Fitness is strongly influenced by rare mutations of large effect in a microbial mutation accumulation experiment

Karl Heilbron, Macarena Toll-Riera, Mila Kojadinovic and R. Craig MacLean

Department of Zoology, University of Oxford, Oxford, Oxfordshire, United Kingdom, OX13PS

Short title: Fitness decay and selection during MA

Keywords: Pseudomonas aeruginosa, hypermutator, whole genome re-sequencing, spontaneous mutation, experimental evolution

Corresponding author:
Karl Heilbron

Department of Zoology
The Tinbergen Building
South Parks Road
Oxford, UK
OX1 3PS

Tel: +44 1865 271265
Email: karl.heilbron@zoo.ox.ac.uk

Copyright 2014.
Abstract

Our understanding of the evolutionary consequences of mutation relies heavily on estimates of the rate and fitness effect of spontaneous mutations generated by mutation accumulation (MA) experiments. We performed a classic MA experiment in which frequent sampling of MA lines was combined with whole genome re-sequencing to develop a high-resolution picture of the effect of spontaneous mutations in a hypermutator (ΔmutS) strain of the bacterium Pseudomonas aeruginosa. After ~644 generations of mutation accumulation, MA lines had accumulated an average of 118 mutations, and we found that average fitness across all lines decayed linearly over time. Detailed analyses of the dynamics of fitness change in individual lines revealed that a large fraction of the total decay in fitness (42.3%) was attributable to the fixation of rare, highly deleterious mutations (comprising only 0.5% of fixed mutations). Furthermore, we found that at least 0.64% of mutations were beneficial and probably fixed due to positive selection. The majority of mutations that fixed (82.4%) were base substitutions and we failed to find any signatures of selection on non-synonymous or intergenic mutations. Short indels made up a much smaller fraction of the mutations that were fixed (17.4%), but we found evidence of strong selection against indels that caused frameshift mutations in coding regions. These results help to quantify the amount of natural selection present in microbial MA experiments and demonstrate that changes in fitness are strongly influenced by rare mutations of large effect.

Introduction

Mutations are the ultimate source of genetic variation which natural selection acts upon. Understanding the rate at which mutations arise and the distribution of fitness effects of spontaneous mutations is therefore of central importance to the study of evolutionary biology.
One of the most widely used methods for determining the rate and fitness effect of spontaneous mutations is the mutation accumulation (MA) experiment. Following the pioneering work of Bateman (Bateman 1959) and Mukai (Mukai 1964), MA experiments involve propagating many replicate lines at very small effective population sizes so that the effect of natural selection is swamped out by that of genetic drift, allowing weakly selected mutations to accumulate randomly. The decline in mean fitness and increase in among-line variance in fitness are then used to indirectly infer mutation rate and effect estimates (Bateman 1959; Mukai 1964; Keightley 1994; García-Dorado 1997; Shaw et al. 2002).

Recently, whole genome re-sequencing of MA lines has been used to directly measure the mutation rate in microorganisms (Lynch et al. 2008; Lee et al. 2012; Ness et al. 2012; Sung et al. 2012a; Sung et al. 2012b; Long et al. 2013). In line with classic mutation rate estimates from reporter gene assays, the emerging consensus is that the genomic mutation rate is remarkably constant across DNA-based microbes, approximately $3 \times 10^{-3}$ mutations/genome/generation (Drake 1991; Lynch 2010). Accurate estimates of the fitness effects of spontaneous mutation, however, have remained elusive (Eyre-Walker and Keightley 2007; Halligan and Keightley 2009).

Because MA experiments rely on making comparisons among lines, they have traditionally focused on studying how fitness changes across as many lines as possible. An alternative approach is to combine whole genome re-sequencing in a smaller number of MA lines of a hypermutator strain in order to allow a greater number of mutations to accumulate, thus increasing our ability to detect and quantify the amount of natural selection that occurs during microbial mutation accumulation experiments. Furthermore, whole genome re-sequencing directly determines the average number of mutations that accumulate between fitness measurements, allowing for improved estimates of the distribution of fitness effects of spontaneous mutations.
Natural selection must occur to some extent during microbial mutation accumulation experiments because colonies must grow big enough to become visible, resulting in an effective population size ($N_e$) greater than 1. Beneficial and deleterious mutations should be subject to effective selection when $N_e s > 1$, where $s$ is the absolute value of the fitness effect of the mutation, and the fluctuating population size of microbial MA experiments may further increase the efficacy of selection (Otto and Whitlock 1997). This may explain why many microbial MA experiments have reported results that are consistent with the fixation of some beneficial mutations as a result of positive selection (Shaw et al. 2000; Joseph and Hall 2004; Perfeito et al. 2007; Dickinson 2008; Trindade et al. 2010; Stevens and Sebert 2011). Studies have begun to combine both MA and whole genome re-sequencing in microorganisms (Lynch et al. 2008; Lee et al. 2012; Ness et al. 2012; Sung et al. 2012a; Sung et al. 2012b; Long et al. 2013), but none have detected a genomic signature of natural selection.

Using detailed fitness measurements and whole genome re-sequencing, we studied the evolutionary dynamics of 8 replicate mutation accumulation lines of a hypermutator strain of the pathogenic bacterium *P. aeruginosa*. MA lines were passaged through 28 single-cell bottlenecks followed by rapid population growth over a period of ~644 generations. Under this regime, we estimate that the effective population size of MA lines had a lower limit of approximately 16, which should be sufficient to prevent natural selection on the vast majority of spontaneous mutations. We determined the evolutionary dynamics of our lines with a high degree of precision by: 1) directly measuring competitive fitness instead of a component of fitness such as growth rate, and 2) measuring fitness at every second bottleneck to capture a small number of mutations between each time point. In line with recent work, we used deep whole genome sequencing to determine the genetic consequences of population bottlenecking, infer the molecular basis of altered fitness, and test for genomic signatures of natural selection during the MA procedure.
Consistent with previous MA experiments, we found that mean fitness decayed linearly over time. Detailed trajectories of fitness in individual lines coupled to whole genome sequencing revealed that rare, strongly deleterious mutations account for nearly half of the total loss of fitness. Furthermore, we found that positive selection resulted in the fixation of beneficial mutations, and that purifying selection was able to remove the majority of frameshift mutations.

Results

Here we present the results from a ~644 generation long mutation accumulation experiment in 8 replicate MA lines. We measured the fitness of each MA line every two days, providing a high-resolution picture of the evolutionary dynamics of heavily bottlenecked bacterial populations. We performed whole genome re-sequencing on multiple time points of each line in order to determine the molecular nature of mutations fixed under conditions of relaxed natural selection.

Whole genome re-sequencing identified 944 mutations in the 8 mutation accumulation lines. Sanger sequencing of a random sample of these mutations confirmed 35/35 mutations (Table S1), indicating a very low false positive rate. As expected, mutations were Poisson distributed across MA lines (One-sample Kolmogorov-Smirnoff test: $P = 0.521$, $D = 0.270$) with an average of 118 mutations fixed per line and an average of 8.4 mutations fixed between each adjacent time point. This equates to a per base pair mutation rate of $2.95 \pm 0.21$ SE $\times 10^{-8}$ mutations/site/generation and a genomic mutation rate of $0.18 \pm 0.01$ SE mutations/genome/generation. Given that the hypermutator strain used in this study increases the mutation rate by approximately 70-fold (TORRES-BARCELO et al. 2013), this estimated genomic mutation rate is in line with the consensus bacterial genomic mutation rate of approximately $3 \times 10^{-3}$ mutations/genome/generation (DRAKE 1991; LYNCH 2010).
Of the 944 mutations, 778 (82.4%) were base substitutions, 164 (17.4%) were short indels (<10 bp), and 2 (0.2%) were large structural variations, consisting of a partial gene duplication event (pvdD) and a 1,880bp intergenic deletion. Insertions were approximately 2.5-fold more common than deletions (118 insertions versus 46 deletions) (Figure 1). As is typical for a ΔmutS hypermutator strain, almost all base substitutions were transitions (774/778 = 99.5%), and G:C → A:T transitions (478) were approximately 60% more common than A:T → G:C transitions (298).

As expected, the average fitness of the hypermutator populations decreased significantly over time (Figure 2; ANOVA: P = 1.68 × 10^-6, F_{1,13} = 67.409), indicating that the average effect of spontaneous mutations was deleterious and that recurrent population bottlenecks inhibited the action of natural selection (mean mutational fitness effect = -0.16%). In fact, in some lines fitness became so low that it was no longer possible to reliably measure (Figure 3). These data are included in Figure 2 to prevent bias, but excluded from subsequent analyses. The average fitness of bottlenecked non-hypermutator control lines did not change significantly over the course of the experiment (ANOVA: P = 0.712, F_{1,118} = 0.137), indicating that the loss of fitness in hypermutator lines was due to mutation accumulation.

**Fitness data**

Unlike the linear decrease observed for average fitness, the evolutionary trajectory of individual lines was much more complex (Figure 3). The net change in the fitness of MA lines ranged from -1% to -27% (mean: -13% ± 9 SD). A large portion of the net decrease in fitness of each line was due to a single drop between adjacent time points (we hereafter refer to a pair of adjacent time points as a “step”). Specifically, on average, 42.2% (± 12.9% SD) of the total decrease in fitness between the first and last time point in an individual MA line (excluding any beneficial steps) was due to the largest deleterious step in that line. Furthermore, the 4 most deleterious steps across all lines accounted for 42.3% of the total fitness decrease throughout the entire experiment. To
determine whether these large drops in fitness were caused by 1) the accumulation of a greater
number of mutations than other steps, or 2) the accumulation of mutations of larger effect, we
performed whole genome sequencing on the four largest deleterious steps across all MA lines, as
well as on an exceptionally large deleterious step which caused the fitness of its MA line to drop to
an undetectable level. These steps did not contain a significantly greater number of mutations than
the remaining steps (mean of 5 largest steps: 9.0 mutations, mean of remainder: 7.9 mutations,
paired \( t \)-test: \( P = 0.285, t_{4} = 1.235 \)). However, these large deleterious steps showed a significantly
higher frequency of mutations in highly conserved core genes than other steps (\( \chi^{2} \) goodness-of-fit
test: \( P = 0.049, \chi^{2}_{1} = 3.882 \); Table S2). Therefore, large drops in fitness are due to mutations in more
important genes rather than due to a greater number of mutations.

Although the average fitness effect of a step was deleterious, there were numerous steps in
which fitness increased (Figure 4). To confirm the presence of steps containing beneficial
mutations, we repeated the competitive fitness assays for the 11 steps with the largest increases in
fitness. Even after false discovery rate correction (Benjamini and Hochberg 1995), fitness increased
significantly (\( P < 0.05 \)) in 6/89 (6.7%) of the measurable steps. Because steps where fitness
increased were rare, it is likely that each of these steps only contained a single beneficial mutation.
This implies that at least 6 beneficial mutations were fixed during the mutation accumulation
experiment, which corresponds to 0.64% of all mutations that were fixed during the experiment.

Signatures of natural selection

Selection on base substitutions in protein coding genes

The vast majority of protein-altering base substitutions were non-synonymous mutations,
but the ratio of the rate of non-synonymous mutations to silent mutations (\( dN/dS = 1.08 \)) did not
differ significantly from the neutral expectation of 1 (Table 1; \( Z \)-test: \( Z = 0.92, P = 0.26 \)). We observed
only a single loss-of-stop mutation, but this was similar to our predicted number of 1.4. Truncation
mutations that introduce a premature stop codon were much more frequent (n= 14), but this was not significantly different from the neutral expectation of 9 truncation mutations (Z-test: Z= 1.63, \( P= 0.10 \)).

Selection on coding and non-coding regions

Protein coding sequence accounts for 89.4% of the \textit{P. aeruginosa} genome and so we expected that if no natural selection has occurred during the MA experiment then approximately 89.4% of mutations will have occurred in protein coding sequences. We found that the percentage of mutations (short indels and base substitutions) that occurred in coding regions was 85.4% (804/942), which was significantly different from the neutral expectation of 89.4% (\( \chi^2 \) goodness-of-fit test: \( P< 0.001 \), \( \chi^2_1 = 15.888 \)). Interesting patterns arose when we analyzed base substitutions and short indels separately.

We found that the percentage of base substitutions in coding regions (89.6%, 697/778) and intergenic regions (10.41%, 81/778) was not significantly different from the neutral expectation (\( \chi^2 \) goodness-of-fit test: \( P= 0.833 \), \( \chi^2_1 = 0.045 \)). This result may be confounded because intergenic regions contain a larger proportion of repetitive DNA than coding regions (intergenic: 7.2%, coding: 3.1%), but when we restricted our analysis to repetitive regions we still observed that the percentage of base substitutions that fell in coding (4.3%) and intergenic (16.5%) repetitive regions did not differ from the neutral expectation (\( \chi^2 \) goodness-of-fit test: \( P= 0.181 \), \( \chi^2_1 = 1.791 \)).

Selection on indels

In contrast to base substitutions, we found significantly fewer indels in coding regions than expected (observed: 107/164= 65.2%; expected: 89.4%; \( \chi^2 \) goodness-of-fit test: \( P< 0.0001 \), \( \chi^2_1 = 100.236 \)). Again, this difference could be confounded because intergenic regions contain a larger proportion of indel-prone repetitive DNA, but we also found significantly fewer indels in repetitive
coding regions (observed: 103/160 = 64.4%; expected: 78.2%; $\chi^2$ goodness-of-fit test: $P<0.0001$, $\chi^2_{1}=17.920$) than expected in the absence of selection. This indicates strong purifying selection against frameshift mutations. In fact, these data suggest that at least 49.6% of frameshift mutations are sufficiently deleterious to be removed by natural selection, even under a regime of intense bottlenecking. Despite selection against frameshift mutations we still found 106 frameshifts in our experiment. Almost all of them (101/106 = 95.3%) overlapped with homopolymeric tracts of C (ranging from 4C to 8C) or G (ranging from 5G to 8G). There were significantly more frameshifts located near the N-terminus of the protein than expected given the distribution of homopolymeric tracts in the *P. aeruginosa* genes (Figure 5; one-sided exact binomial test: $P=0.037$). We found no significant difference for frameshifts near the middle (one-sided exact binomial test: $P=0.453$) or near the C-terminus of the protein (one-sided exact binomial test: $P=0.063$).

*Tests for parallel evolution*

Previous work has shown that exposing replicate microbial populations to a similar selective pressure results in parallel adaptation at a molecular level in both lab experiments ([Wichman et al. 2000; Segrè et al. 2006; Barrick et al. 2009] and clinical populations ([Huse et al. 2010; Lieberman et al. 2011]). To test for parallel evolution at the level of individual genes, we compared the distribution of the number of mutations fixed per gene in the 8 MA lines with the distribution expected based on the lengths of the genes in the *P. aeruginosa* genome (Figure S1; see Materials and Methods for details on calculating the expected distribution). We found no deviation from the expected distribution for synonymous mutations ($\chi^2$ goodness-of-fit test: $P=0.643$, $\chi^2_{2}=0.883$). On the other hand, we found significantly fewer parallel non-synonymous mutations than expected ($\chi^2$ goodness-of-fit test: $P<0.0001$, $\chi^2_{2}=19.302$), which does not support the hypothesis that natural selection was capable of causing parallel evolution on the genomic scale in these MA lines. Rather, longer genes simply had more mutations than smaller genes.
(Figure S2): genes with one or more or mutations were significantly longer than genes without mutations (Kolmogorov-Smirnov test, $P < 0.001$).

It is also possible that parallel evolution could act on levels higher than the gene. We analyzed our mutation data for evidence of over- or under-enrichment of mutations in Clusters of Orthologous Groups (COGs)—genes which share a common function. After false discovery rate correction (Benjamini and Hochberg 1995), we found a significant underrepresentation of mutated genes involved in transcription (Table S3; Fisher's exact test: $P = 0.023$, Fisher's odds ratio $= 0.530$), suggesting that mutations in these genes tend to have highly deleterious effects.

Core genes

We observed that large drops in fitness during the MA experiment were associated with the accumulation of mutations in core genes (Figure 2), and so we sought to determine whether natural selection was effective against mutations in these genes. Surprisingly, there was no significant underrepresentation of mutations in core genes (Fisher’s exact test: $P = 0.611$, Fisher's odds ratio $= 1.051$) despite their potentially large deleterious effects on fitness.

Discussion

Mutations are rare events that often lead to small changes in fitness, and these properties of mutations make it intrinsically difficult to directly study the evolutionary consequences of mutation. Our experiment, which combined a classic mutation accumulation experiment with powerful whole genome re-sequencing technology, found that 42.3% of the decrease in fitness in our lines was driven by 4.5% of the steps with highly deleterious effects on fitness. Given the rarity of large drops in fitness, the most parsimonious explanation is that each one of these drops was driven by a single highly deleterious mutation. Under this assumption, the 42.3% of the decrease in
fitness in our experiment was driven by 0.5% of the mutations fixed, which is consistent with previous work in *Caenorhabditis elegans* (DAVIES et al. 1999). The mean mutational effect, $s = -1.6 \times 10^{-3}$, is similar to previous work in *Saccharomyces cerevisiae* ($s = -6 \times 10^{-3}$) in which whole genome re-sequencing and MA were combined (LYNCH et al. 2008) and, as expected, is approximately 1-2 orders of magnitude smaller than previous microbial MA studies that did not use whole genome re-sequencing and were therefore unable to detect neutral mutations (HALLIGAN AND KEIGHTLEY 2009; TRINDADE et al. 2010). We also found evidence of both positive and negative selection in our MA experiment, demonstrating that the results of our experiment cannot be interpreted as a proxy for the effects of spontaneous mutation alone.

### Beneficial mutations

Previous studies in *Arabidopsis thaliana* (SHAW et al. 2000), *Escherichia coli* (PERFEITO et al. 2007; TRINDADE et al. 2010), *Streptococcus pneumoniae* (STEVENS AND SEBERT 2011), and *S. cerevisiae* (JOSEPH AND HALL 2004; DICKINSON 2008) have also found evidence that beneficial mutations are fixed during mutation accumulation experiments. Our experimental approach allowed us to experimentally demonstrate that it is highly likely that at least 0.64% of the mutations that fixed during our MA experiment were beneficial. For these mutations to have been fixed by drift, the beneficial mutation rate in a non-hypermutator population with a genomic mutation rate of $3 \times 10^{-3}$ mutations/genome/generation would have to have been approximately $5 \times 10^{-6}$ mutations/genome/generation, which is 2-3 orders of magnitude higher than existing estimates (GERRISH AND LENSKI 1998; MIRALLES et al. 1999; IMHOF AND SCHLOTTERER 2001; ROZEN et al. 2002; BARRETT et al. 2006) (but for exceptions see (PERFEITO et al. 2007)). Instead, we argue that positive selection was able to drive the fixation of beneficial mutations in our experiment. Consistent with
this idea, 5 out of 6 of the significantly beneficial mutations that fixed were sufficiently beneficial that $N_e s$ was greater than 1.

**Tests for selection at a molecular level**

In agreement with recent microbial mutation accumulation experiments that have used whole genome re-sequencing, we found no evidence of selection on base substitutions, including non-synonymous mutations (Lynch et al. 2008; Lee et al. 2012; Ness et al. 2012; Sung et al. 2012a; Sung et al. 2012b; Long et al. 2013). Additionally, we found no evidence of positive selection on the same genes in different MA lines. Surprisingly, we found that non-synonymous mutations in highly conserved core genes can have strong deleterious effects on fitness (Figure 2), and yet we found no evidence that these mutations were removed by natural selection. The most striking evidence of selection at a genetic level comes from the lack of short indel mutations in coding regions. We estimate that negative selection prevented the fixation of at least 50% of indels in coding regions. In contrast, we did not find any evidence of an underrepresentation of base substitutions that generated a premature stop codon, implying that the absence of indels in coding regions is due to selection against frameshifts, and not selection against gene loss.

Despite strong selection, we still found that frameshifts comprise 13.2% of all mutations in coding regions. This high incidence of frameshifting could be because 95.3% of frameshifts overlapped with homopolymeric tracts. Homopolymeric tracts are hypermutable: they are highly prone to gaining or losing repeats through slippage, thereby producing indels. Consistent with recent work (Orsi et al. 2010; Lin and Kusell 2012), we found a significant overrepresentation of frameshifts at the 5’ end of genes and underrepresentation at the 3’ end (given the distribution of homopolymeric tracts in the PAO1 genome). Although the reasons for the enrichment of 5’
frameshifts is unclear, possible explanations include: 1) 5′ frameshifts tend to create shorter proteins and thus may be less prone to forming toxic aggregations; 2) intergenic regions in *P. aeruginosa* are very short and 3′ indels may knock out downstream genes; and/or 3) 5′ indels are more likely to destroy gene function, which may be beneficial in some circumstances. For example, Moxon and colleagues (Moxon *et al.* 2006) have proposed that simple sequence repeats (such as homopolymeric tracts) are localized hypermutation targets and a mechanism for adaptation. Moreover, standing genetic variation in homopolymeric tracts has been shown to drive the adaptation of *Campylobacter jejuni* to a novel host (Jerome *et al.* 2011).

**Implications for mutation accumulation experiments**

It is important to emphasize that our experiment differed from most previous MA experiments because we used a hypermutator strain. To what extent is this likely to have biased our results? Hypermutators produce an altered spectrum of spontaneous mutations (e.g. bias towards transitions), which can have important evolutionary implications when strong selection acts on a small number of sites in the genome (Couce *et al.* 2013) (e.g. some cases of high-level antibiotic resistance). In our system, frameshifts experienced much stronger selection than any other class of mutation, and it is possible that using a ΔmutS hypermutator altered the rate of appearance of indel mutations relative to base substitutions (Marvig *et al.* 2013). However, by using a hypermutator we were able to detect a sufficiently large number of mutations to analyze the effects of relatively rare types of mutation, such as indels, which have traditionally been overlooked in MA studies.

**Conclusion**
In conclusion, we find that fitness decays in recurrently bottlenecked populations of hypermutator *P. aeruginosa* because of the fixation of many weakly deleterious mutations and a few highly deleterious mutations. We argue that this pattern of punctuated decay of fitness arises for two reasons. First, most mutations carry little, if any, fitness cost in a laboratory environment, but a substantial fraction of mutations are highly deleterious. Our results suggest that weakly deleterious mutations tend to be intergenic and non-synonymous mutations, while highly deleterious mutations tend to be indels and mutations in core genes. Second, we find that recurrent bottlenecking does not completely compromise the efficacy of natural selection in microbial mutation accumulation experiments, although large deleterious mutations are unlikely to play a substantial role in the evolution of natural populations. We hope that this study will pave the way for future work aimed at understanding: 1) why frameshift mutations are subject to such strong selection, 2) how bacteria adapt to the deleterious effects of spontaneous mutations and 3) how the molecular basis of spontaneous mutation is linked to the fitness effects of mutations in natural populations.

### Materials and Methods

#### Strains

The eight replicate clones used in this study were founded from the *P. aeruginosa* hypermutator strain PAO1ΔmutS, which was created by replacing *mutS*—part of the methyl-directed mismatch repair pathway—with the antibiotic resistance marker *aac1* using the Cre-lox system for gene deletion and antibiotic resistance marker recycling following the methods of Mandsberg *et al.* (Mandsberg *et al.* 2011). Deleting *mutS* increases the mutation rate by ~70 fold in *P. aeruginosa* (Torres-Barcelo *et al.* 2013), primarily by increasing the rate of transitions (Miller...
The reference strain used to assess competitive fitness was PAO1-GFP. This strain was generated by integrating a constitutively expressed GFP marker at the chromosomal tn7 insertion site in *P. aeruginosa* PAO1 using the methods of Choi and Schweizer (Choi and Schweizer 2006).

**Mutation accumulation**

Eight replicate mutation accumulation lines were generated by streaking randomly selected colonies of PAO1ΔmutS onto individual M9KB agar plates (glycerol: 10 g/L, peptone: 20 g/L, M9 salts: 10.5 g/L, agar: 12 g/L, MgSO₄: 2 mL/L). Plates were incubated at 37°C for 18 hours before repeating the process of picking a random colony and streaking it on a fresh plate. This process was repeated daily for 30 days. Each day colonies would form from a single cell which had doubled ~23 times, resulting in an *N* of ~16. Every second day a portion of the randomly selected colony was suspended in a 50% v/v solution of glycerol and frozen at -80°C to be stored for competition assays. To ensure random selection of colonies, the last colony of the streak which was not touching another colony was selected. It is unlikely that random colony selection suffered a detection bias due to missing extremely small colonies; we sampled 14 regions between the visible colonies of our streaked plates and re-streaked them, but did not detect a single instance of colony growth after 10 days.

**Competitive fitness assay**

Fitness of each line at each time point was determined relative to the PAO1-GFP strain. Strains were precultured in M9KB medium from frozen 50% glycerol stocks. Overnight cultures of each strain were mixed in M9KB broth at a ratio of approximately 80% mutant to 20% PAO1-GFP. The exact initial proportions were confirmed via flow cytometry. Mixtures were competed for 18
hours at 37°C, with agitation at 200 RPM, and the final proportion was again measured by flow
cytometry. We define the relative fitness of the mutant as the number of doublings that the mutant
strain undergoes during the 18 hour competition divided by the number of doublings of the wild-
type strain, given by the formula:

\[
    w_{\text{mutant}} = \frac{\log_2 \left( \frac{N_{\text{final,mutant}}}{N_{\text{initial,mutant}}} \right)}{\log_2 \left( \frac{N_{\text{final,wild-type}}}{N_{\text{initial,wild-type}}} \right)}
\]

where \( w_{\text{mutant}} \) is the fitness of the mutant relative to the wild-type and \( N_{ij} \) is the number of either
the mutant or the wild-type cells at either the beginning or the end of the competition. Each
competition assay was performed in two experimental blocks with 3 replicate competitions per
block. In some mutation accumulation lines, fitness became too low to accurately measure (final
mutant proportion <10%) and thus these data have been excluded from all analyses except those
pertaining to Figure 2 and the decay in average fitness over time. The inclusion of these inaccurate
points does not change the statistical significance of any of the results presented.

Flow cytometry

Flow cytometry was used to determine the relative proportions of mutant and wild-type
strains at the beginning and end of the competitive fitness assays. Bacterial cultures were diluted
200-fold in sterile filtered M9 salts that was prepared using deionized water to minimize
background signal in the flow cytometer. Diluted mixtures were run on an Accuri C6 Flow
Cytometer Instrument (BD Accuri, San Jose, CA, USA) until 10,000 cells had been assayed. Events
with a forward scatter value less than 10,000 or a side scatter value less than 8,000 were excluded
to prevent the false detection of small particles in the medium and electrical noise. To discriminate
between GFP-tagged and untagged cells, cells were excited at a wavelength of 488nm and fluorescence emissions between 518-548nm were measured. There was a small overlap in the fluorescence profiles of tagged and untagged cells (i.e. the most fluorescent untagged cells were slightly more fluorescent than the least fluorescent GFP-tagged cells), so pure cultures of PAO1 and PAO1-GFP were used as controls to correct for such spillover.

Whole genome sequencing

Illumina whole genome sequencing was performed on the first and last time point of each line, as well as on the 5 pairs of adjacent time points which showed the largest decrease in fitness. Raw sequencing data was analyzed using an in-house pipeline. Briefly, raw reads were filtered using the NGS QC Toolkit (Patel and Jain 2012) and aligned against the reference genome using BWA (Li and Durbin 2009). Two approaches were used to call variants, GATK’s Unified Genotyper (DePristo et al. 2011) and Samtools’s mpileup (Li et al. 2009). Identified variants were annotated with Snpeff (Cingolani et al. 2012). To detect structural variants we combined two algorithms, Breakdancer (Chen et al. 2009) and Pindel (Ye et al. 2009). Finally, copy number variants (CNVs) were detected using Control-FREEC (Boeva et al. 2012).

All differences between the P. aeruginosa PAO1 reference genome and the first time point of each bacterial line were excluded, leaving only mutations that accumulated throughout the experiment. Sequences from intermediate time points were treated as sequences from end points. All mutations found in intermediate time-points were found at the end points except for one that fell in a mutation hotspot.

Testing for selection on base substitutions
To test for selection on base substitutions in protein coding genes we estimated the expected number of protein altering mutations, under the assumption that synonymous mutations are effectively neutral. Specifically, since almost all base substitutions in our experiment were transitions (99.5%), we calculated the neutral mutation rate of each of the four bases to its partner (A→G, G→A, C→T, T→C) using the observed synonymous mutations in our experiment. Given these mutation rates, we used the nucleotide composition and codon usage of *P. aeruginosa* proteins to estimate the rates of non-synonymous and synonymous mutations (dN/dS ratio), as well as the rates of stop-gain, stop-loss, and intergenic mutations. To test for a deviation from the neutral expectation, we tested the null hypothesis that the proportion of mutations in a given class (non-synonymous, truncation or intergenic) relative to the number of observed synonymous mutations is equal to the predicted ratio calculated using the synonymous mutation rate. This hypothesis was tested using the normal approximation of the binomial distribution (ZAR 2010).

**Repetitive regions**

The RepeatMasker program (SMIT *et al.* 1996-2010) was used to screen the PAO1 genome for simple repeats, interspersed repeats, and low complexity DNA sequences. Homopolymeric tracts of single nucleotide repeats ranging from 4 to 20 bases were identified using the *dreg* program, implemented in the EMBOSS package (RICE *et al.* 2000).

**Magnitude of selection against indels in coding regions**

The percentage of indels in repetitive coding regions removed by natural selection was calculated under the assumption that indels in the repetitive non-coding genome are neutral. The expected number of indels in repetitive coding regions before natural selection was calculated by
dividing the observed number of “neutral” mutations in repetitive non-coding regions by the
total number of repetitive elements that are in non-coding regions (21.8%) and multiplied this value by
the fraction of repetitive elements that are in coding regions (78.2%). The percentage of indels
removed due to natural selection is then 1 – observed/expected. If mutations in non-coding
repetitive regions are not neutral then this method will generate a lower limit estimate.

Core genes

Pre-computed pairwise reciprocal best BLAST hits for 36 Pseudomonas species were
downloaded from the Pseudomonas Genome Database (WINSOR et al. 2011). The core genome for
Pseudomonas aeruginosa PAO1 was defined as the set of PAO1 genes that had a pairwise
reciprocal best BLAST hits in the 35 remaining Pseudomonas species. We find a total of 1435 core
genes.

Clustering of Orthologous Groups (COGs) analysis

A list of P. aeruginosa PAO1 genes with annotated Clusters of Orthologous Groups (COG)
categories (TATUSOV et al. 2000) was downloaded from NCBI. This list was intersected with the list of
genes that had experienced at least one mutation during our experiment. Genes with annotated
mutations and COG categories were compared to the rest of the genes in the PAO1 genome that
were unmutated, but had been assigned a COG category. P-values were computed using Fisher’s
exact test and corrected for multiple testing using the false discovery rate method (BENJAMINI AND
HOCHBERG 1995).
Statistical analysis and simulations

All statistical analyses were conducted in R (version 2.15.0) (R DEVELOPMENT CORE TEAM 2012). All statistical tests are reported as a $P$-value and the value for the test statistic with a subscript indicating the degrees of freedom. All tests use $\alpha=0.05$ and, where applicable, are two-tailed.

Simulations were used to generate the expected distribution of the number of mutations per gene, given the substantial variation in gene length in the *P. aeruginosa* genome (mean: 830bp, 95% confidence interval: 247bp – 2786bp). The lengths of all genes in the *P. aeruginosa* genome were obtained from the Pseudomonas Genome Database (WINSOR et al. 2011). In each simulation, mutations (either synonymous or non-synonymous) were randomly distributed across a simulated genome, using the same number of mutations as was detected in our experiment. The number of mutations per gene was recorded and results were averaged across 100 simulations.

Acknowledgements

We thank Antonio Oliver for providing us with the PAO1ΔmutS strain. We thank the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (funded by Wellcome Trust grant reference 090532/Z/09/Z and MRC Hub grant G0900747 91070) for the generation of the sequencing data.

Literature Cited


Barrett, R. D., R. C. MacLean and G. Bell, 2006 Mutations of intermediate effect are responsible for adaptation in evolving *Pseudomonas fluorescens* populations. Biol Lett 2: 236-238.


Cingolani, P., A. Platts, I. L. Wang, M. Coon, T. Nguyen et al., 2012 A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly (Austin) 6: 80-92.


Figures

Figure 1. Types of mutations accumulated

A The distribution of accumulated mutations according to type of mutation. Indels <10 base pairs long were considered to be “short”. B Further information on the effects of point mutations.
Figure 2. Average fitness decays in mutation accumulation lines

Plotted points show the mean fitness (± SE) of hypermutator lines (black symbols, n=8) and control lines (grey symbols, n=4) that were passaged through 28 daily bottlenecks, which corresponds to ~644 generations of mutation accumulation. The fitness of hypermutator lines rapidly declined, but the fitness of control lines did not change over the course of the experiment (ANOVA: $F_{1,3} = 0.436, P = 0.556$). Note that in some MA lines, fitness decayed to the point where it was not possible to measure fitness reliably, but this data is included to prevent bias.
Figure 3. Fitness trajectories for individual mutation accumulation lines.

The mean (± SE; n=6) fitness of individual hypermutator lines through time. Red data points indicate that fitness is too low to measure accurately. The mean fitness (± SE; n=6) of individual hypermutator lines through time. Red data points indicate that fitness is too low to measure accurately. The y-axis of each plot is scaled differently to maximize the resolution of evolutionary dynamics within a single line.
The distribution of fitness changes for each ‘step’ in the mutation accumulation experiment across all 8 hypermutator lines. Each ‘step’ represents the difference in fitness between successive assays for an MA line (~8.4 mutations accumulated/step). The solid black line depicts no change in fitness and the area between the dashed grey lines is the area in which $N_e < 1$, where $N_e$ is the harmonic mean of population size over time (although this may be an underestimate (OTTO AND WHITLOCK 1997)).
Figure 5. The distribution of indel mutations in proteins.

Comparison between the observed and expected position of frameshifts in coding regions.

Proteins were divided into 3 equal pieces and we counted the number of frameshifts (overlapping with homopolymeric tracts) that fell in each section. Expected frequencies were computed by counting the number of homopolymeric tracts the *P. aeruginosa* PAO1 proteome that fall in each section. The differences between observed and expected values were statistically significant for the N-terminal third of proteins (one-sided exact binomial test: $P = 0.037$).
Table 1. Testing for selection on single base pair substitutions

<table>
<thead>
<tr>
<th>Protein effect</th>
<th>Observed</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-synonymous</td>
<td>480</td>
<td>444.38</td>
</tr>
<tr>
<td>Intergenic</td>
<td>80</td>
<td>84.33</td>
</tr>
<tr>
<td>Stop-gain</td>
<td>14</td>
<td>8.94</td>
</tr>
<tr>
<td>Stop-loss</td>
<td>1</td>
<td>1.41</td>
</tr>
</tbody>
</table>

The number of observed single base pair substitutions relative to the neutral expectation, as determined from the synonymous mutation rate and genome composition of *P. aeruginosa*. The observed number of mutations does not differ from the neutral expectation for any functional category of mutation.