EVOLUTIONARY HIGHWAYS TO PERSISTENT BACTERIAL INFECTION

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

[§]Lead contact

KEYWORDS

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

ABSTRACT

Persistent infections require bacteria to evolve from their naïve colonization state by optimizing fitness in the host. This optimization involves coordinated adaptation of multiple traits, obscuring evolutionary trends and complicating infection management. Accordingly, we screen 8 infection-relevant phenotypes of 443 longitudinal *Pseudomonas aeruginosa* isolates from 39 young cystic fibrosis patients over 10 years. Using statistical modeling, we map evolutionary trajectories and identify trait correlations accounting for patient-specific influences. By integrating previous genetic analyses of 474 isolates, we provide a window into early adaptation to the host, finding: 1) a 2-3 year timeline of rapid adaptation after

colonization, 2) variant "naïve" and "adapted" states reflecting discordance between phenotypic and genetic adaptation, 3) adaptive trajectories leading to persistent infection via 3 distinct evolutionary modes, and 4) new associations between phenotypes and pathoadaptive mutations. Ultimately, we effectively deconvolute complex trait adaptation, offering a framework for evolutionary studies and precision medicine in clinical microbiology.

Bacteria have spent millennia evolving complex and resilient modes of adaptation to new environments, and some species effectively deploy these skills as pathogens during colonization and persistence within human hosts[1–3]. Due to gradual increases in fitness via accumulating genetic and epigenetic changes, it has been difficult to pinpoint overarching drivers of adaptation (from systems-level traits down to individual mutations) that reliably signal fitness [Leon et al. 2018]. Distinct populations may travel along the same predictable path to successful persistence, but other unique sequences of multi-trait adaptation can be equally optimal[4] in a complex, fluctuating environment[5]. This is even more relevant in a clinical context where dynamic selection pressures are applied via therapeutic treatment intended to eradicate infection.

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 competing modes of evolution. We know from laboratory evolution studies in highly controlled conditions that these multiple modes are at work and induce substantial phenotypic adaptation to minimal media within the initial 5,000-10,000 generations[6–8], but only an estimate is available of the timeline of adaptation in the complex CF lung environment[9]. Multiple recent studies have shown a high degree of population heterogeneity in chronic CF infections that could be influenced by competing evolutionary modes, but past consensus has been that select traits converge torwards similar “evolved” states during most CF infections (e.g. loss of virulence and increase in antibiotic resistance)[3,10–12]. This convergence can be complex and drug-driven, as recent studies have shown development of collateral sensitivity to antibiotics (treatment with one drug can induce reciprocal changes in sensitivity to other drugs)[13]; this illustrates that a single selection

pressure can reversibly affect multiple other traits, obscuring evolutionary trends. Bacterial infections of CF airways are thus influenced by strong and competing selective forces from very early in a patient's life, but few studies have focused on the early periods of infection where environmental strains transition to successful pathogens in patient lungs.

Studies have assessed the genetic evolution of human pathogens and identified specific genetic adaptations correlating with colonization and persistence[14–16]. However, only a few have linked genotypic and phenotypic changes[2,9,17,18], as this is especially challenging in natural populations. The genetic signature of adapting phenotypes is obscured over the course of evolution by the continuous accumulation of mutations and acclimatization by environment-based tuning of pathogen activity. Furthermore, it is inherently difficult to identify genotype-phenotype links for complex traits governed by multiple regulatory networks[19,20]. Consequently, we are far from the reliable prediction of phenotypic adaptation by mutations alone during evolution in a complex, dynamic environment[19,21], and we propose that for now, phenotypic characterization is equally important.

To address the complexity of pathogen adaptation in the host environment, we analyzed our phenotypic dataset using statistical methods that account for the environmental effects on patient-specific lineages (Generalized Additive Mixed Models – GAMMs) and assess adaptive paths traversing the evolutionary landscape from a multi-trait perspective (Archetype Analysis – AA). We identify emergent patterns of bacterial phenotypic change across our patient cohort that depart from expected evolutionary paths and estimate the period of initial rapid adaptation during which the bacteria transition from a "naïve" to an "evolved" phenotypic state. We further identify distinct and repeating trajectories of pathogen evolution, and by leveraging our prior genomics study of this isolate collection[16], we propose new associations between these phenotypic phenomena and genetic adaptation. We find that specific traits, such as growth rate and ciprofloxacin resistance, can serve as rough estimators of adaptation in our patients, while multi-trait modeling can map complex, patient-specific trajectories towards distinct evolutionary optimums that enable persistence. Implementation of this trajectory modeling as a diagnostic tool in patient care might enable clinicians to respond more quickly and effectively to evolving pathogens and inhibit the transition to a persistent infection.

RESULTS

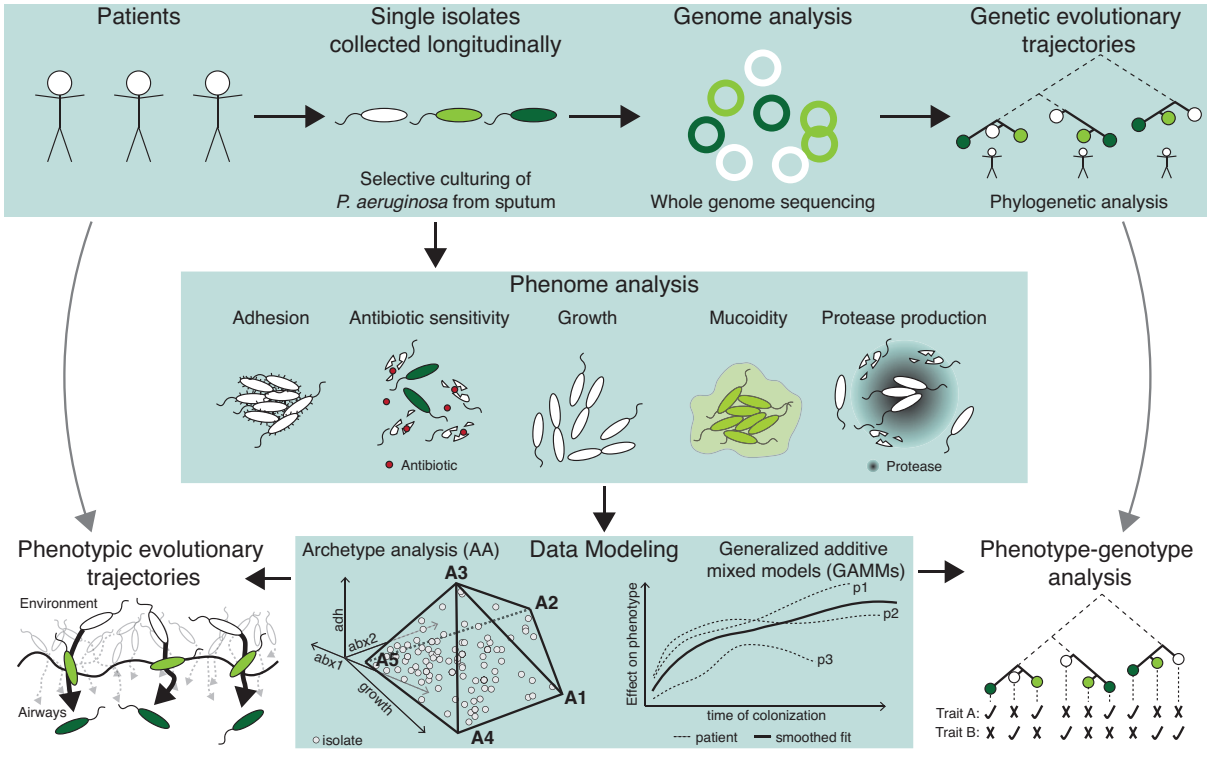


Figure 1. *Study design.* **Upper panel:** Every month, CF patients are seen at the CF clinic at Rigshospitalet in Copenhagen, Denmark. Here they deliver a sputum or endolaryngeal suction sample where selective microbiological culturing is performed[76]. The longitudinally collected isolates have been genome sequenced and analyzed previously[16]. **Middle panel:** Longitudinally collected isolates have been subjected to different phenotypic analyses for this study and are here (**lower panel**) analyzed using two data modelling approaches: Archetype Analysis (AA) and Generalized Additive Mixed Model (GAMM). By integrating these approaches, we map dominant evolutionary trajectories and analyze mechanistic links between phenotypic and genetic adaptation.

Evaluating pathogen adaptation in the early stage of infection

A unique dataset. The 443 clinical *P. aeruginosa* isolates originate from a cohort of 39 youth 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[16]. The "colonization time of an isolate" (ColT) is defined for each specific lineage, approximating the length of time since a given clone type began colonization of the CF airways in the specific patient. Importantly, our colonization time metric does not necessarily start at the true "time zero", since a significant bacterial load is necessary for a positive culture. Our isolate collection also does not capture the complete

105 population structure, but a previous study shows that 75% of our patients have a monoclonal
106 infection persisting for years with mutations accumulating in a highly parsimonious fashion
107 indicating unidirectional evolution[16]. Additionally, a metagenomic study of 4 patients from
108 our cohort indicates that the single longitudinal isolates are representative of the major
109 propagating subpopulation[22].
110
111 To obtain systems-level readouts of pathogen adaptation in the host and thereby assess
112 multi-trait evolutionary trajectories, we present an infection-relevant characterization of our
113 isolate collection entailing high-throughput measurements of 8 phenotypes: growth rates (in
114 Luria-Bertani broth (LB) and Artificial Sputum Medium (ASM)), antibiotic susceptibility (to
115 ciprofloxacin and aztreonam), virulence factors (protease production and mucoidity), and
116 adherence (adhesion and aggregation) (Figure 1 and 2). We define adherence as a shared
117 trend in adhesion and aggregation which we associate with a biofilm-like lifestyle (see
118 Methods for further discussion of limitations of these measures). These phenotypes are
119 generally accepted to change over the course of colonization and infection of CF airways
120 based primarily on studies of chronically-infected patients[10,17,23,24].
121
122 That is, an "evolved" isolate would grow slowly, adhere proficiently, be more likely to exhibit
123 a mucoid and/or hypermutator phenotype, have reduced protease production, and resist
124 antibiotics, in contrast to a "naïve" isolate (Figure 2B). However, simply ordering our
125 measurements by colonization time does not illustrate an overarching adaptive trajectory
126 from naïve to evolved phenotypes (Figure 2C). Instead, we see substantial heterogeneity, with
127 isolates that resemble both naïve and evolved phenotypic states throughout the study period.
128 Given that we are investigating a unique collection from a young patient cohort that we track
129 for a substantial period of colonization, this data fills the critical gap between studies of acute
130 infections and chronic infections[25]. We are surprised to see naïve phenotypes retained in late
131 colonization as well as isolates in early colonization that deviate significantly from PAO1
132 phenotypes. However, a general pattern of heterogeneity is in alignment with previous
133 studies of both *P. aeruginosa* and *Burkholderia* spp. infections[3,11,12].
134

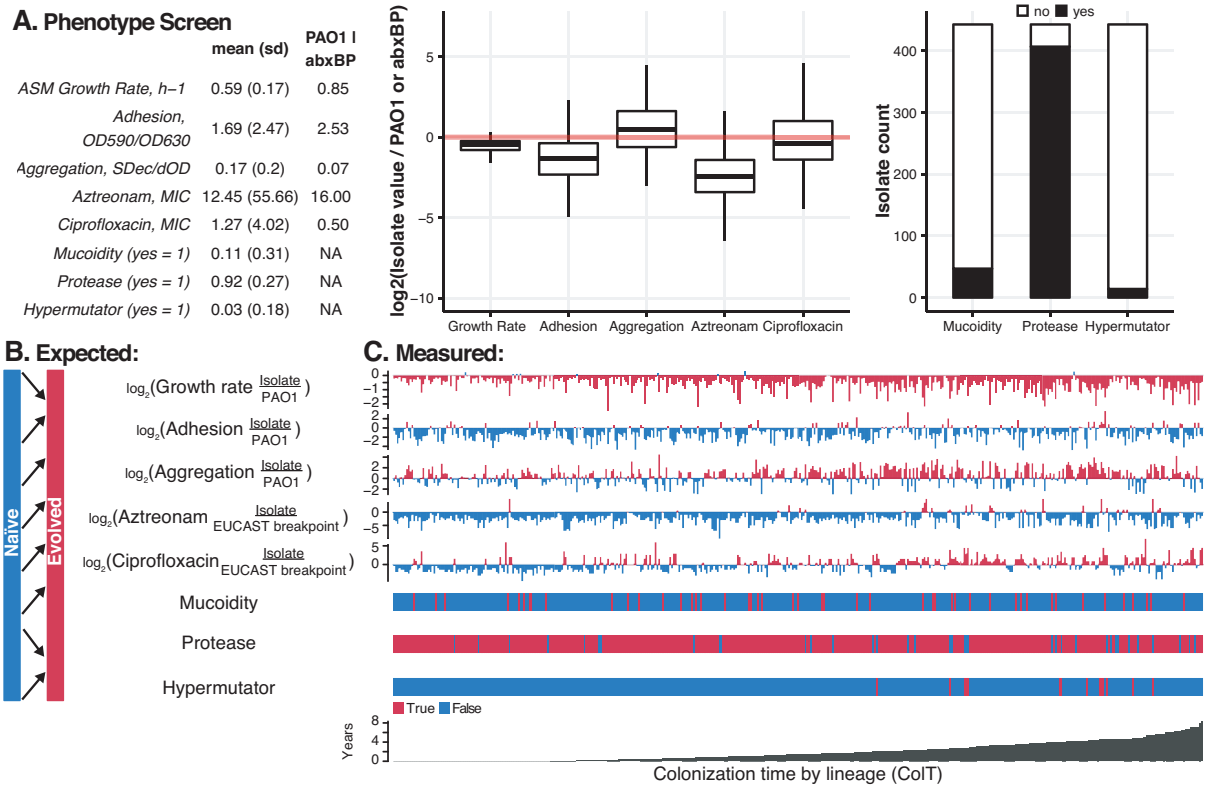


Figure 2. *Phenotypic characterization.* We present summary statistics of our phenotype screen including **(A)** mean and standard deviation of isolate data versus the *P. aeruginosa* PAO1 value or antibiotic breakpoint we use for normalization, respectively, as well as boxplots of continuous variables (showing the median, 1st and 3rd quartile hinges, whiskers extending from the hinges to the most extant value within 1.5x inter-quartile range, and outliers as points). We then compare the **(B)** expected adaptation over time based on field consensus versus **(C)** the measured raw adaptation of our isolate collection over time. After sorting the isolates (x-axis) by the time since colonization of a specific lineage or "colonization time" (ColT), it is still difficult to see consistent patterns of phenotypic change over time. Colors are linked with the expected change of the specific phenotype (B), so that blue denotes a "naïve" phenotype and red denotes an "evolved" phenotype. For growth rate (in artificial sputum medium (ASM)), adhesion, and aggregation, naïve and evolved phenotypes are roughly divided by comparison with the reference isolate PAO1 phenotype. For aztreonam and ciprofloxacin MIC, naïve and evolved phenotypes are based on sensitivity or resistance as indicated by the EUCAST breakpoint values as of March 2017.

A unique modeling approach. Because our data is heterogeneous, we required specialized modeling approaches to account for specific environmental pressures and assess the boundaries of the evolutionary landscape. Previous studies have employed linear mixed models of phenotypic adaptation[26], and employed archetype analysis in the comparison of features of transcriptomic adaptation by *P. aeruginosa*[27]. Similar studies of multi-trait evolutionary trade-offs using polytope fitting have predicted the genetic polymorphism structure in a population[28]. We use related modeling methods to ensure that patient-specific effects are minimized, irregular sampling intervals are smoothed and a multi-trait perspective is prioritized by 1) modeling the dynamic landscape of multi-trait evolution using AA and 2) evaluating temporal correlations of phenotypic adaptation by fitting cross-patient trendlines

using GAMMs (Figure 1). We describe our approach below in brief, with more extended explanation available in both the Methods and Supplements 1 and 2.

With AA, we want to assess multi-trait adaptive paths within the context of the evolutionary landscape. We map these paths (or trajectories) by first fitting idealized extreme isolates ("archetypes") located on the boundaries of the evolutionary landscape and then evaluating every other isolate according to its similarity to these idealized extremes. The archetypes are positioned at the "corners" of the principal convex hull (PCH), the polytope of minimal volume that effectively encapsulates our phenotype dataset[29] (Figure 1, bottom panel). We conceptualize archetypes as the "naïve" and "evolved" states of plausible adaptive trajectories and predict both the optimal number of archetypes and their distinct phenotypic profiles. We illustrate the AA by the 2D projection of our multi-trait model via a "simplex" plot, as shown in Figure 3C[30].

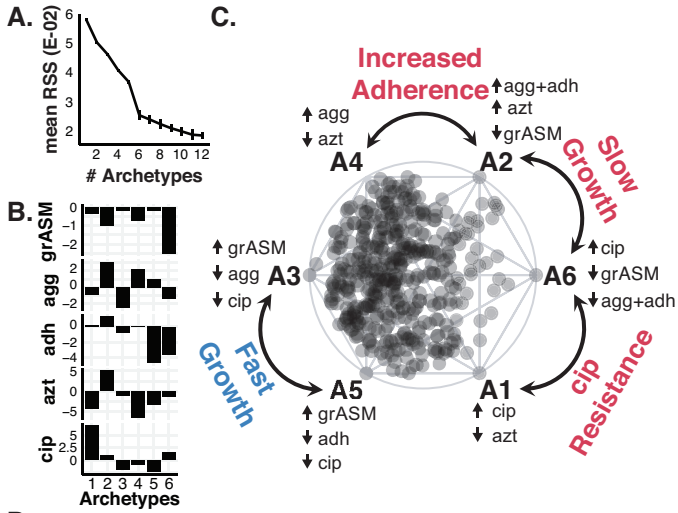
With the GAMMs, we want to predict whether a given phenotype (the "predicted" variable) significantly correlates with other phenotypes or time (the "explanatory" variables). To do this we need to account for the effects of patient specific environments and the effect of sampling time, while fitting trend lines for each trait (Figure 1, bottom panel). This is done by fitting patient and time as random effects; we reduce the risk of overfitting by using a penalized regression spline approach with smoothing optimization via restricted maximum likelihood (REML)[31]. To avoid assumptions of "cause-and-effect" between our variables, we permute through different one-to-one models of all phenotypes, and then reduce our models by combining only the statistically significant individual phenotypes into a multi-variable model. We further remove any phenotype that loses significance in the multi-variable model, assuming that it is correlated with a more impactful phenotype. From this point, all mentions of significance are obtained from the GAMM analyses with p-values < 0.01 based on Wald-type tests as described in[31,32], unless otherwise stated.

Revealing multi-trait adaptation on a cross-patient scale

AA predicted six distinctive archetypes sufficient to describe each isolate within the evolutionary landscape of 5 continuous traits as shown in Figure 3A. We use only growth rate

in ASM due to its correlation with growth rate in LB (Figure 3D). The simplex plot of Figure 3C highlights the standout features of each archetype by annotating according to the highest or lowest values for each phenotype across all archetype trait profiles (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), while two others accounted for slow-growing evolved archetypes (A2 and A6), in accordance with the accepted paradigm[10,24]. 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). This aligns with the infection stage of the patients included in this study. Importantly, we also find two regions in the simplex visualization which represent different focal points of adaptation: 1) an increase in adherence (A2 and A4) and 2) ciprofloxacin resistance (A1 and A6).

We also built a GAMM for each of our six continuous phenotypes to identify whether any of the other traits and time influenced it significantly across our patient cohort (Figure 3D). When evaluating adaptation of the specific phenotypes, we found that the colonization time had a significant impact on both growth rate and sensitivity to ciprofloxacin but did not significantly influence sensitivity to aztreonam (Figure 3C, Figure 4A and 4B), which is a reflection of the regular administration of ciprofloxacin but not aztreonam to our patients[33].



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.

Figure 3. *AA and GAMM models.* We present a summary of the models underpinning our study of pathogen adaptation. **(A)** Screeplot showing the average residual sum of squares (RSS) for 25 iterations of each fit of a given number of archetypes. The "elbow" of the plot indicates that six archetypes are sufficient to model our dataset. **(B)** Characteristic trait profiles describing the 5 distinct phenotype levels that each of our 6 archetypes represents. We use the following abbreviations to represent our normalized data: grASM – growth rate in ASM, agg – Aggregation, adh – Adhesion, azt – aztreonam susceptibility, cip – ciprofloxacin susceptibility. **(C)** Simplex plot of the AA showing the six archetypes (A1-A6) sorted by their characteristic growth rate (A3 and A5 vs A2 and A6), decreased sensitivity towards ciprofloxacin (A1 and A6), and increased aggregation and adhesion (A2 and A4). All further simplex visualizations are also sorted accordingly and can be interpreted using this key, which is annotated with the extreme phenotype values for each archetype. The complete analysis can be found in Supplementary material 1. **(D)** P-values for GAMM models with multiple explanatory variables (columns) for the six predictor variables (rows), after model reduction. P-values are only shown for explanatory variables that showed a significant (p-value<0.01) impact on the predictor in question. The complete analysis can be found in Supplementary material 2.

Phenotypic trends contrast with CF paradigms

An important distinction between AA and GAMMs is that many isolates clearly cluster in AA 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, though 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[4,34–36]. Furthermore, the biofilm-related metric of

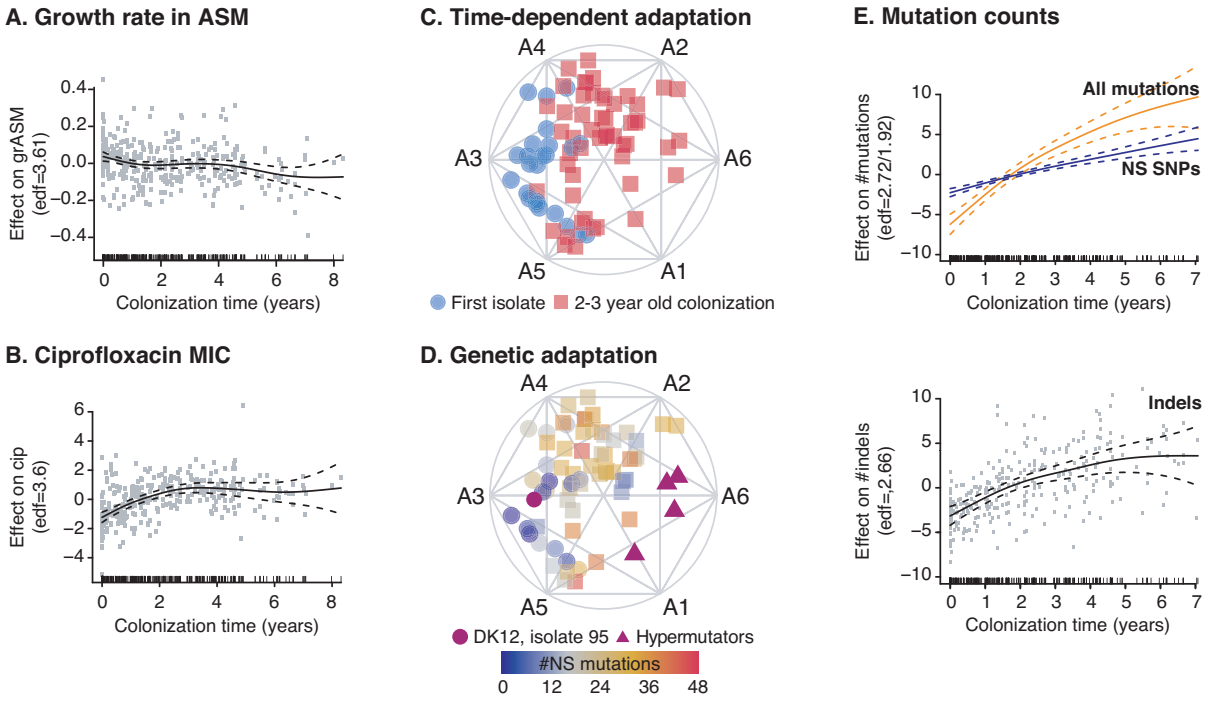
mucoidity does not significantly correlate with any other measured phenotype, despite its use as an important biomarker of chronic infection in the Copenhagen CF Centre[37]. We hypothesize that the rate of adaptation and relative benefit of this phenotype may vary significantly and be sensitive to temporal 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[38]. Together, these results prompt further reassessment of common assumptions regarding the evolutionary objectives of *P. aeruginosa* in CF infections.

Initial adaptation happens within 3 years of colonization

We find that the routes to successful persistence and a transition to chronic infection are initiated early in infection[16,39]. The GAMMs indicate that a substantial change occurs in both growth rate and ciprofloxacin susceptibility during the first 2-3 years (5256 - 7884 bacterial generations[23]) of colonization as shown by the slopes in this period (Figure 4A-B). Using AA, we also see a substantial shift from naïve towards evolved archetypes as shown by the broad distribution of isolates reaching the outer simplex boundaries by year 3 (Figure 4C), further confirming the rapid adaptation shown by the GAMMs. While the first isolate of each patient in our collection may not represent the true start of adaptation given sampling limitations, the window of rapid adaptation is still likely substantially contracted compared to the previous estimate of within 42,000 generations[9]. In fact, our data resembles the rate of fitness improvement found in the laboratory evolution study of *Escherichia coli* that showed change within the first 5,000-10,000 generations[6,7].

Interestingly, the four hypermutator isolates arising in the early adaptation window do not alone define the AA boundary, indicating that the acquisition of a high number of mutations does not explain all extreme phenotypes (Figure 4D, full dataset in Figure S1). To further evaluate parallels between phenotypic and genetic adaptation, we investigated the accumulation of nonsynonymous mutations in coordination with archetypal relationships (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. We observed that most of the first isolates with 0-30 mutations aligned with naïve archetypes, and 2-3-year-old

240 isolates with 9-48 mutations extended to the outer boundaries of adaptation (A2, A6, and A1)
241 (Figure 4C-D). We also observed the persistence of WT-like genotypes with few mutations
242 alongside evolved genotypes (Figure 4D). Thus, we find discordant molecular and phenotypic
243 adaptation from a multi-trait perspective.
244



245

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

246

247 When analyzing the entire dataset using GAMMs, we found a significant, near-linear
248 relationship between colonization time and the number of nonsynonymous SNPs, but

accumulation of all nonsynonymous mutations appears logarithmic with accumulation slowing after 2 years (Figure 4E). This behavior resembles that of the laboratory evolution of *E. coli* (propagated for more than 60,000 generations)[40], though accumulation may slow sooner in the CF lung. When we plot accumulation of indels alone, we see the likely driver of the logarithmic trend. When combined with the discordance found by AA, these findings support the theory that select beneficial mutations (for example, a highly impactful indel) can alone induce important phenotypic changes that improve fitness[41]. However, the likelihood of beneficial mutations presumably decreases over time as theorized previously[42] and other methods of adaptation also contribute, such as acclimation to the CF lung environment via gene expression changes[43,44].

Multi-trait analysis enables complex genotype-phenotype associations

The obscuring of genotype-phenotype links via polygenic effects and the possible pleiotropic effects of single mutations is difficult to resolve, especially when working with complex traits. However, we have a unique multi-dimensional perspective from which to map genotype-phenotype relationships. We previously identified 52 "pathoadaptive genes" - genes mutated more often than expected from genetic drift and thus assumed to confer an adaptive advantage during infection[16,45]. By overlaying nonsynonymous mutations on AA simplex plots, we evaluated the impact of mutation of the following pathoadaptive genes: 1) *mexZ* (the most frequently mutated gene) and other repressors of drug efflux pumps (*nfxB* and *nalD*), 2) mucoidity regulators *mucA* and *algU* and the hypothesized infection-state switching *retS*/*gacAS*/*rsmA* regulatory pathway previously examined from a genetic adaptation perspective[16,46], and 3) ciprofloxacin resistance genes *gyrA* and *gyrB*[47–49]. Isolates with *mexZ* mutations are broadly distributed by AA, so we analyzed *mexZ* mutants in combination with other pump repressor gene mutations. Even double-mutant isolates (grouped by efflux pump associations) 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). We saw no obvious spatial correlations with mutations linked to mucoidity regulation via AA (Figure S2), paralleling mucoidity's lack of significance in our GAMM analyses. However, the isolate distributions of *retS/gacAS/rsmA* and *gyrA/B* mutants were striking in their spatial segregation (Figure 5A-B).

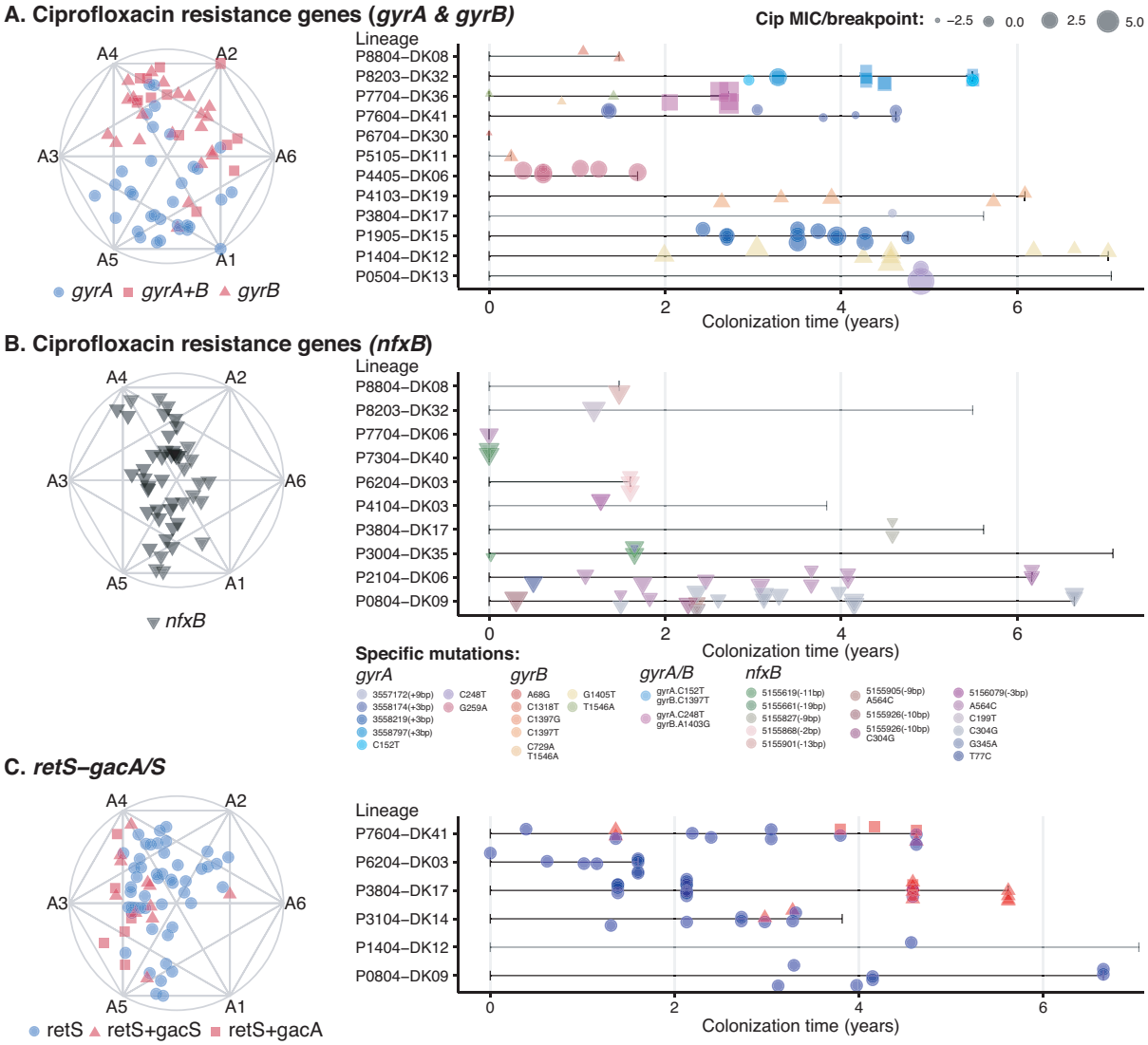


Figure 5. *Mechanistic links between phenotypic changes and mutations in ciprofloxacin resistance genes and the retS/gacAS/rsmA system.* We use AA to illustrate phenotypic separation by isolates affected by distinct mutations in ciprofloxacin resistance genes *gyrA*, *gyrB*, and *nfxB* and the *retS/gacAS/rsmA*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

Differential evolutionary potential via ciprofloxacin resistance mechanisms

The primary drivers of ciprofloxacin resistance in *P. aeruginosa* are theorized to be mutations in drug efflux pump repressor *nfxB* and the gyrase subunits *gyrA* and *gyrB* of the DNA replication system[47–49]. We would therefore expect isolates with mutations in these genes to cluster around archetypes A1 and A6 characterized by high ciprofloxacin minimal inhibitory concentrations (MICs) (Figure 3C). However, AA illustrates a broad distribution of *gyrA/B* mutants among archetypes, and a contrasting narrow distribution of *nfxB* mutants (Figure 5A-B, left panel). In association, we see a range of ciprofloxacin resistance levels associated with affected isolates both across and within patient lineages, and no dominant mutations/mutated regions repeating 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 fewer lineages compared to *gyrA/B* mutations. This likely contributes to *nfxB*'s distinctive band-like distribution via AA which suggests an evolutionary restriction associated with sustaining the mutation.

Interestingly, we noted that isolates with a *gyrB* mutation (22 isolates alone or 14 in concert with *gyrA* mutation) are concentrated closer to "biofilm-linked" archetypes A2 and A4 than isolates with only a *gyrA* mutation (33 isolates). To our knowledge, there is no direct relationship between *gyrB* and the capability to adhere[49]. This positive association of *gyrB* on adhesion was confirmed by GAMM, but when we moved the two SNPs affecting the most isolates in both *gyrA* and *gyrB* (2 lineages each, Figure S4) into lab strain *P. aeruginosa* 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 looked for co-occurring mutations in biofilm-linked genes in the *gyrB*-mutated lineages; for all but one lineage, there was no obvious explanation for increased adhesion. Ultimately, this association underlines the impact that genetic background and the multi-genetic signature of biofilm regulation can have on the identification of links between genotype and phenotype[50].

Infection trajectory reversal via a regulatory switch

The functional model of the *retS*/*gacA*/*gacS*/*rsmA* regulatory system is theorized to be a bimodal switch between acute and chronic infection phenotypes[46,51]. 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, promoting a chronic infection phenotype. However, our previous genetic analysis[16] unexpectedly showed that multiple evolving lineages gained a subsequent mutation in *gacA*/*S* that often 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 six patients with nonsynonymous mutations in this system have isolates which are *retS+gacA*/*S* double mutants (Figure 5C, right panel). While *retS* mutants resemble the evolved archetypes (A1 – 2 and A6), all but one double mutant 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 limited isolates and patients affected, we did not follow up with additional GAMM analyses of the effect of these mutations on different phenotypes.

This unexpected phenotypic reversion to an "acute infection state" does not easily reconcile with theories about persistence via convergence towards a "chronic phenotype". However, over time some patients are colonized by new clone types and/or other pathogens; this could require re-establishment of a colonization mid-infection and thus induce the population to revert towards an acute infection state where fast growth and motility improve its ability to compete.

Infections persist via distinct routes of adaptation

Given the above insights from lineage-based analysis, we further investigate lineage influences by mapping patient-specific adaptive trajectories. We find 3 overarching modes of evolution that *P. aeruginosa* can utilize to persist successfully in individual patients: 1) convergent evolution, 2) directed diversity or 3) general diversity. Figure 6A-D shows examples of adapting lineages employing these modes. We see rapid convergent evolution

towards an endpoint of ciprofloxacin resistance in patient P5304 (Figure 6A). Diverse isolates appear to move in the same general direction of increased adhesion and aggregation in patient P4104 (Figure 6B), which we term "directed diversity", while no directionality is apparent in the diverse isolates of the trajectory of patient P0804 (Figure 6C), which we term "general diversity". In the complex trajectory of patient P1404 (Figure 6D), the genotypic distinction of the young isolate near A4 indicates that the persisting sublineage initiates with the isolate near A3, after which it gains a *gyrB* mutation guiding the trajectory towards ciprofloxacin resistant A1. This mutation is retained during the subsequent shift towards A2, characterized by increased adherence and decreased sensitivity to aztreonam. 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.

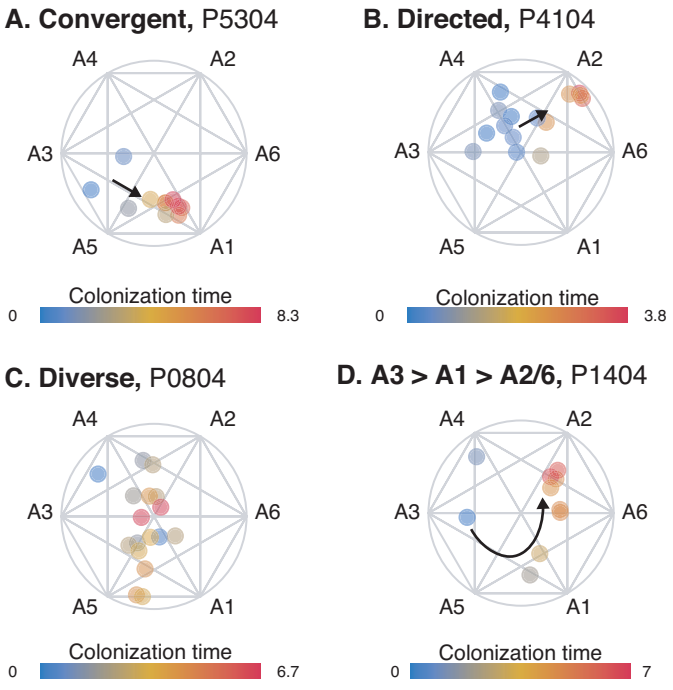


Figure 6. *Evolutionary trajectories guided by different adaptation needs.* We present four different trajectories showing modes of evolution found in multiple patients: **(A)** Convergent evolution driven primarily by changes of a single phenotypic trait (decreased ciprofloxacin sensitivity). **(B)** Directed diversity with early/naïve isolates showing a population moving in a broad and diverse plane from naïve archetypes towards evolved archetypes. **(C)** General diversity where the population has no clear evolutionary trajectory. **(D)** A special case of convergent evolution with one outlier isolate (isolate 96 of DK12) but an otherwise clear trajectory first towards ciprofloxacin resistance and afterwards a gain in adhesive capabilities.

Here, we draw specific examples from patients with high sampling resolution and at least 3 years of infection within our cohort, but to capture the full spectrum of evolutionary modes will require more uniform cross-cohort sampling that also addresses population dynamics as

362 well as the inclusion of more patients. With these expansions, we theorize that distinctive
363 evolutionary trajectories will correlate with infection persistence and patient outcomes.
364

365 DISCUSSION

366 Complex mutation patterns are an inherent byproduct of evolution and result in equally
367 complex adaptive trajectories that lead to persistence. Phenotype represents the cumulative
368 systems-level impacts of these mutation patterns. We therefore emphasize the value of
369 classical phenotype-based investigations as a highly relevant complement to genomics
370 approaches. By integrating these perspectives via our statistical modeling framework, it is
371 possible to identify consistent pan-cohort trends while illuminating complex patient-specific
372 patterns and their genetic drivers. This approach could also be valuable in assessing evolution-
373 based scenarios such as interpretation of laboratory evolution experiments, investigations of
374 long-term microbiome fluctuations and other studies of evolving clonal populations.
375

376 Our study identifies rapid phenotypic adaptation of isolates within the first few years of
377 colonization by both mutational accumulation and acclimation as indicated by the
378 discordance between genotypic and phenotypic adaptation. This resembles the findings from
379 the long-term laboratory evolution of *E. coli*[40]. While specific traits show cross-patient
380 convergence (growth rate and ciprofloxacin resistance), we highlight remarkable diversity
381 both within and across patients. In addition to convergent and directed evolution, we thus
382 emphasize the maintenance of general diversity as a useful evolutionary mode of persistence
383 as supported by prior observations of resilience in diverse populations[52–54]. Among our
384 patient-specific trajectories, we also find varying routes within these categories of evolution
385 that are used by different patient lineages to achieve successful persistence. These important
386 evolutionary findings can further be translated to the clinic. Although early aggressive
387 antibiotic therapy has been shown to substantially delay the transition to chronic infection[33],
388 we provide a valuable estimate of this narrow window based on analysis at high temporal
389 resolution. Furthermore, we provide a quantitative approach to monitoring infection state
390 via patient-specific trajectories which can offer important insights into bacterial response to
391 treatment.
392

Given that individual mutations may have pleiotropic effects and obscure genetic signatures as they accumulate over time[19], our study underlines the necessity of a multi-trait perspective. Our genotype-phenotype associations support the theory that specific mutations confer unique evolutionary restrictions to adaptive trajectories; these restrictions impact the fixation of other mutations or adaptation of other traits, but genetic background and host-specific evolutionary pressures influence the type and degree of restriction[8]. By mapping phenotypic trajectories, we can identify both genetic mechanisms that regulate these highways and complex traits that signal the impact of treatment on individual infections. In the future, we see particular promise in incorporating records of patient treatment and response to our assessment of adaptive trajectories to further guide clinicians and advance precision medicine in clinical microbiology.

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. 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.

DATA AND SOFTWARE AVAILABILITY

We provide our complete phenotype dataset in raw form as a supplemental spreadsheet and include a visualization and summary statistics of normalized data in Figure 2. Data normalization, processing and construction of all models was performed in R as described above and all essential code for reproduction of these steps is provided in R Markdown format in supplemental files 1-2. These files also include code for replicating the model visualizations of Figure 3A-D and Figure 4A-C, E. Code to reproduce various secondary analysis figures is available on request. All genomic information is publicly available as described in [16].

DECLARATION OF INTERESTS

The authors declare no competing financial interests.

REFERENCES

1. Flores-Mireles, A. L., Walker, J. N., Caparon, M. & Hultgren, S. J. Urinary tract infections: Epidemiology, mechanisms of infection and treatment options. *Nat. Rev. Microbiol.* **13,** 269–284 (2015).
2. Rau, M. H. *et al.* Early adaptive developments of *Pseudomonas aeruginosa* after the transition from life in the environment to persistent colonization in the airways of human cystic fibrosis hosts. *Environ. Microbiol.* **12,** 1643–1658 (2010).
3. Lieberman, T. D. *et al.* Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. *Nat. Genet.* **46,** 82–7 (2014).
4. Cohen-Cymberknoh, M., Shoseyov, D. & Kerem, E. Managing cystic fibrosis: Strategies that increase life expectancy and improve quality of life. *Am. J. Respir. Crit. Care Med.* **183,** 1463–1471 (2011).
5. Cooper, T. F. & Lenski, R. E. Experimental evolution with E. coli in diverse resource environments. I. Fluctuating environments promote divergence of replicate populations. *BMC Evol. Biol.* **10,** 11 (2010).

6. Woods, R. J. *et al.* Second-order selection for evolvability in a large Escherichia coli population. *Science* **331,** 1433–6 (2011).
7. Barrick, J. E. *et al.* Genome evolution and adaptation in a long-term experiment with Escherichia coli. *Nature* **461,** 1243–1247 (2009).
8. Leon, D., D'Alton, S., Quandt, E. M. & Barrick, J. E. An evolutionary innovation is contingent on maintaining adaptive potential until competition subsides. *PLoS Genet.* **14,** 249110 (2018).
9. Yang, L. *et al.* Evolutionary dynamics of bacteria in a human host environment. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 7481–6 (2011).
10. Winstanley, C., O'Brien, S. & Brockhurst, M. A. Pseudomonas aeruginosa Evolutionary Adaptation and Diversification in Cystic Fibrosis Chronic Lung Infections. *Trends Microbiol.* **24,** 327–337 (2016).
11. Markussen, T. *et al.* Environmental heterogeneity drives within-host diversification and evolution of Pseudomonas aeruginosa. *MBio* **5,** e01592-14 (2014).
12. Jorth, P. *et al.* Regional Isolation Drives Bacterial Diversification within Cystic Fibrosis Lungs. *Cell Host Microbe* **18,** 307–19 (2015).
13. Imamovic, L. *et al.* Drug-Driven Phenotypic Convergence Supports Rational Treatment Strategies of Chronic Infections. *Cell* **172,** 121–134.e14 (2018).
14. Marvig, R. L., Johansen, H. K., Molin, S. & Jelsbak, L. Genome Analysis of a Transmissible Lineage of Pseudomonas aeruginosa Reveals Pathoadaptive Mutations and Distinct Evolutionary Paths of Hypermutators. *PLoS Genet.* **9,** e1003741 (2013).
15. Smith, E. E. *et al.* Genetic adaptation by Pseudomonas aeruginosa to the airways of cystic fibrosis patients. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 8487–92 (2006).
16. Marvig, R. L., Sommer, L. M., Molin, S. & Johansen, H. K. Convergent evolution and adaptation of Pseudomonas aeruginosa within patients with cystic fibrosis. *Nat. Genet.* **47,** (2015).
17. Silva, I. N. *et al.* Long-Term Evolution of *Burkholderia multivorans* during a Chronic Cystic Fibrosis Infection Reveals Shifting Forces of Selection. *mSystems* **1,** e00029-16 (2016).
18. Sommer, L. M. *et al.* Bacterial evolution in PCD and CF patients follows the same mutational steps. *Sci. Rep.* **6,** 28732 (2016).
19. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From

484 Polygenic to Omnigenic. *Cell* **169,** 1177–1186 (2017).
485 20. Deatherage, D. E., Kepner, J. L., Bennett, A. F., Lenski, R. E. & Barrick, J. E. Specificity
486 of genome evolution in experimental populations of Escherichia coli evolved at
487 different temperatures. *Proc. Natl. Acad. Sci. U. S. A.* **114,** E1904–E1912 (2017).
488 21. Jansson, J. K. & Hofmockel, K. S. The soil microbiome — from metagenomics to
489 metaphenomics. *Curr. Opin. Microbiol.* **43,** 162–168 (2018).
490 22. Sommer, L. M. *et al.* Is genotyping of single isolates sufficient for population structure
491 analysis of Pseudomonas aeruginosa in cystic fibrosis airways? *BMC Genomics* **17,**
492 (2016).
493 23. Yang, L. *et al.* In situ growth rates and biofilm development of Pseudomonas
494 aeruginosa populations in chronic lung infections. *J. Bacteriol.* **190,** 2767–76 (2008).
495 24. Moradali, M. F., Ghods, S. & Rehm, B. H. A. Pseudomonas aeruginosa Lifestyle: A
496 Paradigm for Adaptation, Survival, and Persistence. *Front. Cell. Infect. Microbiol.* **7,**
497 393389–39 (2017).
498 25. Furukawa, S., Kuchma, S. L. & O'Toole, G. A. Keeping their options open: acute versus
499 persistent infections. *J. Bacteriol.* **188,** 1211–7 (2006).
500 26. Andersen, S. B., Marvig, R. L., Molin, S., Krogh Johansen, H. & Griffin, A. S. Long-term
501 social dynamics drive loss of function in pathogenic bacteria. *Proc. Natl. Acad. Sci. U.*
502 *S. A.* **112,** 10756–61 (2015).
503 27. Thøgersen, J., Mørup, M., Damkiær, S., Molin, S. & Jelsbak, L. Archetypal analysis of
504 diverse Pseudomonas aeruginosa transcriptomes reveals adaptation in cystic fibrosis
505 airways. *BMC Bioinformatics* **14,** 279 (2013).
506 28. Sheftel, H., Szekely, P., Mayo, A., Sella, G. & Alon, U. Evolutionary trade-offs and the
507 structure of polymorphisms. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **373,** 20170105
508 (2018).
509 29. Mørup, M. & Hansen, L. K. Archetypal analysis for machine learning and data mining.
510 *Neurocomputing* **80,** 54–63 (2012).
511 30. Seth, S. & Eugster, M. J. A. Probabilistic archetypal analysis. *Mach. Learn.* **102,** 85–113
512 (2016).
513 31. Wood, S. N. *Generalized additive models : an introduction with R*. (CRC Press, 2006).
514 32. Wood, S. N. On p-values for smooth components of an extended generalized additive
515 model. *Biometrika* **100,** 221–228 (2013).

33. Hansen, C. R., Pressler, T. & Høiby, N. Early aggressive eradication therapy for intermittent Pseudomonas aeruginosa airway colonization in cystic fibrosis patients: 15 years experience. *J. Cyst. Fibros.* **7,** 523–30 (2008).

34. Høiby, N. Understanding bacterial biofilms in patients with cystic fibrosis: current and innovative approaches to potential therapies. *J. Cyst. Fibros.* **1,** 249–254 (2002).

35. Bjarnsholt, T. *et al. Pseudomonas aeruginosa* biofilms in the respiratory tract of cystic fibrosis patients. *Pediatr. Pulmonol.* **44,** 547–558 (2009).

36. Pressler, T. *et al.* Chronic Pseudomonas aeruginosa infection definition: EuroCareCF Working Group report. *J. Cyst. Fibros.* **10,** S75–S78 (2011).

37. Pedersen, S. S., Høiby, N., Espersen, F. & Koch, C. Role of alginate in infection with mucoid Pseudomonas aeruginosa in cystic fibrosis. *Thorax* **47,** 6–13 (1992).

38. Heltshe, S. L. *et al.* Longitudinal development of initial, chronic and mucoid Pseudomonas aeruginosa infection in young children with cystic fibrosis. *J. Cyst. Fibros.* (2017). doi:10.1016/j.jcf.2017.10.008

39. Hansen, S. K. *et al.* Evolution and diversification of Pseudomonas aeruginosa in the paranasal sinuses of cystic fibrosis children have implications for chronic lung infection. *ISME J.* **6,** 31–45 (2012).

40. Good, B. H., McDonald, M. J., Barrick, J. E., Lenski, R. E. & Desai, M. M. The dynamics of molecular evolution over 60,000 generations. *Nature* **551,** 45–50 (2017).

41. Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M. & Smoller, J. W. Pleiotropy in complex traits: challenges and strategies. *Nat. Rev. Genet.* **14,** 483–495 (2013).

42. Desai, M. M. & Fisher, D. S. Beneficial mutation selection balance and the effect of linkage on positive selection. *Genetics* **176,** 1759–98 (2007).

43. Dötsch, A. *et al.* The Pseudomonas aeruginosa Transcriptional Landscape Is Shaped by Environmental Heterogeneity and Genetic Variation. *MBio* **6,** e00749 (2015).

44. Rossi, E., Falcone, M., Molin, S. & Johansen, H. K. High-resolution in situ transcriptomics of Pseudomonas aeruginosa unveils genotype independent patho-phenotypes in cystic fibrosis lungs. *Submitted* (2018).

45. Sokurenko, E. V *et al.* Pathoadaptive mutations: gene loss and variation in bacterial pathogens. *Trends Microbiol.* **7,** 191–5 (1999).

46. Goodman, A. L. *et al.* A Signaling Network Reciprocally Regulates Genes Associated with Acute Infection and Chronic Persistence in Pseudomonas aeruginosa. *Dev. Cell* **7,**

548 745–754 (2004).
549 47. Robillard, N. J. & Scarpa, A. L. Genetic and physiological characterization of
550 ciprofloxacin resistance in Pseudomonas aeruginosa PAO. *Antimicrob. Agents*
551 *Chemother.* **32,** 535–9 (1988).
552 48. Nakamura, S., Nakamura, M., Kojima, T. & Yoshida, H. gyrA and gyrB mutations in
553 quinolone-resistant strains of Escherichia coli. *Antimicrob. Agents Chemother.* **33,**
554 254–5 (1989).
555 49. Kugelberg, E., Löfmark, S., Wretlind, B. & Andersson, D. I. Reduction of the fitness
556 burden of quinolone resistance in Pseudomonas aeruginosa. *J. Antimicrob.*
557 *Chemother.* **55,** 22–30 (2005).
558 50. Wolska, K. I., Grudniak, A. M., Rudnicka, Z. & Markowska, K. Genetic control of
559 bacterial biofilms. *J. Appl. Genet.* **57,** 225–38 (2016).
560 51. Ventre, I. *et al.* Multiple sensors control reciprocal expression of Pseudomonas
561 aeruginosa regulatory RNA and virulence genes. *Proc. Natl. Acad. Sci. U. S. A.* **103,**
562 171–6 (2006).
563 52. Wolf, D. M., Vazirani, V. V. & Arkin, A. P. Diversity in times of adversity: probabilistic
564 strategies in microbial survival games. *J. Theor. Biol.* **234,** 227–253 (2005).
565 53. KOH, K. S. *et al.* Minimal increase in genetic diversity enhances predation resistance.
566 *Mol. Ecol.* **21,** 1741–1753 (2012).
567 54. Boles, B. R., Thoendel, M. & Singh, P. K. Self-generated diversity produces ‘“insurance
568 effects”’ in biofilm communities. *Proc. Natl. Acad. Sci.* **101,** 16630–16635 (2004).
569 55. Kirchner, S. *et al.* Use of artificial sputum medium to test antibiotic efficacy against
570 Pseudomonas aeruginosa in conditions more relevant to the cystic fibrosis lung. *J. Vis.*
571 *Exp.* e3857 (2012). doi:10.3791/3857
572 56. O’Toole, G. A. & Kolter, R. Flagellar and twitching motility are necessary for
573 Pseudomonas aeruginosa biofilm development. *Mol. Microbiol.* **30,** 295–304 (1998).
574 57. Hentzer, M. *et al.* Alginate overproduction affects Pseudomonas aeruginosa biofilm
575 structure and function. *J. Bacteriol.* **183,** 5395–401 (2001).
576 58. Ryder, C., Byrd, M. & Wozniak, D. J. Role of polysaccharides in Pseudomonas
577 aeruginosa biofilm development. *Curr. Opin. Microbiol.* **10,** 644–8 (2007).
578 59. Kragh, K. N. *et al.* Role of Multicellular Aggregates in Biofilm Formation. *MBio* **7,**
579 e00237 (2016).

60. Déziel, E., Comeau, Y. & Villemur, R. Initiation of biofilm formation by Pseudomonas aeruginosa 57RP correlates with emergence of hyperpiliated and highly adherent phenotypic variants deficient in swimming, swarming, and twitching motilities. *J. Bacteriol.* **183,** 1195–204 (2001).

61. Caceres, S. M. *et al.* Enhanced in vitro formation and antibiotic resistance of nonattached Pseudomonas aeruginosa aggregates through incorporation of neutrophil products. *Antimicrob. Agents Chemother.* **58,** 6851–60 (2014).

62. Ricaurte, D. E. *et al.* A standardized workflow for surveying recombinases expands bacterial genome-editing capabilities. *Microb. Biotechnol.* **11,** 176–188 (2018).

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

64. Team, R. C. R: A Language and Environment for Statistical Computing. (2017).

65. Wood, S. N., Pya, N. & Säfken, B. Smoothing Parameter and Model Selection for General Smooth Models. *J. Am. Stat. Assoc.* **111,** 1548–1563 (2016).

66. Wood, S. N. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. Ser. B (Statistical Methodol.* **73,** 3–36 (2011).

67. Eugster, M. J. A. & Leisch, F. From Spider-Man to Hero - Archetypal Analysis in *R*. *J. Stat. Softw.* **30,** 1–23 (2009).

68. Eugster, M. J. A. & Leisch, F. Weighted and robust archetypal analysis. *Comput. Stat. Data Anal.* **55,** 1215–1225 (2011).

69. Seth, S. & Eugster, M. J. A. Probabilistic Archetypal Analysis. (2014).

70. Wickham, H. tidyverse: Easily Install and Load the 'Tidyverse'. (2017).

71. van Rij, J., Wieling, M., Baayen, R. H. & van Rijn, H. itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs. (2017).

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

73. Xie, Y. knitr: A General-Purpose Package for Dynamic Report Generation in R. (2017).

74. Zhu, H. kableExtra: Construct Complex Table with 'kable' and Pipe Syntax. (2017).

75. Fernandez, M., Wilson, H. F. & Barnard, A. S. Impact of distributions on the archetypes and prototypes in heterogeneous nanoparticle ensembles. *Nanoscale* **9,** 832–843 (2017).

76. Johansen, H. K., Moskowitz, S. M., Ciofu, O., Pressler, T. & Høiby, N. Spread of colistin resistant non-mucoid Pseudomonas aeruginosa among chronically infected Danish cystic fibrosis patients. *J. Cyst. Fibros.* **7,** 391–397 (2008).

METHODS

The isolate collection

The current isolate library is comprised of 443 longitudinally collected single *P. aeruginosa* isolates distributed within 52 clone types collected from 39 young CF patients treated at the Copenhagen CF Centre at Rigshospitalet (median age at first *P. aeruginosa* isolate = 8.1 years, range = 1.4-24.1 years, median coverage of colonization: 4.6 years, range: 0.2-10.2 years). This collection is a complement to and extension of the collection previously published [16] and captures the period of initial rapid adaptation[6,7,9], with 389 isolates of the previously published collection included here in addition to 54 new isolates. To build a homogeneous collection for our study of evolution, we excluded two patients with a sustained multi-clonal infection. For the GAMM analysis, we excluded isolates belonging to clone types present in a patient at two or fewer time-points, unless the two time-points were sampled more than 6 months apart. The isolates not included in the previous study have been clone typed as a routine step at the Department of Clinical Microbiology at Rigshospitalet. This clone type identification was performed as described previously[16], and the sequencing was carried out as follows: DNA was purified from over-night liquid cultures of single colonies using the DNEasy Blood and Tissue Kit (Qiagen), libraries were made with Nextera XT and sequenced on an Illumina MiSeq using the v2 250x2 kit.

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

Growth rate in Luria-Bertani broth (LB) and Artificial sputum medium (ASM)[55]

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 [56–58]. In this study, we have measured adhesion to peg-lids and estimated the ability to make aggregates – both traits have been linked with an isolate's ability to make biofilm [59,60]. Because of this, we are using these two measures as an estimate of our isolates' ability to make biofilm. However, because we are aware of the complexity of the actual biofilm-forming phenotype, we have chosen to refer to this adhesion/aggregation phenotype as "adherence" and not "biofilm formation".

Adhesion in LB. Adhesion was estimated by measuring attachment to NUNC peg lids. Isolates were re-grown in 96 well plates with 150µl medium where peg lids were used instead of the standard plate lids. The isolates were incubated for 20 hours at 37°C, after which OD_{600nm} was measured and subsequently, the peg lids were washed in a "washing microtiter plate" with 180µl PBS to remove non-adhering cells. The peg lids were then transferred to a microtiter plate containing 160µl 0.01% crystal violet (CV) and left to stain for 15 min. The lids were then washed again three times in three individual "washing microtiter plates" with 180µl PBS to remove unbound crystal violet. To measure the adhesion, the peg lids were transferred to a microtiter plate containing 180µl 99% ethanol, causing the adhering CV stained cells to detach from the peg lid. This final plate was used for measurements using an ELISA reader, measuring the CV density at OD_{590nm}. (Microtiter plates were bought at Fisher Scientific, NUNC Cat no. 167008, peg lids cat no. 445497)

Aggregation in ASM. Aggregation in each well was first screened by visual inspection of wells during growth assays in ASM and by evaluation of noise in the growth curves, resulting in a binary metric of "aggregating" versus "not aggregating". However, to incorporate this trait in our archetype analysis, we needed to develop a continuous metric of aggregation. Based on the above manual assessment, we developed a metric based on the average noise of each strain's growth curves. While we tested several different metrics based on curve variance, the metric that seemed to delineate isolates according to the binary aggregation measure most successfully was based on a sum of the amount of every decrease in OD that was followed by a recovery at the next time point (versus the expected increase in exponential phase and flatline in stationary phase). This value was normalized by the increase in OD across the whole growth curve, to ensure that significant, irregular swings stood out with respect to overall growth. This metric therefore specifically accounts for fluctuation - both a limited number of large fluctuations in OD_{630nm} (often seen 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 [61], 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)[62]. A PAO1 strain carrying a pSEVA658-ssr plasmid[63] 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 37°C

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 *P. aeruginosa* 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 [16] 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[64] using the packages *mgcv*[65,66] for the GAMM analysis and *archetype*[67–69] for the AA. Complementary packages used for analysis are: *tidyverse*[70], *itsadug*[71], *ggthemes*[72], *knitr*[73] and *kableExtra*[74]. We also referred to Thøgersen *et al.*[27] and Fernandez *et al.*[75] in the design of appropriate assessment methods for the final AA model. We include two R markdown documents that explain our modeling steps and further evaluation plots in detail (AA: Supplemental file 1, GAMM: Supplemental file 2), and summarize our methods below in brief.

Data modeling

Archetype analysis (AA). We evaluated several different model fitting approaches by varying the number and type of phenotypes modeled as well as the archetype number and fit method, using RSS-based screeplots of stepped fits of differing archetype numbers, explained sample variance (ESV), isolate distribution among archetypes, convex hull projections of paired phenotypes (all combinations), and parallel coordinate plots as metrics for choosing the best fit parameters and approach to accurately represent our data. Ultimately, we focused on 5 continuous phenotypes correlated with growth (growth rate in ASM), biofilm (adhesion and aggregation), and antibiotic resistance (aztreonam and ciprofloxacin MICs), which also were linked to relevant findings provided by the GAMM models. We used a root sum squared (RSS) versus archetype number screeplot of different fits to determine that a 6 archetype fit would produce the optimal model for this dataset.

We then performed 500 simulations of a 100 iteration fit using the "robustArchetypes" method[68], which reduces the impact of data outliers in fitting the convex hull of the data. We evaluated the mean ESV and the number of isolates with an ESV greater than 80% for the best model from each simulation in this study and differences in archetype characteristics to assess convergence, ultimately selecting the model with the second highest mean ESV (90.32%) and highest number of isolates with an ESV over 80% (87.13%); this model also resembled the other 10 top models of the simulation study. The order of archetypes around the simplex plot boundary obscures the true dimensionality of the isolate distribution by implying the archetypes are equidistant, so relationships between phenotypes are not always obvious. We re-ordered the archetypes in the simplex plot by growth rate and secondarily antibiotic resistance to improve clarity in the complex 6 archetype plot. This reordering was also justified when projecting the archetypes onto a PCA plot of the phenotypes (Supplemental file 1). All simplex plots have also had the 11 isolates with an ESV < 50% removed such that we are not drawing any conclusions from these poorly fit data (they are shown via simplex plot in the supplemental markdown).

Generalized Additive Mixed Models (GAMMs). For all phenotypes, GAMMs were used to identify evolutionary trends over time since first colonization. We correct for the patient environment and inconsistent sampling over time using a smooth random factor. Models

were fitted in the following way: All continuously measured phenotypes included in the Archetype analysis were fitted as a response variable ("predicted" or "dependent" variable in Figure 3D) one-to-one, with both time as an "explanatory" or "independent" variable alone and combined with each of the phenotypes to account for potential time-dependence of the observations. Factorial/binary phenotypes were implemented as categorical functions and continuous phenotypes as smooth functions, allowing for non-parametric fits. Normally only one variable/phenotype of interest is used as the predictor while other alterable variables or factors are used as explanatory variables to explain or predict changes in the predictor. However, this requires a preconceived idea of a "one-way-relationship" where one variable (the predictor) is assumed to be affected by certain other variables (the explanatory variables), but where the explanatory variables cannot be affected by the predictor. By testing all phenotypes against each other, we avoid assumptions regarding the specific direction of relationships between the predictor variable and the explanatory variable. Furthermore, in using the GAMMs we prioritize accuracy of fitting but increase our risk of overfitting as a byproduct. We sought to counteract the risk of overfitting by the default penalization of fits inherent to the method used[65,66] and by model estimation via restricted maximum likelihood (REML) which has been found to be more robust against overfitting[31,66]. When significant relationships were identified in one-to-one models (p-value < 0.05, as based on Wald-type tests as described in[31,32]), all significant explanatory variables were used to build a multi-trait model for the associated predictor. If select explanatory phenotypes were then identified as non-significant (p-value > 0.05) in the multi-trait model, they would be removed in a reduction step. To identify whether a reduced multi-trait model 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*[71] (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 <

858 0.01, unless otherwise stated. For information on deviance explained, R^2, and degrees of
859 freedom for the individual models/variables, we refer to the Supplemental file 2.
860

861 SUPPLEMENTARY INFORMATION

862 Supplemental File 1. Construction and assessment of the archetype model.
863 Supplemental File 2. Construction and assessment of the generalized additive mixed models.
864 Supplemental spreadsheet 1. Phenotype Database
865 Supplemental Information.
866 Figure S1, related to Figure 4. Hypermutators versus normomutators
867 Figure S2, related to Figure 5. *mucA* and *algU* mutants
868 Figure S3, related to Figure 5. mexZ mutants and drug efflux pumps
869 Figure S4, related to Figure 5. Specific mutations in *gyrA/B* by patient and adhesion
870 Figure S5, related to Figure 5. Adhesion and generation time of *gyrA/B* mutants (PAO1)
871 Figure S6, related to Figure 2. Example growth curves
872 Figure S7, related to Figure 2. Development of an aggregation metric
873