0.06
61	SNP	4760143	T	snp_intergenic	NA			NA	NA	X17	min	0.06
62	SNP	5000478	T	snp_nonsynonymous	TAC	593	1	AAC	pctA	X17	min	1.00
63	SNP	5276948	T	snp_nonsynonymous	CAT	75	2	CCT	PA14_59230	X17	min	0.05
64	SUB	5290113	30	small_indel	NA			NA	NA	X17	min	0.11
65	SNP	5353547	T	snp_nonsynonymous	GAC	168	2	GCC	dtd	X17	min	0.06
66	SNP	5464958	A	snp_intergenic	NA			NA	NA	X17	min	0.06
67	SNP	6098930	T	snp_nonsynonymous	TAC	196	1	GAC	PA14_68360	X17	min	0.08
68	SNP	711565	T	snp_intergenic	NA			NA	NA	X18	min	0.07
69	SNP	1131143	A	snp_nonsynonymous	AGC	415	1	CGC	PA14_13150	X18	min	0.06
70	SNP	3840548	A	snp_intergenic	NA			NA	NA	X18	min	0.10
71	SNP	1406896	A	snp_nonsynonymous	ACG	418	1	GCG	wspA	X19	scfm_mucin	0.10
72	SNP	2043176	A	snp_nonsynonymous	AAG	523	1	CAG	wbpM	X19	scfm_mucin	0.06
73	SNP	3725577	A	snp_nonsense	TAG	215	1	AAG	PA14_41740	X19	scfm_mucin	0.08

Continued on next page

Table C.2 – continued from previous page

	Type	Position	Mut	Mutation category	Codon	Number	Pos	Ref	Gene name	Population	Treatment	Frequency
74	DEL	4638173	84	large_deletion	NA			NA	PA14_52260	X19	scfm_mucin	0.22
75	SNP	4638252	T	snp_synonymous	CTT	816	3	CTG	PA14_52260	X19	scfm_mucin	0.15
76	DEL	4638257	60	large_deletion	NA			NA	PA14_52260	X19	scfm_mucin	0.36
77	SUB	4638317	3	small_indel	NA			NA	PA14_52260	X19	scfm_mucin	0.33
78	SNP	5431702	T	snp_nonsynonymous	TGC	956	1	AGC	morA	X19	scfm_mucin	0.72
79	SNP	6041803	A	snp_nonsynonymous	GTC	367	2	GCC	PA14_67630	X19	scfm_mucin	0.06
80	SNP	1500471	T	snp_intergenic	NA			NA	NA	X2	scfm	0.08
81	SNP	1567906	T	snp_synonymous	GCT	124	3	GCG	fruA	X2	scfm	0.05
82	SNP	2395847	T	snp_intergenic	NA			NA	NA	X2	scfm	0.05
83	SNP	2820614	A	snp_nonsynonymous	GAC	214	2	GGC	PA14_32420	X2	scfm	1.00
84	DEL	4286069	8	small_indel	NA			NA	NA	X2	scfm	0.83
85	SNP	5237874	A	snp_nonsense	TGA	223	3	TGG	pilD	X2	scfm	0.85
86	SNP	3561957	T	snp_intergenic	NA			NA	NA	X21	scfm_mucin	0.07
87	SNP	4085389	A	snp_nonsynonymous	TGC	224	1	CGC	lasR	X21	scfm_mucin	0.07
88	SNP	5327784	T	snp_intergenic	NA			NA	NA	X21	scfm_mucin	0.05
89	SNP	5811870	A	snp_synonymous	GTT	310	3	GTG	rnr	X21	scfm_mucin	0.05
90	SNP	193160	T	snp_nonsynonymous	CAG	150	2	CTG	PA14_02140	X22	min_mucin	1.00
91	SNP	1229406	C	snp_nonsynonymous	CTC	82	1	GTC	PA14_14420	X22	min_mucin	0.06
92	SUB	3743840	19	small_indel	NA			NA	aroF-1	X22	min_mucin	0.16
93	SNP	4627181	A	snp_nonsynonymous	TAC	442	1	CAC	relA	X22	min_mucin	1.00
94	SNP	5718287	T	snp_intergenic	NA			NA	NA	X22	min_mucin	0.08
95	SNP	1406957	T	snp_nonsynonymous	GTG	438	2	GCG	wspA	X23	min_mucin	0.36
96	SNP	2432899	A	snp_nonsynonymous	ATC	542	2	AGC	PA14_28100	X23	min_mucin	0.06
97	SNP	2711435	A	snp_nonsynonymous	CTC	522	2	CGC	PA14_31190	X23	min_mucin	0.05
98	SNP	3591204	C	snp_synonymous	GGG	1579	3	GGC	PA14_40260	X23	min_mucin	0.05
99	SNP	4085398	A	snp_nonsynonymous	TTG	221	1	GTG	lasR	X23	min_mucin	0.15
100	SNP	4382138	T	snp_intergenic	NA			NA	NA	X23	min_mucin	0.08
101	DEL	4562942	4	small_indel	NA			NA	mvfR	X23	min_mucin	0.18
102	SNP	4562985	A	snp_nonsynonymous	CAG	50	2	CTG	mvfR	X23	min_mucin	0.27
103	SNP	4566597	T	snp_nonsense	TAG	129	2	TGG	pqsE	X23	min_mucin	0.16
104	SNP	4627016	A	snp_nonsynonymous	TGG	497	1	CGG	relA	X23	min_mucin	0.09
105	SNP	5000184	T	snp_nonsynonymous	TCG	495	1	GCG	pctA	X23	min_mucin	0.14
106	SNP	5431912	A	snp_nonsynonymous	ACC	1026	1	TCC	morA	X23	min_mucin	0.11
107	SNP	1782499	A	snp_nonsynonymous	ATG	118	1	CTG	PA14_20690	X24	min_mucin	0.05
108	SNP	2389215	T	snp_synonymous	CCT	92	3	CCG	PA14_27520	X24	min_mucin	0.05
109	SNP	2449191	A	snp_nonsynonymous	ATG	154	1	CTG	PA14_28340	X24	min_mucin	0.05
110	DEL	2643554	3	small_indel	NA			NA	PA14_30540	X24	min_mucin	0.05

Continued on next page

Table C.2 – continued from previous page

	Type	Position	Mut	Mutation category	Codon	Number	Pos	Ref	Gene name	Population	Treatment	Frequency
111	SNP	2702264	T	snp_nonsynonymous	ATG	423	1	CTG	PA14_31070	X24	min_mucin	0.05
112	SNP	3090780	C	snp_synonymous	CAC	203	3	CAT	PA14_34800	X24	min_mucin	0.06
113	SNP	3627397	C	snp_nonsynonymous	GGC	676	2	GAC	cti	X24	min_mucin	0.05
114	SNP	4085461	A	snp_nonsynonymous	TTC	200	1	ATC	lasR	X24	min_mucin	0.81
115	SNP	4305508	T	snp_nonsynonymous	ACG	182	1	CCG	PA14_48420	X24	min_mucin	0.05
116	SNP	4440627	C	snp_nonsynonymous	CCC	333	2	CGC	PA14_49940	X24	min_mucin	0.06
117	SNP	5261683	A	snp_synonymous	GAA	430	3	GAG	PA14_59070	X24	min_mucin	0.05
118	SNP	5432294	T	snp_nonsynonymous	GTG	1153	2	GAG	morA	X24	min_mucin	1.00
119	SNP	5510964	T	snp_nonsynonymous	CTC	205	2	CAC	prs	X24	min_mucin	0.09
120	SNP	5913712	C	snp_nonsynonymous	CCG	436	1	TCG	PA14_66320	X24	min_mucin	1.00
121	SNP	2029791	T	snp_nonsynonymous	GTA	164	2	GGA	orfH	X25	scfm	0.65
122	DEL	2249637	97	large_deletion	NA			NA	pilZ	X25	scfm	0.09
123	SUB	2998670	20	small_indel	NA			NA	PA14_33760	X25	scfm	0.23
124	SNP	3150313	A	snp_nonsynonymous	CAT	161	3	CAG	pvcA	X25	scfm	0.06
125	SUB	3954240	30	small_indel	NA			NA	PA14_44420	X25	scfm	0.10
126	SNP	4560676	A	snp_synonymous	AGG	397	1	CGG	PA14_51320	X25	scfm	0.06
127	INS	4562937	C	small_indel	NA			NA	mvfR	X25	scfm	0.35
128	SNP	5002588	C	snp_nonsynonymous	GCC	588	2	GAC	pctB	X25	scfm	0.10
129	DEL	5233651	36	small_indel	NA			NA	pilA	X25	scfm	0.42
130	SNP	5951293	A	snp_nonsense	TAG	271	1	CAG	pilM	X25	scfm	0.09
131	DEL	5951627	51	large_deletion	NA			NA	pilM	X25	scfm	0.17
132	DEL	798130	11	small_indel	NA			NA	PA14_09300	X27	scfm	0.22
133	SNP	1863587	T	snp_synonymous	GGT	171	3	GGC	PA14_21490	X27	scfm	0.14
134	SNP	2820245	T	snp_nonsynonymous	GTC	91	2	GCC	PA14_32420	X27	scfm	0.20
135	DEL	3421317	42	small_indel	NA			NA	PA14_38380	X27	scfm	0.14
136	SNP	4085388	T	snp_nonsynonymous	CAC	224	2	CGC	lasR	X27	scfm	0.25
137	SNP	4085418	C	snp_nonsynonymous	AGT	214	2	AAT	lasR	X27	scfm	0.12
138	INS	4085716	C	small_indel	NA			NA	lasR	X27	scfm	0.11
139	INS	4085716	A	small_indel	NA			NA	lasR	X27	scfm	0.11
140	SNP	4353161	A	snp_intergenic	NA			NA	NA	X27	scfm	0.18
141	SNP	711562	T	snp_intergenic	NA			NA	NA	X28	min	0.06
142	AMP	4085564	68	large_amplification	NA			NA	lasR	X28	min	0.29
143	SNP	4563508	A	snp_nonsynonymous	ATA	224	3	ATG	mvfR	X28	min	0.64
144	SNP	717712	T	snp_intergenic	NA			NA	NA	X29	min	0.06
145	SNP	1072841	T	snp_nonsynonymous	TAC	126	1	CAC	PA14_12450	X29	min	0.07
146	SNP	1766531	A	snp_nonsynonymous	ACA	113	1	CCA	PA14_20530	X29	min	0.07
147	SNP	2459451	A	snp_nonsynonymous	AGT	55	1	GGT	PA14_28510	X29	min	0.05

Continued on next page

Table C.2 – continued from previous page

	Type	Position	Mut	Mutation category	Codon	Number	Pos	Ref	Gene name	Population	Treatment	Frequency
148	SNP	3283914	A	snp_nonsynonymous	TCG	373	1	GCG	glgP	X29	min	0.05
149	SNP	4286136	A	snp_intergenic	NA			NA	NA	X29	min	0.49
150	SNP	1253819	T	snp_intergenic	NA			NA	NA	X3	scfm	1.00
151	INS	2030092	G	small_indel	NA			NA	orfH	X3	scfm	1.00
152	SNP	2820358	T	snp_nonsynonymous	TGC	129	1	GGC	PA14_32420	X3	scfm	1.00
153	DEL	4085387	1	small_indel	NA			NA	lasR	X3	scfm	0.52
154	INS	5235882	4	small_indel	NA			NA	pilB	X3	scfm	0.14
155	DEL	5237063	11	small_indel	NA			NA	pilC	X3	scfm	0.75
156	SNP	5431409	C	snp_nonsynonymous	GCA	858	2	GTA	morA	X3	scfm	0.78
157	SNP	2028207	T	snp_nonsynonymous	TGT	42	1	CGT	orfK	X30	min	0.10
158	SNP	2029551	T	snp_nonsynonymous	CTC	84	2	CCC	orfH	X30	min	0.11
159	SNP	4400389	A	snp_nonsynonymous	AAC	390	2	ACC	PA14_49480	X30	min	0.05
160	SNP	4996301	T	snp_nonsynonymous	TCG	498	1	GCG	pctC	X30	min	0.08
161	SNP	1406872	A	snp_nonsynonymous	ATG	410	1	GTG	wspA	X31	scfm_mucin	0.69
162	SNP	2528418	T	snp_synonymous	GGA	528	3	GGC	PA14_29230	X31	scfm_mucin	0.08
163	SNP	2592436	A	snp_synonymous	GCT	583	3	GCG	nuoG	X31	scfm_mucin	0.06
164	SNP	2620539	A	snp_nonsynonymous	TTC	61	2	TGC	PA14_30260	X31	scfm_mucin	0.05
165	SNP	2761531	T	snp_nonsynonymous	ATC	115	2	AGC	PA14_31740	X31	scfm_mucin	0.05
166	SNP	3405151	A	snp_intergenic	NA			NA	NA	X31	scfm_mucin	0.05
167	SNP	3997251	A	snp_synonymous	CGA	223	3	CGC	alc	X31	scfm_mucin	0.07
168	SNP	4059678	C	snp_nonsynonymous	GGC	194	2	GCC	fleN	X31	scfm_mucin	0.21
169	SNP	4085989	A	snp_nonsense	TAG	24	1	CAG	lasR	X31	scfm_mucin	0.68
170	DEL	4637945	1	small_indel	NA			NA	PA14_52260	X31	scfm_mucin	0.11
171	INS	5001844	6	small_indel	NA			NA	pctB	X31	scfm_mucin	0.17
172	SNP	5704504	T	snp_nonsynonymous	AGC	21	1	GGC	PA14_64000	X31	scfm_mucin	0.07
173	DEL	5236903	11	small_indel	NA			NA	pilC	X32	scfm_mucin	0.35
174	SNP	5327726	A	snp_intergenic	NA			NA	NA	X32	scfm_mucin	0.08
175	SNP	5432483	C	snp_nonsynonymous	CCG	1216	2	CTG	morA	X32	scfm_mucin	0.45
176	SNP	6184260	T	snp_nonsynonymous	ACA	187	1	CCA	PA14_69330	X32	scfm_mucin	0.07
177	SNP	183050	A	snp_intergenic	NA			NA	NA	X33	scfm_mucin	0.05
178	SNP	309899	T	snp_nonsynonymous	TCC	51	1	GCC	PA14_03380	X33	scfm_mucin	0.08
179	DEL	1651156	1	small_indel	NA			NA	rhlR	X33	scfm_mucin	0.22
180	SUB	2370048	28	small_indel	NA			NA	PA14_27310	X33	scfm_mucin	0.19
181	SNP	2646906	G	snp_nonsynonymous	AAC	163	3	AAG	PA14_30570	X33	scfm_mucin	0.05
182	SNP	3647839	T	snp_intergenic	NA			NA	NA	X33	scfm_mucin	0.05
183	SNP	3959113	A	snp_nonsynonymous	CAC	117	2	CGC	PA14_44460	X33	scfm_mucin	0.08
184	SNP	4085442	A	snp_nonsynonymous	GTC	206	2	GCC	lasR	X33	scfm_mucin	0.33

Continued on next page

Table C.2 – continued from previous page

	Type	Position	Mut	Mutation category	Codon	Number	Pos	Ref	Gene name	Population	Treatment	Frequency
185	SNP	4571589	A	snp_intergenic	NA			NA	NA	X33	scfm_mucin	0.32
186	SNP	6077079	A	snp_intergenic	NA			NA	NA	X33	scfm_mucin	0.05
187	DEL	822979	23	small_indel	NA			NA	PA14_09570	X34	min_mucin	0.06
188	AMP	1412111	66	large_amplification	NA			NA	wspF	X34	min_mucin	0.09
189	SNP	4085394	A	snp_nonsynonymous	ATC	222	2	ACC	lasR	X34	min_mucin	0.39
190	SNP	4085430	G	snp_nonsynonymous	TCC	210	2	TTC	lasR	X34	min_mucin	0.10
191	SNP	4085452	T	snp_nonsynonymous	AGC	203	1	TGC	lasR	X34	min_mucin	0.06
192	DEL	4085614	11	small_indel	NA			NA	lasR	X34	min_mucin	0.05
193	SNP	4562898	C	snp_nonsynonymous	ACT	21	2	ATT	mvfR	X34	min_mucin	0.11
194	SNP	4563207	A	snp_nonsynonymous	CAG	124	2	CTG	mvfR	X34	min_mucin	0.09
195	SNP	5431466	T	snp_nonsynonymous	CTG	877	2	CAG	morA	X34	min_mucin	0.83
196	SNP	1462264	C	snp_nonsynonymous	AGG	195	2	AAG	map	X35	min_mucin	0.06
197	SNP	1733178	T	snp_nonsynonymous	ATC	383	2	AGC	PA14_20100	X35	min_mucin	0.05
198	SNP	2389199	G	snp_nonsynonymous	TGC	87	2	TAC	PA14_27520	X35	min_mucin	0.06
199	SNP	3752179	A	snp_intergenic	NA			NA	NA	X35	min_mucin	0.05
200	SNP	3752182	C	snp_intergenic	NA			NA	NA	X35	min_mucin	0.06
201	SNP	4085868	A	snp_nonsynonymous	TTC	64	2	TAC	lasR	X35	min_mucin	0.82
202	SNP	4362172	A	snp_intergenic	NA			NA	NA	X35	min_mucin	0.05
203	SNP	4689752	A	snp_nonsynonymous	CTG	157	2	CGG	PA14_52900	X35	min_mucin	0.08
204	DEL	5237535	5	small_indel	NA			NA	pilD	X35	min_mucin	0.46
205	SNP	5432476	T	snp_nonsynonymous	TTC	1214	1	ATC	morA	X35	min_mucin	0.31
206	SNP	5947522	G	snp_nonsynonymous	ACC	576	2	ATC	pilQ	X35	min_mucin	0.16
207	DEL	5947651	1	small_indel	NA			NA	pilQ	X35	min_mucin	0.06
208	SNP	6294638	A	snp_nonsynonymous	GTC	162	2	GGC	glcE	X35	min_mucin	0.06
209	SUB	1057599	12	small_indel	NA			NA	leuS	X37	scfm	0.11
210	SNP	2029863	T	snp_nonsynonymous	GTA	188	2	GCA	orfH	X37	scfm	0.39
211	SNP	2030543	T	snp_nonsynonymous	TAT	415	1	GAT	orfH	X37	scfm	0.23
212	SNP	2819976	A	snp_nonsynonymous	ATA	1	3	ATG	PA14_32420	X37	scfm	0.14
213	SNP	2998689	G	snp_nonsynonymous	GCC	600	2	GGC	PA14_33760	X37	scfm	0.08
214	DEL	3421372	9	small_indel	NA			NA	PA14_38380	X37	scfm	0.41
215	SNP	5233725	A	snp_nonsynonymous	CTG	109	2	CCG	pilA	X37	scfm	0.06
216	DEL	5235648	15	small_indel	NA			NA	pilB	X37	scfm	0.14
217	SNP	5236856	T	snp_nonsense	TAA	259	1	GAA	pilC	X37	scfm	0.05
218	DEL	5950937	15	small_indel	NA			NA	pilN	X37	scfm	0.30
219	AMP	798891	52	large_amplification	NA			NA	PA14_09320	X39	scfm	0.05
220	SNP	916378	C	snp_nonsynonymous	GCC	156	1	ACC	hpcC	X39	scfm	0.15
221	SNP	2029671	T	snp_nonsynonymous	GTC	124	2	GGC	orfH	X39	scfm	0.22

Continued on next page

Table C.2 – continued from previous page

	Type	Position	Mut	Mutation category	Codon	Number	Pos	Ref	Gene name	Population	Treatment	Frequency
222	SNP	2724679	C	snp_nonsynonymous	GCC	84	1	ACC	PA14_31330	X39	scfm	0.07
223	SNP	3190739	T	snp_intergenic	NA			NA	NA	X39	scfm	0.10
224	INS	4085349	G	small_indel	NA			NA	lasR	X39	scfm	0.29
225	SNP	4085414	C	snp_nonsynonymous	ATG	215	3	ATT	lasR	X39	scfm	0.14
226	SNP	4563134	A	snp_nonsynonymous	AAC	100	1	GAC	mvfR	X39	scfm	0.13
227	SNP	5432321	T	snp_nonsynonymous	CTC	1162	2	CGC	morA	X39	scfm	0.67
228	SNP	2028637	C	snp_nonsynonymous	CCC	185	2	CGC	orfK	X4	min	0.09
229	SNP	2038154	A	snp_nonsynonymous	CAT	23	2	CGT	orfL	X4	min	0.29
230	SNP	2040861	T	snp_nonsynonymous	ATC	238	2	AGC	orfN	X4	min	0.08
231	SNP	4085478	T	snp_nonsynonymous	AAT	194	2	AGT	lasR	X4	min	0.22
232	SNP	4085892	C	snp_nonsynonymous	TGC	56	2	TAC	lasR	X4	min	0.05
233	DEL	4085933	18	small_indel	NA			NA	lasR	X4	min	0.13
234	SNP	1297835	A	snp_nonsynonymous	GAG	48	2	GCG	guaB	X40	min	0.05
235	SNP	2158870	C	snp_nonsynonymous	GAG	75	1	AAG	PA14_24690	X40	min	0.06
236	SNP	2280278	A	snp_nonsynonymous	CTC	190	2	CGC	PA14_26110	X40	min	0.05
237	SNP	3633832	G	snp_nonsynonymous	GTC	917	1	ATC	metH	X40	min	0.06
238	DEL	4085745	1	small_indel	NA			NA	lasR	X40	min	0.71
239	SNP	4563636	A	snp_nonsynonymous	GAG	267	2	GTG	mvfR	X40	min	0.18
240	SNP	927898	T	snp_synonymous	TTG	50	1	CTG	PA14_10770	X41	min	0.06
241	SNP	1793545	A	snp_intergenic	NA			NA	NA	X41	min	0.07
242	SNP	2636412	G	snp_nonsynonymous	GCC	79	2	GTC	PA14_30440	X41	min	0.13
243	SNP	2998681	A	snp_nonsynonymous	TCC	603	1	GCC	PA14_33760	X41	min	0.11
244	SNP	4563305	T	snp_nonsynonymous	TGC	157	1	CGC	mvfR	X41	min	0.07
245	SNP	4660487	T	snp_nonsynonymous	CAC	31	2	CGC	rsmA	X41	min	0.07
246	SNP	5516031	T	snp_synonymous	TAT	304	3	TAC	PA14_61850	X41	min	0.07
247	SUB	5981889	16	small_indel	NA			NA	PA14_67010	X41	min	0.19
248	SNP	791864	T	snp_nonsynonymous	TTG	615	2	TGG	pchF	X42	min	0.70
249	SNP	1182133	T	snp_nonsynonymous	ATT	3	3	ATG	narK2	X42	min	0.10
250	SNP	2477846	A	snp_synonymous	AGA	392	3	AGG	PA14_28770	X42	min	0.13
251	SNP	2828100	T	snp_synonymous	CGA	285	3	CGG	PA14_32570	X42	min	0.08
252	SNP	3673535	T	snp_intergenic	NA			NA	NA	X42	min	1.00
253	SNP	5002581	G	snp_nonsynonymous	GAC	586	1	AAC	pctB	X42	min	0.10
254	SNP	6187778	T	snp_synonymous	GCA	179	3	GCG	algP	X42	min	0.06
255	SNP	336179	A	snp_intergenic	NA			NA	NA	X44	scfm_mucin	0.17
256	SNP	772264	T	snp_nonsynonymous	AAT	290	3	AAA	rpoA	X44	scfm_mucin	0.49
257	SNP	1135368	T	snp_intergenic	NA			NA	NA	X44	scfm_mucin	0.10
258	INS	1412532	5	small_indel	NA			NA	wspF	X44	scfm_mucin	0.54

Continued on next page

Table C.2 – continued from previous page

	Type	Position	Mut	Mutation category	Codon	Number	Pos	Ref	Gene name	Population	Treatment	Frequency
259	SNP	2028153	C	snp_nonsynonymous	CCA	24	1	GCA	orfK	X44	scfm_mucin	0.05
260	SNP	2029772	A	snp_nonsynonymous	ACT	158	1	CCT	orfH	X44	scfm_mucin	0.23
261	SNP	2135339	A	snp_intergenic	NA			NA	NA	X44	scfm_mucin	0.07
262	SNP	2484116	A	snp_synonymous	ATA	1236	3	ATC	PA14_28810	X44	scfm_mucin	0.06
263	SNP	3038400	A	snp_nonsynonymous	AGA	274	3	AGC	sfa3	X44	scfm_mucin	0.05
264	SNP	3048512	A	snp_synonymous	GCT	35	3	GCC	PA14_34280	X44	scfm_mucin	0.07
265	SNP	3980465	T	snp_nonsynonymous	ATG	936	1	CTG	PA14_44680	X44	scfm_mucin	0.05
266	SNP	4103065	T	snp_nonsynonymous	TCG	529	1	GCG	PA14_46120	X44	scfm_mucin	0.06
267	SNP	4327170	G	snp_nonsynonymous	GCC	74	1	ACC	PA14_48650	X44	scfm_mucin	0.06
268	SNP	5235200	T	snp_nonsense	TAG	308	1	CAG	pilB	X44	scfm_mucin	0.43
269	SNP	5603910	A	snp_intergenic	NA			NA	NA	X44	scfm_mucin	0.07
270	DEL	5951734	1	small_indel	NA			NA	pilM	X44	scfm_mucin	0.06
271	SNP	426695	T	snp_nonsynonymous	GAC	107	2	GCC	PA14_04830	X45	scfm_mucin	0.05
272	SNP	1500470	T	snp_intergenic	NA			NA	NA	X45	scfm_mucin	0.09
273	SNP	2190492	T	snp_synonymous	CTA	118	3	CTG	PA14_25040	X45	scfm_mucin	0.05
274	SNP	2392315	C	snp_intergenic	NA			NA	NA	X45	scfm_mucin	0.06
275	SNP	2720624	T	snp_nonsynonymous	CAG	197	2	CCG	PA14_31280	X45	scfm_mucin	0.06
276	SNP	2969352	T	snp_nonsynonymous	CAT	4246	3	CAG	PA14_33610	X45	scfm_mucin	0.05
277	SNP	2977306	A	snp_nonsynonymous	AAG	1747	1	GAG	pvdJ	X45	scfm_mucin	0.07
278	SNP	3208352	G	snp_nonsynonymous	CGC	98	1	AGC	PA14_36020	X45	scfm_mucin	0.06
279	SNP	3272202	A	snp_nonsynonymous	TCC	1028	1	GCC	PA14_36730	X45	scfm_mucin	0.06
280	SNP	3716236	T	snp_nonsynonymous	AAG	55	1	CAG	PA14_41630	X45	scfm_mucin	0.06
281	SNP	3781810	A	snp_nonsynonymous	TCG	236	1	CCG	pcrV	X45	scfm_mucin	0.06
282	SNP	3971746	A	snp_nonsynonymous	TCG	617	1	GCG	dnaX	X45	scfm_mucin	0.05
283	SNP	4085413	A	snp_nonsynonymous	TGG	216	1	CGG	lasR	X45	scfm_mucin	0.47
284	DEL	4086002	12	small_indel	NA			NA	lasR	X45	scfm_mucin	0.17
285	SNP	4563126	G	snp_nonsynonymous	GGG	97	2	GTG	mvfR	X45	scfm_mucin	1.00
286	SNP	4653653	C	snp_nonsynonymous	CCG	9	2	CTG	PA14_52440	X45	scfm_mucin	0.05
287	SNP	5432004	A	snp_nonsynonymous	ATA	1056	3	ATG	morA	X45	scfm_mucin	0.53
288	SNP	154655	T	snp_nonsynonymous	GTC	235	2	GCC	PA14_01680	X46	min_mucin	0.05
289	SNP	1406966	T	snp_nonsynonymous	GTC	441	2	GCC	wspA	X46	min_mucin	0.26
290	SNP	1546047	A	snp_synonymous	TCA	15	3	TCC	PA14_18010	X46	min_mucin	0.08
291	SNP	2049951	T	snp_nonsynonymous	GAA	137	2	GCA	PA14_23530	X46	min_mucin	0.05
292	SNP	2529063	T	snp_synonymous	CGA	313	3	CGC	PA14_29230	X46	min_mucin	0.06
293	SNP	2976258	A	snp_nonsynonymous	AAA	1397	3	AAC	pvdJ	X46	min_mucin	0.05
294	SNP	3530433	T	snp_nonsynonymous	TTA	816	3	TTC	cobN	X46	min_mucin	0.05
295	SNP	3611013	A	snp_nonsynonymous	AAC	82	2	ACC	modB	X46	min_mucin	0.06
296	SNP	4085512	T	snp_nonsynonymous	AAA	183	1	GAA	lasR	X46	min_mucin	0.21

Continued on next page

Table C.2 – continued from previous page

	Type	Position	Mut	Mutation category	Codon	Number	Pos	Ref	Gene name	Population	Treatment	Frequency
297	SNP	4627721	T	snp_nonsynonymous	AGG	262	1	GGG	relA	X46	min_mucin	0.15
298	SNP	4696935	T	snp_nonsynonymous	GAA	326	3	GAC	phhR	X46	min_mucin	0.05
299	SNP	5073876	T	snp_synonymous	TCA	156	3	TCC	PA14_56930	X46	min_mucin	0.06
300	SNP	5431844	T	snp_nonsynonymous	GTC	1003	2	GCC	morA	X46	min_mucin	0.37
301	SNP	5432308	C	snp_nonsynonymous	CAA	1158	1	GAA	morA	X46	min_mucin	0.17
302	SNP	5840443	T	snp_intergenic	NA			NA	NA	X46	min_mucin	0.40
303	SNP	5983507	A	snp_nonsynonymous	ATG	30	2	AGG	PA14_67030	X46	min_mucin	0.06
304	SNP	924058	G	snp_synonymous	CTC	431	3	CTG	PA14_10730	X47	min_mucin	0.07
305	SNP	1944948	A	snp_nonsynonymous	GAG	133	2	GCG	PA14_22330	X47	min_mucin	0.05
306	SUB	1947057	17	small_indel	NA			NA	actP	X47	min_mucin	0.12
307	SNP	2603955	G	snp_nonsynonymous	CAC	67	1	TAC	PA14_30050	X47	min_mucin	0.06
308	SNP	4085471	A	snp_nonsynonymous	GAT	196	3	GAG	lasR	X47	min_mucin	0.63
309	INS	4085791	5	small_indel	NA			NA	lasR	X47	min_mucin	0.12
310	SNP	4347219	T	snp_nonsynonymous	AAG	95	1	CAG	PA14_48890	X47	min_mucin	0.06
311	DEL	4563143	28	small_indel	NA			NA	mvfR	X47	min_mucin	0.48
312	SNP	4563369	A	snp_nonsense	TAG	178	2	TTG	mvfR	X47	min_mucin	0.10
313	SNP	4821943	T	snp_synonymous	GCA	163	3	GCC	lepA	X47	min_mucin	0.06
314	SNP	838037	T	snp_synonymous	CTA	319	3	CTG	PA14_09740	X48	min_mucin	0.06
315	SNP	1098982	A	snp_synonymous	TCA	276	3	TCC	PA14_12820	X48	min_mucin	0.05
316	SNP	1335004	A	snp_nonsynonymous	ATA	59	3	ATG	PA14_15700	X48	min_mucin	0.05
317	INS	1791078	31	small_indel	NA			NA	recQ	X48	min_mucin	0.06
318	SNP	2240434	G	snp_nonsynonymous	CGT	9	2	CAT	PA14_25620	X48	min_mucin	1.00
319	SNP	2552036	T	snp_nonsynonymous	ATG	228	1	CTG	PA14_29420	X48	min_mucin	0.05
320	SNP	2557519	T	snp_nonsynonymous	TGT	110	3	TGG	PA14_29490	X48	min_mucin	0.06
321	SNP	2663318	A	snp_nonsynonymous	TTA	725	3	TTC	PA14_30700	X48	min_mucin	0.06
322	SNP	2724310	A	snp_nonsynonymous	TAT	207	1	GAT	PA14_31330	X48	min_mucin	0.05
323	SNP	3780833	A	snp_synonymous	TCT	101	3	TCC	popB	X48	min_mucin	0.06
324	SNP	4085481	C	snp_nonsynonymous	AGC	193	2	ACC	lasR	X48	min_mucin	0.12
325	SNP	4436534	A	snp_synonymous	GTA	368	3	GTG	tpbB	X48	min_mucin	0.05
326	SNP	4618596	T	snp_nonsynonymous	CAT	288	3	CAG	purM	X48	min_mucin	0.06
327	SNP	5115062	T	snp_synonymous	CGA	367	3	CGC	murE	X48	min_mucin	0.07
328	DEL	5511285	1	small_indel	NA			NA	prs	X48	min_mucin	0.71
329	SNP	6279459	C	snp_nonsynonymous	CCC	104	1	ACC	spoT	X48	min_mucin	0.08
330	DEL	1651362	1	small_indel	NA			NA	rhlR	X49	scfm	0.07
331	DEL	1651363	1	small_indel	NA			NA	rhlR	X49	scfm	0.06
332	SNP	2029518	T	snp_nonsynonymous	CTT	73	2	CCT	orfH	X49	scfm	0.08
333	SNP	2030543	C	snp_nonsynonymous	CAT	415	1	GAT	orfH	X49	scfm	0.35

Continued on next page

Table C.2 – continued from previous page

	Type	Position	Mut	Mutation category	Codon	Number	Pos	Ref	Gene name	Population	Treatment	Frequency
334	SNP	2816668	A	snp_synonymous	CCT	225	3	CCG	mexF	X49	scfm	0.05
335	SNP	2818982	T	snp_nonsynonymous	AGC	258	1	GGC	mexT	X49	scfm	0.09
336	SNP	2820109	T	snp_nonsynonymous	TTC	46	1	CTC	PA14_32420	X49	scfm	0.53
337	SNP	2820238	G	snp_nonsynonymous	GTC	89	1	TTC	PA14_32420	X49	scfm	0.07
338	SNP	3237836	A	snp_nonsynonymous	ACC	365	1	CCC	exoY	X49	scfm	0.05
339	SNP	4079152	A	snp_nonsynonymous	TCG	371	1	GCG	PA14_45890	X49	scfm	0.05
340	INS	4637872	5	small_indel	NA			NA	PA14_52260	X49	scfm	0.23
341	SNP	5432161	A	snp_nonsynonymous	ACC	1109	1	GCC	morA	X49	scfm	0.08
342	DEL	4085636	1	small_indel	NA			NA	lasR	X5	min	0.14
343	SNP	4085670	G	snp_nonsynonymous	CCC	130	2	CTC	lasR	X5	min	0.13
344	SNP	4085709	A	snp_nonsynonymous	CTG	117	2	CCG	lasR	X5	min	0.22
345	SNP	4085773	A	snp_nonsense	TGA	96	1	CGA	lasR	X5	min	0.36
346	SUB	4407878	30	small_indel	NA			NA	toxA	X5	min	0.09
347	SUB	5903651	27	small_indel	NA			NA	ilvE	X5	min	0.08
348	SNP	309907	T	snp_nonsynonymous	AAT	53	3	AAG	PA14_03380	X50	scfm	0.09
349	SNP	1115331	G	snp_synonymous	CCC	4	3	CCG	PA14_12990	X50	scfm	0.12
350	DEL	1261781	4	small_indel	NA			NA	pilF	X50	scfm	0.06
351	SNP	3283492	T	snp_synonymous	AGA	513	3	AGG	glgP	X50	scfm	0.05
352	SNP	4085499	T	snp_nonsense	TAG	187	2	TGG	lasR	X50	scfm	0.64
353	SNP	5248330	A	snp_synonymous	GCA	459	3	GCG	PA14_58900	X50	scfm	0.05
354	SNP	5624300	A	snp_synonymous	GTA	65	3	GTC	omlA	X50	scfm	0.05
355	SNP	182973	T	snp_intergenic	NA			NA	NA	X51	scfm	0.05
356	DEL	976008	1	small_indel	NA			NA	PA14_11210	X51	scfm	0.10
357	SUB	1228164	27	small_indel	NA			NA	PA14_14390	X51	scfm	0.14
358	DEL	1261549	1	small_indel	NA			NA	pilF	X51	scfm	0.13
359	SNP	2029395	G	snp_nonsynonymous	GGC	32	2	GTC	orfH	X51	scfm	0.07
360	SNP	2040598	A	snp_nonsense	TAA	150	3	TAC	orfN	X51	scfm	0.10
361	SUB	2969226	30	small_indel	NA			NA	PA14_33610	X51	scfm	0.10
362	SNP	3179559	G	snp_nonsynonymous	GCC	685	1	ACC	PA14_35740	X51	scfm	0.05
363	SNP	3179768	G	snp_nonsynonymous	GAG	754	3	GAC	PA14_35740	X51	scfm	0.06
364	SNP	3421373	T	snp_nonsynonymous	GAC	50	2	GGC	PA14_38380	X51	scfm	0.08
365	DEL	4085696	19	small_indel	NA			NA	lasR	X51	scfm	0.30
366	SNP	4085832	C	snp_nonsynonymous	GGC	76	2	GTC	lasR	X51	scfm	0.18
367	DEL	4562937	1	small_indel	NA			NA	mvfR	X51	scfm	0.05
368	DEL	4562938	11	small_indel	NA			NA	mvfR	X51	scfm	0.21
369	SNP	5235113	C	snp_nonsynonymous	CCC	279	1	ACC	pilB	X51	scfm	0.13
370	SNP	5237595	A	snp_nonsense	TGA	130	3	TGG	pilD	X51	scfm	0.06

Continued on next page

Table C.2 – continued from previous page

	Type	Position	Mut	Mutation category	Codon	Number	Pos	Ref	Gene name	Population	Treatment	Frequency
371	SNP	5432281	A	snp_nonsynonymous	ACC	1149	1	GCC	morA	X51	scfm	0.07
372	SNP	5763244	T	snp_nonsynonymous	TCC	165	1	GCC	PA14_64710	X51	scfm	0.05
373	SNP	1247322	A	snp_nonsynonymous	CAG	59	2	CTG	yajC	X52	min	0.05
374	SNP	2982718	G	snp_nonsynonymous	CGG	1357	2	CAG	pvdD	X52	min	0.05
375	SNP	3027568	T	snp_nonsynonymous	ATC	6	1	CTC	hcp3	X52	min	0.05
376	SNP	5153626	A	snp_nonsynonymous	GAC	244	2	GGC	PA14_57890	X52	min	0.05
377	SNP	463441	T	snp_nonsynonymous	ACC	228	1	CCC	gshB	X53	min	0.06
378	DEL	2030254	3	small_indel	NA			NA	orfH	X53	min	0.12
379	SNP	4085494	T	snp_nonsynonymous	ACC	189	1	GCC	lasR	X53	min	0.28
380	DEL	4563823	1	small_indel	NA			NA	phnB	X53	min	0.12
381	SNP	3593824	G	snp_nonsynonymous	GCC	706	2	GGC	PA14_40260	X54	min	0.06
382	SNP	1913785	G	snp_intergenic	NA			NA	NA	X57	scfm_mucin	0.25
383	SNP	4636294	T	snp_nonsense	TGA	164	1	GGA	PA14_52260	X57	scfm_mucin	0.09
384	SNP	5431406	T	snp_nonsynonymous	ATG	857	2	ACG	morA	X57	scfm_mucin	1.00
385	DEL	5914263	6	small_indel	NA			NA	PA14_66320	X57	scfm_mucin	0.71
386	SUB	6361913	30	small_indel	NA			NA	PA14_71430	X57	scfm_mucin	0.06
387	SNP	103	T	snp_intergenic	NA			NA	NA	X58	min_mucin	0.07
388	SUB	1514309	29	small_indel	NA			NA	potC	X58	min_mucin	0.15
389	SNP	3591151	T	snp_nonsynonymous	GAC	1597	2	GCC	PA14_40260	X58	min_mucin	0.05
390	SNP	3644424	T	snp_intergenic	NA			NA	NA	X58	min_mucin	0.08
391	INS	4085740	6	small_indel	NA			NA	lasR	X58	min_mucin	0.16
392	DEL	4563393	1	small_indel	NA			NA	mvfR	X58	min_mucin	0.46
393	DEL	1631424	93	large_deletion	NA			NA	PA14_18920	X59	min_mucin	0.88
394	SNP	1676526	A	snp_nonsynonymous	TGC	50	1	GGC	PA14_19390	X59	min_mucin	0.06
395	SNP	3138551	T	snp_nonsynonymous	TAC	163	2	TCC	gnd	X59	min_mucin	0.06
396	SNP	3502356	A	snp_synonymous	CCT	7	3	CCG	rbsA	X59	min_mucin	0.05
397	SNP	3780514	A	snp_nonsynonymous	TTC	208	1	GTC	popB	X59	min_mucin	0.07
398	SNP	4130818	A	snp_synonymous	CTT	110	3	CTG	PA14_46440	X59	min_mucin	0.06
399	SNP	4406474	G	snp_nonsynonymous	GGA	42	2	GAA	PA14_49540	X59	min_mucin	0.06
400	INS	4563815	C	small_indel	NA			NA	phnB	X59	min_mucin	0.08
401	INS	4563815	C	small_indel	NA			NA	phnB	X59	min_mucin	0.08
402	SNP	4975216	T	snp_nonsynonymous	TTG	126	1	GTG	PA14_55740	X59	min_mucin	0.05
403	SNP	5431756	C	snp_nonsynonymous	CTC	974	1	ATC	morA	X59	min_mucin	1.00
404	SUB	1228164	29	small_indel	NA			NA	PA14_14390	X6	min	0.08
405	SNP	1660582	G	snp_intergenic	NA			NA	NA	X6	min	0.11
406	DEL	1718824	46	small_indel	NA			NA	PA14_19940	X6	min	0.06
407	SUB	2514981	19	small_indel	NA			NA	PA14_29070	X6	min	0.12

Continued on next page

Table C.2 – continued from previous page

	Type	Position	Mut	Mutation category	Codon	Number	Pos	Ref	Gene name	Population	Treatment	Frequency
408	SNP	2870859	T	snp_nonsynonymous	ACA	798	1	CCA	PA14_32830	X6	min	0.09
409	DEL	4085357	1	small_indel	NA			NA	lasR	X6	min	1.00
410	SNP	6456436	A	snp_intergenic	NA			NA	NA	X6	min	0.05
411	SNP	195681	T	snp_intergenic	NA			NA	NA	X60	min_mucin	1.00
412	SNP	334774	C	snp_intergenic	NA			NA	NA	X60	min_mucin	0.06
413	SNP	1516173	A	snp_nonsynonymous	CTC	253	2	CGC	potA	X60	min_mucin	0.07
414	SNP	2278222	A	snp_intergenic	NA			NA	NA	X60	min_mucin	0.05
415	SNP	2579549	C	snp_nonsynonymous	AAC	281	3	AAG	PA14_29770	X60	min_mucin	0.05
416	SNP	2730434	A	snp_nonsynonymous	GAA	95	3	GAC	PA14_31390	X60	min_mucin	0.05
417	SNP	2902417	T	snp_nonsynonymous	GTC	49	2	GGC	PA14_33150	X60	min_mucin	0.05
418	SNP	2963052	G	snp_synonymous	AAG	2146	3	AAA	PA14_33610	X60	min_mucin	0.05
419	SNP	4289518	A	snp_nonsynonymous	GAT	542	3	GAG	PA14_48170	X60	min_mucin	0.06
420	SNP	4350410	A	snp_nonsynonymous	GAT	85	3	GAG	PA14_48930	X60	min_mucin	0.92
421	SNP	4350459	A	snp_nonsynonymous	ATC	69	2	ACC	PA14_48930	X60	min_mucin	0.88
422	SNP	4350523	C	snp_nonsynonymous	GTG	48	1	ATG	PA14_48930	X60	min_mucin	0.92
423	SNP	4353292	T	snp_synonymous	AGT	2	3	AGC	PA14_49020	X60	min_mucin	0.94
424	INS	4353540	4	small_indel	NA			NA	PA14_49020	X60	min_mucin	0.07
425	SNP	4635074	A	snp_synonymous	ACT	182	3	ACC	PA14_52250	X60	min_mucin	0.07
426	DEL	5376988	71	large_deletion	NA			NA	pilE	X60	min_mucin	0.09
427	SNP	5951196	G	snp_nonsynonymous	CCC	303	2	CTC	pilM	X60	min_mucin	0.07
428	SNP	748276	T	snp_nonsynonymous	TGT	692	1	CGT	rpoB	X61	scfm	0.17
429	SNP	2603059	T	snp_synonymous	GTA	365	3	GTC	PA14_30050	X61	scfm	0.05
430	SNP	3502894	G	snp_nonsynonymous	GCC	155	2	GTC	rbsB	X61	scfm	0.05
431	SNP	3929490	T	snp_nonsynonymous	GAA	370	2	GGA	PA14_44160	X61	scfm	0.06
432	SNP	4060101	C	snp_nonsynonymous	AGC	53	2	AAC	fleN	X61	scfm	0.16
433	SNP	4085383	G	snp_nonsynonymous	CTA	226	1	GTA	lasR	X61	scfm	0.26
434	SNP	4085504	A	snp_nonsynonymous	TTT	185	3	TTG	lasR	X61	scfm	0.08
435	SNP	2031197	A	snp_nonsense	TAA	184	2	TCA	orfE	X62	scfm	0.05
436	SNP	4085391	A	snp_nonsynonymous	TTC	223	2	TCC	lasR	X62	scfm	0.18
437	SNP	4085433	C	snp_nonsynonymous	AGC	209	2	AAC	lasR	X62	scfm	0.61
438	SNP	5981875	T	snp_nonsynonymous	TCG	513	1	GCG	PA14_67010	X62	scfm	0.09
439	SNP	6121696	A	snp_nonsynonymous	CAG	115	2	CCG	rimK	X62	scfm	0.05
440	SNP	2028985	T	snp_nonsynonymous	TTG	301	2	TCG	orfK	X64	min	1.00
441	SNP	3735568	A	snp_nonsynonymous	GTC	97	2	GGC	PA14_41810	X64	min	0.05
442	DEL	4084849	4947	large_deletion	NA			NA	lasI	X64	min	1.00
443	SNP	2228566	T	snp_intergenic	NA			NA	NA	X65	min	0.08
444	SNP	2618110	T	snp_nonsynonymous	GTG	260	2	GCG	clpA	X65	min	0.05

Continued on next page

Table C.2 – continued from previous page

	Type	Position	Mut	Mutation category	Codon	Number	Pos	Ref	Gene name	Population	Treatment	Frequency
445	SNP	3722467	A	snp_synonymous	CGA	2	3	CGC	PA14_41710	X65	min	0.05
446	SNP	4286154	A	snp_intergenic	NA			NA	NA	X65	min	0.26
447	SNP	4660506	A	snp_nonsynonymous	TTC	25	1	GTC	rsmA	X65	min	0.10
448	SNP	501785	T	snp_nonsynonymous	TGG	691	1	CGG	PA14_05640	X66	min	0.05
449	SNP	1817289	T	snp_nonsynonymous	TTC	45	1	GTC	PA14_21030	X66	min	0.07
450	SNP	2028214	C	snp_nonsynonymous	ACG	44	2	ATG	orfK	X66	min	0.29
451	SNP	2029791	A	snp_nonsynonymous	GAA	164	2	GGA	orfH	X66	min	0.21
452	SNP	3102730	T	snp_nonsense	TAG	1982	1	CAG	PA14_34840	X66	min	0.06
453	SNP	3639177	G	snp_nonsynonymous	CGG	436	2	CAG	cysI	X66	min	0.06
454	SNP	3673812	A	snp_nonsense	TAA	535	1	GAA	ppiD	X66	min	0.06
455	DEL	4085332	13	small_indel	NA			NA	rsaL	X66	min	0.12
456	SNP	4085383	A	snp_nonsynonymous	TTA	226	1	GTA	lasR	X66	min	0.08
457	SUB	4402222	22	small_indel	NA			NA	PA14_49500	X66	min	0.06
458	SNP	4562871	C	snp_nonsynonymous	TCC	12	2	TTC	mvfR	X66	min	0.09
459	DEL	4562946	4	small_indel	NA			NA	mvfR	X66	min	0.13
460	SNP	1586038	T	snp_nonsense	TAA	170	3	TAC	algX	X67	scfm_mucin	0.06
461	SNP	3080722	T	snp_synonymous	GAA	49	3	GAG	PA14_34680	X67	scfm_mucin	0.07
462	SNP	3203459	A	snp_synonymous	GTT	82	3	GTG	PA14_35990	X67	scfm_mucin	0.06
463	SNP	3236575	A	snp_intergenic	NA			NA	NA	X67	scfm_mucin	0.05
464	DEL	4085372	1	small_indel	NA			NA	lasR	X67	scfm_mucin	0.08
465	DEL	4085373	1	small_indel	NA			NA	lasR	X67	scfm_mucin	0.08
466	SNP	5124442	A	snp_synonymous	CGA	123	3	CGC	PA14_57510	X67	scfm_mucin	0.05
467	SNP	661309	A	snp_synonymous	TCT	55	3	TCG	PA14_07680	X68	scfm_mucin	0.05
468	SNP	1827299	T	snp_nonsynonymous	GAC	203	2	GCC	PA14_21130	X68	scfm_mucin	0.05
469	SNP	2890678	T	snp_nonsynonymous	TAT	80	1	GAT	sdaA	X68	scfm_mucin	0.05
470	SNP	3060089	T	snp_nonsynonymous	ATG	26	1	CTG	PA14_34390	X68	scfm_mucin	0.06
471	SNP	3185909	C	snp_nonsynonymous	CGC	136	2	CAC	PA14_35800	X68	scfm_mucin	0.05
472	DEL	4085423	1	small_indel	NA			NA	lasR	X68	scfm_mucin	1.00
473	SNP	4563049	C	snp_nonsynonymous	ATC	71	3	ATG	mvfR	X68	scfm_mucin	0.07
474	SNP	4563533	A	snp_nonsynonymous	AGA	233	1	GGA	mvfR	X68	scfm_mucin	0.67
475	SNP	4870444	A	snp_synonymous	CGT	508	3	CGC	PA14_54920	X68	scfm_mucin	0.09
476	SNP	5536793	G	snp_noncoding	NA			NA	NA	X68	scfm_mucin	0.33
477	SNP	5536795	G	snp_noncoding	NA			NA	NA	X68	scfm_mucin	0.28
478	SNP	1406636	T	snp_nonsynonymous	TTG	331	2	TCG	wspA	X69	scfm_mucin	0.77
479	SNP	1884757	C	snp_nonsynonymous	AAG	815	3	AAC	PA14_21700	X69	scfm_mucin	0.21
480	SNP	2294479	A	snp_nonsynonymous	AGA	453	3	AGC	PA14_26280	X69	scfm_mucin	0.05
481	SNP	2633524	T	snp_nonsynonymous	ATA	195	3	ATG	PA14_30410	X69	scfm_mucin	0.06

Continued on next page

Table C.2 – continued from previous page

	Type	Position	Mut	Mutation category	Codon	Number	Pos	Ref	Gene name	Population	Treatment	Frequency
482	SNP	3592852	C	snp_nonsynonymous	AGC	1030	2	ACC	PA14_40260	X69	scfm_mucin	0.07
483	SNP	4085596	C	snp_nonsynonymous	GAG	155	1	AAG	lasR	X69	scfm_mucin	0.07
484	SNP	4221815	A	snp_nonsynonymous	ATC	412	2	AGC	PA14_47380	X69	scfm_mucin	0.06
485	SNP	2418822	A	snp_synonymous	ATA	274	3	ATC	PA14_27940	X7	scfm_mucin	0.05
486	SNP	2853168	T	snp_nonsynonymous	GTC	1534	2	GGC	PA14_32790	X7	scfm_mucin	0.05
487	SNP	2943382	A	snp_nonsynonymous	GAC	232	2	GGC	PA14_33440	X7	scfm_mucin	0.06
488	SNP	3310838	T	snp_nonsynonymous	GTG	88	2	GCG	PA14_37150	X7	scfm_mucin	0.06
489	SNP	3534006	T	snp_nonsense	TAG	279	2	TCG	PA14_39650	X7	scfm_mucin	0.05
490	SNP	3600151	A	snp_nonsynonymous	TTT	107	3	TTG	lasA	X7	scfm_mucin	0.06
491	SNP	4055709	C	snp_nonsynonymous	GGC	495	1	CGC	PA14_45590	X7	scfm_mucin	0.06
492	SNP	4085716	G	snp_nonsynonymous	CCC	115	1	ACC	lasR	X7	scfm_mucin	0.78
493	SNP	4085818	A	snp_nonsense	TAG	81	1	CAG	lasR	X7	scfm_mucin	0.09
494	SNP	4085379	A	snp_nonsynonymous	GTG	227	2	GCG	lasR	X70	min_mucin	0.41
495	SNP	5430581	A	snp_nonsynonymous	CAC	582	2	CGC	morA	X70	min_mucin	1.00
496	SNP	744880	G	snp_intergenic	NA			NA	NA	X71	min_mucin	0.30
497	SNP	1302902	T	snp_synonymous	CGA	274	3	CGC	PA14_15360	X71	min_mucin	0.10
498	SNP	2033774	A	snp_nonsynonymous	GAA	181	2	GCA	orfJ	X71	min_mucin	0.16
499	SNP	2866661	T	snp_intergenic	NA			NA	NA	X71	min_mucin	0.06
500	DEL	3764737	60	large_deletion	NA			NA	PA14_42220	X71	min_mucin	0.08
501	SNP	4085479	C	snp_nonsynonymous	GGT	194	1	AGT	lasR	X71	min_mucin	0.35
502	SNP	4562892	C	snp_nonsynonymous	GCT	19	2	GGT	mvfR	X71	min_mucin	0.53
503	SUB	2626160	19	small_indel	NA			NA	PA14_30320	X72	min_mucin	0.21
504	SNP	2870083	T	snp_intergenic	NA			NA	NA	X72	min_mucin	0.05
505	SNP	3106954	A	snp_synonymous	GCA	196	3	GCG	PA14_34900	X72	min_mucin	0.06
506	SNP	3914653	T	snp_synonymous	GTA	147	3	GTG	lpdG	X72	min_mucin	0.05
507	SNP	4085521	A	snp_nonsynonymous	TGG	180	1	CGG	lasR	X72	min_mucin	1.00
508	INS	4201892	6	small_indel	NA			NA	cyoE	X72	min_mucin	0.95
509	SNP	5236660	A	snp_nonsense	TGA	193	3	TGG	pilC	X72	min_mucin	0.90
510	SNP	6246917	T	snp_intergenic	NA			NA	NA	X72	min_mucin	1.00
511	DEL	3095107	30	small_indel	NA			NA	PA14_34820	X73	scfm	0.05
512	SUB	1228164	28	small_indel	NA			NA	PA14_14390	X75	scfm	0.15
513	SNP	2027756	T	snp_nonsynonymous	GTC	304	2	GGC	wzz	X75	scfm	0.06
514	SNP	2029772	T	snp_nonsynonymous	TCT	158	1	CCT	orfH	X75	scfm	0.26
515	DEL	2343030	1	small_indel	NA			NA	PA14_26930	X75	scfm	0.05
516	DEL	2343031	1	small_indel	NA			NA	PA14_26930	X75	scfm	0.05
517	SUB	2587770	33	small_indel	NA			NA	nuoL	X75	scfm	0.17
518	SNP	2820748	C	snp_nonsynonymous	CTC	259	1	TTC	PA14_32420	X75	scfm	0.47

Continued on next page

Table C.2 – continued from previous page

	Type	Position	Mut	Mutation category	Codon	Number	Pos	Ref	Gene name	Population	Treatment	Frequency
519	DEL	3421001	1	small_indel	NA			NA	PA14_38380	X75	scfm	0.26
520	SNP	4085526	A	snp_nonsynonymous	ATC	178	2	ACC	lasR	X75	scfm	0.48
521	SNP	5233548	A	snp_nonsynonymous	ATC	168	2	ACC	pilA	X75	scfm	0.08
522	INS	5951396	5	small_indel	NA			NA	pilM	X75	scfm	0.10
523	SNP	2038049	T	snp_intergenic	NA			NA	NA	X76	min	1.00
524	SNP	3344707	A	snp_intergenic	NA			NA	NA	X76	min	0.05
525	SNP	4009844	T	snp_nonsynonymous	ATG	82	1	CTG	moaA2	X76	min	0.06
526	SNP	4085808	T	snp_nonsynonymous	CAG	84	2	CTG	lasR	X76	min	0.57
527	SNP	4399309	T	snp_nonsynonymous	GTA	30	2	GGA	PA14_49480	X76	min	0.11
528	SNP	4571445	A	snp_nonsense	TAG	9	1	GAG	pqsA	X76	min	0.45
529	SNP	4085670	C	snp_nonsynonymous	CGC	130	2	CTC	lasR	X77	min	0.52
530	SNP	4085516	A	snp_nonsynonymous	GAT	181	3	GAG	lasR	X78	min	0.06
531	DEL	4085721	10	small_indel	NA			NA	lasR	X78	min	0.34
532	SNP	187671	T	snp_intergenic	NA			NA	NA	X79	scfm_mucin	0.06
533	DEL	609230	1	small_indel	NA			NA	PA14_07030	X79	scfm_mucin	0.05
534	SNP	1578415	A	snp_intergenic	NA			NA	NA	X79	scfm_mucin	0.06
535	SNP	2061164	A	snp_nonsynonymous	GAA	37	2	GCA	PA14_23700	X79	scfm_mucin	0.06
536	SNP	3233921	A	snp_nonsynonymous	ATC	251	2	AGC	hcnC	X79	scfm_mucin	0.05
537	SNP	4817673	A	snp_nonsynonymous	TAC	292	1	GAC	era	X79	scfm_mucin	0.05
538	SNP	5301958	A	snp_nonsynonymous	AAG	24	1	CAG	PA14_59580	X79	scfm_mucin	0.06
539	SNP	5357056	T	snp_nonsynonymous	ATT	411	3	ATG	PA14_60130	X79	scfm_mucin	0.06
540	INS	2032480	T	small_indel	NA			NA	PA14_23400	X8	scfm_mucin	1.00
541	SNP	2198185	A	snp_nonsynonymous	CAC	512	2	CGC	topA	X8	scfm_mucin	0.05
542	SNP	2763101	T	snp_synonymous	GGT	273	3	GGG	PA14_31760	X8	scfm_mucin	0.05
543	SNP	2980187	A	snp_synonymous	CGA	513	3	CGG	pvdD	X8	scfm_mucin	0.05
544	SNP	3445440	A	snp_nonsynonymous	TCG	97	1	GCG	PA14_38610	X8	scfm_mucin	0.06
545	SNP	3534723	G	snp_nonsynonymous	GCG	40	2	GTG	PA14_39650	X8	scfm_mucin	0.06
546	SNP	3580057	T	snp_nonsynonymous	GTC	84	2	GCC	PA14_40180	X8	scfm_mucin	0.05
547	SNP	3864536	A	snp_nonsynonymous	TCG	405	1	GCG	kdpB	X8	scfm_mucin	0.06
548	INS	4085804	T	small_indel	NA			NA	lasR	X8	scfm_mucin	1.00
549	SNP	4114838	A	snp_synonymous	CTA	61	3	CTC	PA14_46260	X8	scfm_mucin	0.06
550	SNP	4562909	A	snp_nonsynonymous	ACG	25	1	GCG	mvfR	X8	scfm_mucin	0.42
551	SNP	2350747	T	snp_intergenic	NA			NA	NA	X81	scfm_mucin	0.06
552	SNP	2565889	A	snp_intergenic	NA			NA	NA	X81	scfm_mucin	0.05
553	DEL	4085659	13	small_indel	NA			NA	lasR	X81	scfm_mucin	0.21
554	INS	4085750	C	small_indel	NA			NA	lasR	X81	scfm_mucin	0.11
555	SNP	5488376	A	snp_synonymous	GGA	78	3	GGG	PA14_61520	X81	scfm_mucin	0.05

Continued on next page

Table C.2 – continued from previous page

	Type	Position	Mut	Mutation category	Codon	Number	Pos	Ref	Gene name	Population	Treatment	Frequency
556	SNP	3064441	A	snp_nonsynonymous	TGC	82	1	AGC	PA14_34450	X82	min_mucin	0.14
557	SNP	4085470	C	snp_nonsynonymous	GTA	197	1	ATA	lasR	X82	min_mucin	0.85
558	SNP	4563049	A	snp_nonsynonymous	ATA	71	3	ATG	mvfR	X82	min_mucin	0.16
559	SNP	5300930	A	snp_intergenic	NA			NA	NA	X82	min_mucin	0.05
560	SNP	5432911	T	snp_nonsynonymous	TGC	1359	1	AGC	morA	X82	min_mucin	1.00
561	DEL	5914039	1	small_indel	NA			NA	PA14_66320	X82	min_mucin	1.00
562	SNP	4059675	C	snp_nonsynonymous	AGA	195	2	AAA	fleN	X83	min_mucin	1.00
563	SNP	4085373	G	snp_nonsynonymous	ACT	229	2	ATT	lasR	X83	min_mucin	0.22
564	DEL	4085750	1	small_indel	NA			NA	lasR	X83	min_mucin	0.07
565	DEL	4085751	1	small_indel	NA			NA	lasR	X83	min_mucin	0.07
566	SNP	1220720	T	snp_synonymous	GTA	181	3	GTC	PA14_14300	X84	min_mucin	0.06
567	SNP	1725430	T	snp_intergenic	NA			NA	NA	X84	min_mucin	0.05
568	SNP	1817364	A	snp_nonsynonymous	AAC	70	1	CAC	PA14_21030	X84	min_mucin	0.07
569	SNP	2297166	T	snp_nonsynonymous	GTG	72	2	GGG	PA14_26340	X84	min_mucin	0.06
570	SNP	2598429	G	snp_nonsynonymous	CCG	92	2	CGG	nuoB	X84	min_mucin	0.05
571	SNP	2783630	T	snp_intergenic	NA			NA	NA	X84	min_mucin	0.06
572	SNP	2858857	A	snp_synonymous	CTA	3430	3	CTG	PA14_32790	X84	min_mucin	0.06
573	SNP	3590877	C	snp_synonymous	ACG	1688	3	ACC	PA14_40260	X84	min_mucin	0.05
574	SNP	3590901	G	snp_synonymous	GAC	1680	3	GAT	PA14_40260	X84	min_mucin	0.05
575	SNP	4563588	A	snp_nonsynonymous	CAG	251	2	CTG	mvfR	X84	min_mucin	0.16
576	SNP	4627412	A	snp_nonsynonymous	TAC	365	1	CAC	relA	X84	min_mucin	0.21
577	SNP	4627699	A	snp_nonsynonymous	TTG	269	2	TCG	relA	X84	min_mucin	0.28
578	SNP	4627718	A	snp_nonsynonymous	TGG	263	1	CGG	relA	X84	min_mucin	0.21
579	SNP	5243305	A	snp_nonsense	TAG	342	2	TCG	ndh	X84	min_mucin	0.05
580	SNP	5431736	T	snp_nonsynonymous	CTC	967	2	CGC	morA	X84	min_mucin	1.00
581	SNP	6187751	C	snp_synonymous	GCG	188	3	GCA	algP	X84	min_mucin	0.07
582	SNP	772261	T	snp_nonsynonymous	TTT	289	3	TTG	rpoA	X85	scfm	0.05
583	SNP	1412130	A	snp_nonsynonymous	AGG	30	1	TGG	wspF	X85	scfm	0.28
584	SNP	2133454	T	snp_synonymous	GGT	1137	3	GGC	gdhB	X85	scfm	0.36
585	SNP	2820815	G	snp_nonsynonymous	CGG	281	2	CTG	PA14_32420	X85	scfm	0.31
586	SNP	4563632	T	snp_nonsense	TAG	266	1	AAG	mvfR	X85	scfm	0.07
587	SNP	5234891	T	snp_nonsynonymous	TAT	205	1	CAT	pilB	X85	scfm	0.35
588	SNP	336186	C	snp_intergenic	NA			NA	NA	X87	scfm	0.26
589	SNP	841359	T	snp_intergenic	NA			NA	NA	X87	scfm	0.05
590	DEL	1651096	12	small_indel	NA			NA	rhlR	X87	scfm	0.12
591	DEL	1651651	10	small_indel	NA			NA	rhlR	X87	scfm	0.12
592	SNP	2030064	C	snp_nonsynonymous	ACT	255	2	ATT	orfH	X87	scfm	0.57

Continued on next page

Table C.2 – continued from previous page

	Type	Position	Mut	Mutation category	Codon	Number	Pos	Ref	Gene name	Population	Treatment	Frequency
593	SNP	2819976	C	snp_nonsynonymous	ATC	1	3	ATG	PA14_32420	X87	scfm	0.22
594	SNP	2820750	G	snp_nonsynonymous	TTG	259	3	TTC	PA14_32420	X87	scfm	0.13
595	SNP	3090423	A	snp_nonsense	TAA	84	3	TAC	PA14_34800	X87	scfm	0.05
596	SNP	4085481	A	snp_nonsynonymous	ATC	193	2	ACC	lasR	X87	scfm	0.14
597	INS	4563546	CGC	small_indel	NA			NA	mvfR	X87	scfm	0.11
598	INS	5355068	A	small_indel	NA			NA	dcd2	X87	scfm	0.07
599	INS	5355068	T	small_indel	NA			NA	dcd2	X87	scfm	0.07
600	SNP	5432293	A	snp_nonsynonymous	AAG	1153	1	GAG	morA	X87	scfm	0.44
601	DEL	6097093	14	small_indel	NA			NA	arcC	X87	scfm	0.13
602	SNP	2647340	A	snp_nonsynonymous	TCG	19	1	GCG	PA14_30570	X88	min	0.06
603	SNP	3369029	T	snp_synonymous	CTT	198	3	CTG	PA14_37780	X88	min	0.05
604	SNP	5603838	T	snp_nonsynonymous	GAG	18	2	GCG	PA14_62780	X88	min	0.05
605	SNP	2030955	T	snp_nonsynonymous	TTT	103	3	TTG	orfE	X89	min	0.05
606	DEL	4085947	1	small_indel	NA			NA	lasR	X89	min	0.25
607	SNP	5648631	C	snp_nonsynonymous	CGC	16	1	GGC	phaJ3	X89	min	0.07
608	SNP	799171	A	snp_nonsynonymous	AAC	324	1	CAC	PA14_09320	X9	scfm_mucin	0.06
609	SNP	3648808	A	snp_synonymous	GCA	22	3	GCC	PA14_40910	X9	scfm_mucin	0.08
610	SNP	3755372	T	snp_synonymous	GGA	151	3	GGC	PA14_42090	X9	scfm_mucin	0.05
611	SNP	3857875	T	snp_nonsynonymous	ACG	167	1	CCG	PA14_43310	X9	scfm_mucin	0.06
612	SNP	3919581	A	snp_synonymous	TTG	217	1	CTG	sdhB	X9	scfm_mucin	0.05
613	DEL	4086037	29	small_indel	NA			NA	lasR	X9	scfm_mucin	0.72
614	SNP	4753398	T	snp_nonsynonymous	TAT	371	1	GAT	PA14_53590	X9	scfm_mucin	0.06
615	SNP	5158614	A	snp_synonymous	CCA	87	3	CCC	ptsN	X9	scfm_mucin	0.07
616	SNP	3264400	T	snp_nonsynonymous	TGC	73	1	GGC	PA14_36650	X91	scfm_mucin	0.06
617	SNP	4085456	T	snp_nonsense	TGA	201	3	TGC	lasR	X91	scfm_mucin	0.92
618	SNP	3709356	T	snp_nonsynonymous	TGC	152	1	CGC	PA14_41560	X92	scfm_mucin	0.06
619	SNP	3902587	T	snp_nonsynonymous	TAT	49	1	GAT	panB	X92	scfm_mucin	0.07
620	SNP	4135248	G	snp_nonsynonymous	CCG	345	2	CTG	pdxB	X92	scfm_mucin	0.06
621	SNP	4578885	T	snp_intergenic	NA			NA	NA	X92	scfm_mucin	0.05
622	SNP	5511470	T	snp_nonsynonymous	CTC	21	2	CGC	PA14_61780	X92	scfm_mucin	0.05
623	SNP	6187491	G	snp_nonsynonymous	GCG	275	2	GTG	algP	X92	scfm_mucin	0.06
624	DEL	1409027	130	large_deletion	NA			NA	wspC	X93	scfm_mucin	1.00
625	SNP	2392040	T	snp_nonsynonymous	ACC	85	1	GCC	PA14_27550	X93	scfm_mucin	0.06
626	SNP	2747983	A	snp_synonymous	TTT	144	3	TTC	PA14_31620	X93	scfm_mucin	0.06
627	SNP	3288866	T	snp_nonsynonymous	AAG	482	1	CAG	ligD	X93	scfm_mucin	0.06
628	DEL	4080853	33393	large_deletion	NA			NA	PA14_45910	X93	scfm_mucin	1.00
629	DEL	5948001	12	small_indel	NA			NA	pilQ	X93	scfm_mucin	0.10

Continued on next page

Table C.2 – continued from previous page

	Type	Position	Mut	Mutation category	Codon	Number	Pos	Ref	Gene name	Population	Treatment	Frequency
630	DEL	5951963	17	small_indel	NA			NA	pilM	X93	scfm_mucin	0.93
631	SNP	1960879	T	snp_nonsynonymous	ATC	55	1	CTC	PA14_22530	X94	min_mucin	0.07
632	SNP	2148729	C	snp_nonsynonymous	ATC	52	3	ATG	PA14_24590	X94	min_mucin	0.05
633	SNP	3444236	T	snp_intergenic	NA			NA	NA	X94	min_mucin	0.05
634	SNP	3513313	C	snp_nonsynonymous	TCC	74	2	TTC	PA14_39480	X94	min_mucin	0.06
635	SNP	3551730	T	snp_intergenic	NA			NA	NA	X94	min_mucin	0.06
636	SNP	4085518	T	snp_nonsynonymous	AAG	181	1	GAG	lasR	X94	min_mucin	0.15
637	DEL	4085719	11	small_indel	NA			NA	lasR	X94	min_mucin	0.15
638	SNP	4563320	A	snp_nonsynonymous	ATG	162	1	GTG	mvfR	X94	min_mucin	0.17
639	DEL	4627665	3	small_indel	NA			NA	relA	X94	min_mucin	0.06
640	SNP	5510730	T	snp_nonsynonymous	ATG	127	2	ACG	prs	X94	min_mucin	0.18
641	SNP	6280299	C	snp_nonsynonymous	CTC	384	1	TTC	spoT	X94	min_mucin	0.13
642	SNP	2125064	A	snp_synonymous	GGT	55	3	GGC	PA14_24390	X95	min_mucin	0.05
643	SNP	2209792	C	snp_nonsynonymous	CCG	368	2	CTG	PA14_25250	X95	min_mucin	0.06
644	SNP	2756718	A	snp_nonsynonymous	AAG	1053	2	AGG	PA14_31690	X95	min_mucin	0.07
645	SNP	2790069	T	snp_intergenic	NA			NA	NA	X95	min_mucin	0.07
646	SNP	3441712	C	snp_intergenic	NA			NA	NA	X95	min_mucin	0.05
647	SNP	3462136	T	snp_synonymous	CGA	22	3	CGC	pqqB	X95	min_mucin	0.05
648	SNP	3671564	A	snp_nonsynonymous	GAA	181	3	GAC	nppD	X95	min_mucin	0.05
649	SNP	4637225	A	snp_nonsynonymous	GAC	474	2	GGC	PA14_52260	X95	min_mucin	0.30
650	SNP	5157173	T	snp_nonsynonymous	CTC	238	2	CGC	rpoN	X95	min_mucin	0.05
651	SNP	5429053	C	snp_nonsynonymous	CCC	73	1	ACC	morA	X95	min_mucin	0.18
652	SNP	5429519	A	snp_nonsynonymous	CAG	228	2	CTG	morA	X95	min_mucin	0.54
653	SNP	5432474	G	snp_nonsynonymous	GGG	1213	2	GAG	morA	X95	min_mucin	1.00
654	SNP	5925945	A	snp_intergenic	NA			NA	NA	X95	min_mucin	0.05
655	SNP	6194783	C	snp_nonsynonymous	GGC	241	2	GTC	algR	X95	min_mucin	0.17
656	SNP	6275712	C	snp_nonsynonymous	CCC	225	1	ACC	crc	X95	min_mucin	0.18

Table C.3: Results from multivariate permutational analysis (PerMANOVA). Partitioning of genetic variance based on frequencies of variants in genes and Euclidean distances between populations. Df (degrees of freedom); SS (sum of squares); pseudo F (F-value by permutation); *p*-values based on 999 permutations, values in bold indicate statistical significance with $p<0.05$.

Source	df	SS	pseudo-F	R2	*p*-value
Media	1	1.32	2.048	0.023	**0.016**
Mucin	1	3.29	5.106	0.058	**0.001**
Media x mucin	1	2.16	3.349	0.038	**0.001**
Residual	77	49.65	-	0.880	-
Total	80	56.42	-	1.000	-

C.2 Supporting Figures

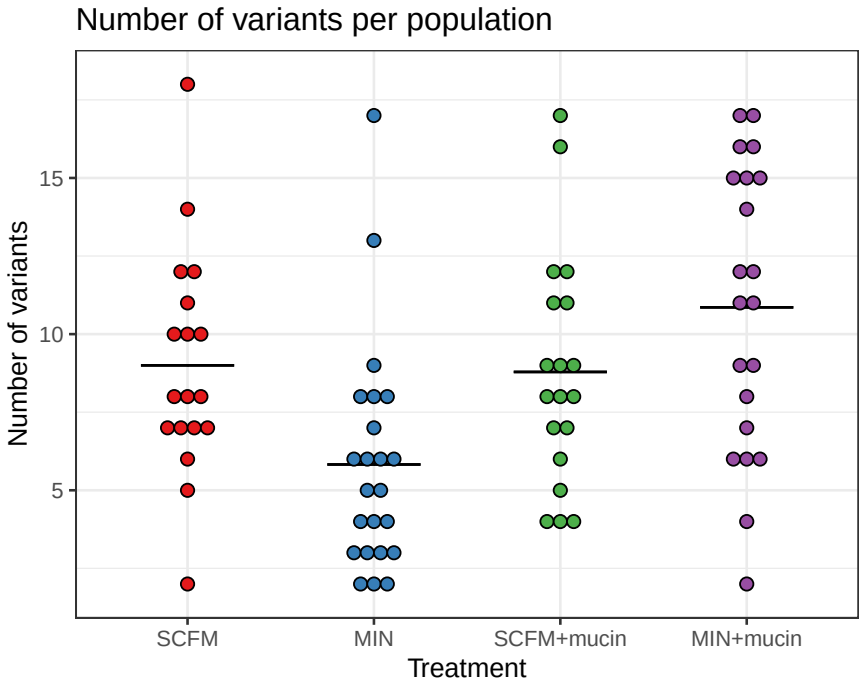

Figure C.1: Number of variants per population. Number of variants discovered in each population, including low frequency SNPs. Horizontal bars represent treatment means.

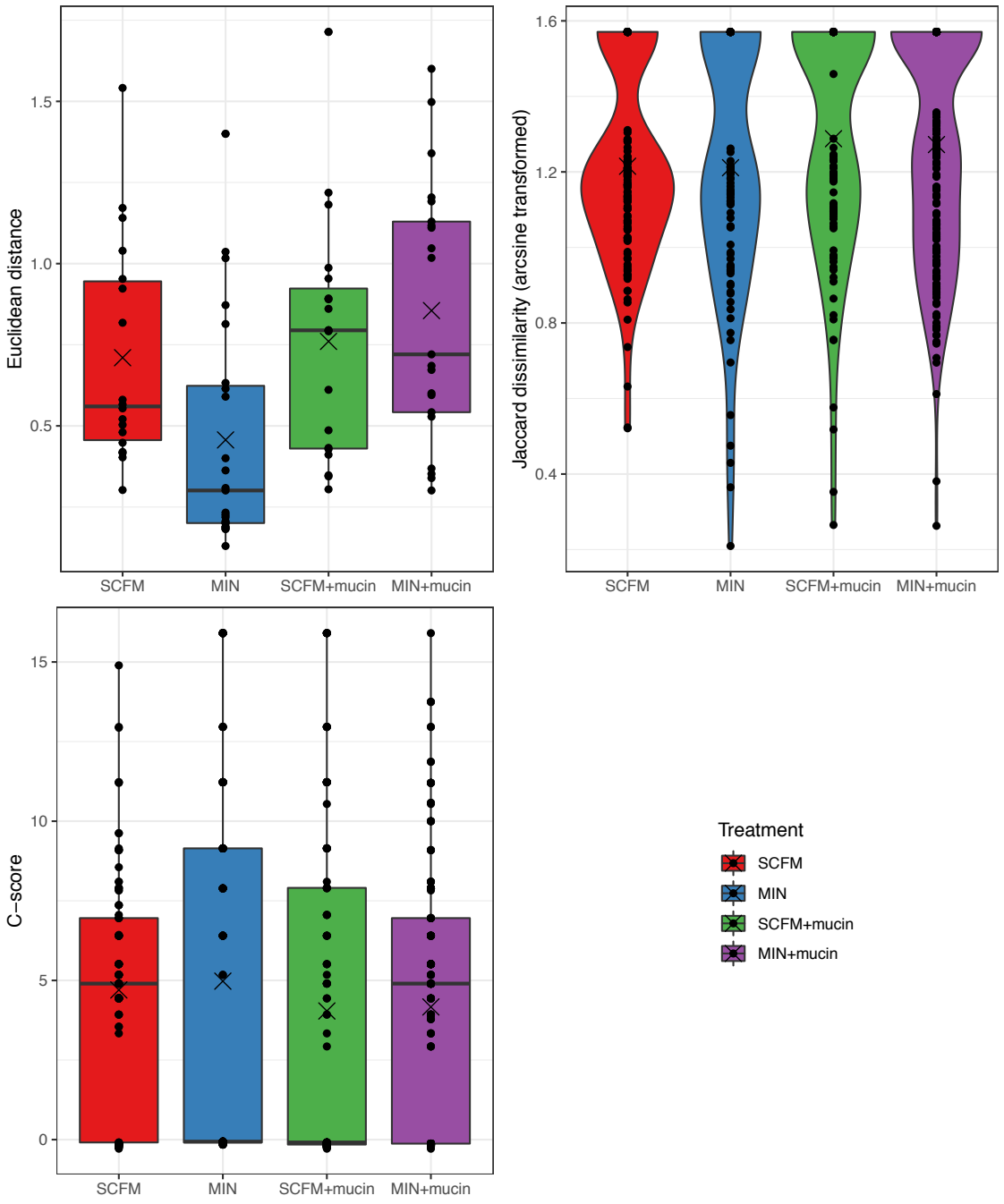

Figure C.2: Genetic divergence among populations. A. Euclidean distance is the distance each population is from the centroid of the treatment group, based on a principal coordinates analysis (PCoA). B. Jaccard dissimilarity, arcsine transformed for visualization, of all pairwise comparisons between populations within a treatment group. C. C-scores, magnitude of genetic constraint, for all pairwise comparisons between populations within a treatment group. In all panels, horizontal bars denote treatment median and crosses denote treatment mean.

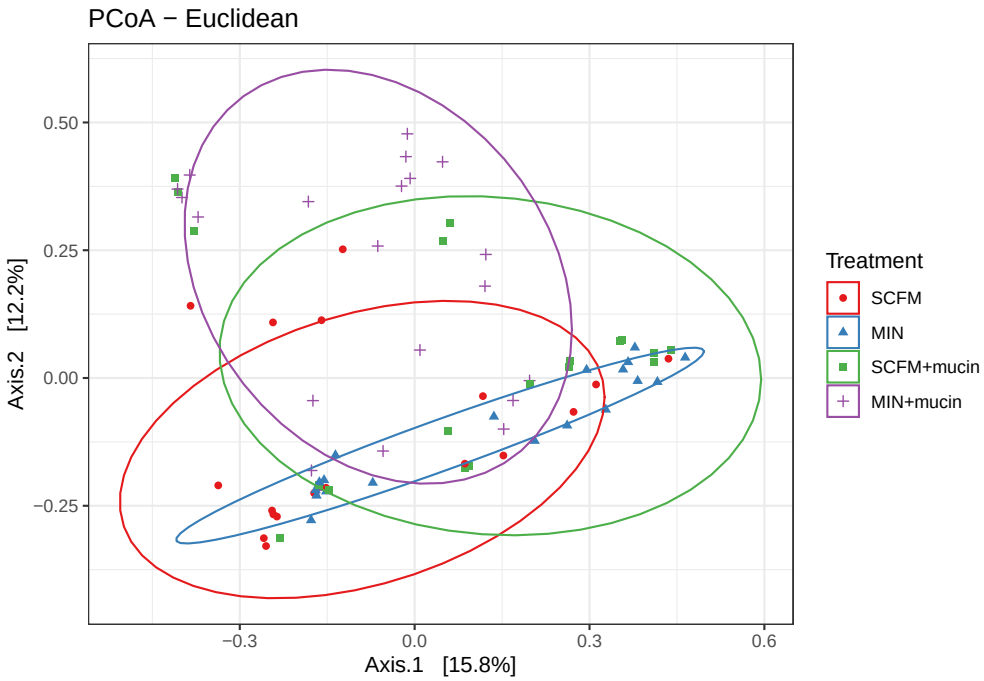

Figure C.3: Diversity among populations. The first two axes of a principal coordinates analysis (PCoA), based on a Euclidean distance matrix. Each point represents a population, and ellipses represent a 90 percent confidence interval around a multivariate t-distribution.

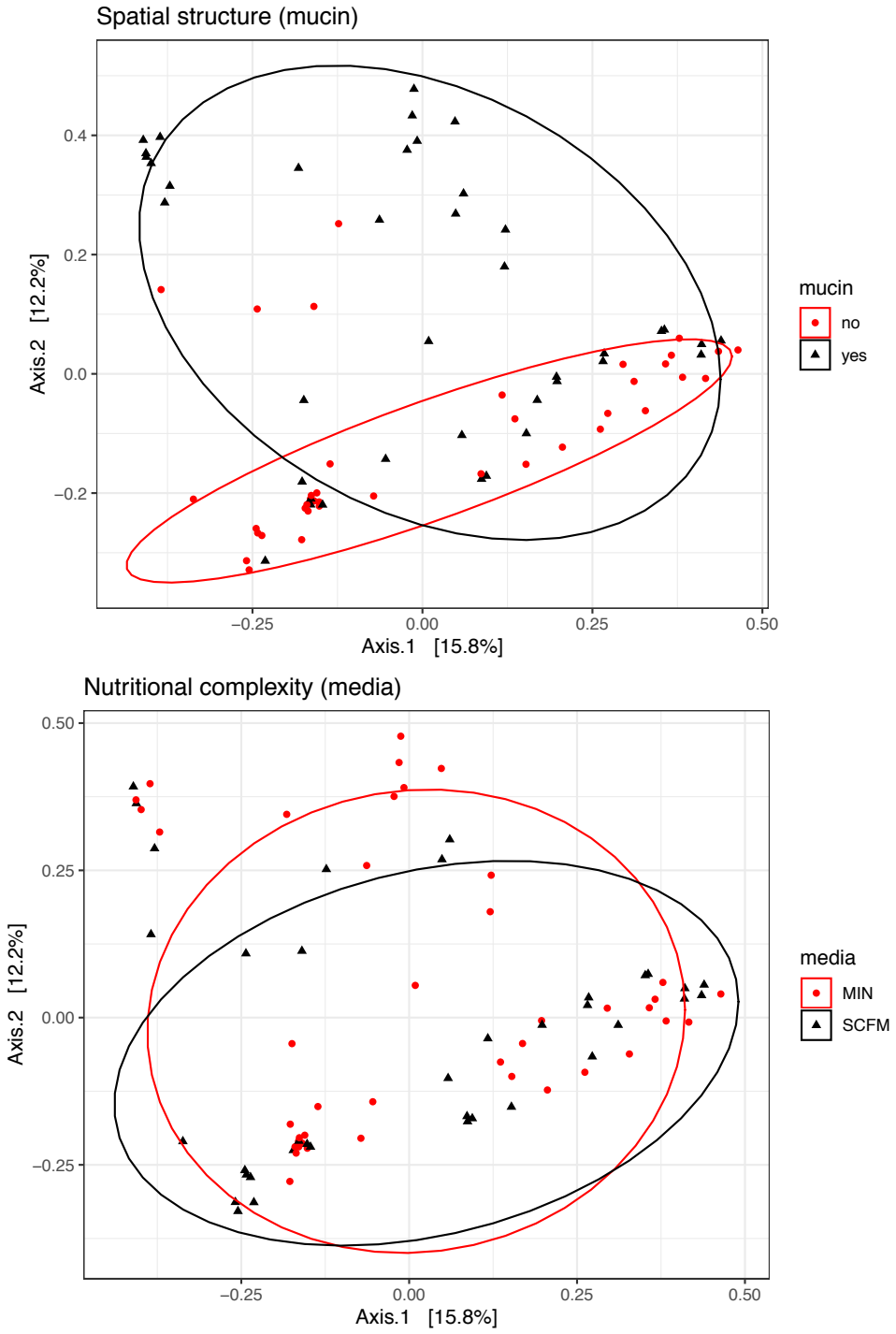

Figure C.4: Diversity among populations with effects of media and mucin separately. The first two axes of a principal coordinates analysis (PCoA), based on a Euclidean distance matrix. Populations are classified by media (SCFM or MIN; A), and mucin (presence or absence; B). Each point represents a population, and ellipses represent a 90 percent confidence interval around a multivariate t-distribution.

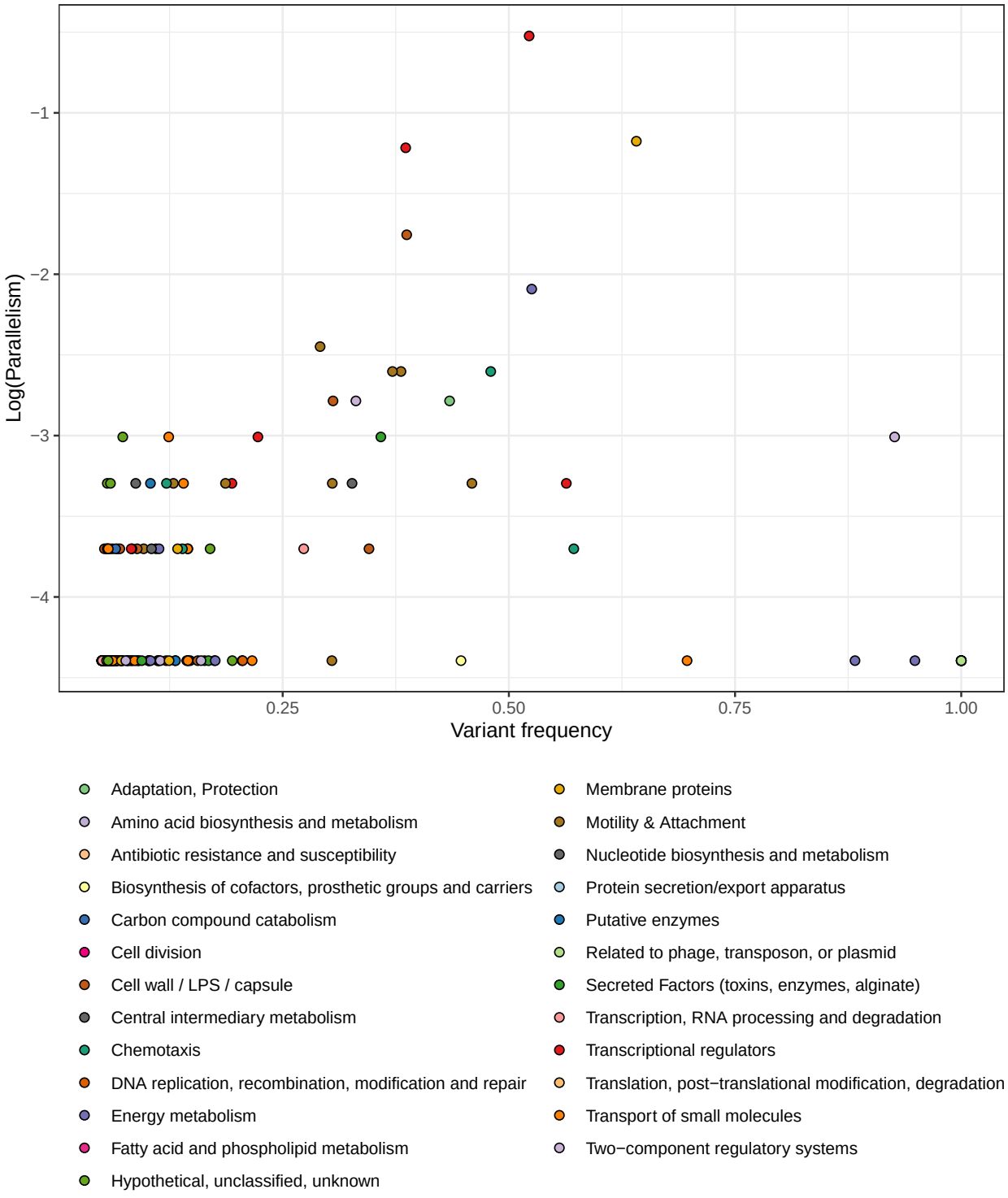

Figure C.5: Gene level parallelism versus frequency. Gene level parallelism for a gene is defined as the proportion of populations with mutations in that gene. Variant frequency is determined by taking the mean of the frequencies of variants in that gene.