REPEATED EVOLUTION IN DIVERSE STRAINS OF

PSEUDOMONAS AERUGINOSA

KEATON SINCLAIR

Thesis submitted to the University of Ottawa in partial fulfillment of the requirements for the
Master of Science in Biology

Department of Biology
Faculty of Science
University of Ottawa

© Keaton Sinclair, Ottawa, Canada, 2026

Table of Contents

Abstract iii

Résumé iv

Acknowledgments v

1.0 Introduction 1

2.0 Materials and Methods 8

- 2.1 Bacterial strains and evolution experiment 8
- 2.2 Isolate selection 8
- 2.3 Fitness competition 9
- 2.4 Genetic distance 10
- 2.5 Genome assembly and annotation 10
- 2.6 Jaccard similarity and statistical analysis 12

3.0 Results 14

- 3.1 Change in relative fitness in descendant isolates of each strain 14
- 3.2 Mutation identification and repeated evolution 15
- 3.3 Correlation between repeated evolution and change in relative fitness 17
- 3.4 Correlation between repeated evolution and genetic relatedness 17

4.0 Discussion 18

5.0 Conclusion 22

Tables and Figures 23

References 31

Abstract

Evolution can often repeat itself across different populations and species, but the reasons aren't fully understood. The likelihood of repeated evolution between populations is affected by both the similarity of their environments and their genetic backgrounds. Studying genetically different populations grown in the same environment could show the extent of genetic background on repeated evolution. We aimed to study this using genetically different strains of the Gram-negative bacteria *Pseudomonas aeruginosa,* an opportunistic pathogen known to grow in different environments around the world. Six independently evolved populations from each of 8 strains of *P. aeruginosa* were grown for 750 generations in M9 minimal salts media. For each strain, isolates taken from the initial ancestral populations (0th generation) were competed against their respective descendants (750th generation) in competitive fitness assays. We extracted DNA from each ancestor and descendant isolate, performed genome assembly, then compared detected mutations both within and between strains. Measures of genetic distance between strains were collected from *P. aeruginosa* phylogenetic data. We hypothesized that repeated evolution would be inversely correlated to the change in relative fitness and to the genetic distance between strains. We found an inverse correlation between repeated evolution and genetic distance, but no significant correlation between repeated evolution and relative fitness. This suggests that repeated evolution is more likely to occur in genetically similar populations.

Résumé

L'évolution peut souvent se répéter au sein de différentes populations et espèces, mais les raisons ne sont pas entièrement comprises. La probabilité d'une évolution répétée entre populations est influencée à la fois par la similarité de leurs environnements et par leurs contextes génétiques. L'étude de populations génétiquement différentes cultivées dans le même environnement pourrait montrer l'importance du contexte génétique sur l'évolution répetée. Nous avions pour objectif d'étudier ce phénomène en utilisant des souches génétiquement différentes de la bactérie Gram-négative *Pseudomonas aeruginosa*, un pathogène opportuniste connu pour se développer dans différents environnements à travers le monde. Six populations issues de chacune des huit souches de *P. aeruginosa*, ayant évolué indépendamment, ont été cultivées pendant 750 générations dans un milieu à sels minimaux M9. Pour chaque souche, des isolats issus des populations ancestrales initiales (0e génération) ont été mis en compétition avec leurs descendants respectifs (750e génération) lors d'essais de fitness compétitifs. Nous avons extrait l'ADN de chaque isolat ancêtre et descendant, procédé à l'assemblage du génome, puis comparé les mutations détectées au sein des souches et entre elles. Les mesures de la distance génétique entre les souches ont été collectées à partir des données phylogénétiques de *P. aeruginosa*. Nous avions comme hypothèses que l'évolution répétée serait inversement corrélée à la variation du fitness relative et à la distance génétique entre les souches. Nous avons constaté une corrélation inverse entre l'évolution répétée et la distance génétique, mais aucune corrélation significative entre l'évolution répétée et le fitness relative. Cela suggère que l'évolution répétée est plus susceptible de se produire dans des populations génétiquement similaires.

Acknowledgements

I'd like to begin by thanking my excellent supervisor Dr. Rees Kassen for his guidance and support throughout this project. I'm deeply grateful for his understanding and for the opportunity to work as part of the Kassen lab. Additionally, a big thank you to my thesis advisory committee members Dr. Allyson MacLean, Dr. Alex Wong, and Dr. Julien Martin for their help in all parts of the project, especially their insight proper statistical handling, evolutionary theory, and best practices for presenting my work.

I'd also like to thank my colleagues and friends in the Kassen lab; I can't imagine having done any of this without them, from start to finish. A massive thank you to Alex Hicks for allowing me to build upon the huge amount of work she did, and especially for her incredible support both inside and outside the lab. Thank you to Dr. Partha Chakraborty for being there to assist with many challenges in the last few years. To Angela Alonso, thank you for always reaching out to support and for the great conversations. Thank you to Aaron Hinz and Sonal Shewaramani for helping me set up the experiment, and troubleshoot the technical side of things. Thank you to Caio Rosa and Justin Gupta for never failing to show me cool and different perspectives, in academia and beyond.

Thank you all my friends both here and abroad for being a pillar of support I could turn to whenever I needed. Thank you to all the Yule Lads, to Luke and Patrick for always sharing in my nonsense and being there for me at any distance, and to Harrison, Tyler, David, for the many fun times and great laughs. Thank you Matt and Niki for your support and excellent hospitality, to Kehinde for always humoring me, and to Emil for showing me how it's done (skål!).

Thank you to Alexa and Emily for the nice chats and for being inspirations of strength. Thank you to my teammates for keeping me sharp and on the ball, I'm proud of what we've built. Thank you to Erica for being an uplifting presence when I needed it and for helping me cross the finish line.

To my family, I cannot thank you enough for your unconditional love and support from the beginning. Thank you to my wonderful parents, Craig and Michelle, for having my back every step of the way, always making sure I was well fed and equipped to follow my path. Thank you to Jennings and Cory for making me feel supported and accommodated no matter where I was, as well as to Brenan and Kara, for being such cool influences and teaching me how to be a little cooler myself. To Gram and Gramps, thank you for all your support and lovely conversations over the years, they never failed to center me. Thank you to my many loving aunts and uncles as well, for helping keep things light and in perspective. I love you all immensely.

1.0 Introduction

One of the most enduring questions in evolutionary biology pertains to the phenomenon of repeatability, the repeated evolution of the same genotype or phenotype in independently evolved populations. Instances of repeated evolution can give us insights into the environmental selection pressures present, the availability of beneficial adaptations to a population, or how independent populations may reach the same phenotype or genotype through some shared adaptive pathways (Wood, Burke, & Rieseberg, 2005).

Repeated evolution can occur along these pathways and is thus an important concept in our understanding of a species' evolutionary past and future; the repetition creates shared patterns which can be used to trace genomic history and, if shown to be consistent enough, might help to predict how species may change in the future (Bolnik et al, 2018). Examples of repeated evolution include the development of flight in both birds and bats, or the independent development of photosynthesis across different plant species (Christin et al. 2012).

A case where repeated evolution of a mutation or trait occurs can be categorized as either being 'parallel' if the founding genotype is the same for all populations, or 'convergent' if the founding genotypes are different (Kassen, 2024). While there is no widely accepted standard for classifying repeated evolution instances as convergent or parallel, comparisons between genetically distinct populations (such as species) are evidently considered convergence. For instance, the examples given above can be understood as cases of convergent evolution, since repeated evolution occurred across populations with different source genotypes.

Broadly speaking, repeated evolution is more likely to occur when the number of beneficial mutations available to selection is constrained or otherwise biased (Bailey et al. 2017; Lenormand et al., 2016). The factors that may introduce biases include: population size, epistatic and pleiotropic effects, mutational and environmental heterogeneity, initial distance from the optimum fitness, and the relatedness of populations. Previous work has found that population size has a positive correlation with the probability of repeated evolution (Frickel et al., 2018; Schenk et al. 2022), likely by providing a larger number of mutations from which a beneficial mutation, particularly large effect mutations, may arise and be selected for. Large effect mutations increase the probability of repeated evolution by ‘channeling’ adaptation into fewer pathways, which is further amplified by clonal interference as the rare large effect mutations outcompete the more numerous mutations with smaller effects.

Just as the environment can alter the fitness impact of a mutation, the genetic background itself can influence whether a mutation is beneficial, neutral, or deleterious (Weinreich, Watson, & Chao 2005). Both epistasis and pleiotropy are known to increase the likelihood of repeated evolution (Chevin, Martin, & Lenormand, 2010**;** Nosil, Crespi, & Sandoval, 2002). High epistatic and pleiotropic interactions both increase the likelihood that a mutation has deleterious effects, effectively constraining the supply of beneficial mutations. Different mutational rates along a genome (mutational heterogeneity) also play a role in affecting repeated evolution (Bailey et al. 2017), as do differences in environmental selection pressures (Oke et al. 2017).

Since mutations provide the genetic diversity upon which selection may act, this mutational heterogeneity increases the probability of parallel or convergent evolution between populations possessing those genes (Conte et al. 2012; Schluter et al. 2004). How far the populations are from the "peak" fitness of their environments affects their path of adaptation and their likelihood of having the paths "overlap" (i.e. repeated evolution) (Fisher, 1999). For populations at different distances from the same peak, the fitter population closer to the peak will have access to fewer adaptive pathways available than the populations further from the peak. This can be helpfully displayed using Fisher's geometric model, which is used to visualize adaptation to a given environment by relating a subset of all possible genotypes of an organism to its corresponding fitness values in a particular environment (Orr, 2005).

In the geometric model, each genotype is assumed to have a corresponding fitness value which, when plotted together, creates a hypothetical fitness landscape for that species in that environment. The topography of a fitness landscape is determined by the extent of the interactions between mutations that affect fitness; higher levels of epistasis lead to more rugged landscapes with multiple fitness peaks, while lower levels lead to fewer peaks. Within this framework, the probability of parallel evolution is inversely correlated with the distance to a fitness optimum, since high-fitness populations have access to fewer higher-fitness adaptations than low-fitness populations (Van Cleve & Weissman, 2015). Studies have shown that repeated evolution is influenced by genetics (Witt & Huerta-Sánchez, 2019), environment (Bailey, Rodrigue, & Kassen, 2015), and the complex interaction between them (Stern, 2013; Turner, Marshall, & Cooper, 2018).

Studies on the genetic influence found that, across the high-altitude environments of the world, populations of both humans and domesticated animals exhibited some convergent evolution in their adaptation to the hypoxic conditions, even identifying the gene *EPAS1* as being commonly involved in the adaptive pathway (Witt & Huerta-Sánchez, 2019). By contrast, an experiment by Bailey, Rodrigue, & Kassen (2015) showed that repeated evolution in replicate populations of *Pseudomonas aeruginosa* is significantly influenced by and is variable across different growth environments. They found that parallelism was more likely between replicate populations that were evolved in the same environment, and a marginally significant negative correlation between number of beneficial mutations and the frequency of parallelism.

In another case, Thöming et al. (2020) kept environment constant across 414 evolved clinical isolates of *P. aeruginosa* in lysogeny broth, monitoring the development of biofilm formation, virulence, and motility phenotypes. They found evidence for repeated evolution as three main biofilm phenotypes, which shared virulence and motility characteristics, had evolved across multiple isolates. Phylogenetic relatedness is not often incorporated into studies of repeated evolution, so there is an opportunity to use strains of *P. aeruginosa* from more dissimilar environmental niches and examine how phylogenetic relatedness influences repeatability compared to other factors like distance from the optimum (fitness).

Despite repeated evolution being a well-studied topic, the extent to which genetic relatedness between the populations influences repeated evolution is not fully understood. It is broadly assumed that closely related populations will experience higher rates of repeated evolution between them since they are more likely to have similar available adaptive pathways, though this hypothesis has not been tested directly.

An experiment by Bollback & Huelsenbeck (2009) using three species of bacteriophage was able to address some of this knowledge gap: the frequency of parallel evolution was positively correlated with genetic relatedness, no convergent evolution was observed between diverged species, and there was no apparent fixed order to the appearance or fixation of adaptations. A similar experimental design using a higher number of genetic backgrounds may be able to catch convergent evolution events as well as parallel evolution, allowing for a broader understanding of how genetic relatedness between populations influences the likelihood of repeated evolution, in addition to the extent of its importance relative to the other factors. It could also incorporate measures of distance to the optimum fitness, another influence on repeated evolution which was not included in the studies reviewed here.

We sought to distinguish between the effects of distance to the optimum and genetic background on the likelihood of repeated evolution in a constant environment. To do this, we measured the relationship between repeated evolution and relative fitness of the evolved descendants, and between genetic relatedness of six strains from three different environmental niches.

For this study we used *Pseudomonas aeruginosa,* a species of Gram-negative bacteria which can opportunistically infect wounds and humans with compromised immune systems. The species is widespread globally and thrives in moist environments from soil to living tissue, leading to strains adapting to varying conditions. These different strains provide some diverse genetic backgrounds which can be used to examine how that diversity affects the likelihood of repeated evolution. We grew six independently evolved populations from each of eight different strains of *Pseudomonas aeruginosa* for 750 generations.

As part of a broader experiment by Alexandra Hicks (2023) which studied the effect of genetic background on phenotypic diversification, three replicate populations were evolved lysogeny broth (LB) and three in minimal salts (M9) media. The populations grown in M9 media were the primary focus of this study, as it is less nutritionally complex than LB and therefore more likely to have a single adaptive peak. We randomly selected 40 isolates from the populations grown in minimal media: four isolates from the evolved (750th generation) populations of all eight strains, and one isolate from the ancestral population of each strain.

We sequenced the isolates and recorded the genetic differences between the descendants and their respective ancestors, then compared the presence and absence of SNP mutations between descendants to obtain a measure of repeated evolution for each comparison. We performed a relative fitness assay via flow cytometry to obtain fitness values for each descendant. Finally, we used the genetic distance values between strains shown in a phylogenetic tree created by Alex Hicks, based on a core genome alignment from Dettman & Kassen (2021).

Within the context of a fitness landscape, the distribution of mutation effect sizes depends on the shape of the fitness landscape and the fitness of the initial genotype (it's "starting position"). At a given position on a smooth, single-peak landscape, the fitness difference between positions and number of neighboring positions with higher fitness is inversely proportional with the distance to the optimum. Near the peak there will be fewer mutations that increase fitness, and the magnitude of the increase will be smaller, which may "channel" into the same mutational pathways and result in higher likelihood of repeated evolution.

We therefore hypothesized that frequency of repeated evolution will be inversely correlated to change in fitness. The isolates with the smaller changes in relative fitness will have higher similarity between them than between isolates with high change in relative fitness.

Genetic distance is a measure of the relatedness between two samples, where zero represents complete relatedness (genetically identical), and higher distances represent increased genetic divergence (Nei 1972). Samples which are highly related to each other will have similar genetic backgrounds and thus access to similar mutational pathways when under selection, while highly genetically divergent samples will tend to have fewer shared mutational pathways. It follows that when evolved in the same minimal media environment, isolates with lower genetic distance between them will have an increased likelihood of developing the same adaptations. In other words, the frequency of repeated evolution would be inversely correlated to genetic distance.

Repeated evolution was assessed for each possible pairing of samples by comparing their number of shared mutations against the number of total mutations, which results in a Jaccard similarity value (J) for every pairwise comparison. We expected to see higher Jaccard similarity values between samples from the same strain and from two highly related strains, and lower Jaccard values between two distantly related strains.

In summary, we expected that repeated evolution would be highest amongst strains with the smallest increase in fitness, and that closely related strains would show a higher frequency of repeated evolution between them than distantly related strains.

2.0 Material and methods

2.1 Bacterial strains and evolution experiment

Data was generated via eight strains of *Pseudomonas aeruginosa*: two lab reference strains (PAO1 and PA14), three environmental strains (Jp54, Jp1140, and Pae111), and three clinical strains (JD313, JD322, and JD328) taken from the airways of cystic fibrosis patients suffering chronic infections. The samples used in this experiment were created from populations made by Alex Hicks (Hicks, 2023), which were evolved for 750 bacterial generations over a 110-day period (about 6.64 generations per day) in both lysogeny broth (LB) and M9 minimal salts (referred to as MIN) liquid media; the isolates in this experiment were taken exclusively from those populations grown in MIN media.

2.2 Isolate selection

Every isolate was created by streaking a small volume of inoculated liquid media onto an LB agarose plate and incubating at 37°C for 24 hours. Resulting colonies that were separate from all others on the plate were chosen as isolates, then inoculated at 1:100 dilution into 1.5 mL of liquid LB media and grown at 37°C overnight, without shaking. This process was repeated to create one isolate from the ancestor population (0th generation) and one isolate from each of four independently evolved populations (750th generation), resulting in eight ancestral and 32 descendant isolates. Eight additional isolates were made from yellow fluorescent protein (YFP)-marked ancestor populations for use in the flow cytometry fitness competition. All isolate samples were stored at -80°C in 20% glycerol mixtures using 1.5 mL Eppendorf tubes.

2.3 *Fitness competition*

Flow cytometry was used to perform competitive fitness assays on descendant isolates versus their corresponding YFP-marked ancestor. Frozen isolate cultures of YFP-marked ancestor and the descendant isolates were inoculated in 1.5mL liquid MIN media, then incubated at 37°C for overnight without shaking (static conditions). The cultures of descendants and their YFP-marked ancestors from clinical strains were combined to create the initial (t0) mixtures; a ratio of 30:70 (higher proportion of ancestor) was used for clinical strains, while a 50:50 ratio was used for the mixtures of all other strains. This higher proportion was used for the clinical ancestors because they would completely perish at lower ratios; this different starting ratio is compensated for in the selection calculation shown below. The t24 mixes were created by co-culturing the t0 mixtures and incubating overnight at 37°C in static conditions. Both t0 and t24 mixtures were stored frozen in 20% glycerol mixtures.

The Gallios Flow Cytometer (Beckman Coulter) and Kaluza for Gallios acquisition software (Beckman Coulter) were used to perform flow cytometry on the t0 and t24 mixtures, with filter-sterilized minimal (M9) salts acting as the buffer solution. Resulting flow data was processed using the Kaluza Analysis Software (Beckman Coulter) to acquire the cell counts of the mixtures. The selection coefficient (s) was calculated by comparing the change in cell counts between each descendant isolate and their ancestor after 24 hours, divided by the number of generations expected in that time period (~6.64), shown below:

$$s = \frac{\left(\ln\left(\frac{evolved\ final}{evolved\ initial}\right) - \ln\left(\frac{ancestor\ final}{ancestor\ initial}\right)\right)}{number\ of\ generations}$$

Each fitness measurement was taken twice per isolate, and the entire process was performed a total of three times in three separate days.

2.4 Genetic distance

Values for the genetic distance between each strain were taken from the phylogenetic tree constructed by Alex Hicks (Hicks, 2023), which included all strains used in this experiment against a PA7 strain outgroup.

2.5 Genome assembly and annotation

For each descendant and ancestor isolate, genomic DNA was extracted from overnight cultures using the DNeasy Blood & Tissue Kit (Qiagen). DNA samples were sent to Genome Quebec for 150-bp paired-end Illumina sequencing using the Illumina Novoseq6000. The average coverage depth across all isolates was approximately 450x, with a quality score of Q36.

Of the eight strains used, only PAO1 and PA14 have complete genome sequences available to use as assembly references. To maintain a consistent approach, all ancestor isolates were first assembled against the PAO1 reference genome (Stover et al. 2000) from the Pseudomonas Genome Database (Windsor et al. 2016) using the SPAdes Genome Assembler version 3.15.4 (Prjibelski et al. 2020). After filtering out all contigs shorter than 300-bp, the resultant assemblies were ordered using the MeDuSa Multi-Draft Based Scaffolder version 1.6 (Bosi et al. 2015) against a PAO1 reference. The resulting scaffold was then used as the reference for a second round of MeDuSa contig ordering. The ancestor isolate sequences were annotated using Prokka version 1.14.5 (Seemann 2014), then used as annotation references for their respective descendants in the Breseq pipeline version 0.38.3 (Deatherage & Barrick, 2014).

Two of the clinical strains, JD322 & JD328, were particularly difficult to assemble in this stage, as the descendant isolates aligned poorly against their respective ancestors. After a lengthy period of troubleshooting, we found that all descendant isolates of JD322 and JD328 aligned closely with the ancestor isolate of Jp1140. This indicated that JD322 and JD328 descendant isolates were contaminated and eliminated by Jp1140 descendant isolates at an unknown point earlier in the experiment, so we decided to exclude data from the contaminated strains from further analysis. Additionally, three descendant isolates, one each from PA14, Jp54, and Jp1140, were removed from analysis due to probable contamination of DNA sample, resulting in 21 descendant isolates being used for subsequent analyses.

Single nucleotide polymorphisms (SNPs), insertions, deletions, and larger-scale mutations were validated by checking the read mapping at the called location. Many called mutations were in highly repetitive regions with low coverage scores (<20x) and thus were filtered out. Due to the lack of a well-resolved reference database for all strains except PAO1 and PA14, mutations were manually verified via BLASTN searches on the Pseudomonas Genome Database. Pseudomonas Ortholog Groups (POGs) were used to assist gene identification based on results which returned >99% sequence similarity for a 50-bp sequence taken from the area affected by a mutation. We also noted the predicted functions affected in the mutated genes using Clusters of Orthologous Groups (COGs) classifications (Tatusov et al. 2000) from the Pseudomonas Genome Database. Particular attention was given to genes affecting biofilm production, motility, and virulence, as those influence the motility of *P. aeruginosa* and higher motility was expected to be favored in MIN media.

In total, 115 mutations were found, of which 61 were SNPs, 38 were small indels, and 15 were large-scale deletions, insertions, or mutation regions. Our analyses focused primarily on SNPs which occurred in or around the coding section of genes for ease of identification. Using Roary version 3.11.2 (Page et al., 2015) we classified genes as belonging to the core or pan genome of the strains, based on which genes were or were not present in all ancestral isolates respectively. This allowed us to determine the amount and type of mutations that occurred in core and pan genomic genes. A total of 9397 genes were identified, of which 4948 were found in all six ancestral isolates.

2.6 Jaccard similarity and statistical analysis

A list of 24 genes affected by the 61 SNP mutations were used to assess the level of repeated evolution between each pair of isolates. A gene was considered affected if at least one SNP occurred somewhere in the coding region, irrespective of exact location or effect. The number of affected and unaffected genes for each pair of isolates was used to calculate a Jaccard similarity coefficient (*J*) as described by the formula:

$$J(A,B) \quad = \frac{(A \cap B)}{(A \; U \; B)}$$

The symbols *A* and *B* each represent a list of genes affected by mutation in descendant isolates *A* and *B,* respectively. Each list comprises the tally of mutations found in any of 24 genes (the set of all affected genes across all isolates) for that isolate.

By dividing the number of mutations shared between the isolates by the total number of mutations across all isolates, we obtain a value of *J* between 0 (no similarity) and 1 (complete similarity). The pairwise comparisons were assessed among six strains, resulting in 15 comparison groups: A permutation and resampling technique was used to accommodate for the non-independence of calculated *J* values using these comparison groups as a basis.

From the list of 24 genes affected by SNPs, the mutations found in each isolate were randomly resampled a number of times equal to how many mutations that isolate originally had. The *J* values were then recalculated for each pairwise of isolates to give a new mean score for each comparison group. We repeated this 50,000 times to generate a distribution of the *J* scores for the comparison groups, from which we obtained our P value. Because there are 15 comparison groups, Bonferroni correction of the standard P value of 0.05 resulted in the significance threshold being set to 0.0033. Jaccard similarity was calculated in this way for every pair of descendant isolates using the vegan package (Oksanen et al., 2022) in R and RStudio (R core team, 2023).

3.0 Results

Twenty-four populations from six different ancestral strains of *P. aeruginosa* were evolved independently in minimal media for 750 generations, four populations per strain. The six strains can be broadly categorized into three groups based on their source environment: PA14 and PAO1 are lab reference strains (as they are commonly used in lab experimentation); Jp54, Jp1140, and Pae111 are strains taken from environmental sources, and JD313 is a clinical strain associated with chronic infections in human patients with compromised immune systems. An isolate was taken from each population of the 750th generation and used in a competitive fitness assay against an isolate of its ancestor. All twenty-four descendant and six ancestral isolates were genome sequenced and assembled; the ancestors were assembled using the *P. aeruginosa* PAO1 genome as a reference (NC_002516; 6264404 bp; 5,697 genes), then the descendants were assembled against their respective ancestors.

3.1 Change in relative fitness in descendant isolates of each strain

To estimate the distance of each isolate from the fitness optimum in MIN media, we performed a competitive fitness assay of the ancestors versus their descendants. After 750 generations of growth in MIN media, the average relative fitness of all strains increased by between 15% to 35% (Figure 1). Jp54 showed the average change at 35%, followed by Jp1140 at 20%, PA14 at 19%, PAO1 at 16%, Pae111 at 14%, and JD313 at 11%. The niche of origin for each strain did not seem to affect the change in relative fitness, as both the highest and lowest change occurred in Jp54 and Pae111 respectively, both considered ‘Environmental’ strains.

3.2 Mutation identification and repeated evolution

A total of 115 mutations were detected across all 21 independent isolate genomes, the number of mutations per isolate ranging from 3 to 10. Of the 115 detected mutations, there were 61 single-nucleotide polymorphisms (SNPs), 38 small indels, 9 large deletions, 4 large substitutions, 1 large insertion, and 1 small, mutated region which could not be fully resolved and categorized. Of the 61 SNPs, the majority of 28 were non-synonymous, followed by 16 synonymous, 13 intergenic, and 4 non-sense.

We identified 47 different genes affected by mutation, including 24 SNPs, using the Pseudomonas aeruginosa Database (Winsor et al. 2016) (Figure 2) and PseudoCAP for functional annotation. While genes such as *aer* (chemotaxis), *lasR* (transcriptional regulation), and *pscP* (protein transport/secretion) were mutated across multiple strains (PA14, Jp1140, and JD313, respectively), no mutation in a gene occurred across all strains. The most frequently affected functions were chemotaxis, adaptation, and transcriptional regulation (Table 1).

We constructed a core genome using the genes found in common amongst all six ancestral isolates (Table 2). We used Roary version 3.11.2 (Page et al., 2015) to identify a total of 9397 genes present across the six ancestor isolates, of which 4948 genes were present in all ancestor isolates and were classified as part of the core genome, while all other genes were designated as part of the pan genome.

Approximately 83% of all detected mutations occurred in the core genome (95 of 115), and 85% of all mutations occurred in coding genes (98 of 115). Single nucleotide polymorphisms (SNPs) and small indels (those affecting less than 50bp) accounted for 85% of detected mutations, the rest being insertions, deletions, or substitutions that affected 50 bp or more. In combination with our fitness data showing that all strains had an average minimum relative fitness increase of 15%, we believe that many of the mutations we found had a beneficial effect on the descendants.

To assess the level of repeated evolution, we calculated the Jaccard similarity index (*J*) between every pair of descendant isolates, making 210 comparisons in total. Our analysis focused on the most abundant mutation (SNPs) which occurred in the regions shared by all strains (core genome). We categorized those repeated evolution events into either “parallel evolution” or “convergent evolution”, depending on the strain-strain relationship of the isolates being compared. Repeated evolution between isolates from the same strain was considered parallel evolution, whereas it constituted convergent evolution when the isolates belonged to different strains (Figure 3).

3.3 Correlation between repeated evolution and change in relative fitness

We wanted to examine the relationship between the mean Jaccard similarity index and the change in relative fitness for all 21 isolates. We expected that the two metrics would be inversely correlated, but we found no significant correlation between relative fitness and repeated evolution ($P= 0.256$) (Figure 4).

3.4 Correlation between repeated evolution and genetic relatedness

We examined how phylogenetic relatedness affects the frequency of repeated evolution by comparing the Jaccard similarity of each pair of isolates to their relative genetic distance (Figure 4). The genetic distance values were taken from a phylogenetic tree created by Alex Hicks (Hicks, 2023) based on a core genome alignment from Dettman & Kassen (2021). There was an inverse correlation between repeated evolution and genetic distance (Figure 5, $p < 0.001$), suggesting that repeated evolution tends to be less frequent with decreasing relatedness. This pattern remains when comparing the mean Jaccard similarity against categorical variables created from genetic distance measures, where genetic distances of zero were grouped into 'parallel', and those greater than zero were grouped into 'convergent'.

4.0 Discussion

Tracking the frequency of repeated evolution can inform us about selective pressures that species are under and the mutational pathways that are available to them. This concept can also be applied in the other direction, whereby knowing the environmental and genotypic conditions can give us an idea of which genes are likely to mutate. Experiments that evolved replicate populations in different environments have shown the significant and varied influence environmental conditions can have on the extent of parallel evolution (Bailey, Rodrigue, & Kassen, 2015), and even heterogeneity within a single environment can alter the frequency of parallel evolution (Stuart et al. 2017).

Here we studied the complementary aspect by assessing repeated evolution in genetically distinct populations that had independently evolved in the same environmental conditions. We used whole-genome sequencing and de-novo assembly to calculate instances of repeated evolution across isolates from six *P. aeruginosa* strains and assessed the proportion of mutations which occurred in core- or pan-genomic genes. We incorporated endpoint relative fitness data and phylogenetic distance to examine how change in fitness and genetic relatedness influence the frequency of repeated evolution, respectively.

Overall, we found no significant correlation between repeated evolution and relative fitness, and an inverse correlation between repeated evolution and genetic distance (Figure 4 & Figure 5). This result suggests that genetic distance can in part characterize the probability of repeated evolution, while distance to the optimum does not.

The majority of mutations occurred in coding genes and within the core genome, with SNPs being the most common type (Table 2 & Figure 1). Additionally, we found evidence of repeated evolution occurring both within and between strains (Figure 3), though repeated evolution was both more common albeit variable in parallel comparisons than in convergent comparisons (Figure 6).

The only between-strain comparison with a significant Jaccard index was Pae111 and PA14, while there were three significant within-strain Jaccard indexes for JD313, Pa14, and Jp54. In particular, the within-strain Jaccard index for Jp54 was 0.600, a remarkably high level of repeated evolution. This, in addition to the high level of fitness change seen in Jp54 (Figure 1), suggests that the ancestor began further from the optimum fitness of the growth media environment. While shared ancestry of all strains likely creates a bias in the frequency of repeated evolution we observed, most comparisons did not have a significant Jaccard index (only 4 of 21 comparison groups), suggesting that the strains are sufficiently disparate to accommodate this bias.

Due to the lower complexity of the minimal media environment, we expected that the fitness landscape would be similarly less complex and rugged, and thus have a single peak for genotypes adapted to the selection pressure of nutritional scarcity (i.e. higher motility and/or lower biofilm production). In such a system, mutations are more likely to be beneficial in genotypes near the bottom of the fitness landscape than in those near the peak, since genotypes at the peak have access to fewer higher fitness genotypes, decreasing the likelihood that mutations are beneficial (de Visser & Krug 2014). Therefore, relative fitness increase is expected to be smaller in descendant isolates whose ancestor began close to peak fitness. Since fewer beneficial

mutations are available towards the peak of the landscape, genotypes are likely to converge towards similar adaptations, resulting in a higher frequency of repeated evolution towards the peak.

Additionally, genotypes that are distantly related to each other likely fall into different areas of the landscape, resulting in fewer instances of repeated evolution as they evolve since they have access to entirely different neighboring genotypes. From this we develop two hypotheses: that repeated evolution will be inversely correlated to the change in relative fitness, and that repeated evolution will be inversely correlated to genetic distance.

All descendant isolates gained fitness compared to their ancestor after 750 generations of independent evolution as expected, although the amount varied between and within strain (Figure 2). Some variation between strains is expected due to differences in their respective niches; the lab reference strains PA14 and PAO1 are commonly used in laboratory conditions, the environmental strains Jp54, Jp1140, and Pae111 are sourced from seawater and soil, and clinical strain JD313 is found in the lungs of cystic fibrosis patients (Oliver & Mena 2010). However, these expectations are consistently reflected in the data (Figure 1), as average relative fitness for Jp54 reaches 35%, while all other strains showed a fitness increase between 14% and 20%.

Within our framework, this would suggest that the Jp54 ancestral population began the experiment the least well-adapted to minimal media, while those of PAO1 and Pae111 began the most well-adapted. Isolates within PAO1 and Jp1140 had higher variation in their average relative fitness than isolates of all other strains, with one isolate from a PAO1 population showing an increase of only ~2% after 750 generations of growth.

This within-strain variation may reflect the presence of strong epistatic influence preventing the further adaptation of certain genotypes.

Finding no significant correlation between Jaccard similarity and selection coefficient may signify that the actual fitness landscape of the minimal media is more rugged than initially assumed. Ruggedness is determined by the extent of epistatic interactions between genes and mutations, which results in more fitness peaks and a higher range of possible fitness values (Østman, 2013). It's also possible that all isolates started far from peak fitness, which would make measures of distance from the optimum less predictive of repeated evolution.

The substitution rate can differ between genes across bacterial genomes (Matic 2019), which may also contribute to some of the variance in repeated evolution frequency. In our study, the average number of fixed mutations per isolate ranged from 2.75 (in PAO1) to 6.00 (in PA14). The genome size of the former strain is 6.3 Mbp (Stover et al., 2000) while genome size of the latter is 6.5 Mbp (Ozer, Allen, & Hauser, 2014), and isolates from both strains were taken from populations grown for 750 generations. This translates to a substitution rate of 5.82×10^{-10} for PAO1, and 1.23×10^{-9} for PA14, with the rates for all other strains lying within that range. We found no significant correlation between parallelism and rate of substitution in our strains ($p > 0.5$), though the rates of substitution among strains may not fully reflect the actual mutation rates since the isolates were grown under selection.

Additionally, although the substitution rate for PA14 is nearly 50x greater than that of PAO1, all rates are near the range previously found for wild-type strains (Dettman, Sztepanacz, & Kassen, 2016), and an order of magnitude lower than what would be expected in a highly mutable mutator strain. The source populations of the bacterial isolates were large (>10,000,000

colony forming units/mL, see Hicks (2023)), so genetic drift is not expected to have had a significant influence on the observed mutations.

However, genetic hitchhiking, where second-site mutations rise to high frequency in the same genetic background as a strongly beneficial mutation, may have an influence on the observed mutation frequency and variation. (Smith & Haigh, 1974). Genetic pleiotropy, where a single mutation may influence two or more seemingly unrelated traits, may increase the frequency of repeated evolution by effectively reducing the number of available beneficial mutations, and by increasing the advantageous effect size of any one mutation.

Originally the experiment included two additional clinical strains (JD322 and JD328) with four isolates each, but genome assembly revealed that all isolates of both strains had been contaminated by Jp1140 at an unknown point earlier in the experiment, so the data were removed from further analysis. Additionally, 3 isolates (one each from PA14, Jp54, and Jp1140) were found during genome assembly to have been contaminated by an unknown microorganism and were similarly removed from analysis. This resulted in fewer fitness, mutational, and genetic distance data points included in our analysis, which reduced our ability to draw stronger conclusions regarding our hypotheses.

5.0 Conclusion

The phenomenon of repeated evolution is a product of complex interactions between genetic and environmental factors which can be difficult to untangle and examine in isolation. The current understanding primarily focuses on the significant influence exerted by environmental pressures, population size, and pleiotropic effects. However, there is relatively much less knowledge about the other factors at play, such as genetic relatedness and epistasis. By keeping factors like population size and environmental heterogeneity relatively consistent across disparate strains of *Pseudomonas aeruginosa,* we were able to examine how genetic relatedness and distance to the fitness optimum may influence the likelihood of repeated evolution. This work thus can serve as a step towards a more complete understanding of repeated evolution and the probability of it occurring between populations.

The assembled genomes of the six different strains provide more information about the underlying genetic variation present in the species. Future work may examine the genomes of populations evolved in a more nutritionally rich environment with a potentially more rugged fitness landscape (such as lysogeny broth media) and incorporate strains from a wider variety of the phylogenetic tree to further examine the relationship between genetic relatedness and repeated evolution.

Tables and Figures

<table>
<tr><th>Function</th><th>Times affected</th><th>Example mutations</th></tr>
<tr><td>Chemotaxis</td><td>19</td><td>aer, cheR1*, PA14_27000 | PA2867*</td></tr>
<tr><td>Adaptation, Protection</td><td>16</td><td>htrB1*, pctA*, lasR | PA1424 -> PA1433*</td></tr>
<tr><td>Transcriptional regulators</td><td>13</td><td>mexT, pilR, rpoB</td></tr>
<tr><td>Putative enzymes</td><td>10</td><td>wapH, cdrA*, pchF*</td></tr>
<tr><td>Transferase activity, transferring glycosyl groups</td><td>9</td><td>spuD, ssg, PA14_30470</td></tr>
<tr><td>Motility & Attachment</td><td>8</td><td>pscP, pelA -> pelG | pelE*, pvdT -> PA2421*</td></tr>
<tr><td>Secreted Factors
(toxins, enzymes, alginate)</td><td>8</td><td>dipA, PA0041, phzA1</td></tr>
<tr><td>Carbon compound catabolism</td><td>7</td><td>crc, cdrA*</td></tr>
<tr><td>DNA replication, recombination, modification & repair</td><td>7</td><td>dnaX, PA3465, fiuA*</td></tr>
<tr><td>Cell wall / LPS / capsule</td><td>6</td><td>orfK, orfN, htrB1*</td></tr>
</table>

Table 1. The ten functions most affected by coding mutations across all isolates. If a mutation was detected in the coding region of a gene, their corresponding functions were considered affected (based on functional classifications assigned by PseudoCAP). Asterisks were used to mark mutations which affected more than one function.

Mutation Type	Count	Coding	Non-coding	Core	Pan
Synonymous SNP	19	19	0	14	5
Non-synonymous SNP	28	28	0	24	4
Non-sense SNP	4	4	0	4	0
Intergenic SNP	10	0	10	8	2
Small Indel	38	34	4	31	7
Large Deletion	10	9	1	10	0
Large Insertion	2	0	1	1	0
Large Substitution	4	4	0	3	1
Total	115	98	16	95	19

Table 2. All mutations by gene location in relative to core genome and mutation location relative to a gene coding region. Genes were classified as part of "core genome" if they were present in all six ancestor genomes, otherwise they were classified as part of the "pan genome". "Small" indels are insertions or deletions affecting less than 50bp, while "large" mutations are those that affected 50bp or more.

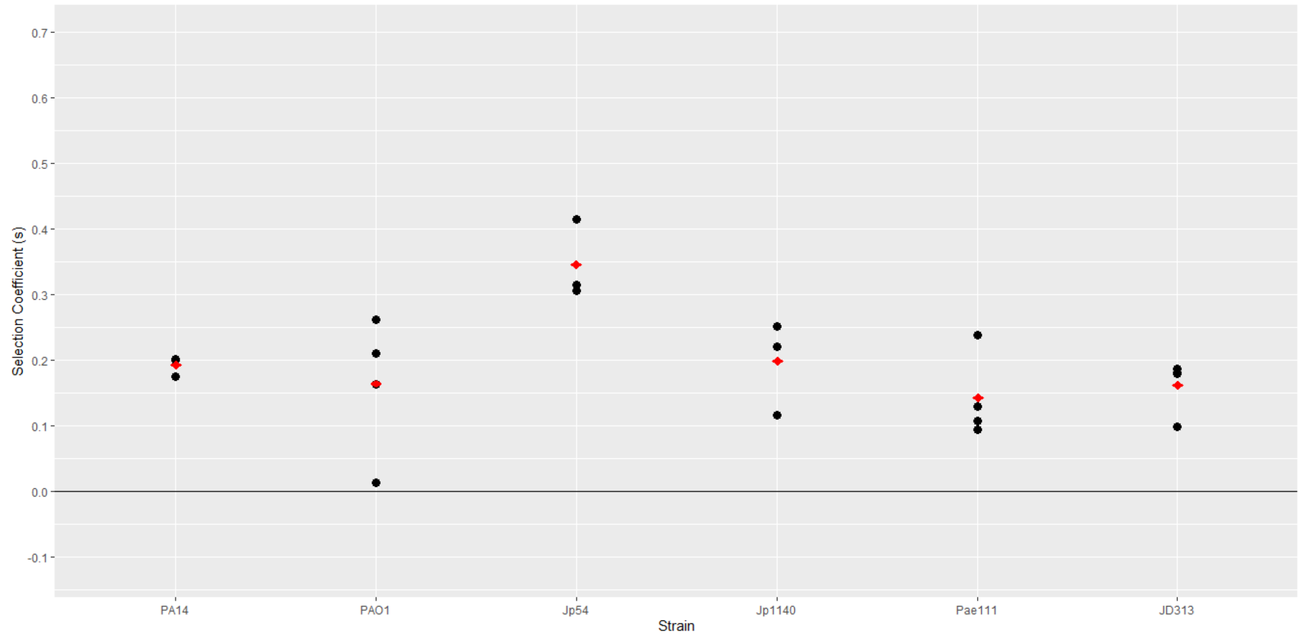

Figure 1. Relative fitness of evolved isolates versus their ancestors, across six strains of *P. aeruginosa*. Black dots represent the average selection coefficient (s) for isolates, and red diamonds represent the strain average. There are 6 replicate measures per isolate for all strains. Strains PAO1, Pae111, and JD313, had 4 isolates each, totaling 24 replicates per strain. Strains PA14, Jp54, and Jp1140 had 3 isolates each, totaling 18 replicates per strain.

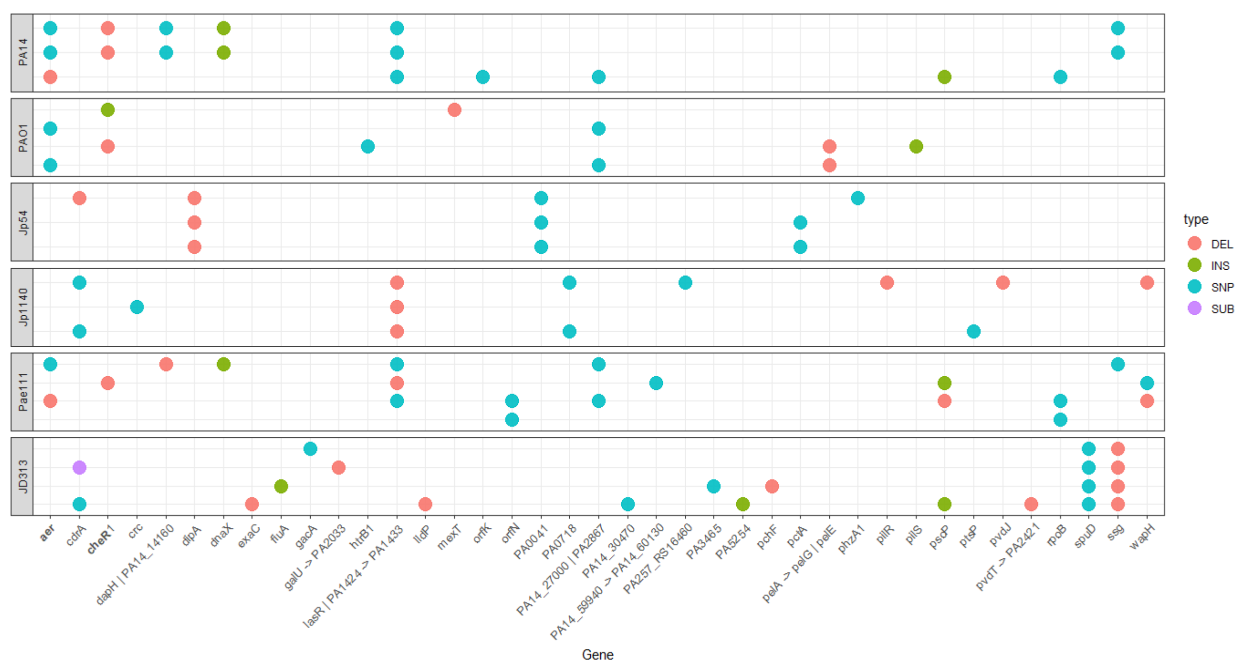

Figure 2. All gene-level mutations in coding regions across all *P. aeruginosa* isolates, grouped by strain. Out of 115 detected mutations, 98 occurred in coding regions, the majority of which were SNPs and small indels.

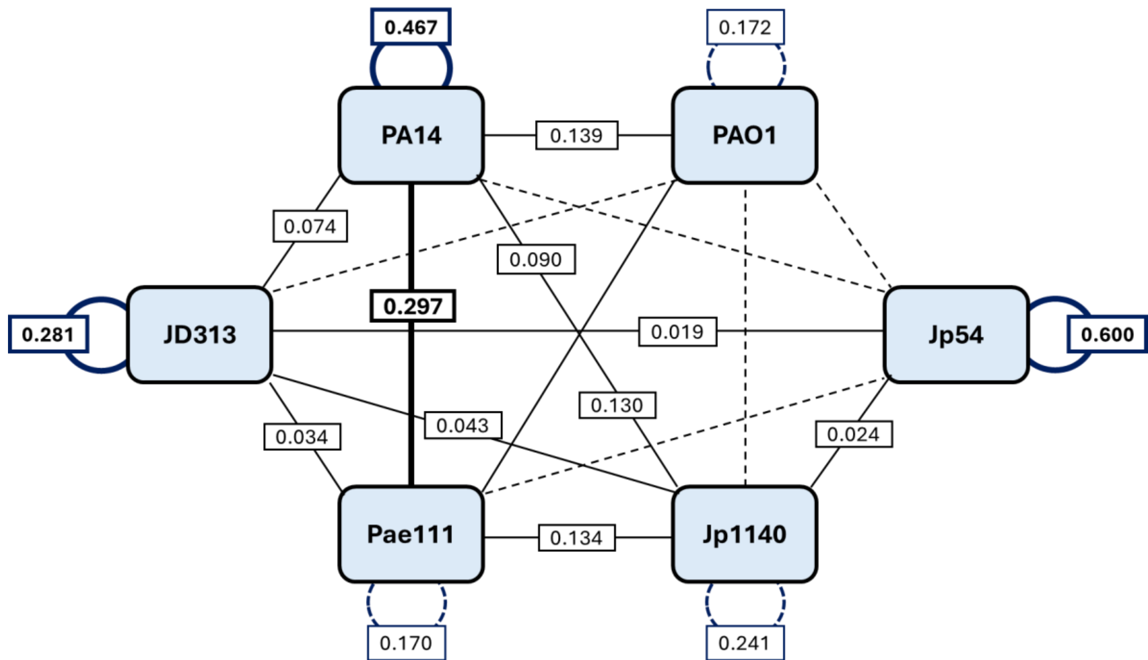

Figure 3. Average Jaccard index scores between and within each *P. aeruginosa* strain, where higher values indicate higher levels of similarity. Each comparison between two strains is represented by a line and the average Jaccard index score for that comparison (circular lines representing within-strain comparison). Dashed lines between strains represent average Jaccard indexes of zero, while bold lines indicate values higher than expected after Bonferroni correction for 21 comparison groups ($p = 0.0024$).

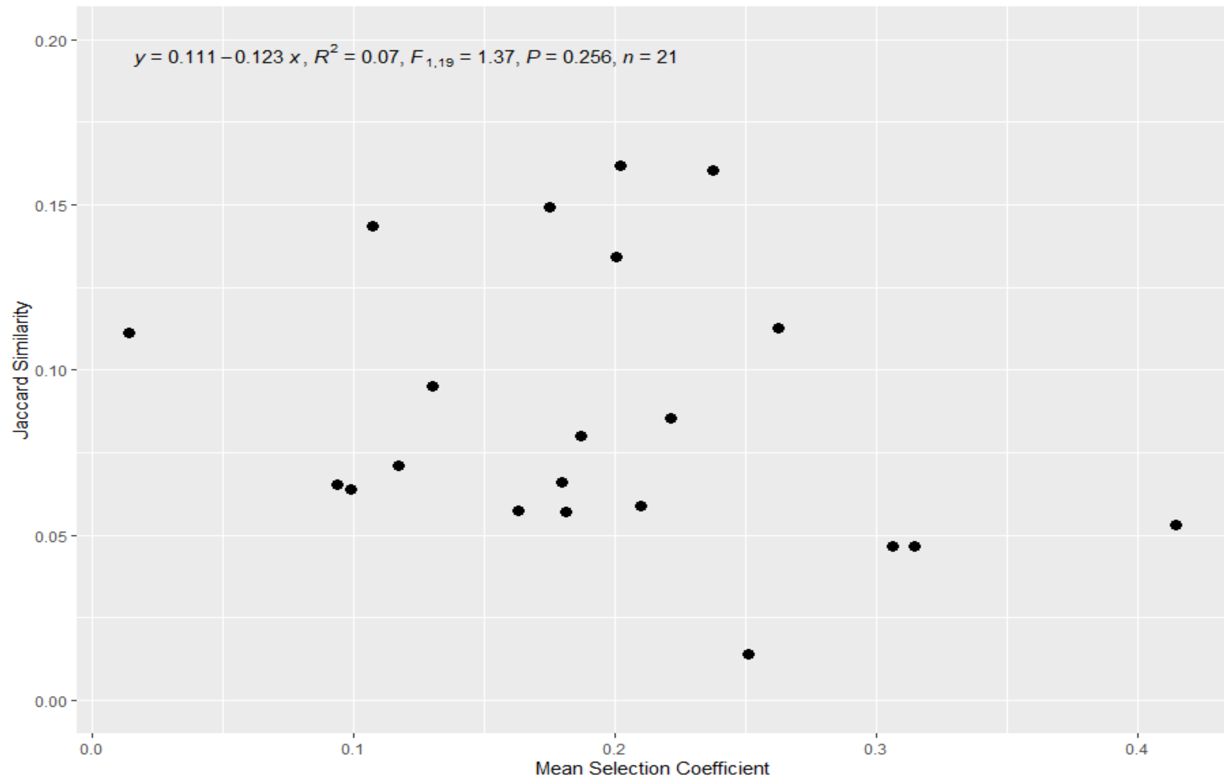

Figure 4. The extent of repeated evolution relative to change in fitness. The mean Jaccard index value and selection coefficient for each isolate is used to estimate the degree of repeated evolution and change in relative fitness, respectively.

Figure 5. Frequency of repeated evolution as related to genetic distance between *P. aeruginosa* isolates. Genetic distance values are on the level of strain, so isolates from the same strain will have a genetic distance of zero (parallel evolution grouping).

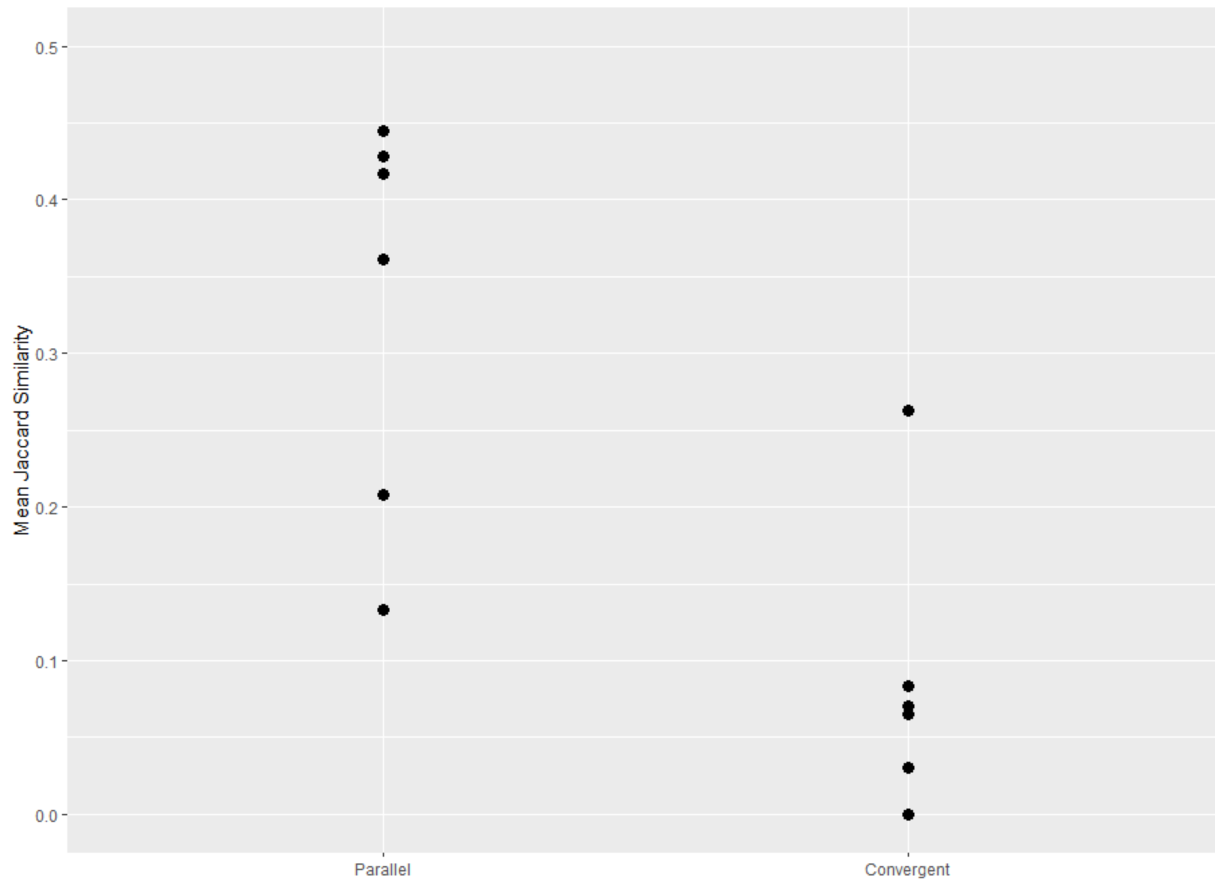

Figure 6. Mean frequency of repeated evolution in parallel (within-strain) and convergent (between-strain) comparisons for all strains. Comparisons were parallel when genetic distance = 0, and convergent where genetic distance > 0.

References

Andrew J. Page, Carla A. Cummins, Martin Hunt, Vanessa K. Wong, Sandra Reuter, Matthew T. G. Holden, Maria Fookes, Daniel Falush, Jacqueline A. Keane, Julian Parkhill, "Roary: Rapid large-scale prokaryote pan genome analysis*", Bioinformatics*, 2015;31(22):3691-3693 doi:10.1093/bioinformatics/btv421

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Amiel, E., Lovewell, R. R., O'Toole, G. A., Hogan, D. A., & Berwin, B. (2010). Pseudomonas aeruginosa evasion of phagocytosis is mediated by loss of swimming motility and is independent of flagellum expression. *Infection and immunity*, 78(7), 2937–2945. https://doi.org/10.1128/IAI.00144-10

Bains M, Fernández L, Hancock RE. (2012) Phosphate starvation promotes swarming motility and cytotoxicity of Pseudomonas aeruginosa. Appl Environ Microbiol. 78(18):6762-8. doi: 10.1128/AEM.01015-12. Epub 2012 Jul 6. PMID: 22773629; PMCID: PMC3426718.

Bailey, S. F., Blanquart, F., Bataillon, T., & Kassen, R. (2017). What drives parallel evolution? How population size and mutational variation contribute to repeated evolution. *BioEssays*, *39*(1), 1-9.

Bailey, S. F., Rodrigue, N., & Kassen, R. (2015). The effect of selection environment on the probability of parallel evolution. Molecular biology and evolution, 32(6), 1436-1448.

Bolnick, D. I., Barrett, R. D., Oke, K. B., Rennison, D. J., & Stuart, Y. E. (2018). (Non) parallel evolution. *Annual Review of Ecology, Evolution, and Systematics*, *49*(1), 303-330.

Bollback, J. P., & Huelsenbeck, J. P. (2009). Parallel genetic evolution within and between bacteriophage species of varying degrees of divergence. *Genetics*, *181*(1), 225-234.

Chevin, L. M., Martin, G., & Lenormand, T. (2010). Fisher's model and the genomics of adaptation: restricted pleiotropy, heterogenous mutation, and parallel evolution. *Evolution*, *64*(11), 3213-3231.

Christin, P. A., Edwards, E. J., Besnard, G., Boxall, S. F., Gregory, R., Kellogg, E. A., ... & Osborne, C. P. (2012). Adaptive evolution of C4 photosynthesis through recurrent lateral gene transfer. *Current Biology*, *22*(5), 445-449.

Conte, G. L., Arnegard, M. E., Peichel, C. L., & Schluter, D. (2012). The probability of genetic parallelism and convergence in natural populations. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1749), 5039-5047.

Deatherage, D.E., Barrick, J.E. (2014) Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods Mol. Biol. 1151: 165–188.

Dettman, J. R., & Kassen, R. (2021). Evolutionary genomics of niche-specific adaptation to the cystic fibrosis lung in Pseudomonas aeruginosa. *Molecular biology and evolution*, *38*(2), 663-675.

Dettman, J. R., Rodrigue, N., Aaron, S. D., & Kassen, R. (2013). Evolutionary genomics of epidemic and nonepidemic strains of Pseudomonas aeruginosa. *Proceedings of the National Academy of Sciences*, *110*(52), 21065-21070.

Dettman, J. R., Sztepanacz, J. L., & Kassen, R. (2016). The properties of spontaneous mutations in the opportunistic pathogen Pseudomonas aeruginosa. *BMC genomics*, *17*, 1-14.

Emanuele Bosi, Beatrice Donati, Marco Galardini, Sara Brunetti, Marie-France Sagot, Pietro Lió, Pierluigi Crescenzi, Renato Fani, Marco Fondi, MeDuSa: a multi-draft based scaffolder, *Bioinformatics*, Volume 31, Issue 15, August 2015, Pages 2443–2451, https://doi.org/10.1093/bioinformatics/btv171

Filipow, N., Mallon, S., Shewaramani, S., Kassen, R., & Wong, A. (2024). The impact of genetic background during laboratory evolution of Pseudomonas aeruginosa in a cystic fibrosis-like environment. *Evolution*, *78*(3), 566-578.

Fisher, R. A. (1999). The genetical theory of natural selection: a complete variorum edition. *Oxford University Press*.

Folkesson, A., Jelsbak, L., Yang, L., Johansen, H.K., Ciofu, O., Høiby, N., Molin, S. (2012) Adaptation of *Pseudomonas aeruginosa* to the cystic fibrosis airway; an evolutionary perspective. *Nat Rev Microbiol* 10: 841-851

Frickel, J., Feulner, P. G., Karakoc, E., & Becks, L. (2018). Population size changes and selection drive patterns of parallel evolution in a host–virus system. *Nature communications*, *9*(1), 1706.

Gillespie, J. H. (1984). Molecular Evolution Over the Mutational Landscape. Evolution, 38(5), 1116–1129. https://doi.org/10.2307/2408444

Hicks, Alexandra. *Effect of Genetic Background on Diversification of Pseudomonas aeruginosa*. Diss. *Université d'Ottawa/University of Ottawa*, 2023.

Hoffman, L. R., Kulasekara, H. D., Emerson, J., Houston, L. S., Burns, J. L., Ramsey, B. W., & Miller, S. I. (2009). Pseudomonas aeruginosa lasR mutants are associated with cystic fibrosis lung disease progression. *Journal of Cystic Fibrosis*, 8(1), 66-70.

Huynh, T. T., McDougald, D., Klebensberger, J., Al Qarni, B., Barraud, N., Rice, S. A., Kjelleberg, S., & Schleheck, D. (2012). Glucose starvation-induced dispersal of Pseudomonas aeruginosa biofilms is cAMP and energy dependent. *PloS one*, 7(8), e42874. https://doi.org/10.1371/journal.pone.0042874

Kassen, R. (2014). *Experimental evolution and the nature of biodiversity (1st Edition)*. Roberts.

Kassen, R. (2024). *Experimental evolution and the nature of biodiversity (2nd Edition)*. Oxford University Press.

Lenormand, T., Chevin, L. M., Bataillon, T., Ramsey, G., & Pence, C. (2016). Parallel evolution: what does it (not) tell us and why is it (still) interesting. *Chance in evolution*, *196*.

Lyczak, J.B., Cannon, C.L., Pier, G.B. (2000) Establishment of *Pseudomonas aeruginosa* infection: lessons from a versatile opportunist. *Microbes Infect* 2(9):1051-1060

Matic, I. (2019). Mutation rate heterogeneity increases odds of survival in unpredictable environments. *Molecular cell*, 75(3), 421-425.

Moradali, M. F., Ghods, S., & Rehm, B. H. (2017). Pseudomonas aeruginosa lifestyle: a paradigm for adaptation, survival, and persistence. *Frontiers in cellular and infection microbiology*, 7, 39.

Nei, M. (1972). Genetic distance between populations. *The American Naturalist*, *106*(949), 283-292.

Nosil, P., Crespi, B. J., & Sandoval, C. P. (2002). Host-plant adaptation drives the parallel evolution of reproductive isolation. *Nature*, *417*(6887), 440-443.

Oke, K. B., Rolshausen, G., LeBlond, C., & Hendry, A. P. (2017). How parallel is parallel evolution? A comparative analysis in fishes. *The American Naturalist*, *190*(1), 1-16.

Oliver, A., & Mena, A. (2010). Bacterial hypermutation in cystic fibrosis, not only for antibiotic resistance. Clinical Microbiology and Infection, 16(7), 798-808.

Orr, H. A. (2005). The probability of parallel evolution. *Evolution,* 59(1), 216-220.

Østman, B. (2013). Effects of epistasis and pleiotropy on fitness landscapes. *Evolutionary Biology: Exobiology and Evolutionary Mechanisms*, 37-46.

Ozer, E. A., Allen, J. P., & Hauser, A. R. (2014). Characterization of the core and accessory genomes of Pseudomonas aeruginosa using bioinformatic tools Spine and AGEnt. *BMC genomics*, *15*, 1-17.

Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes De Novo Assembler. Curr Protoc Bioinformatics. 2020 Jun;70(1):e102. doi: 10.1002/cpbi.102. PMID: 32559359.

Schenk, M. F., Zwart, M. P., Hwang, S., Ruelens, P., Severing, E., Krug, J., & de Visser, J. A. G. (2022). Population size mediates the contribution of high-rate and large-benefit mutations to parallel evolution. *Nature Ecology & Evolution*, *6*(4), 439-447.

Schick, A., & Kassen, R. (2018). Rapid diversification of Pseudomonas aeruginosa in cystic fibrosis lung-like conditions. *Proceedings of the National Academy of Sciences*, 115(42), 10714-10719.

Schluter, D., Clifford, E. A., Nemethy, M., & McKinnon, J. S. (2004). Parallel evolution and inheritance of quantitative traits. *The American Naturalist*, *163*(6), 809-822.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics, 30(14), 2068-2069.

Smith, E. E., Buckley, D. G., Wu, Z., Saenphimmachak, C., Hoffman, L. R., D'Argenio, D. A., Miller S.I, Ramsey B.W, Speert D.P., Moskowitz S.M., Burns J.L., Kaul R., & Olson, M. V. (2006). Genetic adaptation by Pseudomonas aeruginosa to the airways of cystic fibrosis patients. *Proceedings of the National Academy of Sciences,* 103(22), 8487-8492.

Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*, *23*(1), 23-35.

Snitkin, E. S., & Segre, J. A. (2015). Pseudomonas aeruginosa adaptation to human hosts. *Nature genetics,* 47(1), 2-3.

Stern, D. L. (2013). The genetic causes of convergent evolution. *Nature Reviews Genetics*, *14*(11), 751-764.

Stover, K.C., X.Q. Pham, A.L. Erwin, S.D. Mizoguchi, P. Warrener, M.J. Hickey, F.S.L. Brinkman, W. O. Hufnagle, D.J. Kowalik, M. Lagrou, R.L. Garber, L. Goltry, E. Tolentino, S. Westbrock-Wadman, Y. Yuan, L.L. Brody, S.N. Coulter, K.R. Folger, A. Kas, K. Larbig, R. Lim, K. Smith, D. Spencer, G.K.-S. Wong, Z. Wu, I. Paulsen, J. Reizer, M.H. Saier, R.E.W. Hancock, S. Lory, and M.V. Olson (2000). Complete genome sequence of Pseudomonas aeruginosa PAO1: an opportunistic pathogen. *Nature*. 406: 959-964. PMID:10984043.

Stuart, Y., Veen, T., Weber, J. *et al.* (2017). Contrasting effects of environment and genetics generate a continuum of parallel evolution. *Nat Ecol Evol* **1**, 0158 https://doi.org/10.1038/s41559-017-0158

Thöming, J. G., Tomasch, J., Preusse, M., Koska, M., Grahl, N., Pohl, S., Willger, S. D., Kaever, V., Müsken, M., & Häussler, S. (2020). Parallel evolutionary paths to produce more than one Pseudomonas aeruginosa biofilm phenotype. *NPJ biofilms and microbiomes*, 6, 2. https://doi.org/10.1038/s41522-019-0113-6

Turner, C. B., Marshall, C. W., & Cooper, V. S. (2018). Parallel genetic adaptation across environments differing in mode of growth or resource availability. *Evolution letters*, *2*(4), 355–367. https://doi.org/10.1002/evl3.75

Weinreich, D. M., Delaney, N. F., DePristo, M. A., & Hartl, D. L. (2006). Darwinian evolution can follow only very few mutational paths to fitter proteins. S*cience*, *312*(5770), 111-114.

Weinreich, D. M., Watson, R. A., & Chao, L. (2005). Perspective: sign epistasis and genetic constraint on evolutionary trajectories. *Evolution*, 59(6), 1165-1174.

Winsor GL, Griffiths EJ, Lo R, Dhillon BK, Shay JA, Brinkman FS (2016). Enhanced annotations and features for comparing thousands of Pseudomonas genomes in the Pseudomonas genome database. *Nucleic Acids Res*. (2016) doi: 10.1093/nar/gkv1227 (Database issue). PMID:26578582

Witt, K. E., & Huerta-Sánchez, E. (2019). Convergent evolution in human and domesticate adaptation to high-altitude environments. *Philosophical Transactions of the Royal Society B*, *374*(1777), 20180235.

Wood, T. E., Burke, J. M., & Rieseberg, L. H. (2005). Parallel genotypic adaptation: when evolution repeats itself. *Genetica,* 123(1-2), 157–170. https://doi.org/10.1007/s10709-003-2738-9

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding, and selection in evolution.

Van Cleve, J., & Weissman, D. B. (2015). Measuring ruggedness in fitness landscapes. *Proceedings of the National Academy of Sciences*, 112(24), 7345-7346.

de Visser, J., Krug, J. (2014). Empirical fitness landscapes and the predictability of evolution. *Nat Rev Genet 15*, 480–490. https://doi.org/10.1038/nrg3744