Genomics of Diversification of *Pseudomonas aeruginosa* in Cystic Fibrosis Lung-like Conditions

Alana Schick, Sonal Shewaramani*, and Rees Kassen

Biology Department and Centre for Advanced Research in Environmental Genomics, University of Ottawa, Ottawa, ON K1N 6N5, Canada

*Corresponding author: E-mail: sshewara@uottawa.ca.

Accepted: 12 May 2022

Abstract

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. 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 human lung ecology: 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 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 positive selection (drivers). A significantly lower proportion of driver mutations in spatially structured populations suggests that reduced dispersal 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 CF lung adaptation can arise as a result of adaptation to a novel environment and do not necessarily require antimicrobial treatment, immune system suppression, or competition from other microbial species to occur.

Keywords: *Pseudomonas aeruginosa*, evolve-and-resequence experiments, nutrient adaptation, spatial structure, diversification.

Significance

Pseudomonas aeruginosa is a common cause of persistent and difficult to eradicate chronic infections in patients with cystic fibrosis (CF). In laboratory environments designed to capture key selective drivers in the CF lung, *P. aeruginosa* strains rapidly diversified into a range of phenotypically distinct types. Genome sequencing revealed that this phenotypic diversification is underlain by extensive genetic variation and genetic divergence was more pronounced in populations that evolved in ecologically complex environments. Notably, laboratory populations harbored mutations in many of the same genes and genetic pathways seen in *P. aeruginosa* isolates obtained from chronically infected CF patients, shedding light on the range of genetic routes to adaptation in CF-like conditions.

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 and is associated with increased morbidity and mortality, irrespective of lung function (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 (Poole 2005; Smith et al. 2006; Mowat et al. 2011). The transition to chronic infection is also marked—and, indeed, may be caused—by the rapid diversification of *P. aeruginosa* into phenotypically and genetically distinct clones, some of which coexist for years within the same host (Foweraker et al. 2005; Ashish et al. 2013; Workentine et al. 2013; Wright et al. 2013; Markussen et al. 2014). The causes of diversification, and the contribution of this diversity to the long-term persistence of *P. aeruginosa* infections, remain poorly understood.

Two ecological factors shown to promote diversification of *P. aeruginosa* in the CF lung and recapitulate the phenotypic variation seen among clinical isolates from CF patients are nutritional complexity and dispersal rate (Schick and Kassen 2018). CF airways are nutritionally complex, whereas there is reduced dispersal among subpopulations due to the thick mucous layer covering the lung epithelia and/or the spatially compartmentalized nature of the lung. Our previous work (Schick and Kassen 2018) 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; Schluter 2000; Kassen 2009; Flynn et al. 2016). 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). This rapid and repeated diversification of *P. aeruginosa*, documented in both clinical isolates and laboratory experiments, suggests that 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.

However, the genetic properties of diversification in the CF lung remain unclear. In particular, the extent to which phenotypic variation is matched by comparable levels of genetic variation is not well understood. Theory suggests that there should be a close correspondence between the two. 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 (Poltak and Cooper 2011; Traverse et al. 2013; Ellis et al. 2015; Schick et al. 2015; Flynn et al. 2016) that can coexist for prolonged periods of time or even indefinitely (Behringer et al. 2018; Leale and Kassen 2018). Third, lower rates of dispersal among subpopulations associated with spatial segregation allow beneficial mutations arising independently in different subpopulations to persist longer than in a well-mixed system, especially when there are high levels of environmental heterogeneity (Park and Krug 2007; Campos et al. 2008). Genetic divergence should thus be closely linked to phenotypic disparity, being most pronounced in ecologically complex environments, such as those with abundant ecological opportunity and spatial structure, precisely the conditions thought to characterize the CF airway.

The extent to which genetic changes associated with divergence in the CF lung repeatedly evolve also remains poorly understood. 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; Turner et al. 2018). In large populations, rare, large-effect beneficial mutations are more likely to arise and outcompete independently evolved beneficial mutations of smaller effect (Lanfear et al. 2014). For a given population size, however, the probability of parallelism under divergent selection should be lower than under directional selection, because the former results in the evolution of genetically distinct niche specialists, whereas the latter is expected to lead to a single, well-adopted genotype (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 subpopulations that evolve more or less independently of each other. Within patients, therefore, 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, populations were evolved in four different environments that vary in the degree to which they resemble the CF lung along dimensions of nutritional complexity and viscosity. The synthetic CF sputum (SCFM) is a defined medium resembling

the nutritional components of the CF lung (Palmer et al. 2007). Mucin increases culture medium viscosity. In CF patients, highly viscous sputum is thought to support the growth of *P. aeruginosa* in biofilms and mucin-supplemented media supports this mode of growth (Sriramulu et al. 2005; Fung et al. 2010). Thus, the most CF lung-like environment was composed of SCFM and mucin. 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. Our results, which are based on data from a large number of end-point populations rather than time series of data from fewer populations, provide insight into the spectrum, and quantity of genetic changes associated with diversification in response to 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.

Results

Genome Sequencing of Evolved Populations

Comparing whole-genome sequence data from the 81 evolved populations against the ancestral Pa14 genome revealed a total of 656 unique genetic variants (supplementary table S1, Supplementary Material online). The identity and genomic location of all mutations recovered in our analysis (single-nucleotide polymorphisms [SNPs], small insertions and deletions [small indels], large deletions and large amplifications) are summarized in supplementary figures S1, S2 and table S1, Supplementary Material online. Overall, an average of 8.10 variants per population was identified. Upon further examination, an average of 8.67, 5.48, 8.32, and 10.29 variants per population were detected in the SCFM, MIN, SCFM + mucin, and MIN + mucin environments, respectively (supplementary fig. S3, Supplementary Material online).

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) (figure 1 and supplementary fig. S4 Supplementary Material online). Variants were distributed across the genome (supplementary fig. S1, Supplementary Material online), 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 were the most common variant class in the experiment, being nearly three times more common than the next most abundant class, small indels (supplementary fig. S2*A*, Supplementary Material online). Approximately 20% of all detected mutations were indels and structural variants, with small indels occurring at a range of frequencies from rare to fixed (supplementary fig. *S4B*, Supplementary Material online). Indels were observed most often in genes commonly mutated in *P. aeruginosa* isolates from CF patients, particularly *lasR* (discussed in more detail below). Overall, there was little variation among environments in the distribution of SNPs (supplementary fig. S2*B*, Supplementary Material online). 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 detected among environments, which varied substantially across treatments due to fewer variants being recovered in MIN relative to the other three environments (supplementary figs. S2*B* and S3, Supplementary Material online; ANOVA, $F_{(3,77)} = 6.24$, $P < 0.001$).

Positive Selection and the Prevalence of Genetic Hitchhiking

We aimed to determine the extent to which 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, such as the stochastic effects of drift. In all environments, population sizes were $\sim 10^8$ CFU/ml, producing ~6.7 generations/day and 220 generations over the course of the experiment (Schick and Kassen 2018). The large population sizes 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" (Smith and Haigh 1974). 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, so that we can identify which genes are experiencing directional selection. This task is relatively straightforward when mutations are few in number, as their fitness effects can be evaluated directly through competition experiments (e.g., Khan et al. 2011; Schick et al. 2015; Dillon and Cooper 2016; Buskirk et al. 2017). 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 synonymous mutations are neutral with respect to fitness (although reports to the contrary suggest synonymous mutations can occasionally be adaptive; see Bailey et al. 2014; Agashe et al. 2016; Kristofich et al. 2018; Lebeuf-Taylor et al.

2019; Bailey et al. 2021). Of the 445 SNPs we observed in protein-coding regions, 371 and 74 were nonsynonymous and synonymous, respectively, which corresponds to a frequency of 16.6% synonymous SNPs. Given that 25.1% of all SNPs in protein-coding regions are synonymous in Pa14 (Yang et al. 2011), nonsynonymous SNPs are vastly overrepresented relative to synonymous SNPs in our experiment ($\chi^2 = 9.926$, $P = 0.0017$), 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 time to become sufficiently common to be detectable via sequencing—can therefore be calculated by subtracting the number of expected nonsynonymous SNPs under neutrality 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. This number on its own reveals rather little, although it is notable that this value is not too far off the 20% estimated for evolving populations of yeast (Buskirk et al. 2017). More interesting is how our estimate of the fraction of driver mutations varies across the treatments in our experiment, likely due to the different selection pressures populations are experiencing. Disaggregating by environment reveals a striking result: environments lacking mucin harbor a larger proportion of putative drivers than those with mucin (table 1; $\chi^2 = 7.638$, $P < 0.008$). As population sizes and mutation rates are similar across environments (Schick and Kassen 2018), these results are unlikely to reflect lower mutation supply rates in the presence of mucin. If anything, 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, each with lower effective population sizes (N_e) relative to a well-mixed system. Under such conditions, a large number of low-frequency mutations can arise to appreciable frequency in different subpopulations, whereas access to beneficial mutations of large effect will be restricted because only those mutations with selection coefficients $>1/N_e$ will have an appreciable chance of fixing (Eyre-Walker and Keightley 2007; Lanfear et al 2014). In short, the spatial structure generated by mucin reduces the supply rate of beneficial mutations by generating many smaller subpopulations that have lower effective population sizes than would be the case in a larger, well-mixed population.

Our results indicate that although populations in our study are experiencing different selection pressures, there are a few, strongly selected sites within each environment. In addition, we were able to detect hitchhiker mutations whose frequencies range widely due to clonal interference within the populations. However, it is also worth noting that our model for estimating the number of hitchhiker and driver mutations in our evolved populations assumes that these populations are only experiencing positive selection. It is possible that a larger proportion of nonsynonymous mutations than we have estimated are under strong selection, as a large fraction of nonsynonymous mutations has likely already been removed from the populations due to purifying selection.

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 (fig. 1). The paucity of high-frequency synonymous SNPs suggests the majority of this class of mutation do not contribute to adaptation (with the possible exception of two high-frequency synonymous SNPs: *gdhB*, an NAD-dependent glutamate dehydrogenase, in SCFM and PA14_49020, a putative Pf5 repressor C protein in MIN + mucin), 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. Most synonymous SNPs are thus likely hitchhikers whose frequencies, by definition, must be equal to or less than those of their respective driver mutations. It follows, then, that the frequencies of nonsynonymous driver SNPs should generally be 0.16 or more. Accordingly, going forward, we restrict our analyses to those mutations that are likely to be under selection, that is, putative nonsynonymous driver mutations, by excluding synonymous SNPs and nonsynonymous SNPs with a read frequency <0.16.

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

Table 1
Estimated Proportion of Nonsynonymous SNPs that are Driver Mutations

Treatment	Observed Nonsynonymous SNPs	Observed Synonymous SNPs	Estimated Number of Nonsynonymous Hitchhikers SNPs	Estimated Number of Nonsynonymous Driver SNPs	Estimated Proportion of Nonsynonymous Driver SNPs (%)
All	371	74	222	149	28.38
SCFM	84	12	36	48	45.28
MIN	72	7	21	51	51.00
SCFM + mucin	86	23	69	17	12.88
MIN + mucin	129	32	96	33	17.65

NOTE.—The estimated number of nonsynonymous 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, Yang et al. 2011). 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 (i.e., including SNPs in noncoding regions).

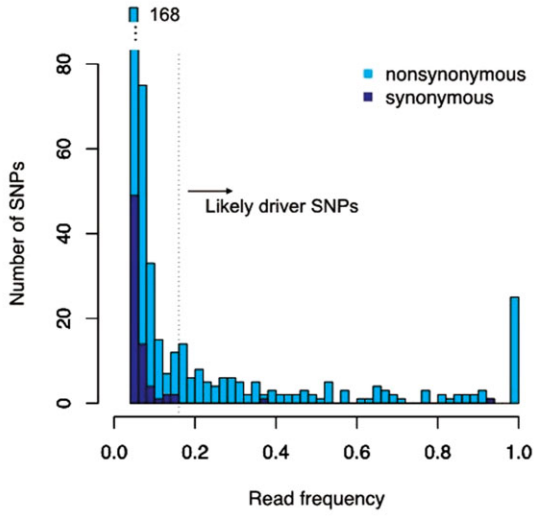

FIG. 1.—Frequency spectra for nonsynonymous and synonymous SNPs. Distribution of read frequencies for all SNPs found across all evolved populations, nonsynonymous (light blue) and synonymous (dark blue). The majority of nonsynonymous SNPs present at frequencies >0.16 (the range of frequencies of most synonymous SNPs) are likely to be driver mutations.

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). Here, mean heterozygosity is used as a measure of genetic diversity in the evolved populations, and is calculated as $2pq$, where p is the estimated allele frequency and $q = 1 - p$ (Blanton 2018). These mean heterozygosity values were then regressed 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 >0 for all treatments (t-test, $P < 0.0001$ for all treatments; fig. 2*A*), 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, $P = 0.006$; fig. 2*B*). Although 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 if relying only on genome sequence data when mutation supply rates are high.

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 (Smith et al. 2006; Mowat et al. 2011) 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 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 nonsynonymous SNPs that we expect to be hitchhiking, as discussed above. Although this approach likely inflates our estimates of parallelism, it should do so in

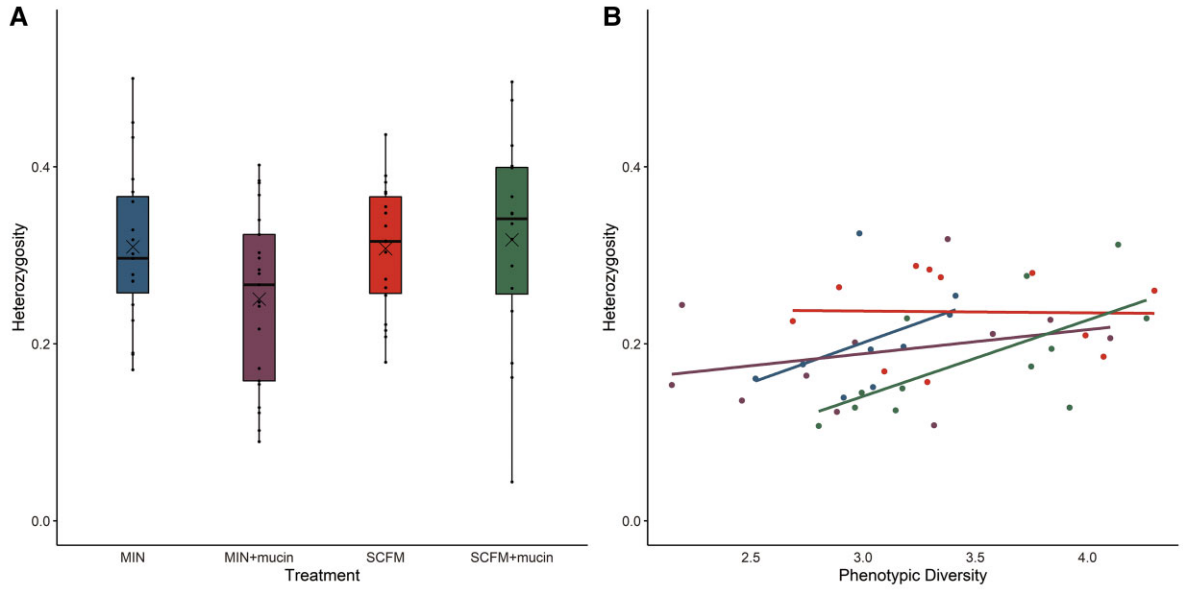

Fig. 2.—Genetic diversity within populations. (*A*) Population mean heterozygosity, calculated using frequencies of variants at all loci with a variant present in a given population, excluding synonymous and low-frequency (<0.16) variants. Horizontal bars represent treatment median and crosses represent 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).

a way that is common across treatments and so should not introduce any systematic bias. Our results are summarized visually in figure 3 and supplementary figure S5, Supplementary Material online and presented more quantitatively in table 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 distance, 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 Materials and Methods. Briefly, dispersion is a measure of the mean distance from the centroid of a principal coordinates analysis (PCoA) based on Euclidean distance. Essentially, larger values signify more divergent populations and therefore less parallelism. On the other hand, Jaccard distance is a measure of how dissimilar sets are; it is the complement of the Jaccard index (J). A Jaccard distance of 1 for any two populations signifies that they are completely dissimilar, having no gene variants in common, whereas a Jaccard distance of 0 indicates that the two populations completely share mutations. Therefore, larger values signify more divergent populations and less parallelism. Finally, *C*-scores measure observed repeatability relative to expectations under randomness (Yeaman et al. 2018), with smaller values signifying less constraint and therefore less parallelism.

For all metrics reported in table 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 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.

Gene-level Parallelism and Environmental Specificity

Gene-level parallelism, the proportion of populations with mutations in that gene, was then disaggregated by treatment and compared with parallelism across all populations. The spectrum of parallel genetic changes evolved in each environment, shown in figure 4, 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* (a motility regulator) in MIN + mucin, or *pilA* (a fimbrial protein) and PA14_32420 (a putative oxidoreductase) 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 nonsynonymous mutations in a given gene, within than

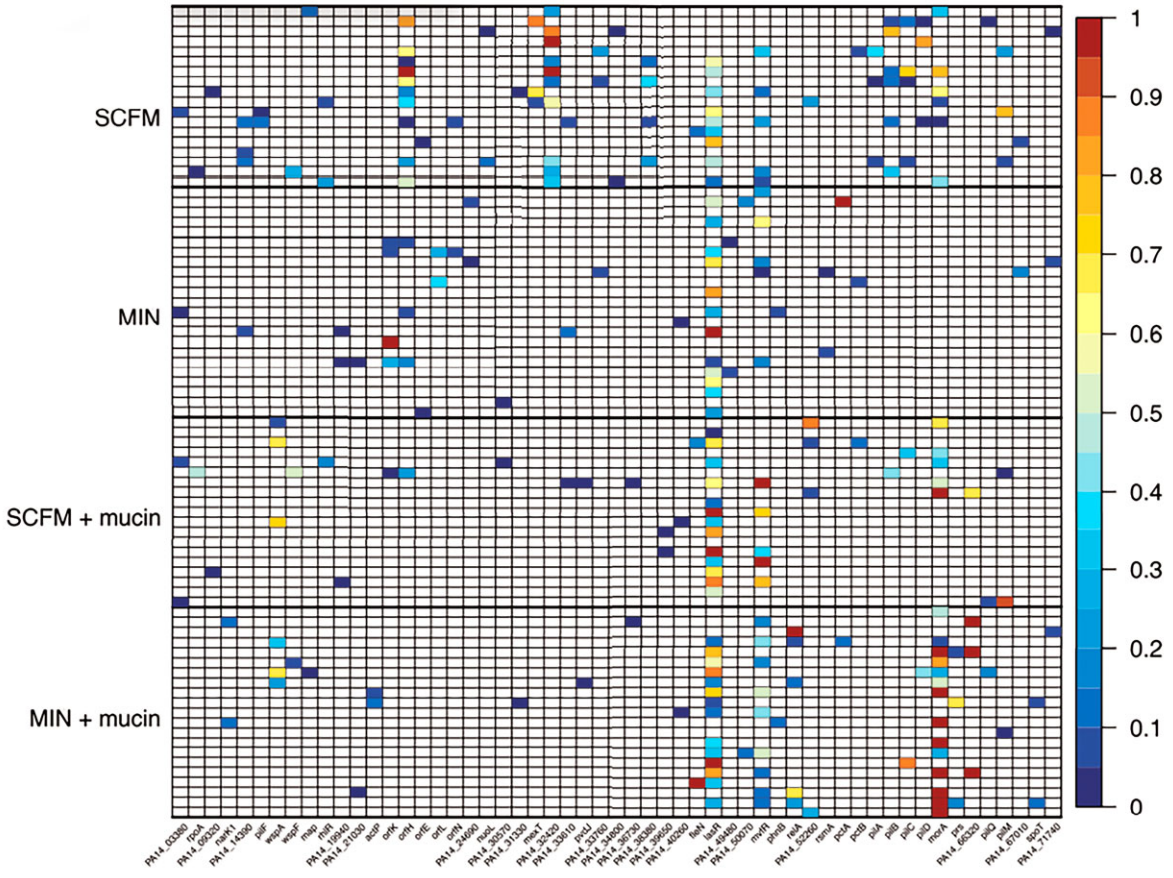

Fig. 3.—Variant frequencies for multiple use genes. Frequency of variants for genes that have variants in more than one population, after excluding synonymous SNPs ($N = 55$ genes). Each row represents a population, with the presence of a mutation in each gene indicated by a colored box (high-frequency mutations in red and low-frequency mutations in blue). Genes are arranged by genomic position.

Table 2

Parallelism and Constraint

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

Note.—Population level patterns of parallelism within each treatment. Three measures were used to quantify parallelism: variance in dispersion of Euclidean distances between populations (larger values signify more divergence), Jaccard distance (values closer to 1 signify more divergence), and *C*-score (smaller values signify more divergence; Yeaman et al. 2018). For all metrics, SE is standard error and treatments are ranked from most divergent (4) to least divergent (1), equivalent to least parallel (4) to most parallel (1). All three metrics were calculated excluding synonymous and low-frequency nonsynonymous 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.

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 <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

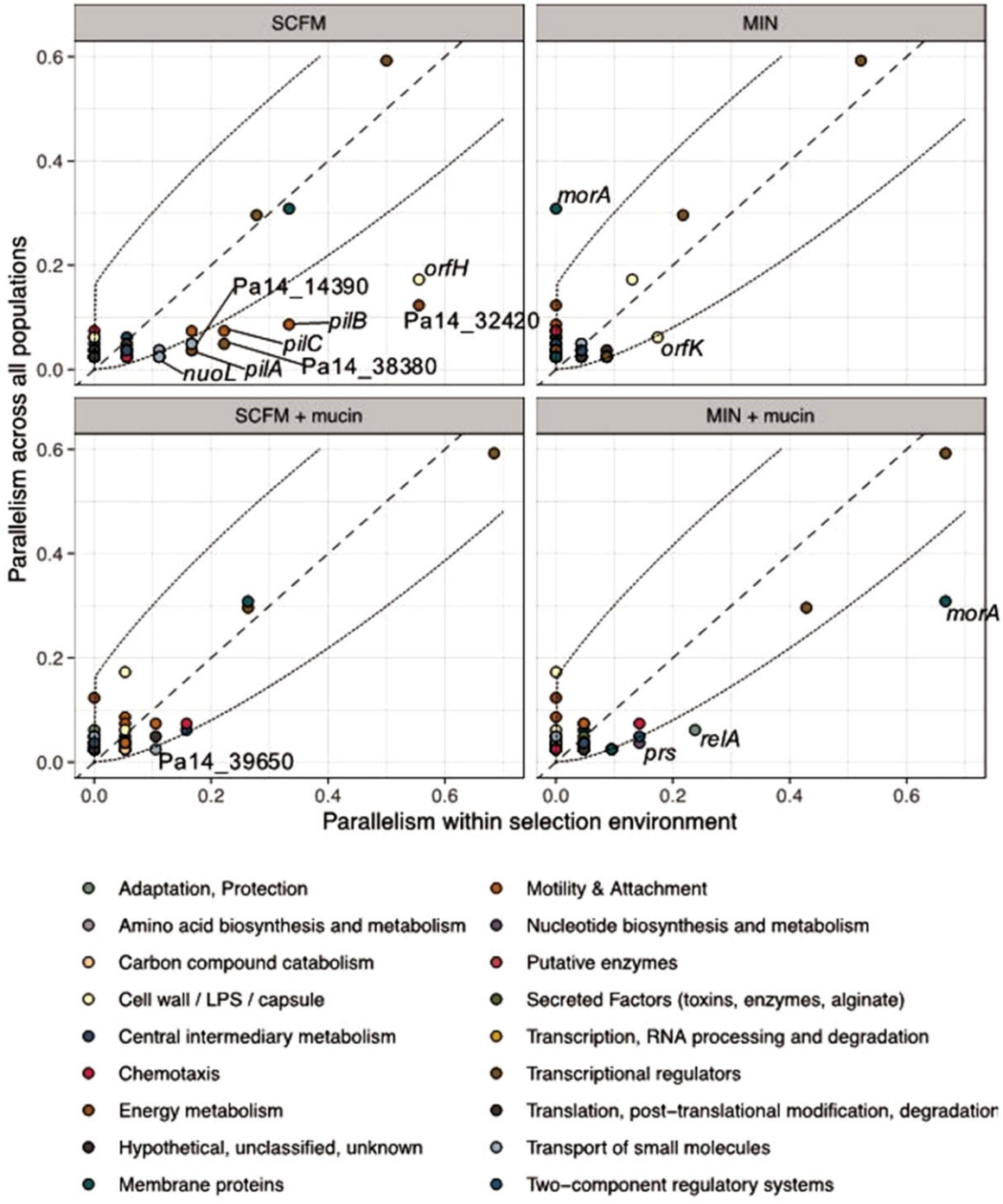

Fig. 4.—Gene-level parallelism. Parallelism within selection environment versus parallelism across all populations, for all genes with variants in more than one population, after excluding synonymous SNPs. Parallelism for a gene is defined as the proportion of populations with mutations in that gene. 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 labeled.

2012). Both the SCFM + mucin and MIN environments were found to have only a single gene mutated in more populations than expected (Pa14_39650 and *orfK*, respectively). Grouping genes by functional classification, assigned according to functional annotations for the ancestral strain *Pa14* using the PseudoCAP database from the *Pseudomonas*

genome project (Winsor et al. 2005, 2011, 2016) revealed no clear patterns. This result is unsurprising given the limited number of treatment-specific genes ($n = 13$) and the large number of functional classes. Together, these results suggest that, whereas some genes are repeatedly mutated in multiple replicate populations, others are not and so, parallel genetic evolution may not be dependable as a marker for CF-specific adaptation in this system.

Genomic 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 microbiota (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 (Smith et al. 2006; D'Argenio et al. 2007; Marvig et al. 2015; Freschi et al. 2018). Mutations in these genes have also been identified in other experimental studies (Wong et al. 2012; Azimi et al. 2020). Few genes, however, show strong associations with specific components of the CF lung, with the possible exception of *mexT*, which arose multiple times in SCFM, and *morA*, a motility regulator, 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 because 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, favored across a range of environmental conditions. 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 distinct 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; Wilderet al. 2009; Ciofu et al. 2010; LaFayette et al. 2015; Feltner et al. 2016). Although we do not know for certain the effect 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. Indeed, 97% of the observed mutations occur in a predicted ligand-binding site of the *lasR* gene (Bottomley et al. 2007), further indicating that these are likely loss-of-function mutants. That *lasR* mutations evolved repeatedly in both SCFM and minimal media suggest that growth advantages associated with amino acid metabolism are not solely 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* mutant may derive simply from reducing the metabolic cost of expression. This is a hypothesis that awaits further tests.

The second notable result 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 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 (Toprak et al. 2011; Wong et al. 2012; Graves et al. 2019; Azimi et al. 2020), and the genetic causes are often due to mutations in housekeeping genes like *rpoB* and *gyrA* (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.

Discussion

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 in each environment ($\sim 10^8$ CFU/ml) and consistent with that seen in other microbial evolution experiments (Wong et al. 2012; Schick et al. 2015; Good et al. 2017). Because population densities in our experiment are, by design, similar to those in chronically infected patients (Palmer et al. 2005; Stressmann et al. 2011), 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 coinfect a single patient (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). On average, populations evolved in the presence of mucin carried more mutations than populations not evolved with mucin. 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 (Schick and Kassen 2018). 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 toward 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 number 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 (*morA*, *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 laboratory 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 nonessential 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 distinct, 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 by no means capture all the dimensions of life in the CF lung; they have evolved in the absence of an active immune system, competing microbiota, 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 favor 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 dimensions, then these additional stressors could serve to preserve genetic diversity in the CF airway. An additional potential source of bias could come from sequencing whole populations from previously frozen stocks, as opposed to fresh ones, if some genotypes are more likely to recover more (or less) quickly after freezing. Although freezing can lead to the loss of very rare alleles, it has little effect on those above a frequency of ~3% of reads (Sprouffske et al. 2016) suggesting that, whereas we may have modestly underestimated the number and average heterozygosity in our populations, our inferences about treatment effects remain unchanged. Importantly, 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.

Materials and Methods

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, *P. aeruginosa* strain 14 (Pa14), were propagated in daily batch culture for ~220 generations in one of four environments: SCFM, 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.

Whole-Genome Sequencing

Populations were revived overnight from frozen in liquid Luria Bertani broth at 37 °C. 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 were performed by Genome Quebec at McGill University on the Illumina HiSeq 4000 platform, using paired-end sequencing of 2 × 100 base-pair reads.

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 five-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 *P. aeruginosa* reference genome UCBPP-PA14 109 (Winsor et al. 2016) using the bwa-mem algorithm of BWA version 0.7.12 (Li 2013) and Samtools version 1.3.1 (Li et al. 2009). Picard Tools version 2.9.2 (Broad Institute) was then used to mark PCR duplicates and add read group information. Variants were called using Breseq version 0.30.0 (Deatherage and Barrick 2014), a tool specifically designed for detecting mutations in microbial genomes in polymorphism mode with default parameters. 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 with 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 four populations to have evidence of cross-contamination due to the identical mutations present in them. Additionally, nine populations were found 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 (Sanders et al. 2006; Wiegand et al. 2008; Oliver and Mena 2010). Populations with evidence of cross-contamination and putative hypermutators present were excluded from subsequent analysis. Finally, mutations at frequencies at or <5% were manually checked for false positives and any false positives were also excluded from the analysis.

Statistical Analysis

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 nonneutral, 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 as depicted in Equation 1 below:

$$\text{Estimated proportion of driver nonsynonymous SNPs} = \frac{\text{Observed number of nonsynonymous SNPs} - (3 \times \text{Observed number of synonymous SNPs})}{\text{Total number of SNPs}}$$

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 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 \lt 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.

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 using a single-factor ANOVA. To quantify genetic divergence among evolved populations, we performed a 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 analog 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 as a measure of parallelism within treatment, discussed in the section below.

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 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 the standard error of dispersion. To determine the significance, we performed an ANOVA with Euclidean distance as the response variable and treatment as the explanatory variable. 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 signifying higher repeatability and therefore more parallelism. To determine the significance of both Jaccard and *C*-score metrics, we performed an exact test (permutation test) by randomizing treatment labels (number of permutations = 10,000) 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 <0.05, we considered that gene to be treatment specific.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This work was supported by a Natural Sciences and Engineering Research Council (Canada) Discovery Grant to R.K.

Data Availability

All sequence reads files are deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive, BioProject PRJNA823717.

Literature Cited

Agashe D, et al. 2016. Large-effect beneficial synonymous mutations mediate rapid and parallel adaptation in a bacterium. Mol Biol Evol. 33(6):1542–1553.

Ashish A, et al. 2013. Extensive diversification is a common feature of *Pseudomonas aeruginosa* populations during respiratory infections in cystic fibrosis. J Cystic Fibrosis 12(6):790–793.

Azimi S, et al. 2020. Allelic polymorphism shapes community function in evolving *Pseudomonas aeruginosa* populations. ISME J. 14(8): 1929–1942.

Bailey SF, Alonso Morales LA, Kassen R. 2021. Effects of synonymous mutations beyond codon bias: the evidence for adaptive synonymous substitutions from microbial evolution experiments. Genome Biol Evol. 13(9):evab141.

Bailey SF, 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 SF, Hinz A, Kassen R. 2014. Adaptive synonymous mutations in an experimentally evolved *Pseudomonas fluorescens* population. Nat Commun. 5:4076.
Behringer MG, et al. 2018. *Escherichia coli* cultures maintain stable subpopulation structure during long-term evolution. Proc Natl Acad Sci U S A. 115(20):E4642–E4650.
Blanton RE. 2018. Population genetics and molecular epidemiology of eukaryotes. Microbiol Spectr. 6(6):1–18.
Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics 30(15): 2114–2120.
Bottomley MJ, Muraglia E, Bazzo R, Carfì A. 2007. Molecular insights into quorum sensing in the human pathogen *Pseudomonas aeruginosa* from the structure of the virulence regulator LasR bound to its autoinducer. J Biol Chem. 282(18):13592–13600.
Breidenstein EBM, de la Fuente-Núñez C, Hancock REW. 2011. *Pseudomonas aeruginosa*: all roads lead to resistance. Trends Microbiol. 19(8):419–426.
Burrows Lori L. 2012. *Pseudomonas aeruginosa* twitching motility: type IV Pili in action. Annu Rev Microbiol. 66(1):493–520. doi:10.1146/micro.2012.66.issue-1.
Buskirk SW, Peace RE, Lang GI. 2017. Hitchhiking and epistasis give rise to cohort dynamics in adapting populations. Proc Natl Acad Sci U S A. 114(31):8330–8335.
Caballero JD, et al. 2015. Selective sweeps and parallel pathoadaptation drive *Pseudomonas aeruginosa* evolution in the cystic fibrosis lung. MBio 6(5):e00981-15.
Campos PRA, Neto PSCA, de Oliveira VM, Gordo I. 2008. Environmental heterogeneity enhances clonal interference. Evolution 62(6):1390–1399.
Ciofu O, Mandsberg LF, Bjarnsholt T, Wassermann T, Høiby N. 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(4):1108–1119.
Darch SE, et al. 2015. Recombination is a key driver of genomic and phenotypic diversity in a *Pseudomonas aeruginosa* population during cystic fibrosis infection. Sci Rep. 5:7649.
D'Argenio DA, et al. 2007. Growth phenotypes of *Pseudomonas aeruginosa* LasR mutants adapted to the airways of cystic fibrosis patients. Mol Microbiol. 64(2):512–533.
Deatherage DE, Barrick JE. 2014. Identification of mutations in laboratory evolved microbes from next-generation sequencing data using breseq. Methods Mol Biol. 1151:165–188.
Dettman JR, et al. 2012. Evolutionary insight from whole-genome sequencing of experimentally evolved microbes. Mol Ecol. 21: 2058–2077.
Dillon MM, Cooper VS. 2016. The fitness effects of spontaneous mutations nearly unseen by selection in a bacterium with multiple chromosomes. Genetics 204(3):1225–1238.
Ellis CN, Traverse CC, Mayo-Smith L, Buskirk SW, Cooper VS. 2015. Character displacement and the evolution of Niche complementarity in a model biofilm community. Evolution 69(2):283–293.
Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. Nat Rev Genet. 8(8):610–618.
Fargier E, et al. 2012. MexT functions as a redox-responsive regulator modulating disulfide stress resistance in *Pseudomonas aeruginosa*. J Bacteriol. 194(13):3502–3511.
Feltner JB, et al. 2016. LasR variant cystic fibrosis isolates reveal an adaptable quorum-sensing hierarchy in *Pseudomonas aeruginosa*. MBio 7(5):e01513-16.
Flynn KM, et al. 2016. Evolution of ecological diversity in biofilms of *Pseudomonas aeruginosa* by altered cyclic diguanylate signaling. J Bacteriol. 198(19):2608–2618.
Foweraker JE, Laughton CR, Brown DF, Bilton D. 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. J Antimicrob Chemother. 55(6):921–927.
Freschi L, 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 Microbiol Lett. 365(14):1–11.
Fung C, et al. 2010. Gene expression of *Pseudomonas aeruginosa* in a mucin-containing synthetic growth medium mimicking cystic fibrosis lung sputum. J Med Microbiol. 59(Pt 9): 1089–1100.
Futuyma DJ, Moreno G. 1988. The evolution of ecological specialization. Annu Rev Ecol Syst. 19:207–233.
Good BH, McDonald MJ, Barrick JE, Lenski RE, Desai MM. 2017. The dynamics of molecular evolution over 60,000 generations. Nature 551(7678):45–50.
Graves JL Jr, et al. 2019. Experimental evolution of gallium resistance in *Escherichia coli*. Evol Med Public Health 2019(1):169–180.
Hershberg R. 2017. Antibiotic-independent adaptive effects of antibiotic resistance mutations. Trends Genet. 33(8):521–528.
Hoffman LR, et al. 2009. *Pseudomonas aeruginosa* LasR mutants are associated with cystic fibrosis lung disease progression. J Cystic Fibrosis 8(1):66–70.
Kassen R. 2009. Toward a general theory of adaptive radiation: insights from microbial experimental evolution. Ann New York Acad Sci. 1168:3–22.
Kassen R. 2014. Experimental evolution and the nature of biodiversity. Greenwood Village (CO): Roberts.
Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF. 2011. Negative epistasis between beneficial mutations in an evolving bacterial population. Science (New York, N.Y.) 332(6034):1193–1196.
Kristofich J, et al. 2018. Synonymous mutations make dramatic contributions to fitness when growth is limited by a weak-link enzyme. PLoS Genet. 14(8):e1007615.
LaFayette SL, et al. 2015. Cystic fibrosis-adapted *Pseudomonas aeruginosa* quorum sensing LasR mutants cause hyperinflammatory responses. Sci Adv. 1(6):e1500199.
Lanfear R, Kokko H, Eyre-Walker A. 2014. Population size and the rate of evolution. Trends Ecol Evol. 29(1):33–41.
Leale AM, Kassen R. 2018. The emergence, maintenance, and demise of diversity in a spatially variable antibiotic regime. Evol Lett. 2(2): 134–143.
Lebeuf-Taylor E, McCloskey N, Bailey SF, Hinz A, Kassen R. 2019. The distribution of fitness effects among synonymous mutations in a gene under directional selection. ELife 8(July):e45952.
Li H, et al. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25(16):2078–2079.
Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv:1303.3997 [*q-Bio*], May. http://arxiv.org/abs/1303.3997.
Markussen T, et al. 2014. Environmental heterogeneity drives within-host diversification and evolution of *Pseudomonas aeruginosa*. MBio 5(5):e01592-14.
Martin G, Lenormand T. 2006. The fitness effect of mutations across environments: a survey in light of fitness landscape models. Evolution 60(12):2413–2427.
Marvig RL, Sommer LM, Molin S, Johansen HK. 2015. Convergent evolution and adaptation of *Pseudomonas aeruginosa* within patients with cystic fibrosis. Nat Genet. 47(1):57–64.
Mowat E, et al. 2011. *Pseudomonas aeruginosa* population diversity and turnover in cystic fibrosis chronic infections. Am J Respir Crit Care Med. 183(12):1674–1679.

Oliver A, Mena A. 2010. Bacterial hypermutation in cystic fibrosis, not only for antibiotic resistance. Clin Microbiol Infect. 16(7):798–808.
Palmer KL, Aye LM, Whiteley M. 2007. Nutritional cues control *Pseudomonas aeruginosa* multicellular behavior in cystic fibrosis sputum. J Bacteriol. 189(22):8079–8087.
Palmer KL, Mashburn LM, Singh PK, Whiteley M. 2005. Cystic fibrosis sputum supports growth and cues key aspects of *Pseudomonas aeruginosa* physiology. J Bacteriol. 187(15):5267–5277.
Park S-C, Krug J. 2007. Clonal interference in large populations. Proc Natl Acad Sci U S A. 104(46):18135–18140.
Broad Institute. n.d. 'Picard Tools'. http://broadinstitute.github.io/picard/.
Poltak SR, Cooper VS. 2011. Ecological succession in long-term experimentally evolved biofilms produces synergistic communities. ISME J. 5(3):369–378.
Poole K. 2005. Efflux-mediated antimicrobial resistance. J Antimicrob Chemother. 56(1):20–51.
Rajan S, Saiman L. 2002. Pulmonary infections in patients with cystic fibrosis. Semin Respir Infect. 17(1):47–56.
Sanders LH, Rockel A, Lu H, Wozniak DJ, Sutton MD. 2006. Role of *Pseudomonas aeruginosa* DinB-encoded DNA polymerase IV in mutagenesis. J Bacteriol. 188(24):8573–8585.
Schaedel C, et al. 2002. Predictors of deterioration of lung function in cystic fibrosis*: CFTR genotype in ung function in CF. Pediatric Pulmonol. 33(6):483–491.
Schick A, Bailey SF, Kassen R. 2015. Evolution of fitness trade-offs in locally adapted populations of *Pseudomonas fluorescens*. Am Nat. 186(S1):S48–S59.
Schick A, Kassen R. 2018. Rapid diversification of *Pseudomonas aeruginosa* in cystic fibrosis lung-like conditions. Proc Natl Acad Sci U S A. 115(42):10714–10719.
Schluter D. 2000. Ecological character displacement in adaptive radiation. Am Nat. 156:S4–S16.
Smith EE, et al. 2006. Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. Proc Natl Acad Sci U S A. 103(22):8487–8492.
Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. Genet Res. 23(1):23–35.
Sprouffske K, Aguilar-Rodríguez J, Wagner A. 2016. How archiving by freezing affects the genome-scale diversity of *Escherichia coli* populations. Genome Biol Evol. 8(5):1290–1298.
Sriramulu DD, Lünsdorf H, Lam JS, Römling U. 2005. Microcolony formation: a novel biofilm model of *Pseudomonas aeruginosa* for the cystic fibrosis lung. J Med Microbiol. 54(7):667–676.
Stressmann FA, et al. 2011. Does bacterial density in cystic fibrosis sputum increase prior to pulmonary exacerbation?' J Cystic Fibrosis 10(5):357–365.
Toprak E, et al. 2011. Evolutionary paths to antibiotic resistance under dynamically sustained drug stress. Nat Genet. 44(1):101–105.
Traverse CC, Mayo-Smith LM, Poltak SR, Cooper VS. 2013. Tangled bank of experimentally evolved *Burkholderia* biofilms reflects selection during chronic infections. Proc Natl Acad Sci U S A. 110(3):E250–E259.
Turner CB, Marshall CW, Cooper VS. 2018. Parallel genetic adaptation across environments differing in mode of growth or resource availability. Evol Lett. 2(4):355–367.
Wiegand I, et al. 2008. Mutator genes giving rise to decreased antibiotic susceptibility in *Pseudomonas aeruginosa*. Antimicrob Agents Chemother. 52(10):3810–3813.
Wilder CN, Allada G, Schuster M. 2009. Instantaneous within-patient diversity of *Pseudomonas aeruginosa* quorum-sensing populations from cystic fibrosis lung infections. Infect Immunity 77(12):5631–5639.
Winsor GL, et al. 2005. *Pseudomonas aeruginosa* genome database and pseudoCAP: facilitating community-based, continually updated, genome annotation. Nucl Acids Res. 33(Database issue):D338–D343.
Winsor GL, et al. 2011. Pseudomonas genome database: improved comparative analysis and population genomics capability for pseudomonas genomes. Nucl Acids Res. 39(Database issue):D596–D600.
Winsor GL, et al. 2016. Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the Pseudomonas genome database. Nucl Acids Res. 44(D1):D646–D653.
Wong A, Rodrigue N, Kassen R. 2012. Genomics of adaptation during experimental evolution of the opportunistic pathogen *Pseudomonas aeruginosa*. PLoS Genet. 8(9):e1002928.
Workentine ML, et al. 2013. Phenotypic heterogeneity of *Pseudomonas aeruginosa* populations in a cystic fibrosis patient. PLoS ONE 8(4):e60225.
Wright EA, Fothergill JL, Paterson S, Brockhurst MA, Winstanley C. 2013. Sub-inhibitory concentrations of some antibiotics can drive diversification of *Pseudomonas aeruginosa* populations in artificial sputum medium. BMC Microbiol. 13:170.
Yang L, et al. 2011. Evolutionary dynamics of bacteria in a human host environment. Proc Natl Acad Sci U S A. 108(18):7481–-7486.Yeaman S, Gerstein AC, Hodgins KA, Whitlock MC. 2018. Quantifying how constraints limit the diversity of viable routes to adaptation. PLoS Genet. 14(10):e1007717.

Associate editor: Kenneth Wolfe