# EVOLUTIONARY HIGHWAYS TO PERSISTENT INFECTION BY *PSEUDOMONAS AERUGINOSA*

Jennifer A Bartell[1,*], Lea M Sommer[2,*], Janus A J Haagensen[1], Anne Loch[1], Rocio Espinosa[1], Søren Molin[1], and Helle Krogh Johansen[2,3]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark.

[2]Department of Clinical Microbiology, Rigshospitalet, 2100 Copenhagen Ø, Denmark.

[3]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.

*Corresponding authors and equal contribution

## SUMMARY

Persistent infections require bacteria to evolve from their naïve colonization state by optimizing fitness in the host. Bacteria may follow the same adaptive path, but many distinct paths could enable equally successful persistence. Here, we map the development of persistent infection over 10 years by screening 8 infection-relevant phenotypes of 443 longitudinal Pseudomonas aeruginosa isolates from 39 young cystic fibrosis patients. Using Archetype Analysis to map the multi-trait evolutionary continuum and Generalized Additive Mixed Models to identify trait correlations accounting for patient-specific influences, we find: 1) a 2-3 year timeline of rapid substantial adaptation after colonization, 2) variant "naïve" and "adapted" states reflecting discordance between phenotypic and molecular adaptation and linked by distinct evolutionary trajectories, and 3) new phenotypes associated with pathoadaptive mutations. Our results underline the environmental

influences affecting evolution of complex natural populations, while providing a clinically accessible approach for tracking patient-specific pathogen adaptation to guide treatment.

## KEYWORDS

Evolution; Persistent Infection; Antibiotic Resistance; Cystic fibrosis; *Pseudomonas aeruginosa*; Clinical Isolates; Longitudinal Analysis; Data Modeling; Genotype-Phenotype

## INTRODUCTION

Bacteria have spent millennia evolving complex and resilient modes of adaptation to new environments, and many species can also effectively deploy these skills as pathogens during colonization and persistence within human hosts (Flores-Mireles et al., 2015; Lieberman et al., 2014; Rau et al., 2010). As they gradually increase their fitness via accumulating genetic and epigenetic changes, distinct pathogen populations may travel along the same predictable path to successful colonization. However, many other unique sequences of multi-trait adaptation could enable equally successful persistence (Cohen-Cymberknoh et al., 2011). This makes it difficult to pinpoint specific traits that signal the state of pathogen fitness and associated risk of an incurable chronic infection, complicating patient treatment intended to inhibit persistence. Meanwhile, recurrent, treatment-resistant infections are increasing problems worldwide (Flores-Mireles et al., 2015; Kline and Bowdish, 2016; May, 2014; O'Neill, 2016).

Even for a well-studied model system of bacterial persistence and chronic infection such as the airway infections of cystic fibrosis (CF) patients, evolutionary trajectories remain difficult to map due in part to the competing modes of evolution at play in these patients. While the general scientific consensus of the field is that select traits converge to similar states during most infections (such as loss of virulence and increase in antibiotic resistance), studies have also shown a high degree of population heterogeneity (Jorth et al., 2015; Lieberman et al., 2014; Markussen et al., 2014; Winstanley et al., 2016). This heterogeneity could be influenced by two distinct modes of evolution: 1) parallel, independent evolution caused by spatial segregation for long-term chronic infected patients (Jorth et al., 2015; Markussen et al., 2014) or 2) diversification, or "bet hedging", creating resilient populations (Yachi et al.,

1999) from the early colonization stage, where greater mixing of the environment and bacterial populations is possible (Johansen et al., 2012).

In this study, we map the development of persistent bacterial infections by phenotypically screening a collection of 443 longitudinal clinical *Pseudomonas aeruginosa* isolates from 39 young CF patients, measuring how 8 infection-relevant traits adapt during the initial 10 years of colonization. While laboratory evolution studies have measured phenotypic adaptation to a new, minimal media environment which generally occurs within the initial 5,000-10,000 generations, only estimates are available for the period of significant adaptation in complex host environments like the CF lung (based on growth rates and genetic adaptation) (Barrick et al., 2009; Woods et al., 2011; Yang et al., 2011). Thus, we provide new insights into the *in vivo* transition from initial colonization to persistent chronic infection. Previous studies have provided information on the genetic evolution of single clonal lineages of human pathogens and identified specific genetic adaptations that correlated with the ability to colonize and persist (Marvig et al., 2013, 2015; Smith et al., 2006). While some studies have paired these findings with phenotypic observations (Rau et al., 2010; Silva et al., 2016; Sommer et al., 2016; Yang et al., 2011) in order to associate genetic changes with phenotypic changes, this is especially challenging in natural populations. Genotype-phenotype links are eroded over the course of evolution by environment-based tuning of pathogen activity, or “acclimation”, and accumulation of mutations, or “genetic adaptation”. Therefore, the genotype alone cannot provide a complete predictive picture of the adaptation process (Jansson and Baker, 2016; Jansson and Hofmockel, 2018), and we therefore propose that phenotypic characterization is equally important for the understanding of evolution.

To map adaptation of the pathogen lineages infecting our patients, we analyzed our phenotypic dataset using generalized additive mixed models (GAMMs) and archetype analysis (AA), reassessing current theories of phenotypic evolution in the CF airways. We identify emergent patterns of bacterial phenotypic change across our patient cohort that depart from expected evolutionary paths and estimate the length of the period of initial rapid adaptation letting the bacteria transition from a “naïve” to a “chronic” phenotypic state. We further identify distinct and repeating trajectories of pathogen evolution, and by

leveraging our prior molecular study of this isolate collection, we propose mechanistic links between these phenotypic phenomena and genetic adaptation. These findings support the promise of using select phenotypic traits to track pathogen adaptation across a patient population, monitor patient-specific infection states, tailor the use of antibiotics, and eventually inhibit the transition to a persistent and chronic infection.

# RESULTS

First, we present our 8-phenotype screen and associated summary statistics. To contextualize our interpretation of this data, we then describe our data-driven modeling approach and validation. Finally, we use our models to identify and present significant evolutionary trends that contribute to persistence of *P. aeruginosa* in our CF patient cohort.

## Evaluating pathogen adaptation in the early stage of infection

*A unique dataset.* The 443 clinical *P. aeruginosa* isolates in this study originate from a cohort of 39 children with CF (median age at first *P. aeruginosa* isolate = 8.1 years) treated at the Copenhagen CF Centre at Rigshospitalet and capture the early period of adaptation, spanning 0.2-10.2 years of colonization by a total of 52 clone types. Of these isolates, 373 were previously characterized in a molecular study of adaptation (Marvig et al., 2015). When we discuss time in this study, we generally refer to the "colonization time of an isolate" (ColT). This is defined for each specific lineage as the time since colonization was first identified in a patient until the isolate in question was sampled. This enables us to have an approximate reference for when the bacteria initially transferred from one environment to another and adaptation to the CF airways could begin. It is important to note that the first time *P. aeruginosa* is identified in a patient sample is not likely to be the true "time zero" of adaptation, since a significant bacterial load is necessary for reliable culturing.

In contrast to our previous study, here we focus on phenotypic characterization including measurement of 8 phenotypes that encompass growth rates, antibiotic susceptibility, virulence factors, and adherence (Figure 1 and 2); we define adherence as a shared trend in adhesion and aggregation which we associate with a biofilm-like lifestyle. These phenotypes are generally accepted to change over the course of colonization and infection of CF patient

120 airways (Jiricny et al., 2014; López-Causapé et al., 2017; Silva et al., 2016; Winstanley et al.,
121 2016; Yang et al., 2008) and also lend themselves to high-throughput screening. That is, an
122 evolved isolate would grow slowly, adhere proficiently, be more likely to exhibit a mucoid
123 and/or hypermutator phenotype, have reduced protease production, and resist antibiotics,
124 in contrast to a naïve isolate. However, a visual inspection of our measurements ordered by
125 the colonization time does not indicate an overarching adaptive trajectory from naïve to
126 evolved phenotypes (Figure 2). Instead, we find isolates that resemble both "naïve" and
127 "evolved" phenotypic states throughout the study period.
128
129 *Modeling multi-trait evolution.* To analyze our dataset capturing the scale, complexity, and
130 noise of pathogen adaptation in a population of patients, we need to use modeling methods
131 that minimize patient-specific effects, smooth irregular sampling intervals and enable the
132 mapping of multi-trait evolution from start to end state. Our framework balances model
133 complexity and precision – also known as "the bias-variance tradeoff". We therefore 1)
134 modeled the inherent dynamic continuum of multi-trait evolution using AA and 2) evaluated
135 temporal correlations between phenotypic adaptation across patients by fitting cross-
136 patient trendlines using GAMMs (Figure 1).
137
138 With AA, we relate each isolate according to its similarity to the isolates with the most
139 extreme phenotypes in our collection. These extremities of our data, or "archetypes", are
140 positioned at the "corners" of the principal convex hull (PCH), the polyhedron of minimal
141 volume that still fully encapsulates our phenotype dataset in a multi-dimensional trait space
142 (Mørup and Hansen, 2012) that represents the pathogen evolutionary landscape of our
143 young patient cohort. We conceptualize archetypes as the initial and final states of plausible
144 adaptive trajectories across this landscape and predict both the number of archetypes and
145 their distinct phenotypic profiles that best represent our data. In contrast to ordination
146 approaches similar to Principal Component Analysis (PCA) which describes samples using
147 ambiguous, difficult to interpret dimensions of major variance, AA describes each isolate
148 contained within the principal convex hull in relation to its similarity to each archetype.
149 Thus, the characteristics of each isolate are directly interpretable along an evolutionary
150 continuum from "naïve" to "evolved" archetypes, comparable to the naïve and evolved
151 phenotypic states as described above. To visualize this continuum, we relied on 2D

projections of our multivariate fits using a "simplex" plot as shown in Figure 3C (Seth and Eugster, 2016). Though this visualization obscures the true dimensionality of the isolate distribution by implying the archetypes are equidistant, we partly compensated for this by using PCA and biological insight gained from the GAMMs to guide archetype placement such that we maintain interpretability from an evolutionary perspective.

With the GAMMs, we want to predict whether a given phenotype (the "predicted" or "dependent" variable) significantly correlates with other phenotypes (the "explanatory" or "independent" variables). We do this by accounting for time as well as patient-specific environments as random effects via this flexible mixed model approach enabling both linear and nonlinear fits. We ultimately prioritize accuracy in our fits rather than forcing linear relationships that do not effectively capture natural evolutionary dynamics that we expect to vary from patient to patient. This accuracy and flexibility invariably increase the risk of overfitting. However, we counteract this by both the default penalization of fits inherent to the method used and by model estimation via restricted maximum likelihood (REML) (Wood, 2006). Furthermore, to avoid assumptions of "cause-and-effect" relationships between our variables, we implement a feature reduction approach; we permute through different one-to-one models of all phenotypes, and then combine the statistically significant individual phenotypes into a multi-feature model. We further remove any phenotype that loses significance in the multi-feature model, assuming that it is correlated with a more impactful phenotype. From this point, all mentions of significant relationships or correlations are obtained from the GAMM analyses with p-values $< 0.01$ based on Wald-type tests as described in (Wood, 2006, 2013), unless otherwise stated.

## Effective mapping of phenotypic trends

*Archetype analysis*. AA predicted six distinctive archetypes sufficient to describe each isolate within the evolutionary landscape (Figure 3A). The simplex plot of Figure 3C shows annotation of the archetypes by the standout features of each archetype that contribute to its identification as an extremal corner of the dataset – we therefore only annotate by the highest or lowest values for each phenotype across all fitted archetype trait profiles and neglect moderate values (Figure 3B). This simplex key illustrates that two archetypes

resembled naïve and un-evolved isolates with fast growth, antibiotic susceptibility, and low adherence (Archetype A3 and A5). Meanwhile, two archetypes accounted for slow-growing evolved archetypes (A2 and A6). Two regions in the simplex visualization represent different focal points of adaptation, namely an increase in adherence (A2 and A4) versus ciprofloxacin resistance (A1 and A6). A substantial portion of isolates in our study resemble the "naïve" archetypes more closely than the "evolved" archetypes as indicated by their localization in the simplex plot (Figure 3C, most isolates cluster on the left near the "naïve" archetypes).

*Generalized additive mixed models.* The GAMM analysis showed that we could statistically support relationships between traits across patients (Figure 3D). We find that the growth rates in Artificial Sputum Medium (ASM) and Lysogeny Broth (LB) are significantly correlated and therefore only refer to the growth rate in ASM from this point – this is closer to the *in vivo* conditions of the CF airway environment. When evaluating adaptation of the specific phenotypes over time we found that the survival time of a lineage in a patient's lungs had a significant correlation with both growth rate and sensitivity to ciprofloxacin but did not correlate with sensitivity to aztreonam (Figure 3C, Figure 4A and 4B). This difference reflects that the Copenhagen CF Centre regularly administers ciprofloxacin to the CF patients but not aztreonam (Hansen et al., 2008).

An important distinction between AA and GAMMs is that many isolates clearly cluster in the AA simplex plot according to phenotypes whose adaptation is not significantly influenced by time of colonization as shown by GAMMs. This contrast shows the importance of combining these approaches to understand our data. As an example, both adhesion and aggregation do not correlate with colonization time for this population of young patients. Furthermore, the biofilm-related metric of mucoidity does not significantly correlate with any other measured phenotype, despite its use as an important biomarker of chronic infection in the Copenhagen CF Centre (Pedersen et al., 1992).

## Initial adaptation happens within 3 years of colonization

We suspect that the routes to successful persistence and a transition to chronic infection are initiated early in infection (Hansen et al., 2012; Marvig et al., 2015). During the initial

period of colonization spanning the first 2-3 years, the GAMMs indicate that a substantial change occurs in both growth rate and ciprofloxacin susceptibility, shown by the sharp slopes in this period (Figure 4A-B). Using AA, we also see a substantial shift of isolate distribution from "naïve" towards "evolved" archetypes in this adaptation window (Figure 4C). Furthermore, the adaptation at 2-3 years of colonization is quite diverse, reaching the outer boundaries of the simplex plot and confirming the rapid adaptation shown by the GAMMs. Interestingly, the four hypermutator isolates arising in this window do not alone define the trait boundaries of the AA; other normo-mutator isolates are located in equal proximity to the "evolved" archetypes (Figure 4D, full dataset in Figure S1).

To evaluate whether the rapid phenotypic adaptation occurred in parallel with genetic adaptation, we investigated the accumulation of nonsynonymous mutations (Figure 4D-E). We used the isolates representing the first *P. aeruginosa* culture from a patient as the reference point for identification of accumulating mutations. Using AA, we observed that most of the first isolates with 0-30 mutations aligned with "naïve" archetypes, and 2-3-year-old isolates with 9-48 mutations extended to the outer boundaries of adaptation (A2, A6, and A1) (Figure 4C-D). However, we also observed the persistence of WT-like genotypes with few mutations alongside evolved genotypes (Figure 4D). When analyzing the entire dataset using GAMMs, we found a significant, near-linear relationship between colonization time and the number of non-synonymous SNPs, but this near-linear trend was not present when evaluating the total number of mutations (Figure 4E). We theorize that this difference is driven by the apparent logarithmic accumulation of indels, where indel accumulation appears to slow around year 2 of colonization as shown in Figure 4E.

## Multi-trait analysis enables complex genotype-phenotype associations

The obscuring of genotype-phenotype links via polygenic mutations and their pleiotropic effects is rarely easy to deconvolute. As our models are unbiased by any genetic information, we have a unique perspective from which to map genotype-phenotype relationships. We previously identified 52 "pathoadaptive genes", which are genes mutated more often than expected from genetic drift and thus are assumed to confer an adaptive advantage during infection (Marvig et al., 2015; Sokurenko et al., 1999). By overlaying

nonsynonymous mutations on AA simplex plots, we evaluated mutation impacts on the following pathoadaptive genes: 1) top ranked *mexZ* and other repressors of drug efflux pumps (*nfxB*/*nalD*), 2) mucoidity regulators *mucA*/*algU* and the hypothesized infection-state switching *retS*/*gacAS*/*rsmA* regulatory pathway which we previously examined from a genetic adaptation perspective (Goodman et al., 2004; Marvig et al., 2015), and 3) quinolone resistance genes *gyrA*/*gyrB* (Kugelberg et al., 2005; Nakamura et al., 1989; Robillard and Scarpa, 1988) given the rapid adaptation of ciprofloxacin susceptibility. We saw no obvious spatial correlations with mutations linked to mucoidity regulation in the AA model (Figure S2) which parallels mucoidity's lack of significance in our GAMM analyses. Isolates with *mexZ* mutations are prevalent and also distributed throughout the simplex plot, so we analyzed *mexZ* mutants in combination with other pump repressor gene mutations. We found that even double-mutant isolates grouped by their effect on different pairings of efflux pumps showed diverse phenotypes via AA, though we noted a unique distribution of the many isolates impacted by a mutation in *nfxB* (Figure S3, Figure 5B). The isolate distributions of *gyrA/B* and *retS/gacAS/rsmA* mutants were also striking in their spatial segregation according to AA (Figure 5A-B).

*Ciprofloxacin resistance genes.* The primary drivers of ciprofloxacin resistance in *P. aeruginosa* are theorized to be mutations in drug efflux pump repressor *nfxB* and mutations in *gyrA* and *gyrB* encoding the two gyrase subunits of the DNA replication system (Kugelberg et al., 2005; Nakamura et al., 1989; Robillard and Scarpa, 1988). We would therefore expect isolates with mutations in these resistance genes to cluster around archetypes A1 and A6, which are characterized by high ciprofloxacin minimal inhibitory concentrations (MICs) (Figure 3C). However, AA illustrates a much broader distribution of *gyrA/B* mutants among archetypes, and a third distinct distribution of *nfxB* mutants (Figure 5A-B, left panel). In association with this AA diversity, we see a range of ciprofloxacin resistance levels associated with affected isolates both across and within patient lineages, and no dominant mutations/mutated regions which repeat across lineages (Figure 5A-B, right panel). The incidence of resistance due to these distinct mechanisms was equal at 78% of affected isolates (54 out of 69 resistant gyrase mutants vs 37 out of 47 resistant *nfxB* mutants based on the European Committee on Antimicrobial Susceptibility Testing (EUCAST) breakpoint). However, the persistence of these respective mutations in affected lineages was dissimilar.

Generally, *nfxB* mutation occurred earlier in lineage evolution and persisted in far fewer lineages compared to *gyrA/B* mutations, which likely contributes to *nfxB*'s band-like distribution via AA compared to the broader distribution of gyrase mutants towards adapted archetypes.

Interestingly, when we further consider the gyrase-mutated isolate plot, we also see that isolates with a *gyrB* mutation (33 isolates alone or 14 in concert with *gyrA*) are concentrated closer to "biofilm-linked" archetypes A2 and A4 than isolates with only a *gyrA* mutation (33 isolates). This positive association of *gyrB* on adhesion was confirmed by GAMM, but when evaluating it by moving the two SNPs affecting the most isolates in both *gyrA* and *gyrB* (2 lineages each, Figure S4) into a laboratory *P. aeruginosa* strain (PAO1), we did not find the same association (Figure S5-6) (p-values > 0.05, ANOVA with Tukey correction, $F(4,10)=0.233$). We then evaluated the presence of co-occurring mutations in biofilm-linked genes in the *gyrB*-mutated lineages. In all but one lineage, there was no obvious genetic explanation for the increased adhesion. Ultimately, this genotype-phenotype link was indecipherable due to the complexity of mutation patterns and the multi-genetic signature of biofilm regulation (Wolska et al., 2016).

*retS/gacAS/rsmA.* The functional model of the *retS/gacA/gacS/rsmA* regulatory system is a "bimodal" switch between acute and chronic infection phenotypes (Goodman et al., 2004; Ventre et al., 2006). Posttranscriptional regulator *rsmA* activates an acute infection phenotype characterized by planktonic growth and inhibits a non-motile "biofilm" lifestyle. *retS* mutants are preserved in many lineages because they repress *rsmA* via the *gacA/S* two-component system and thus promote a chronic infection phenotype. However, our previous genetic analysis (Marvig et al., 2015) unexpectedly showed that multiple evolving lineages gained a subsequent mutation in *gacA/S* that at times appeared years after the *retS* mutation. Despite the complexity of this regulatory system, we show a clear phenotypic separation between clinical isolates that are *retS* mutants versus *retS+gacA/S* mutants via our AA model (Figure 5C, left panel). In this study, three of eight patients with nonsynonymous mutations in this system have isolates which are *retS+gacA/S* double mutants (Figure 5C, right panel). While *retS* mutants resemble more "evolved" archetypes (A1 – 2 and A6), all but one isolate clusters around the "naïve" archetypes (A3 – A5).

According to patient-specific trajectories, this reversion happens after an initial migration towards "evolved" archetypes. Because of the low sampling number and because we only see double mutants in three patients, we did not follow up with additional GAMM analyses of the effect of these mutations on different phenotypes.

## Infections persist via distinct routes of adaptation

As our trait associations with specific mutations highlight the importance of lineage-based analysis, we here further investigate lineage influences by mapping patient-specific adaptive trajectories, which may present clinically useful information for treatment management. By performing a patient-specific analysis using AA, we find that *P. aeruginosa* infections can persist successfully in individual patients despite different phenotypic starting points (A3-5) and/or end points (A1, 2, and 6). Figure 6A-C shows three adapting lineages that follow distinct trajectories and each persist for at least three years in a patient, while Figure 6D shows a patient with diverse isolates that do not appear to follow a clear adaptive trajectory. In both Figure 6A and C, we see a rapid evolution towards an endpoint of ciprofloxacin resistance. In Figure 6A, the colonization initiates with two isolates, but we determined that the isolate near A4 is genotypically distinct from the remaining 11 isolates of that lineage based on mutational differences. The persisting sublineage seems to initiate with the isolate near A3, after which it gains a *gyrB* mutation that guides the trajectory towards A1 and subsequent mutations then push the lineage phenotype towards A2, characterized by increased adherence and decreased sensitivity to aztreonam. This trajectory towards A2 is also seen in Figure 6B, which begins as a broad band of isolates moving from A3/A4 towards A2/A6. However, the isolates seem directed towards A2 rather than A6 or a mix thereof over infection. These results illustrate the diverse adaptive trajectories followed by *P. aeruginosa* in our patient cohort, which connect distinct start and endpoints of adaptation yet enable years of persistence.

In summary, the results of our investigations together illustrate how a multi-trait analysis perspective can identify unique emergent characteristics of evolving bacteria, but also highlight the strong influence of lineage-specific trajectories, the historic contingency of mutations, and the impact it can have on the phenotypes expressed.

## DISCUSSION

By integrating phenotypic and molecular characterizations of our unique isolate collection with well-suited data modeling, we illuminate specific evolutionary priorities in early infection. We overcome remarkable genetic and phenotypic diversity to 1) observe rapid early adaptation and its discordance with genetic adaptation, 2) associate novel phenotypes with pathoadaptive genes, and 3) retrieve meaningful mappings of distinct patient-specific trajectories. Phenotypic traits represent systems-level impacts of many different molecular markers and are shown in this study to adapt along parallel evolutionary paths. We therefore propose a "new" model of investigation or, more appropriately, we re-emphasize the value of classical phenotype-based investigations. Specifically, instead of focusing on genetic readouts of adaptation, where a specific mutation or gene may or may not be consistently linked with a specific phenotype; we suggest that mapping changes of carefully selected traits provides a better basis for predictions of the next steps of colonization and infection. Here, we specifically use clinically feasible phenotypic screens, modeling and visualization methods to evaluate trait adaptation, and map patient-specific evolutionary trajectories with the potential for integration with clinical diagnostics.

We deconvolute pathogen evolution in the host by a unique integration of methods. Previous studies employed linear mixed models of phenotypic adaptation (Andersen et al., 2015), and employed archetype analysis in the comparison of features of transcriptomic adaptation by *P. aeruginosa* (Thøgersen et al., 2013), and most recently, prediction of the polymorphism structure in a population based on evolutionary trade-offs in a multi-trait fitness landscape (Sheftel et al., 2018). However, by integrating the two approaches, we illuminate complex patterns and facilitate the deconvolution of trait adaptation to decipher the major evolutionary highways in our patient cohort. For example, we do not see significant cross-patient selection for adherence using GAMMs, but we see selection for adherence in a few specific patients via AA. That this is not a major trend in our data is surprising when we consider that a biofilm lifestyle is expected to be beneficial to persistence in chronically infected patients (Bjarnsholt et al., 2009; Cohen-Cymberknoh et al., 2011; Høiby, 2002; Pressler et al., 2011). However, it leads us to hypothesize that the

rate of adaptation and relative benefit of this phenotype may vary significantly and be more sensitive to temporal stresses such as antibiotic treatment. In support of our findings, others have recently shown that the longitudinal relationship between mucoidity and a clinical diagnosis of chronic infection is not as direct as previously expected (Heltshe et al., 2017). Together, these results prompt further reassessment of common assumptions regarding the evolutionary objectives of *P. aeruginosa* in CF infections.

We map remarkable evolutionary dynamics in the early stages of patient colonization, where we estimate the initial window of rapid adaptation to be within 5256 - 7884 bacterial generations (Yang et al., 2008). While the first isolate of each patient in our collection may not represent the true start of adaptation given sampling limitations, we see general alignment of our "first" isolates via archetype distribution; the window of rapid adaptation is therefore still likely substantially contracted compared to the previous estimate of within 42,000 generations (Yang et al., 2011). In fact, our data resembles the rate of fitness improvement found in the laboratory evolution study of *E. coli*, which was shown to change significantly within the first 5,000-10,000 generations (Barrick et al., 2009; Woods et al., 2011). With regards to the corresponding genetic adaptation during this period, the cross-patient trend fit by GAMMs reflects an expected accumulation of mutations over time. However, AA demonstrates patient-specific differences; specific lineages show different numbers of mutations after having adapted over 2-3 years (with a range of 9-48 mutations, not including hypermutators). Furthermore, the isolates with the highest numbers of mutations, the hypermutators, do not define the boundaries of phenotypic adaptation, which supports the idea that molecular and phenotypic adaptation can be discordant. Select beneficial mutations (for example, a highly impactful indel) can alone induce important phenotypic changes that improve fitness, especially via pleiotropic effects (Solovieff et al., 2013) in accordance with the theory that the likelihood of beneficial mutations decreases over time (Desai and Fisher, 2007). Our logarithmic gain of indels replicates the findings of the laboratory evolution study of *E. coli* which has been propagating for more than 60,000 generations (Good et al., 2017). This observation suggests that other methods of adaptation may contribute in equal degree to adaptation via mutation, such as acclimation to the CF lung environment via gene expression changes (Dötsch et al., 2015; Rossi et al., 2018).

403 Our study highlights important limitations to genotype-phenotype associations and
404 underlines the usefulness of a multi-trait perspective; individual mutations may have
405 pleiotropic effects and obscure genetic signatures as they accumulate over time. For
406 example, we show an unexpected phenotypic reversion to an "acute infection state" via
407 historically contingent mutations in the *retS*/*gacAS*/*rsmA* system. This does not easily
408 reconcile with theories about preservation of a persistent colonization via convergence
409 towards a "chronic phenotype". However, over time some patients are colonized by new
410 clone types and/or other pathogens; this could require re-establishment of a colonization
411 mid-infection and thus induce the population to revert towards an acute infection state
412 where fast growth and motility improve its ability to compete. In evaluating primary
413 ciprofloxacin resistance genes *nfxB* and *gyrA/B*, we see substantial diversity in ciprofloxacin
414 susceptibility for isolates which share the same mutation, unique mutations fixing in each
415 lineage, and differential extents of adaptation boundaries using AA. This last observation led
416 us to note that *nfxB*-mutated isolates are associated with earlier incidence and more limited
417 lineage persistence versus isolates with *gyrA/B* mutations. This may be caused by a
418 differential fitness impact of *nfxB* mutation that limits tolerance of this mutation to a
419 narrower range of genotypic backgrounds than that of *gyrA/B* mutation, or it could be
420 associated with differential treatment regimens that influence the relative benefit of a given
421 ciprofloxacin resistance mechanism; further study is required to tease out these
422 mechanisms by both evaluating patient treatment histories and adding additional lineages
423 to the study. We also discovered an association between *gyrB* mutation and increased
424 adhesion; to our knowledge, there is no direct link between *gyrB* and the capability to
425 adhere (Kugelberg et al., 2005) and we did not see increased adhesion in our engineered
426 *gyrB* mutants. It is possible that a mutation in *gyrB* enables or is simply more tolerant of
427 other specific mutations that result in increased adhesion in comparison with *gyrA*
428 mutation. In general, our results support the theory that specific mutations confer unique
429 evolutionary restrictions to adaptive trajectories that impact other traits, but genetic
430 background and host-specific evolutionary pressures influence the type and degree of
431 restriction. This process could underpin the diversity of phenotypic adaptation we observe
432 in this study.
433

Complex mutation patterns are an inherent byproduct of evolution and result in equally complex, varied adaptive trajectories that lead to persistence. Clinicians need improved methods to predict and prevent the transition from colonization to persistent and chronic infection. We see particular promise in incorporating records of patient treatment and response to our assessment of adaptive trajectories to further guide clinicians in patient-tailored treatment management. At the moment, our models can assist by illustrating distinct evolutionary highways to pathogen persistence in individual patients. In this study, we find 3 overarching modes of evolution: 1) directed diversity, where a given patient's trajectory is characterized by diverse isolates moving in the same general direction such as increased adhesion and aggregation (Figure 6B) (Andersen et al., 2015), 2) convergent evolution, where the adaptation is constrained by strong selective pressures driving the phenotypic change in one particular direction such as resistance to ciprofloxacin (Figure 6C) (Imamovic et al., 2018), or 3) general diversity, where colonization appears to initiate with diverse "naïve" isolates and this diversity is maintained or expanded with no interpretable adaptive trajectory over time (Figure 6D). We theorize that these evolutionary modes and their dynamics may correlate with infection persistence and patient outcomes.

In conclusion, our study identifies rapid adaptation of isolates by both mutational accumulation and acclimation within the first few years of colonization. While specific traits show cross-patient convergence, we also highlight remarkable diversity both within and across patients, emphasizing the maintenance of diversity as a useful mode of persistence. We identify unique highways of evolution that are used by pathogens to persist in the lungs. By mapping these phenotypic trajectories, we can identify both genetic mechanisms that regulate these highways and complex traits that signal the impact of treatment on individual infections.

## ACKNOWLEDGMENTS

HKJ was supported by The Novo Nordisk Foundation as a clinical research stipend (NNF12OC1015920), by Rigshospitalets Rammebevilling 2015-17 (R88-A3537), by Lundbeckfonden (R167-2013-15229), by Novo Nordisk Fonden (NNF15OC0017444), by RegionH Rammebevilling (R144-A5287) and by Independent Research Fund Denmark /

Technology and Production Sciences (FTP-4183-00051). JAB was funded by a postdoctoral fellowship from the Whitaker Foundation. LMS … SM …

We thank Katja Bloksted, Ulla Rydahl Johansen, Helle Nordbjerg Andersen, Sarah Buhr Bendixen, Camilla Thranow, Pia Poss, Bonnie Horsted Erichsen, Rakel Schiøtt and Mette Pedersen for excellent technical assistance. We also thank Prof. Anders Stockmarr, Prof. Nina Jakobsen and Prof. Morten Mørup (DTU Compute) and Dr. Kevin D'Auria (Counsyl) for helpful discussions.

## AUTHOR CONTRIBUTIONS

SM and HKJ jointly supervised the study. JAH, SM, and HKJ conceived and designed the experiments. JAH performed all phenotypic screening with assistance from AL. REP performed the genetic engineering of isolate mutations. JAB and LMS conceived and performed all computational analysis and wrote the manuscript. JAH, SM and HKJ helped write the manuscript and provided revisions.

## DECLARATION OF INTERESTS

The authors declare no competing financial interests.

## REFERENCES

Andersen, S.B., Marvig, R.L., Molin, S., Krogh Johansen, H., and Griffin, A.S. (2015). Long-term social dynamics drive loss of function in pathogenic bacteria. Proc. Natl. Acad. Sci. U. S. A. *112*, 10756–10761.

Aparicio, T., de Lorenzo, V., and Martínez-García, E. (2018). CRISPR/Cas9-Based Counterselection Boosts Recombineering Efficiency in *Pseudomonas putida*. Biotechnol. J. *13*, 1700161.

Arnold, J.B. (2017). ggthemes: Extra Themes, Scales and Geoms for 'ggplot2.'

Barrick, J.E., Yu, D.S., Yoon, S.H., Jeong, H., Oh, T.K., Schneider, D., Lenski, R.E., and Kim, J.F. (2009). Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature *461*, 1243–1247.

493 Bjarnsholt, T., Jensen, P.Ø., Fiandaca, M.J., Pedersen, J., Hansen, C.R., Andersen, C.B.,
494 Pressler, T., Givskov, M., and Høiby, N. (2009). *Pseudomonas aeruginosa* biofilms in the
495 respiratory tract of cystic fibrosis patients. Pediatr. Pulmonol. *44*, 547–558.
496 Caceres, S.M., Malcolm, K.C., Taylor-Cousar, J.L., Nichols, D.P., Saavedra, M.T., Bratton, D.L.,
497 Moskowitz, S.M., Burns, J.L., and Nick, J.A. (2014). Enhanced in vitro formation and
498 antibiotic resistance of nonattached Pseudomonas aeruginosa aggregates through
499 incorporation of neutrophil products. Antimicrob. Agents Chemother. *58*, 6851–6860.
500 Cohen-Cymberknoh, M., Shoseyov, D., and Kerem, E. (2011). Managing cystic fibrosis:
501 Strategies that increase life expectancy and improve quality of life. Am. J. Respir. Crit. Care
502 Med. *183*, 1463–1471.
503 Desai, M.M., and Fisher, D.S. (2007). Beneficial mutation selection balance and the effect of
504 linkage on positive selection. Genetics *176*, 1759–1798.
505 Déziel, E., Comeau, Y., and Villemur, R. (2001). Initiation of biofilm formation by
506 Pseudomonas aeruginosa 57RP correlates with emergence of hyperpiliated and highly
507 adherent phenotypic variants deficient in swimming, swarming, and twitching motilities. J.
508 Bacteriol. *183*, 1195–1204.
509 Dötsch, A., Schniederjans, M., Khaledi, A., Hornischer, K., Schulz, S., Bielecka, A., Eckweiler,
510 D., Pohl, S., and Häussler, S. (2015). The Pseudomonas aeruginosa Transcriptional Landscape
511 Is Shaped by Environmental Heterogeneity and Genetic Variation. MBio *6*, e00749.
512 Eugster, M.J.A., and Leisch, F. (2009). From Spider-Man to Hero - Archetypal Analysis in *R*. J.
513 Stat. Softw. *30*, 1–23.
514 Eugster, M.J.A., and Leisch, F. (2011). Weighted and robust archetypal analysis. Comput.
515 Stat. Data Anal. *55*, 1215–1225.
516 Fernandez, M., Wilson, H.F., and Barnard, A.S. (2017). Impact of distributions on the
517 archetypes and prototypes in heterogeneous nanoparticle ensembles. Nanoscale *9*, 832–
518 843.
519 Flores-Mireles, A.L., Walker, J.N., Caparon, M., and Hultgren, S.J. (2015). Urinary tract
520 infections: Epidemiology, mechanisms of infection and treatment options. Nat. Rev.
521 Microbiol. *13*, 269–284.
522 Good, B.H., McDonald, M.J., Barrick, J.E., Lenski, R.E., and Desai, M.M. (2017). The dynamics
523 of molecular evolution over 60,000 generations. Nature *551*, 45–50.
524 Goodman, A.L., Kulasekara, B., Rietsch, A., Boyd, D., Smith, R.S., and Lory, S. (2004). A

525 Signaling Network Reciprocally Regulates Genes Associated with Acute Infection and
526 Chronic Persistence in Pseudomonas aeruginosa. Dev. Cell *7*, 745–754.
527 Hansen, C.R., Pressler, T., and Høiby, N. (2008). Early aggressive eradication therapy for
528 intermittent Pseudomonas aeruginosa airway colonization in cystic fibrosis patients: 15
529 years experience. J. Cyst. Fibros. *7*, 523–530.
530 Hansen, S.K., Rau, M.H., Johansen, H.K., Ciofu, O., Jelsbak, L., Yang, L., Folkesson, A., Jarmer,
531 H.Ø., Aanæs, K., von Buchwald, C., et al. (2012). Evolution and diversification of
532 Pseudomonas aeruginosa in the paranasal sinuses of cystic fibrosis children have
533 implications for chronic lung infection. ISME J. *6*, 31–45.
534 Heltshe, S.L., Khan, U., Beckett, V., Baines, A., Emerson, J., Sanders, D.B., Gibson, R.L.,
535 Morgan, W., and Rosenfeld, M. (2017). Longitudinal development of initial, chronic and
536 mucoid Pseudomonas aeruginosa infection in young children with cystic fibrosis. J. Cyst.
537 Fibros.
538 Hentzer, M., Teitzel, G.M., Balzer, G.J., Heydorn, A., Molin, S., Givskov, M., and Parsek, M.R.
539 (2001). Alginate overproduction affects Pseudomonas aeruginosa biofilm structure and
540 function. J. Bacteriol. *183*, 5395–5401.
541 Høiby, N. (2002). Understanding bacterial biofilms in patients with cystic fibrosis: current
542 and innovative approaches to potential therapies. J. Cyst. Fibros. *1*, 249–254.
543 Imamovic, L., Ellabaan, M.M.H., Dantas Machado, A.M., Citterio, L., Wulff, T., Molin, S.,
544 Krogh Johansen, H., and Sommer, M.O.A. (2018). Drug-Driven Phenotypic Convergence
545 Supports Rational Treatment Strategies of Chronic Infections. Cell *172*, 121–134.e14.
546 Jansson, J.K., and Baker, E.S. (2016). A multi-omic future for microbiome studies. Nat.
547 Microbiol. *1*, 16049.
548 Jansson, J.K., and Hofmockel, K.S. (2018). The soil microbiome — from metagenomics to
549 metaphenomics. Curr. Opin. Microbiol. *43*, 162–168.
550 Jiricny, N., Molin, S., Foster, K., Diggle, S.P., Scanlan, P.D., Ghoul, M., Johansen, H.K.,
551 Santorelli, L.A., Popat, R., West, S.A., et al. (2014). Loss of Social Behaviours in Populations
552 of Pseudomonas aeruginosa Infecting Lungs of Patients with Cystic Fibrosis. PLoS One *9*,
553 e83124.
554 Johansen, H.K., Moskowitz, S.M., Ciofu, O., Pressler, T., and Høiby, N. (2008). Spread of
555 colistin resistant non-mucoid Pseudomonas aeruginosa among chronically infected Danish
556 cystic fibrosis patients. J. Cyst. Fibros. *7*, 391–397.

557 Johansen, H.K., Aanaes, K., Pressler, T., Nielsen, K.G., Fisker, J., Skov, M., Høiby, N., and von
558 Buchwald, C. (2012). Colonisation and infection of the paranasal sinuses in cystic fibrosis
559 patients is accompanied by a reduced PMN response. J. Cyst. Fibros. *11*, 525–531.
560 Jorth, P., Staudinger, B.J., Wu, X., Hisert, K.B., Hayden, H., Garudathri, J., Harding, C.L.,
561 Radey, M.C., Rezayat, A., Bautista, G., et al. (2015). Regional Isolation Drives Bacterial
562 Diversification within Cystic Fibrosis Lungs. Cell Host Microbe *18*, 307–319.
563 Kline, K.A., and Bowdish, D.M.E. (2016). Infection in an aging population. Curr. Opin.
564 Microbiol. *29*, 63–67.
565 Kragh, K.N., Hutchison, J.B., Melaugh, G., Rodesney, C., Roberts, A.E.L., Irie, Y., Jensen, P.Ø.,
566 Diggle, S.P., Allen, R.J., Gordon, V., et al. (2016). Role of Multicellular Aggregates in Biofilm
567 Formation. MBio *7*, e00237.
568 Kugelberg, E., Löfmark, S., Wretlind, B., and Andersson, D.I. (2005). Reduction of the fitness
569 burden of quinolone resistance in Pseudomonas aeruginosa. J. Antimicrob. Chemother. *55*,
570 22–30.
571 Lieberman, T.D., Flett, K.B., Yelin, I., Martin, T.R., McAdam, A.J., Priebe, G.P., and Kishony, R.
572 (2014). Genetic variation of a bacterial pathogen within individuals with cystic fibrosis
573 provides a record of selective pressures. Nat. Genet. *46*, 82–87.
574 López-Causapé, C., Sommer, L.M., Cabot, G., Rubio, R., Ocampo-Sosa, A.A., Johansen, H.K.,
575 Figuerola, J., Cantón, R., Kidd, T.J., Molin, S., et al. (2017). Evolution of the Pseudomonas
576 aeruginosa mutational resistome in an international Cystic Fibrosis clone. Sci. Rep. *7*.
577 Markussen, T., Marvig, R.L., Gómez-Lozano, M., Aanæs, K., Burleigh, A.E., Høiby, N.,
578 Johansen, H.K., Molin, S., and Jelsbak, L. (2014). Environmental heterogeneity drives within-
579 host diversification and evolution of Pseudomonas aeruginosa. MBio *5*, e01592-14.
580 Marvig, R.L., Johansen, H.K., Molin, S., and Jelsbak, L. (2013). Genome Analysis of a
581 Transmissible Lineage of Pseudomonas aeruginosa Reveals Pathoadaptive Mutations and
582 Distinct Evolutionary Paths of Hypermutators. PLoS Genet. *9*, e1003741.
583 Marvig, R.L., Sommer, L.M., Molin, S., and Johansen, H.K. (2015). Convergent evolution and
584 adaptation of Pseudomonas aeruginosa within patients with cystic fibrosis. Nat. Genet. *47*.
585 May, M. (2014). Drug development: Time for teamwork. Nature *509*, S4–S5.
586 Mørup, M., and Hansen, L.K. (2012). Archetypal analysis for machine learning and data
587 mining. Neurocomputing *80*, 54–63.
588 Nakamura, S., Nakamura, M., Kojima, T., and Yoshida, H. (1989). gyrA and gyrB mutations in

589 quinolone-resistant strains of Escherichia coli. Antimicrob. Agents Chemother. *33*, 254–255.
590 O'Neill, J. (2016). Tackling drug-resistant infections globally: Final report and
591 recomendations. Https://Amr-Review.Org/.
592 O'Toole, G.A., and Kolter, R. (1998). Flagellar and twitching motility are necessary for
593 Pseudomonas aeruginosa biofilm development. Mol. Microbiol. *30*, 295–304.
594 Pedersen, S.S., Høiby, N., Espersen, F., and Koch, C. (1992). Role of alginate in infection with
595 mucoid Pseudomonas aeruginosa in cystic fibrosis. Thorax *47*, 6–13.
596 Pressler, T., Bohmova, C., Conway, S., Dumcius, S., Hjelte, L., Høiby, N., Kollberg, H.,
597 Tümmler, B., and Vavrova, V. (2011). Chronic Pseudomonas aeruginosa infection definition:
598 EuroCareCF Working Group report. J. Cyst. Fibros. *10*, S75–S78.
599 Rau, M.H., Hansen, S.K., Johansen, H.K., Thomsen, L.E., Workman, C.T., Nielsen, K.F., Jelsbak,
600 L., Høiby, N., Yang, L., and Molin, S. (2010). Early adaptive developments of *Pseudomonas*
601 *aeruginosa* after the transition from life in the environment to persistent colonization in the
602 airways of human cystic fibrosis hosts. Environ. Microbiol. *12*, 1643–1658.
603 Ricaurte, D.E., Martínez-García, E., Nyerges, Á., Pál, C., de Lorenzo, V., and Aparicio, T.
604 (2018). A standardized workflow for surveying recombinases expands bacterial genome-
605 editing capabilities. Microb. Biotechnol. *11*, 176–188.
606 van Rij, J., Wieling, M., Baayen, R.H., and van Rijn, H. (2017). itsadug: Interpreting Time
607 Series and Autocorrelated Data Using GAMMs.
608 Robillard, N.J., and Scarpa, A.L. (1988). Genetic and physiological characterization of
609 ciprofloxacin resistance in Pseudomonas aeruginosa PAO. Antimicrob. Agents Chemother.
610 *32*, 535–539.
611 Rossi, E., Falcone, M., Molin, S., and Johansen, H.K. (2018). High-resolution in situ
612 transcriptomics of Pseudomonas aeruginosa unveils genotype independent patho-
613 phenotypes in cystic fibrosis lungs. Submitted.
614 Ryder, C., Byrd, M., and Wozniak, D.J. (2007). Role of polysaccharides in Pseudomonas
615 aeruginosa biofilm development. Curr. Opin. Microbiol. *10*, 644–648.
616 Seth, S., and Eugster, M.J.A. (2014). Probabilistic Archetypal Analysis.
617 Seth, S., and Eugster, M.J.A. (2016). Probabilistic archetypal analysis. Mach. Learn. *102*, 85–
618 113.
619 Sheftel, H., Szekely, P., Mayo, A., Sella, G., and Alon, U. (2018). Evolutionary trade-offs and
620 the structure of polymorphisms. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *373*, 20170105.

621 Silva, I.N., Santos, P.M., Santos, M.R., Zlosnik, J.E.A., Speert, D.P., Buskirk, S.W., Bruger, E.L.,
622 Waters, C.M., Cooper, V.S., and Moreira, L.M. (2016). Long-Term Evolution of *Burkholderia*
623 *multivorans* during a Chronic Cystic Fibrosis Infection Reveals Shifting Forces of Selection.
624 MSystems *1*, e00029-16.
625 Smith, E.E., Buckley, D.G., Wu, Z., Saenphimmachak, C., Hoffman, L.R., D'Argenio, D.A.,
626 Miller, S.I., Ramsey, B.W., Speert, D.P., Moskowitz, S.M., et al. (2006). Genetic adaptation by
627 Pseudomonas aeruginosa to the airways of cystic fibrosis patients. Proc. Natl. Acad. Sci. U. S.
628 A. *103*, 8487–8492.
629 Sokurenko, E. V, Hasty, D.L., Dykhuizen, D.E., Sokurenko, E. V., Hasty, D.L., Dykhuizen, D.E.,
630 Sokurenko, E. V., Hasty, D.L., and Dykhuizen, D.E. (1999). Pathoadaptive mutations: gene
631 loss and variation in bacterial pathogens. Trends Microbiol. *7*, 191–195.
632 Solovieff, N., Cotsapas, C., Lee, P.H., Purcell, S.M., and Smoller, J.W. (2013). Pleiotropy in
633 complex traits: challenges and strategies. Nat. Rev. Genet. *14*, 483–495.
634 Sommer, L.M., Alanin, M.C., Marvig, R.L., Nielsen, K.G., Høiby, N., von Buchwald, C., Molin,
635 S., and Johansen, H.K. (2016). Bacterial evolution in PCD and CF patients follows the same
636 mutational steps. Sci. Rep. *6*, 28732.
637 Team, R.C. (2017). R: A Language and Environment for Statistical Computing.
638 Thøgersen, J., Mørup, M., Damkiær, S., Molin, S., and Jelsbak, L. (2013). Archetypal analysis
639 of diverse Pseudomonas aeruginosa transcriptomes reveals adaptation in cystic fibrosis
640 airways. BMC Bioinformatics *14*, 279.
641 Ventre, I., Goodman, A.L., Vallet-Gely, I., Vasseur, P., Soscia, C., Molin, S., Bleves, S.,
642 Lazdunski, A., Lory, S., and Filloux, A. (2006). Multiple sensors control reciprocal expression
643 of Pseudomonas aeruginosa regulatory RNA and virulence genes. Proc. Natl. Acad. Sci. U. S.
644 A. *103*, 171–176.
645 Wickham, H. (2017). tidyverse: Easily Install and Load the 'Tidyverse.'
646 Winstanley, C., O'Brien, S., and Brockhurst, M.A. (2016). Pseudomonas aeruginosa
647 Evolutionary Adaptation and Diversification in Cystic Fibrosis Chronic Lung Infections. Trends
648 Microbiol. *24*, 327–337.
649 Wolska, K.I., Grudniak, A.M., Rudnicka, Z., and Markowska, K. (2016). Genetic control of
650 bacterial biofilms. J. Appl. Genet. *57*, 225–238.
651 Wood, S.N. (2006). Generalized additive models□: an introduction with R (CRC Press).
652 Wood, S.N. (2011). Fast stable restricted maximum likelihood and marginal likelihood

653 estimation of semiparametric generalized linear models. J. R. Stat. Soc. Ser. B (Statistical
654 Methodol. *73*, 3–36.
655 Wood, S.N. (2013). On p-values for smooth components of an extended generalized additive
656 model. Biometrika *100*, 221–228.
657 Wood, S.N., Pya, N., and Säfken, B. (2016). Smoothing Parameter and Model Selection for
658 General Smooth Models. J. Am. Stat. Assoc. *111*, 1548–1563.
659 Woods, R.J., Barrick, J.E., Cooper, T.F., Shrestha, U., Kauth, M.R., and Lenski, R.E. (2011).
660 Second-order selection for evolvability in a large Escherichia coli population. Science *331*,
661 1433–1436.
662 Xie, Y. (2017). knitr: A General-Purpose Package for Dynamic Report Generation in R.
663 Yachi, S., Loreau, M., and Singh, P.K. (1999). Biodiversity and ecosystem productivity in a
664 fluctuating environment: the insurance hypothesis. Proc. Natl. Acad. Sci. U. S. A. *96*, 1463–
665 1468.
666 Yang, L., Haagensen, J.A.J., Jelsbak, L., Johansen, H.K., Sternberg, C., Høiby, N., and Molin, S.
667 (2008). In situ growth rates and biofilm development of Pseudomonas aeruginosa
668 populations in chronic lung infections. J. Bacteriol. *190*, 2767–2776.
669 Yang, L., Jelsbak, L., Marvig, R.L., Damkiær, S., Workman, C.T., Rau, M.H., Hansen, S.K.,
670 Folkesson, A., Johansen, H.K., Ciofu, O., et al. (2011). Evolutionary dynamics of bacteria in a
671 human host environment. Proc. Natl. Acad. Sci. U. S. A. *108*, 7481–7486.
672 Zhu, H. (2017). kableExtra: Construct Complex Table with 'kable' and Pipe Syntax.
673

674 # FIGURE TITLES AND LEGENDS

675 ## Figure 1. Workflow and study setup.

676 **Upper panel:** Every month, CF patients are seen at the CF clinic at Rigshospitalet in
677 Copenhagen, Denmark. Here they deliver a sputum or endolaryngeal suction sample where
678 selective microbiological culturing is performed (Johansen et al., 2008). The longitudinally
679 collected isolates have been genome sequenced and analyzed in a previous study by Marvig
680 *et al.* (Marvig et al., 2015). **Middle panel:** Longitudinally collected isolates have been
681 subjected to different phenotypic analyses for this study and are here (**lower panel**)
682 analyzed using two data modelling approaches: Archetype Analysis (AA) and Generalized
683 Additive Mixed Model (GAMM). By integrating these approaches, we map dominant

684 evolutionary trajectories and analyze mechanistic links between phenotypic and genetic
685 adaptation.
686
687 ## Figure 2. Phenotypic characterization.
688 We present **(A)** summary statistics of our phenotype screen as well as compare the **(B)**
689 expected adaptation over time based on field consensus versus **(C)** a simple evaluation of
690 our raw measurements by colonization time. After sorting the isolates (x-axis) by the time
691 since colonization of a specific lineage or "colonization time" (ColT), it is still difficult to see
692 consistent patterns of phenotypic change over time. Colors are linked with the expected
693 change of the specific phenotype (B), so that blue denotes a "naïve" phenotype and red
694 denotes an "evolved" phenotype. For growth rate (in artificial sputum medium (ASM)),
695 adhesion, and aggregation, naïve and evolved phenotypes are roughly divided by
696 comparison with the reference isolate PAO1 phenotype. For aztreonam and ciprofloxacin
697 MIC, naïve and evolved phenotypes are based on sensitivity or resistance as indicated by
698 the EUCAST breakpoint values as of March 2017, respectively.
699
700 ## Figure 3. AA and GAMM models.
701 We present a summary of the models underpinning our study of pathogen adaptation. **(A)**
702 Screeplot showing the average residual sum of squares (RSS) for 25 iterations of each fit of a
703 given number of archetypes. The "elbow" of the plot indicates that six archetypes are
704 sufficient to model our dataset. **(B)** Characteristic trait profiles describing the 5 distinct
705 phenotype levels that each of our 6 archetypes represents. We use the following
706 abbreviations to represent our normalized data: grASM – growth rate in ASM, agg –
707 Aggregation, adh – Adhesion, azt – aztreonam susceptibility, cip – ciprofloxacin
708 susceptibility. **(C)** Simplex plot of the AA showing the six archetypes (A1-A6) sorted by their
709 characteristic growth rate (A3 and A5 vs A2 and A6), decreased sensitivity towards
710 ciprofloxacin (A1 and A6), and increased aggregation and adhesion (A2 and A4). All further
711 simplex visualizations are also sorted accordingly and can be interpreted using this key,
712 which is annotated with the extreme phenotype values for each archetype. The complete
713 analysis can be found in Supplementary material 1. **(D)** P-values for GAMM models with
714 multiple explanatory variables (columns) for the six predictor variables (rows), after model

715 reduction. P-values are only shown for explanatory variables that showed a significant (p-
716 value<0.01) impact on the predictor in question. The complete analysis can be found in
717 Supplementary material 2.
718

719 ## Figure 4. Rapid early adaptation.

720 We present specific GAMM and AA models to illustrate the rapid adaptation of growth rate
721 and ciprofloxacin over time and contrast these patterns with genetic adaptation via the
722 accumulation of nonsynonymous mutations. Here, we use GAMMs to illustrate the
723 significant impact of the explanatory variable colonization time on **(A)** growth rate in ASM,
724 **(B)** ciprofloxacin sensitivity in ASM, **(E)** the accumulation of all mutations (orange) and
725 nonsynonymous SNPs (blue) and indels (insertions and deletions). We use simplex
726 visualizations of AA to show **(C)** "naïve" trait alignment of the first isolate of the twenty
727 patients where we have analyzed the first *P. aeruginosa* isolate ever cultured at the CF clinic
728 (blue circles) in contrast to "adapted" isolates that have been cultured at year 2-3 of
729 colonization (red squares, all patients of the dataset). We contrast this trait-based
730 ordination with **(D)** genetic adaptation, shown by a color overlay of the number of non-
731 synonymous mutations present in each isolate. Isolate 95 (purple circle) of the DK12 clone
732 type has a very high number of mutations (>100) because one isolate in that lineage (isolate
733 96) is very different from the remaining 11 isolates. For the GAMM analysis shown in Figure
734 4E, we filtered out the mutations from the errant DK12 96 single isolate that affected the
735 whole lineage. Hypermutators are marked by purple triangles. **(A/B/E notation)** GAMMs are
736 illustrated by solid smoothed trendlines, dashed two standard error bounds, and gray points
737 as residuals. Y-axes are labelled by the predictor variable on which the effect of colonization
738 time of the clone type ("ColT") has been estimated as well as the estimated degrees of
739 freedom (edf) (for the E upper panel the edf is ordered as all mutations/NS SNPs). Residuals
740 have not been plotted in the upper panel of **(E)** for clarity reasons. X-axes are the ColT in
741 years and patients are included as random smooths together with ColT. A rug plot is also
742 visible on the x-axis to indicate the density of observations over time.
743

744 Figure 5. Mechanistic links between phenotypic changes and mutations in
745 ciprofloxacin resistance genes and the *retS/gacAS/rsmA* system.
746 We use AA to illustrate phenotypic separation by isolates affected by distinct mutations in
747 ciprofloxacin resistance genes *gyrA*, *gyrB*, and *nfxB* and the *retS/gacAS/rsmA* regulatory
748 system. **(A-B, left panel)** As visualized by AA simplex plots, the diversity of trait profiles
749 associated with isolates with mutations in DNA gyrase (*gyrA/B*) is in stark contrast to the
750 constrained band of *nfxB*-mutated isolates. Mutations in DNA gyrase and *nfxB* do not co-
751 occur in the same isolate but co-occur in different isolates of 2 lineages (patient P8804,
752 genotype DK08 and patient P8203, genotype DK32). The differences in time of appearance
753 during the colonization period and persistence of *gyrA/B* mutant isolates versus *nfxB*
754 mutant isolates is shown in the lineage timelines plotted in the right column for *gyrA/B* **(A,**
755 **right panel)** versus *nfxB* **(B, right panel)**. Furthermore, *gyrB*-mutated isolates cluster more
756 closely with A2 and A4 than *gyrA* mutated isolates, indicating a potential association with
757 adhesion; GAMMs predicts that *gyrB* mutation has a significant impact on adhesion (GAMM,
758 p-value << 0.01). **(C, left panel)** Mutations in the *retS/gacAS/rsmA* system shows a clear
759 phenotypic change when *retS* is mutated alone (blue circles) or in combination with *gacA* or
760 *gacS* (red squares and circles). The associated lineage plot **(C, right panel)** shows the
761 appearance of double mutations (*retS* + *gacA/S*) after a colonization period by *retS* mutated
762 isolates in three patient lineages. **(A/B/C – lineage plot notation)** Lineage length is based on
763 the span of time for which we have collected isolates and is indicated by gray bracketed
764 lines, with only isolates affected by a mutation of interest plotted using shape to indicate
765 mutation type. Symbol color indicates the specific mutation location in the affected gene
766 and **(A/B only)** symbol size indicates the level of resistance to ciprofloxacin. Multiple
767 isolates may be collected at the same sampling date based on differences in colony
768 morphology, which explains the visible vertical overlap of isolates for some lineages.
769

770 Figure 6. Evolutionary trajectories guided by different adaptation needs.
771 We present four different trajectories showing modes of evolution found in multiple
772 patients: **(A)** A special case of directed evolution with one outlier isolate (isolate 96 of DK12)
773 but an otherwise clear trajectory first towards ciprofloxacin resistance and afterwards a gain
774 in adhesive capabilities. **(B)** Diverse early/naïve isolates showing a population moving in a

775 broad and diverse plane from naïve archetypes towards evolved archetypes. **(C)** A
776 constrained convergence driven primarily by changes of a single phenotypic trait (decreased
777 ciprofloxacin sensitivity). **(D)** A diverse population with no clear evolutionary trajectory.
778

779 # METHODS

780 ## The isolate collection

781 The current isolate library is comprised of 443 longitudinally collected single *P. aeruginosa*
782 isolates distributed within 52 clone types collected from 39 young CF patients treated at the
783 Copenhagen CF Centre at Rigshospitalet (median age at first *P. aeruginosa* isolate = 8.1
784 years, range = 1.4-24.1 years, median coverage of colonization: 4.6 years, range: 0.2-10.2
785 years). This collection is a complement to and extension of the collection previously
786 published (Marvig et al., 2015) and captures the period of initial rapid adaptation (Barrick et
787 al., 2009; Woods et al., 2011; Yang et al., 2011), with 389 isolates of the previously
788 published collection included here in addition to 54 new isolates. For our study of
789 phenotypic evolution over time using GAMMs, we only included isolates from clone types
790 that were capable of creating a persistent mono-clonal colonization or infection, and
791 therefore two patients with a sustained multi-clonal infection were excluded. However, we
792 included four patients (P9904, P0405, P5504, and P2204) that show clone type substitution
793 during the collection period. We also excluded isolates belonging to clone types present in a
794 patient at two or fewer time-points, unless the two time-points were sampled more than 6
795 months apart. We also excluded any isolate with any missing phenotype measurement from
796 our panel of phenotype screens. 389 of the isolates have been clone typed as a part of our
797 prior phylogenetics study (Marvig et al., 2015) and the remaining isolates have been clone
798 typed as a step of the routine analysis at the Department of Clinical Microbiology at
799 Rigshospitalet. This clone type identification was performed as described previously (Marvig
800 et al., 2015), and the sequencing was carried out as follows: DNA was purified from over-
801 night liquid cultures of single colonies using the DNEasy Blood and Tissue Kit (Qiagen),
802 libraries were done with Nextera XT and sequenced on an Illumina MiSeq using the v2
803 250x2 kit.
804

## Ethics approval and consent to participate

The local ethics committee at the Capital Region of Denmark (Region Hovedstaden) approved the use of the stored *P. aeruginosa* isolates: registration number H-4-2015-FSP.

## Phenotypic characterizations

For all phenotypes, four technical replicates were produced for each isolate. For all but the antibiotic MIC tests, phenotypic analysis was carried out by stabbing from a 96 well plate pre-frozen with overnight cultures diluted with 50% glycerol at a ratio of 1:1.

### Growth rate in Lysogeny broth (LB) and Artificial sputum medium (ASM)

Isolates were re-grown from frozen in 96 well plates in 150ul media (LB or ASM) and incubated for 20h at 37°C with $OD_{630nm}$ measurements every 20 min on an ELISA reader. Microtiter plates were constantly shaking at 150 rpm. LB growth rates were first assessed by manual fitting of a line to the exponential phase of the growth curve. This dataset was then used to confirm the accuracy of R code that calculated the fastest growth rate from each growth curve using a "sliding window" approach where a line was fit to a 3-9 timepoint interval based on the level of noise in the entire curve (higher levels of noise triggered a larger window to smooth the fit). To develop an automated method of analyzing the ASM growth curves, which are much more noisy and irregular than the LB growth curves across the collection, we used standardized metrics for identifying problematic curves that we then also evaluated visually. Curves with a maximum OD increase of less than 0.05 were discarded as non-growing. Curves with linear fits with an $R^2$ of less than 0.7 were discarded as non-analyzable, and a small number of outlier curves (defined as curves analyzed for growth rates of 1.5 times the mean strain growth rate) were also discarded. Examples of our analyzed curves are shown in Figure S7 and all visualizations are available upon request.

## "Adherence" measures

The ability to form biofilm is a complex trait that is impacted by multiple factors, such as the production of polysaccharides, motility, and the ability to adhere (Hentzer et al., 2001; O'Toole and Kolter, 1998; Ryder et al., 2007). In this study, we have measured adhesion to peg-lids and estimated the ability to make aggregates – both traits have been linked with an

836 isolate's ability to make biofilm (Déziel et al., 2001; Kragh et al., 2016). Because of this, we
837 are using these two measures as an estimate of our isolates' ability to make biofilm.
838 However, because we are aware of the complexity of the actual biofilm-forming phenotype,
839 we have chosen to refer to this adhesion/aggregation phenotype as "adherence" and not
840 "biofilm formation".
841
842 *Adhesion in LB.* Adhesion was estimated by measuring attachment to NUNC peg lids.
843 Isolates were re-grown in 96 well plates with 150µl medium where peg lids were used
844 instead of the standard plate lids. The isolates were incubated for 20 hours at 37°C, after
845 which $OD_{600nm}$ was measured and subsequently, the peg lids were washed in a "washing
846 microtiter plate" with 180µl PBS to remove non-adhering cells. The peg lids were then
847 transferred to a microtiter plate containing 160µl 0.01% crystal violet (CV) and left to stain
848 for 15 min. The lids were then washed again three times in three individual "washing
849 microtiter plates" with 180µl PBS to remove unbound crystal violet. To measure the
850 adhesion, the peg lids were transferred to a microtiter plate containing 180µl 99% ethanol,
851 causing the adhering CV stained cells to detach from the peg lid. This final plate was used
852 for measurements using an ELISA reader, measuring the CV density at $OD_{590nm}$. (Microtiter
853 plates were bought at Fisher Scientific, NUNC Cat no. 167008, peg lids cat no. 445497)
854
855 *Aggregation in ASM.* Aggregation in each well was first screened by visual inspection of
856 wells during growth assays in LB and ASM and by evaluation of noise in the growth curves,
857 resulting in a binary metric of "aggregating" versus "not aggregating". However, to
858 incorporate this trait in our archetype analysis, we needed to develop a continuous metric
859 of aggregation. Based on the above manual assessment, we developed a metric based on
860 the average noise of each strain's growth curves. While we tested several different metrics
861 based on curve variance, the metric that seemed to delineate isolates according to the
862 binary aggregation measure most successfully was based on a sum of the amount of every
863 decrease in OD that was followed by a recovery at the next time point (versus the expected
864 increase in exponential phase and flatline in stationary phase). This value was normalized by
865 the increase in OD across the whole growth curve, to ensure that significant, irregular
866 swings stood out with respect to overall growth. This metric therefore specifically accounts
867 for fluctuation - both a limited number of large fluctuations in $OD_{630nm}$ (often seen in ASM

during stationary phase) as well as smaller but significant fluctuations across the entire curve (i.e. sustained irregular growth). While an imperfect assay of aggregation compared to available experimental methods (Caceres et al., 2014), this high-throughput aggregation estimate showed a significant relationship with adhesion when analyzed with GAMMs (Figure 3D), supporting its potential as a measure of adherence-linked behavior. We show examples of the measurement and comparison with binary aggregation data in Figures S7-8.

### Protease production

Protease activity was determined using 20x20 cm squared LB plates supplemented with 1.5% skim milk. From a "master microtitre plate", cells were spotted onto the square plate using a 96 well replicator. Colonies were allowed to grow for 48h at 37°C before protease activity, showing as a clearing zone in the agar, was read as presence/absence.

### Mucoidity

Mucoidity was determined using 20x20 cm squared LB plates supplemented with 25 ug/ml ampicillin. From a "master microtitre plate", cells were spotted onto the square plate using a 96 well replicator. Colonies were allowed to grow for 48h at 37°C before microscopy of colony morphologies using a 1.25x air Leica objective. By this visual inspection, it was determined if a colony was mucoid or non-mucoid.

### MIC determination of ciprofloxacin and aztreonam

MICs were determined by E-tests where a suspension of each isolate (0.5 McFarland standard) was inoculated on 14 cm-diameter Mueller-Hinton agar plates (State Serum Institute, Hillerød, Denmark), where after MIC E-Test Strips were placed on the plate in accordance with the manufacturer's instructions (Liofilchem®, Italy). The antimicrobial concentrations of the E-tests were 0.016-256µg/ml for aztreonam and 0.002-32µg/ml for ciprofloxacin.

### Construction of *gyrA/B* mutants

Four *P. aeruginosa* PAO1 mutants carrying point mutations in *gyrA* and *gyrB* were constructed: PAO1::$gyrA^{G259A}$, PAO1::$gyrA^{C248T}$, PAO1::$gyrB^{C1397T}$ , and PAO1::$gyrB^{G1405T}$. A

recombineering protocol optimized for *Pseudomonas* was adapted from Ricaurte *et al.* (2017)(Ricaurte et al., 2018). A PAO1 strain carrying a pSEVA658-ssr plasmid(Aparicio et al., 2018) expressing the recombinase *ssr* was grown to exponential phase with 250 rpm shaking at 37°C. Bacteria were then induced with 3-methylbenzoate and electroporated with recombineering oligonucleotides. Cells were inoculated in 5 ml of glycerol-free Terrific Broth (TB) and allowed to recover overnight at 37C with shaking. $Cip^R$ colonies were identified after streaking on a Cip-LB plate (0.25 mg $L^{-1}$) and sent for sequencing after colony PCR.

Each recombineering oligonucleotide contained 45 base pair homology regions flanking the nucleotide to be edited. Oligonucleotides were designed to bind to the lagging strand of the replichore of both genes and to introduce the mismatch in each mutation: G259A and C248T in *gyrA*, and C1937T and G1405T in *gyrB*, respectively. The recombineering nucleotides used are the following: (Rec_gyrA_G259A - G*C*ATGTAGCGCAGCGAGAACGGCTGCGCCATGCGCACGATGGTGTtGTAGACCGCGGTGTCGCC GTGCGGGTGGTACTTACCGATCACG*T*C; Rec_gyrA_C248T - A*G*CGAGAACGGCTGCGCCATGCGCACGATGGTGTCGTAGACCGCGaTGTCGCCGTGCGGGTGGT ACTTACCGATCACGTCGCCGACCAC*A*C; Rec_gyrB_C1397T - C*C*GATGCCACAGCCCAGGGCGGTGATCAGCGTACCGACCTCCTGGaAGGAGAGCATCTTGTCGA AGCGCGCCTTTTCGACGTTGAGGAT*C*T; Rec_gyrB_G1405T C*C*TCGCGGCCGATGCCACAGCCCAGGGCGGTGATCAGCGTACCGAaCTCCTGGGAGGAGAGCAT CTTGTCGAAGCGCGCCTTTTCGACG*T*T).

## Modeling of phenotypic evolution

To identify patterns of phenotypic adaptation while limiting necessary model assumptions that might bias our predictions, we chose to implement generalized additive mixed models (GAMMs), where the assumptions are that functions are additive and the components are smooth. These models allow us to account for patient-specific effects, thereby enabling us to identify trends in phenotypic adaptation across different genetic lineages and different host environments. Furthermore, to be able to simultaneously assess multiple phenotypes of each isolate from a systems perspective, we implemented archetype analysis (AA), where

each isolate is mapped according to its similarity to extremes, or archetypes, fitted on the boundaries of the multi-dimensional phenotypic space. This modeling approach allows us to predict the number and characteristics of these archetypes and furthermore identify distinctive evolutionary trajectories that emerge from longitudinal analysis of fitted isolates for each patient.

For all analyses, the time of infection is defined within each lineage as the time since the clone type of interest was first discovered in the patient in question. This is biased in the sense that the time since colonization can only be calculated from the first sequenced isolate of a patient. However, we have collected and sequenced the first isolate that has ever been cultured in the clinic for 20 out of the 39 patients.

Normalization of phenotypic values were carried out the following way for both AA and GAMM: ciprofloxacin and aztreonam MICs were normalized by dividing the raw MICs with the breakpoint values from EUCAST: ciprofloxacin breakpoint value: >0.5 µg/ml, aztreonam breakpoint value: >16 µg/ml (EUCAST update 13. March 2017). This results in values above one equaling resistance and equal to or below one equaling sensitive. The response and the explanatory variables were log2 transformed to get a better model fit for ciprofloxacin MIC, aztreonam MIC, Adhesion, and Aggregation. For the AA, Adhesion, Aggregation and growth rate in ASM was further normalized (before log2 transformation) by scaling the values by the values of the laboratory strain PAO1 such that zero was equivalent to the PAO1 phenotype measurement or the EUCAST MIC breakpoint. PAO1 was chosen to be the reference point of “wild type” phenotypes.

Because the mutations identified in our collection are based on our previous study (Marvig et al., 2015) where mutations were called within the different clone types, we added a second filtering step to identify mutation accumulation within patients. The second filtering step removed mutations present in all isolates of a lineage (a clone type within a specific patient) from the analysis.

All statistics were carried out in R (Team, 2017) using the packages *mgcv* (Wood, 2011; Wood et al., 2016) for the GAMM analysis and *archetype* (Eugster and Leisch, 2009, 2011;

962 Seth and Eugster, 2014) for the AA. Complementary packages used for analysis are:
963 *tidyverse* (Wickham, 2017), *itsadug* (van Rij et al., 2017), *ggthemes* (Arnold, 2017), *knitr* (Xie,
964 2017) and *kableExtra* (Zhu, 2017). We also referred to Thøgersen et al. (Thøgersen et al.,
965 2013) and Fernandez et al. (Fernandez et al., 2017) in the design of appropriate assessment
966 methods for the final AA model. We include two R markdown documents that explains our
967 modeling steps and further evaluation plots in detail (AA: Supplemental file 1, GAMM:
968 Supplemental file 2), and summarize our methods below in brief.
969
970 ## Data modeling
971 *Archetype analysis (AA).* We evaluated several different model fitting approaches by varying
972 the number and type of phenotypes modeled as well as the archetype number and fit
973 method, using RSS-based screeplots of stepped fits of differing archetype numbers,
974 explained sample variance (ESV), isolate distribution among archetypes, convex hull
975 projections of paired phenotypes (all combinations), and parallel coordinate plots as metrics
976 for choosing the best fit parameters and approach to accurately represent our data.
977 Ultimately, we focused on 5 continuous phenotypes correlated with growth (growth rate in
978 ASM), biofilm (adhesion and aggregation), and antibiotic resistance (aztreonam and
979 ciprofloxacin MICs), which also were linked to relevant findings provided by the GAMM
980 models. We used a root sum squared (RSS) versus archetype number screeplot of different
981 fits to determine that a 6 archetype fit would produce the optimal model for this dataset.
982
983 We then performed 500 simulations of a 100 iteration fit using the "robustArchetypes"
984 method (Eugster and Leisch, 2011), which reduces the impact of data outliers in fitting the
985 convex hull of the data. We evaluated the mean ESV and the number of isolates with an ESV
986 greater than 80% for the best model from each simulation in this study and differences in
987 archetype characteristics to assess convergence, ultimately selecting the model with the
988 second highest mean ESV (90.32%) and highest number of isolates with an ESV over 80%
989 (87.13%); this model also resembled the other 10 top models of the simulation study. The
990 order of archetypes around the simplex plot boundary is not reliant on the similarity of
991 archetype characteristics, so relationships between phenotypes are not always obvious. We
992 re-ordered the archetypes in the simplex plot by growth rate and secondarily antibiotic
993 resistance to improve clarity in the complex 6 archetype plot. This reordering was also

994 justified when projecting the archetypes onto a PCA plot of the phenotypes (Supplemental
995 file 1). All simplex plots have also had the 11 isolates with an ESV < 50% removed such that
996 we are not drawing any conclusions from these poorly fit data (they are shown via simplex
997 plot in the supplemental markdown).
998
999 *Generalized Additive Mixed Models (GAMMs).* For all phenotypes, GAMMs were used to
1000 identify evolutionary trends over time since first colonization. We correct for the patient
1001 environment and inconsistent sampling over time using a smooth random factor. Models
1002 were fitted in the following way: All continuously measured phenotypes included in the
1003 Archetype analysis were fitted as a response variable ("predicted" or "dependent" variable
1004 in Figure 3D) one-to-one, with both time as a "explanatory" or "independent" variable alone
1005 and combined with each of the phenotypes to account for potential time-dependence of
1006 the observations. Factorial/binary phenotypes were implemented as categorical functions
1007 and continuous phenotypes as smooth functions, allowing for non-parametric fits. Normally
1008 only one variable/phenotype of interest would be used as the predictor while other
1009 alterable variables or factors would be used as explanatory variables to explain or predict
1010 changes in the predictor. However, this requires a preconceived idea of a "one-way-
1011 relationship" where one variable (the predictor) is assumed to be affected by certain other
1012 variables (the explanatory variables), but where the explanatory variables cannot be
1013 affected by the predictor. By testing all phenotypes against each other, we avoid
1014 assumptions regarding the specific direction of relationships between the predictor variable
1015 and the explanatory variable. Furthermore, in using the GAMMs we prioritize accuracy of
1016 fitting but increase our risk of overfitting as a byproduct. We sought to counteract the risk
1017 of overfitting by the default penalization of fits inherent to the method used (Wood, 2011;
1018 Wood et al., 2016) and by model estimation via restricted maximum likelihood (REML)
1019 which has been found to be more robust against overfitting (Wood, 2006, 2011). When
1020 significant relationships were identified in one-to-one models (p-value < 0.05, as based on
1021 Wald-type tests as described in (Wood, 2006, 2013)), all significant explanatory variables
1022 were used to build a multi-trait model for the associated predictor. If select explanatory
1023 phenotypes were then identified as non-significant (p-value > 0.05) in the multi-trait model,
1024 they would be removed in a reduction step. To identify whether a reduced multi-trait model
1025 resulted in a better fit than the initial multi-trait model, a Chi-square test was carried out on

the models using the compareML function of the R package *itsadug* (van Rij et al., 2017) (Figure 3D). The specific models and additional information can be found in Supplemental file 2.

In demonstration of the utility of this approach, the multi-trait models of our 5 primary predictor phenotypes show that at least one explanatory phenotype has a statistically significant impact on the predictor phenotype. For all of the predictor phenotypes, multiple explanatory traits preserved significant impacts after model reduction steps (Figure 3D and Supplemental file 2). All mentions of significant relationships or correlations in the main text are obtained from the GAMM analyses with Wald-type test statistics presenting p-values < 0.01, unless otherwise stated. For information on deviance explained, R^2, and degrees of freedom for the individual models/variables, we refer to the Supplemental file 2.

## SUPPLEMENTARY INFORMATION

Supplemental File 1. Construction and assessment of the archetype model.

Supplemental File 2. Construction and assessment of the generalized additive mixed models.

Supplemental Information.

Figure S1. Hypermutators versus normomutators

Figure S2. *mucA* and *algU* mutants

Figure S3. mexZ mutants and drug efflux pumps

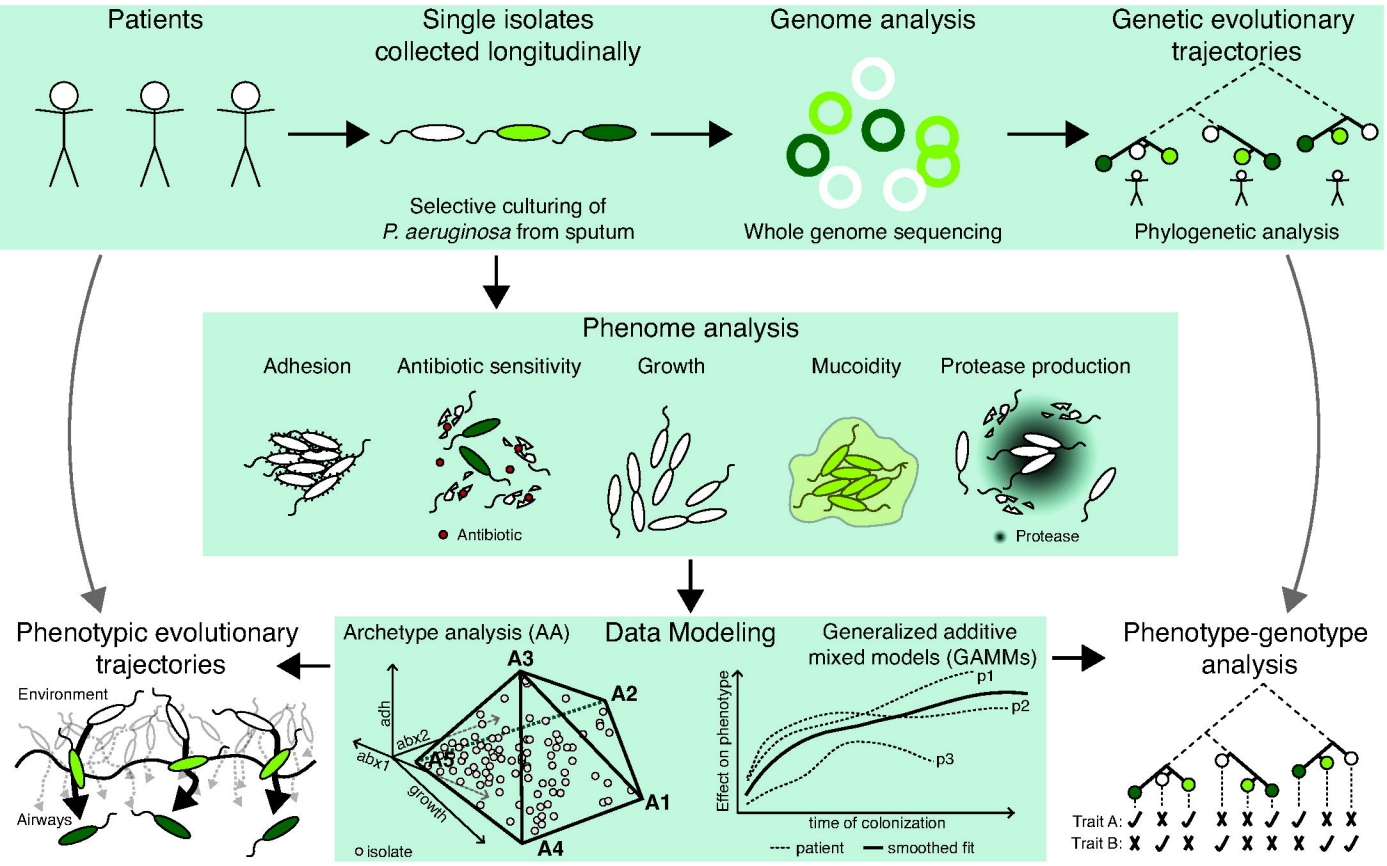
Figure S4. Specific mutations in *gyrA/B* by patient and adhesion

Figure S5. Adhesion of *gyrA/B* mutants (PAO1)

Figure S6. Generation time of *gyrA/B* mutants (PAO1)

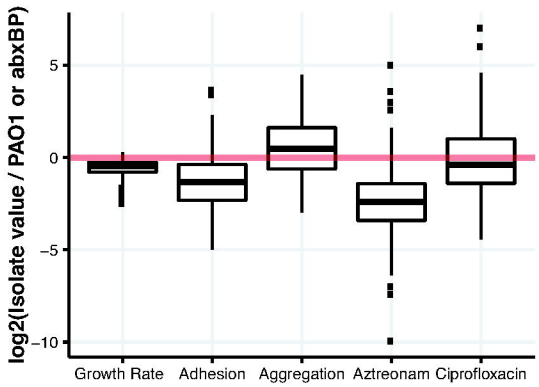
Figure S7. Example growth curves

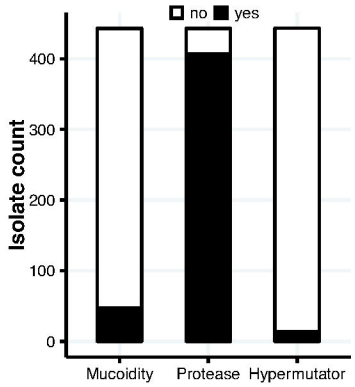
Figure S8. Development of an aggregation metric

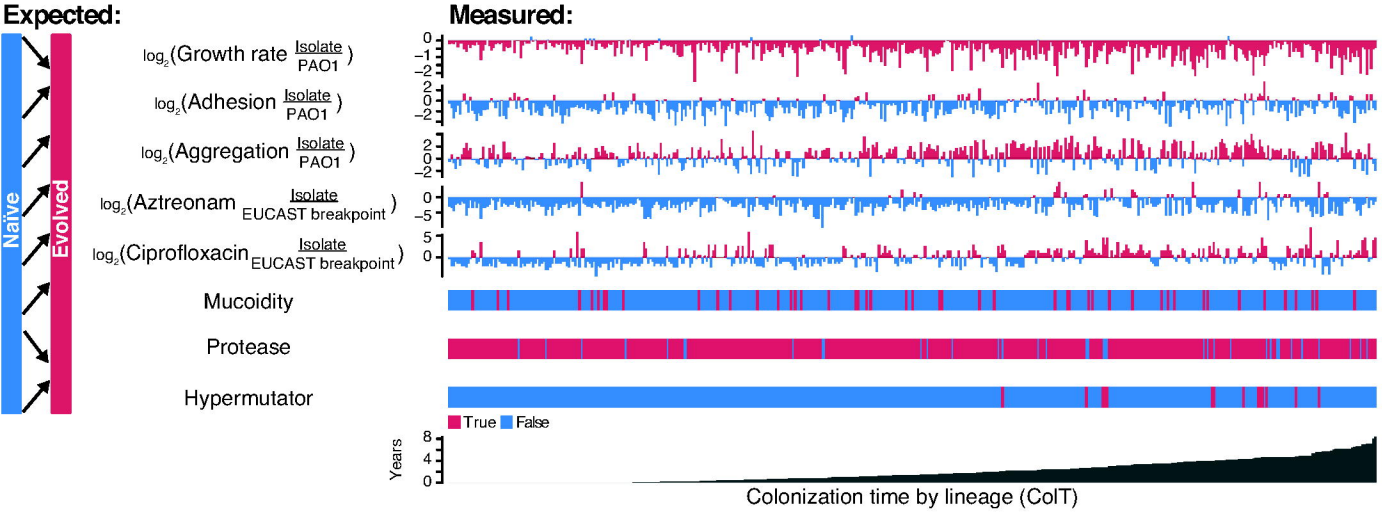


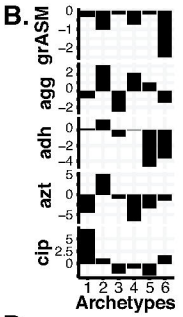
**Phenotype Screen**

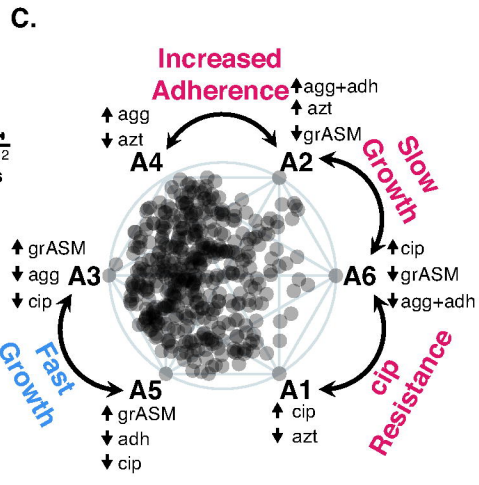
| | mean (sd) | PAO1 \| abxBP |
|---|---|---|
| *ASM Growth Rate, h−1* | 0.59 (0.17) | 0.85 |
| *Adhesion, OD590/OD630* | 1.69 (2.47) | 2.53 |
| *Aggregation, SDec/dOD* | 0.17 (0.2) | 0.07 |
| *Aztreonam, MIC* | 12.45 (55.66) | 16.00 |
| *Ciprofloxacin, MIC* | 1.27 (4.02) | 0.50 |
| *Mucoidity (yes = 1)* | 0.11 (0.31) | NA |
| *Protease (yes = 1)* | 0.92 (0.27) | NA |
| *Hypermutator (yes = 1)* | 0.03 (0.18) | NA |



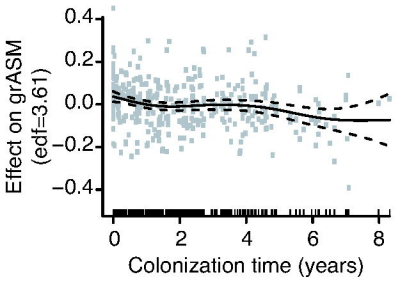

D.

**Predicted variable** / **Explanatory variables**

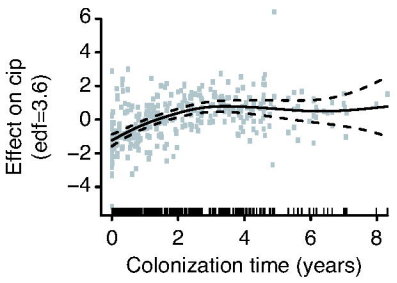
| | ColT | grASM | grLB | azt | adh | agg | prot | mut |
|---|---|---|---|---|---|---|---|---|
| **grASM** | 0.0062 | - | 2.0E-16 | - | NA | 6.8E-15 | - | - |
| **grLB** | 0.0003 | 2.0E-16 | - | NA | - | - | - | - |
| **azt** | - | - | - | - | NA | - | - | 0.0003 |
| **cip** | 2.7E-07 | - | - | 0.0072 | - | - | - | 0.0122 |
| **adh** | - | NA | 0.0063 | NA | - | 0.0001 | - | - |
| **agg** | - | 3.0E-08 | - | - | 0.0053 | - | 2.3E-05 | - |

*cip and mucoid as explanatory variables did not show any significant impact on any of the predictors.

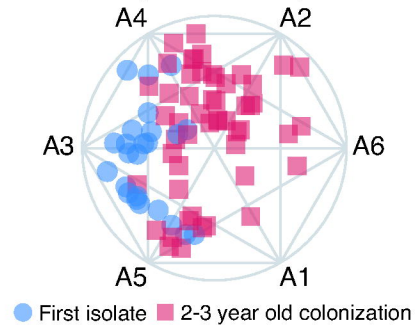
**A. Growth rate in ASM**



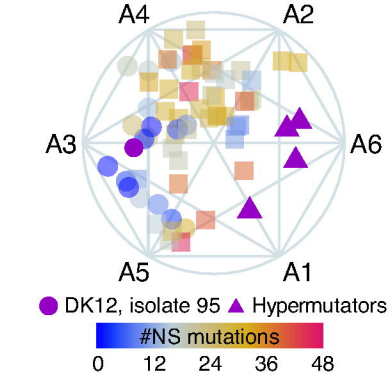

**B. Ciprofloxacin MIC**



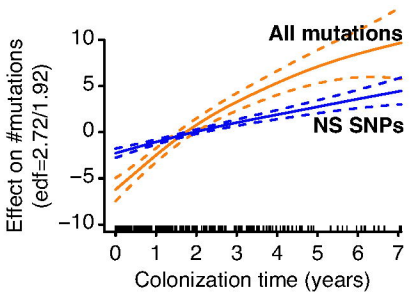

**C. Time-dependent adaptation**

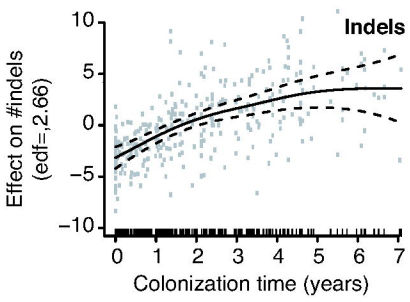


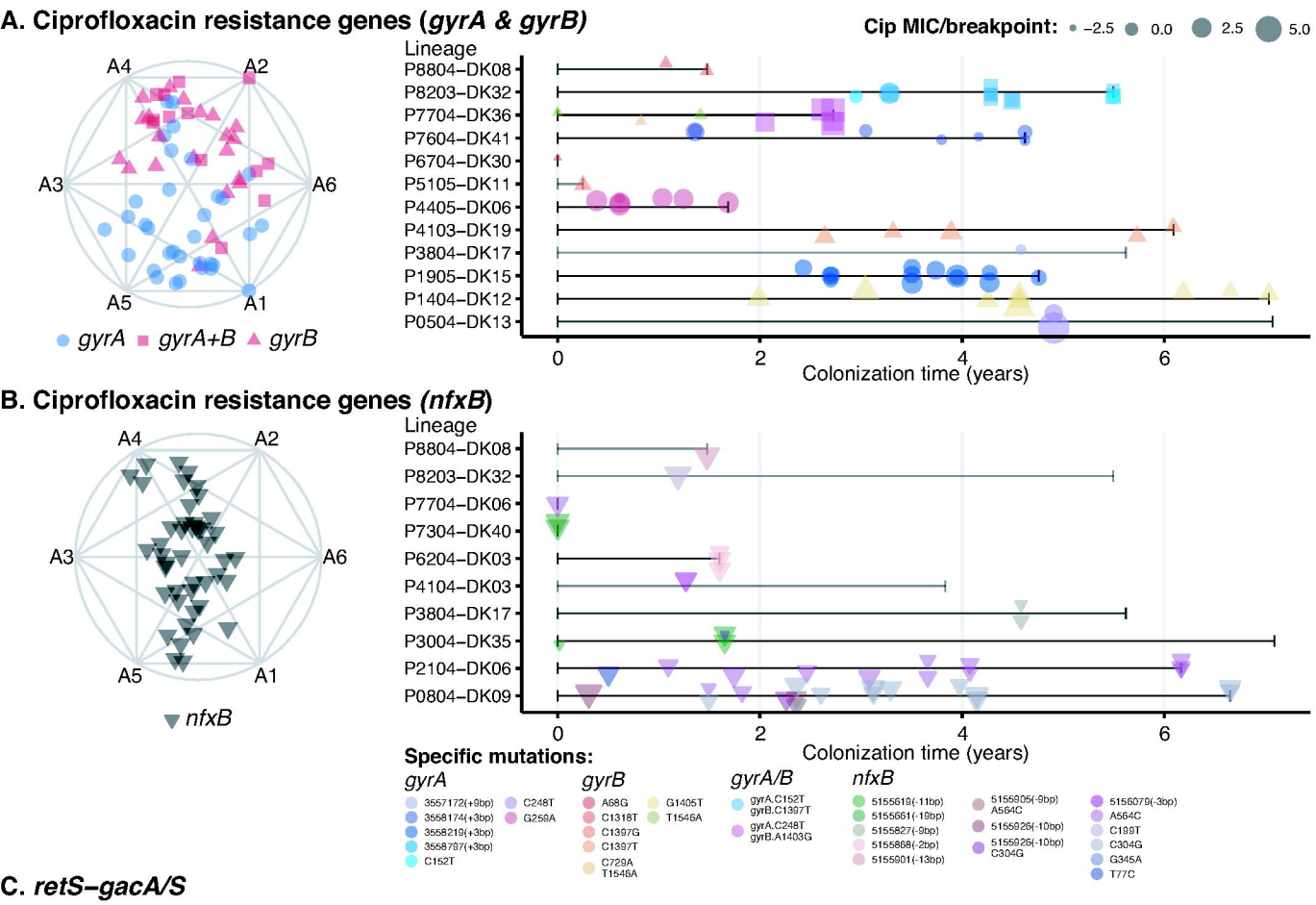

**D. Genetic adaptation**



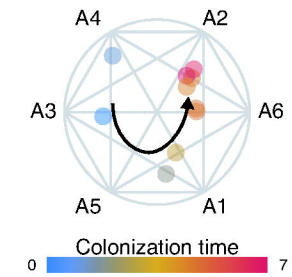

**E. Mutation counts**



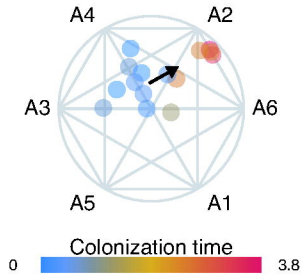





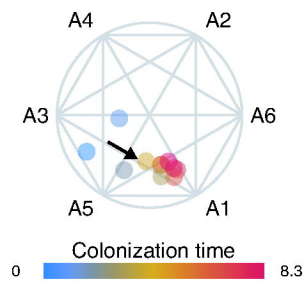
**A. A3/4 > A1 > A2/6,** P1404




**B. A3/4 > A2/6,** P4104




**C. A3/4 > A1/6,** P5304




**D. Diverse,** P0804