Unraveling Genomic Diversity in *Pseudomonas aeruginosa* Cystic Fibrosis Lung Infection and Its Impact on Antimicrobial Resistance

Jelly Vanderwoude[1], Sheyda Azimi[1,2], Timothy D. Read[3] & Stephen P. Diggle[1*]

[1]Center for Microbial Dynamics and Infection, School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA; [2]School of Biology, Georgia State University, Atlanta, GA, USA; [3]Division of Infectious Diseases, Department of Medicine, Emory University School of Medicine, Atlanta, GA, USA

*Correspondence: stephen.diggle@biosci.gatech.edu

Preprint servers: This manuscript has been submitted as a preprint to bioRxiv
Classification: Biological Sciences, Microbiology

Keywords: population heterogeneity; antibiotic resistance; hypermutation; evolution

ABSTRACT

Pseudomonas aeruginosa is the predominant bacterial pathogen in the cystic fibrosis (CF) lung, establishing a chronic, drug-resistant infection often via a single adapting and diversifying strain. In this study, we collected a bio-repository of 300 isolates of *P. aeruginosa* from four CF patients to unravel the extent of within-patient genomic diversity in CF lung infection; elucidate the evolutionary processes by which this diversity is generated; and understand how diversification impacts antimicrobial resistance. We found that populations of *P. aeruginosa* in the CF lung displayed patient specific genomic diversity ranging from 1 to 199 median pairwise SNPs, and 31 to 4,592 total SNPs within a population. Diversification was driven by a combination of *de novo* mutations, recombination, and variations in gene content, with the highest levels of population diversity associated with the presence of DNA mismatch repair mutants. Populations with higher levels of genomic diversity displayed increased variation in antibiotic resistance phenotypes; however, even populations with lower genomic diversity displayed substantial diversity in resistance profiles. Intriguingly, we found evidence that mutations in DNA mismatch repair can facilitate the evolution of seemingly undesirable traits *in vivo*, such as increased sensitivity to certain antimicrobials. Overall, our findings highlight that (i) *P. aeruginosa* genomic population diversity in chronic CF lung infection is highly variable between patients; (ii) hypermutator populations can evolve increased sensitivity to antimicrobials even under apparent antibiotic selection; and that (iii) current methods of single-isolate sampling and susceptibility testing should be re-evaluated in the study of chronic *P. aeruginosa* infection.

SIGNIFICANCE

Chronic lung infection with *Pseudomonas aeruginosa* is the leading cause of morbidity and mortality in cystic fibrosis (CF) patients. Upon infection, *P. aeruginosa* rapidly acquires mutations, which can result in diverse, treatment-resistant populations. However, within-patient diversity is largely overlooked by clinicians and researchers. We sequenced the genomes of 300 CF lung-derived isolates of *P. aeruginosa* to investigate genomic diversity and its role in the evolution of antimicrobial resistance (AMR). Genomic diversity was variable, with higher levels of diversity correlating with increased variation in AMR. In two patients, mutations in DNA mismatch repair genes drove the evolution of increased susceptibility to multiple antimicrobials. Based on these findings, we advocate for deeper population sampling to improve our understanding of CF lung infection.

INTRODUCTION

Pseudomonas aeruginosa is the dominant bacterial pathogen in chronic infections of the airways of adult individuals with cystic fibrosis (CF), a genetic disorder that results in thickened

mucus, persistent lung infection, and progressive decline in lung function. *P. aeruginosa* has multiple intrinsic and acquired mechanisms of antimicrobial resistance (AMR), with clinical strains frequently displaying multi-drug resistance (MDR). While antibiotic treatment can be effective against early-stage, transient *P. aeruginosa* infections, antibiotic regimens ultimately fail to eradicate chronic infections of *P. aeruginosa* from the CF lung (1). This may be a result of the high degree of phenotypic and genomic heterogeneity that naturally evolves in *P. aeruginosa* populations inhabiting CF airways, allowing the population to exploit various pathways of resistance and the emergence of rare clones that evade treatment and re-establish infection afterwards (2, 3). Current understanding suggests that most individuals with CF are initially infected by a single environmental or transmissible epidemic strain of *P. aeruginosa*, which then diversifies in the CF lung over the course of many years of infection (4). Maintaining diversity in populations can be advantageous for bet-hedging in a complex infection environment where there are a multitude of external stressors such as competing microbiota, antibiotic exposure, and host immune response. Mutations in DNA mismatch repair (MMR) mechanisms act as a catalyst for this diversification, potentially providing an evolutionary advantage in an environment that demands rapid adaptation for survival, though oftentimes at a fitness cost (5).

While the within-host adaptations of *P. aeruginosa* to the CF lung have been investigated and described, primarily via longitudinal single-isolate sampling (6-16), the extent of within-patient *P. aeruginosa* genomic diversity in the CF lung and the complex pathways exploited during diversification in order to develop AMR are not fully understood. Longitudinal sampling of single or small subsets of isolates from a population only reflects a fraction of the total evolutionary pathways exhibited within a population. Single-isolate sampling may also result in significant underestimation of the diversity of antimicrobial susceptibility profiles in a chronic infection; however, current methods of susceptibility testing are not designed to measure phenotypes for mixed populations. As population diversity may impact infection outcomes via heteroresistance (17), microbial social interactions (18), or the ability of a population to survive evolutionary bottlenecks (1), this warrants a shift in our sampling and susceptibility testing of chronic microbial infections to reflect our understanding of them as complex, dynamic populations. Deeper sampling also allows for better resolution of evolutionary pathways and population structure, to see which genetic variants fix in a population and how members of the same strain co-exist.

The few studies that have investigated population diversity in this infection context have been limited to descriptions of phenotypic diversity (19-22), genetic analyses via pooled population sequencing (23, 24), or samples acquired during end-stage lung disease (25). As a result, there remain gaps in our knowledge of the level of genomic diversity present in *P. aeruginosa* populations inhabiting the airways of individuals in the early ($\%FEV_1 > 70$) to intermediate ($\%FEV_1 \leq 70, \geq 40$) stages of lung disease. Here we investigated genomic diversity of chronic *P. aeruginosa* populations from the sputa of four individuals with CF, and explored the evolutionary mechanisms responsible for driving this diversity. Because diversity in antimicrobial susceptibilities may have adverse effects on the outcomes of patients with chronic infections, we further explored how genomic diversification impacts the evolution of antimicrobial resistance in these populations. We found that (i) *P. aeruginosa* within-patient genomic population diversity in chronic CF lung infection is highly variable; (ii) diversity is largely generated by *de novo* mutations and hypermutation, but also in part by predicted recombination and gene loss events; and (iii) antibiotic treatment does not necessarily select for increased resistance in hypermutator populations.

METHODS

Cohort selection and strain isolation. We selected four adult individuals aged 24-31 years for this study from a cohort of CF patients at Emory University in Atlanta who had been chronically infected with *P. aeruginosa* for 10-15 years at the time of sampling. From each patient, we

collected and processed a single expectorated sputum sample. We processed sputum by supplementing each sample with 5 ml synthetic cystic fibrosis medium (SCFM) (26) and autoclaved glass beads, homogenizing the mixture via vortexing for 2 mins, centrifuging the homogenized sputum mixture for 4 mins at 4200 rpm, removing the supernatant, and conducting a 10x serial dilution of cell pellet re-suspended in phosphate buffered saline to streak on *Pseudomonas* isolation agar (PIA) plates. These plates were incubated at 37°C overnight, then at room temperature for up to 72 h. From each expectorated sputum sample, we randomly isolated 75 *P. aeruginosa* colonies for a total of 300 isolates.

Whole genome sequencing. To conduct sequencing, we first grew all 300 isolates overnight in 15 ml conical tubes in lysogeny broth (LB) at 37°C with shaking at 200 rpm. We extracted DNA from these cultures using the Promega Wizard Genomic DNA Purification Kit according to the manufacturer's instructions. We prepared sequencing libraries using the Nextera XT DNA Library Preparation Kit and used the Illumina Novaseq platform to obtain 250 bp paired-end reads for a mean coverage of 70x. 28 samples either failed or did not meet the minimum sequencing coverage or quality requirements, so we re-sequenced these using the Illumina MiSeq platform for 250 bp paired-end reads and combined the reads from both sequencing runs to analyze these 28 samples. We randomly selected one isolate from each patient to serve as the reference strain for the other 74 isolates isolated from that patient. For these reference isolates, we additionally obtained Oxford Nanopore long read sequences through the Microbial Genome Sequencing Center (GridION Flow Cell chemistry type R9.4.1 with Guppy high accuracy base calling v4.2.2) at 35x coverage.

Creating annotated reference assemblies. We used Unicyler v0.5.0 (27) to create long-read assemblies for the four reference isolates. We then conducted one round of long-read polishing on these assemblies using Medaka v1.0.3 (28), which produced preliminary consensus sequences. We conducted quality control on all 300 Illumina reads using the Bactopia v1.6.5 (29) pipeline. We conducted two further short-read assembly polishing steps on the long-read assemblies by aligning the quality-adjusted short reads of each of the four reference isolates to its respective consensus sequence using Polypolish v0.5.0 (30) and Pilon v1.24 (31). We validated the final consensus sequences by mapping the Illumina reads of each reference to its respective assembly using Snippy v4.6.0 (32) and confirming that 0 variants were called. We used (i) Prokka v1.14.6 (33) and (ii) RATT v1.0.3 (34) to (i) annotate our reference strains using a *P. aeruginosa* pan-genome database, and to (ii) transfer gene annotations from PAO1 to their respective positions in each of the reference strains, respectively.

Multi-locus sequence typing. Our multi-locus sequence typing was implemented in Bactopia v1.6.5 (29), which employs the PubMLST.org schema (35).

Variant calling and enrichment analysis. We used Snippy v4.6.0 (32) to call variants from the other 296 isolates against their respective reference strain and create a core genome alignment. Using PhyML v3.3.20211231 (36), we created a maximum likelihood phylogeny. Then, using VCFtools v0.1.16 (37) and Disty McMatrixface v0.1.0 (38), we generated a pairwise SNP matrix for each patient. For Disty, we only considered alleles in the core genome and chose to ignore ambiguous bases pairwisely (-s 0). We then employed SnpEff and SnpSift v4.3t (39) to sort the variants by predicted effect. We then conducted an enrichment analysis to determine which functional categories of genes were differentially impacted by mutations than would be expected by random chance. We used an in-house Python script to retrieve the PseudoCAP functional group of each gene where a non-synonymous SNP or microindel was identified. We accounted for the varying lengths of genes in each functional category in our analysis, based off their lengths and prevalence in the PAO1 genome. We used a chi-squared goodness of fit test to conduct the enrichment analyses for Patients 1-3 to determine which functional categories were disproportionately impacted by non-synonymous variants. We used the R package XNomial

v1.0.4 (40) to conduct an exact multinomial goodness of fit test using Monte-Carlo simulations for Patient 4 because the SNP frequencies of Patient 4 did not meet the assumptions for a chi-squared test. If the chi-squared value for a particular PseudoCAP functional category was in the top 30 percentile of all chi-squared values (top 8 of 27 total categories) in the analyses of at least three patients, we noted this as as an enrichment.

Predicting putative recombination events. We input the core genome alignment from each patient to Gubbins v3.3.0 (41) to predict potential recombinant regions in each population.

Analyzing gene content and variability. We used the Bactopia v1.6.5 (29) pipeline to process all paired-end Illumina reads. Following quality control, Bactopia implements Shovill v1.1.0 (42) to prepare *de novo* assemblies and Prokka v1.14.6 (33) to annotate these assemblies. We then used Roary v3.13.0 (43) to conduct a pangenome-style analysis and create a core genome alignment from the annotated assemblies. Our parameters for Roary were as follows: 90% identity threshold for BlastP, 98% prevalence for consideration in the core genome, and paralog splitting turned off. We further filtered gene families from our output that had a median length of <200 bp. We then implemented IQ-TREE v2.1.2 (44) to create a maximum likelihood phylogeny of the core genome alignment.

Antimicrobial resistance phenotyping and resistome genotyping. To assess antimicrobial susceptibility profiles, we followed the guidelines and standards provided by the Clinical and Laboratory Standards Institute (CLSI) *Performance Standards for Antimicrobial Susceptibility Testing M100S*, 30th edition. We first grew all isolates overnight in LB in 24-well microtiter plates at 37°C with shaking at 200 rpm. We diluted cultures to a Macfarland standard of 0.5 (OD_{600} ~0.06) and streaked a lawn on 100x15 mm Petri dishes with 20 ml Mueller-Hinton agar using pre-sterilized cotton swabs. We then stamped amikacin (AK), meropenem (MEM), piperacilin-tazobactam (TZP), ciprofloxacin (CIP), tobramycin (TOB), and ceftazidime (CAZ) on each plate and incubated for 17 h at 37°C. We measured the zone of inhibition at 17 h and classified the values as resistant, intermediate, or susceptible per the established CLSI interpretive criteria. We used *P. aeruginosa* strain ATCC 27853 as a quality control. We tested all isolates in biological triplicates. We assessed genotypes relevant to resistance by uploading the *de novo* assemblies to the Resistance Gene Identifier (RGI) v6.1.0 web portal, which predicts resistomes using the Comprehensive Antibiotic Resistance Database (CARD) v3.2.6 (45). We excluded loose and nudge hits from this analysis. We ran a Mann-Whitney U test to compare the means of antimicrobial susceptibilities between hypermutators and normomutators.

Visualizations. We conducted graphical analyses in R v4.3.0. We employed a modified Python script from PIRATE v1.0.5 (46) to visualize the genomes.

Data availability. The sequences in this study will be made available in the NCBI SRA database upon publication.

RESULTS

A single *P. aeruginosa* strain dominates each of four chronically infected CF patients. In order to investigate the extent of within-patient genomic diversity of *P. aeruginosa* in chronic CF lung infection, we sequenced 75 isolates from a single sputum sample of each of the four individuals of this cohort for a total of 300 isolates. We prepared the sequences of all 300 isolates using *de novo* assembly and annotation. We assembled the genomes in 20 to 444 contigs (mean = 53 contigs). Genomes in this dataset ranged in size from 5,888,197 to 6,746,489 nucleotides, with 5,209 to 5,970 genes. The median genome sizes of isolates sourced from Patients 1-4, respectively, were 6,222,786, 6,331,110, 6,742,689, and 6,308,671 nucleotides, with 5,523, 5,571, 5,964, and 5,567 genes, respectively (**Table S1**). A phylogenetic tree of the core genome alignment revealed that the populations infecting Patients 1, 2, and 4

clustered closely with PAO1, while that of Patient 3 more closely resembled PA14 (**Fig. S1**). Strain typing of the isolates showed that there was a single, dominant strain type in each population— ST870, ST2999, ST1197, and ST274 for Patients 1-4, respectively (**Table 1**). For the rest of the text, we will simply refer to each population by its respective patient number.

***P. aeruginosa* population diversity is patient-specific.** We assessed the genomic diversity in these populations according to the number of single nucleotide polymorphisms (SNPs), insertions and deletions (indels), and gene content variations. We found that genomic diversity varied significantly. The total number of unique SNPs discovered across 75 isolates for Patient 1 was 4,592 (maximum number of pairwise SNPs = 611, median number of pairwise SNPs = 199, mean = 208); for Patient 2 was 1,972 (max. = 326, median = 145, mean = 118); for Patient 3 was 1,638 (max. = 150, median = 76, mean = 87); and for Patient 4 was 31 (max. = 8, median = 1, mean = 3) (**Fig. 1**). Across the population of Patient 1 we found 498 unique microindels, 307 for Patient 2, 330 for Patient 3, and 14 for Patient 4 (**Table 2**).

Within a given patient, we found that variations in gene content were limited, with the exception of one patient. Genes were categorized as core for a given patient infection if they were present in >98% of isolates (≥74), soft core at 95-98% (72-73), shell at 15-95% (12-72), and cloud at <15% (≤11). Patients 2, 3, and 4 had similar core and accessory genome content, with core genes comprising 96-97% of the genome, soft core genes comprising 0.2-0.8%, shell genes comprising 0.6-2%, and cloud genes comprising 1-3%. However, the population of Patient 1 had a significantly larger accessory genome— core (84%), soft core (2%), shell (12%), and cloud (2%) (**Table 3; Fig. 2**).

***P. aeruginosa* diversity is primarily driven by *de novo* mutations, especially mutations in DNA mismatch repair.** We next wanted to further understand the processes by which *P. aeruginosa* diversified in our cohort. We first sought to predict putative recombination events. In Patients 1-4, 527 (11.5%), 19 (<1%), 86 (5.25%), and 0 SNPs were predicted to be in 31, 3, 17, and 0 recombinant regions, respectively. These data show that *de novo* mutation was a much more prominent driver of intra-specific diversity than recombination in our particular cohort. As expected, we found that the infections with the highest SNP diversity harbored strains with DNA MMR mutations. Patients 1 and 2 harbored DNA MMR mutants (hypermutators); however, we found no hypermutators in Patients 3 or 4. The phylogeny of Patient 1 indicates that a non-synonymous SNP in *mutS* (Ser31Gly) evolved first in the population, after which a frameshift deletion in *mutS* (Ser544fs) piggybacked. In total, *mutS* mutants comprise 61.3% of this population. In Patient 2, a non-synonymous SNP in *mutL* resulting in a pre-mature stop codon (Glu101*) evolved first, found in 41.3% of the population. Two of these *mutL* mutants further independently acquired a single non-synonymous mutation in *mutS* (Phe445Leu, Ala507Thr) (**Fig. 1**).

In Patient 1, there were two distinct branches of the phylogenetic tree, one with hypermutators and the other composed of normomutators (38.7%). Interestingly, there was a significant amount of genetic diversity within both the normomutators (mean SNP distance = 156.9 SNPs, median = 91 SNPs) and hypermutators (mean = 174.6 SNPs, median = 197 SNPs). There was a distinct small cluster of normomutator isolates that significantly diverged from the others. Of the hypermutators, these further diverged into those with one DNA MMR mutation (39.1%) and those with two MMR mutations (60.9%). In Patient 2, there was largely a lack of genetic diversity in the normomutators (mean = 0.36 SNPs, median = 0 SNPs), with one clone dominating 48% of the population. Hypermutation appears to have been responsible for the large majority of all the genetic diversity in this population (mean = 211.2 SNPs, median = 224 SNPs). In Patient 3, there were three major clusters, comprising 58.7%, 26.7%, and 14.7% of the total population (mean = 61.9, 55.5, and 65.4 SNPs; median = 62, 61, and 64 SNPs, respectively). In Patient 4, there was one dominant clone encompassing 66.6% of the

population, with a small number of SNPs (mean = 4 SNPs, median = 3 SNPs) differentiating the other 33.3% of the population (**Fig. 1**).

Protein export/ secretion systems and transcriptional regulators are hotspots for *de novo* mutations. To further characterize the mutational adaptations of *P. aeruginosa* to the CF lung environment and identify mutational hotspots, we categorized non-synonymous SNPs and microindels that occurred within coding regions of genes according to the PseudoCAP functional categories and conducted an enrichment analysis. We found that protein secretion and export apparatuses and transcriptional regulators were enriched for such mutations (**Fig. S2**). Additionally, two of the four genes impacted by non-synonymous mutations in all four patients in this study were related to protein secretion, *fha1* and *pscP* (**Table S2**). We found that phage/transposon/plasmid genes were less likely to be impacted by such mutations. Non-coding RNAs were also less likely to be impacted by mutations than other functional categories (see **Table S3** for all supporting statistical values), which is unsurprising given that small non-coding RNAs are known to hold important regulatory functions in bacteria (47). 57 genes were impacted by non-synonymous mutations in 3 of the 4 patients, which included genes with previously described functions in alginate biosynthesis, primary metabolism, antibiotic resistance and efflux, iron uptake, biofilm formation, stress response, amino acid biosynthesis, type IV pili, lipopolysaccharide, quorum sensing, and virulence (**Table S4**). A full list of all SNPs discovered in this dataset can be found in **Tables S5-S8**.

Prophages and determinants of pathogenicity are commonly variable genes within a population. We further investigated the potential pathways of adaptation employed by *P. aeruginosa* by identifying the genes that displayed variation within a population. Our analysis sought to identify variable genes, not whether genes had been specifically lost or acquired in the population. There were 934 variable genes in Patient 1. Many of these are involved in previously described pathogenic functions, such as type VI secretion, type IV pili, phenazine biosynthesis, quorum sensing, fimbrial biogenesis, flagellar motility, triclosan efflux, biofilm signaling regulators, chemotaxis, aerotaxis, and iron acquisition. Unsurprisingly, these also included integrative and conjugative elements, as well as prophages including integrases, recombinases, and tranposases. In Patient 2, there were 185 variable genes, the most notable being a large enrichment of truncated pseudogenes in isolate 74. There were 171 variable genes in Patient 3, notably all 10 genes of the *pil2* operon, 10 integrative and conjugative element genes, as well as *dsbA2*, *katA,* and other genes encoding hypothetical proteins. In Patient 4, there were 196 variable genes notable variable genes, including 6 genes in the *hpa* operon, *ampC* and *ampR*, *bfmRS*, *ccoN*, *aph*, *fepBCDG*, *fvbA*, *acoABCRX*, and *mpaR*. There were 13 genes that were variable in all 4 patients, and 33 genes that were variable in 3 of 4 patients. The phenazine biosynthetic pathway was the most frequently variable across multiple patients, with *phzC1*, *phzD1*, *phzF2*, *phzE*, and *phzG1* all being variable in 3 or more patients. A full list of all genes in this study can be found in **Table S9**.

Genomic diversity correlates strongly with variability of antimicrobial susceptibility profiles in a population. Because we suspect that diversity in AMR is likely underestimated in clinical settings, we sought to determine whether the described genomic diversity correlates to diversity in AMR phenotypes. We hypothesized that genetically diverse populations would also display more variation in AMR. Using a standard disc diffusion assay, we assessed susceptibilities to six antimicrobials commonly prescribed in CF treatment: amikacin, meropenem, piperacilin-tazobactam, ciprofloxacin, tobramycin, and ceftazidime (**Fig. 3; Fig. S3; Tables S10-S13**). We found that patients with higher genomic diversity displayed more variability in antimicrobial resistance phenotypes, though there was still significant phenotypic diversity in the least genetically diverse population (**Fig. S4**).

Hypermutation can drive the evolution of both increased susceptibility and resistance to antimicrobials in the same population. As our cohort had two populations with DNA MMR

mutants, we used this opportunity to ascertain how hypermutation drives the evolution of AMR. In Patient 1, AMR genotypes cluster by DNA MMR genotype. Hypermutators were significantly more resistant to amikacin than normomutators (U = 315.5, p = .00013), although this difference could not be attributed to any hits in the CARD database. Hypermutators were also significantly more resistant to beta-lactams piperacilin-tazobactam (U = 457.5, p = .023) and ceftazidime (U = 428, p = .0095), although there was no significant difference in the resistance profiles of hyper- and normomutators in regards to the beta-lactam meropenem (U = 630, p =.69). Some normomutators in this population acquired a SNP in *ampC* (461 A > G, Asp154Gly), which increased susceptibility to piperacilin-tazobactam and ceftazidime. Of the isolates with one DNA MMR mutation, some lost *ampC* entirely, also leading to increased susceptibility to these antimicrobials. Of the isolates with both DNA MMR mutations, some had acquired a SNP in *ampC* (1066 G > A, Val356Ile), which increased their resistance to these antimicrobials (**Fig. 4; Fig. 5**).

Interestingly, hypermutator isolates in this population were on average 10 times more susceptible to ciprofloxacin (U = 218, p < .00001) and >13 times more susceptible to tobramycin (U = 379.5, p = .0018) than normomutators. Isolates with both DNA MMR mutations in this population were additionally 36 times more susceptible to tobramycin than normomutators (U = 172.5, p < .00001). The altered ciprofloxacin phenotype may be explained in part by SNPs in *gyrA* (248 T > C, Ile83Thr) or *norM* (61 G > A, Ala21Thr). However, there are isolates in this population whose phenotypes are not explained by either of these genotypes. The increased susceptibility to tobramycin is strongly linked to the aforementioned SNP in *norM*. We saw evidence of one of these hypermutators reversing this increased susceptibility to tobramycin by acquisition of the aminoglycoside nucleotidyltransferase *ant(2")-Ia*. There was a normomutator isolate with an outlier tobramycin susceptibility phenotype. Interestingly, 12 isolates from Patient 1 had improved growth in the presence of tobramycin, a phenotype which could not be explained by any hits in the database. All of the normomutator isolates had a truncated *mexF*, although this did not appear to impact any of the tested phenotypes (**Fig. 4; Fig. 5**).

In Patient 2, hypermutators displayed increased susceptibility to meropenem (U = 194, p < .00001), piperacilin-tazobactam (U = 121.5, p < .00001), and ciprofloxacin (U = 213.5, p < .00001) than normomutators. This appears to be caused by a SNP in *mexB* (2257 T > C, Trp753Arg). However, there were outliers that could not be explained by this genotype. Hypermutators were also more susceptible to amikacin (U = 479, p = .029) and more resistant to ceftazidime (U = 417.5, p = .0045), although these strains harbored no apparent genes or SNPs associated with these phenotypes in the CARD database. There was no statistically significant difference between the tobramycin susceptibility profiles of hyper- and normomutators in this population (U = 634.5, p = .61). One hypermutator isolate in Patient 2 had an unusual density of truncated pseudogenes, 10 of which are involved in resistance mechanisms and 9 of which specifically play roles in resistance-nodulation-cell division efflux—*mexY*, *mexQ*, *mexN*, *cpxR*, *muxB*, *muxC*, *mexI*, *mexB*, *mexD*, and *cprR*. Although RGI denoted these genes as missing due to truncation, this isolate was equally or more resistant to every antimicrobial tested relative to other DNA MMR mutants in the population, suggesting that many of these genes are still functional (**Fig. 4; Fig. 5**).

In the two normomutator populations, there was significantly decreased resistome diversity. In Patient 3, a SNP in *ampC* (716 T > C, Val239Ala) resulted in increased resistance to ceftazidime. Some of the isolates with this SNP additionally were missing *nalC*, and displayed increased susceptibility to meropenem and increased resistance to ciprofloxacin relative to other isolates within this group. In Patient 4, variability in susceptibilities to amikacin, ciprofloxacin, and tobramycin were attributable to a truncation in *mexY*. Surprisingly, isolates that were missing a hit to *aph(3')-IIb* were more resistant to aminoglycosides amikacin and tobramycin. However, no hits in the CARD database explained the variations in beta-lactams meropenem, piperacilin-tazobactam, or ceftazidime (**Fig. 4**). Seeing as some of these relationships are

unexpected, it is likely that there are other genetic variations not cataloged in the CARD database that are influencing these phenotypes.

DISCUSSION

The primary goal of this study was to better understand genomic diversification in *P. aeruginosa* CF lung populations and how this drives the evolution of AMR. We found that (i) genomic diversity is highly variable, even in patients of similar age and duration of chronic *P. aeruginosa* infection, ranging from a few dozen to multiple thousands of SNPs in any one population at a single sampling point; (ii) hypermutation and *de novo* mutations were the primary drivers of within-patient diversity in these populations, although there was evidence that recombination and modifications in gene content hold secondary roles in driving diversity; and (iii) hypermutation can lead to either increased antibiotic resistance or susceptibility, and sometimes both in the same population.

Previous studies have reported that *P. aeruginosa* isolates sourced from the same sample differ by anywhere between 1 to over 800 SNPs (23, 25, 48, 49). Jorth *et al.* found in populations of 96 *P. aeruginosa* inhabiting explanted end-stage lungs of three CF patients, a total of 328, 3,169, and 1,653 SNPs (25), and Kordes *et al*. reported 770 SNPs across 47 isolates sampled from different compartments of an explanted CF lung (50). Although the maximum pairwise distance for our most diverse population (611 SNPs) was not dissimilar to previous findings, the 4,592 SNPs we identified across a population of 75 sampled isolates far exceeds previous reports. Taken together, we provide evidence that extensive heterogeneity is not an exclusive feature of end-stage lung disease and suggest that the maintenance of genetic diversity could play a key role in pathogen survival throughout the duration of chronic infection, even in early to intermediate disease stages. The data suggest that some of this diversity was driven by recombination. There may have been a (now extinct) co-resident strain of *P. aeruginosa* that exchanged genomic DNA through homologous recombination, or these may be false positives due to unusually high SNP density in genes. We observed populations with varying degrees of heterogeneity in our cohort, with low genomic diversity presumably a function of either a recent bottleneck event or the emergence of a genotype with a significant fitness advantage. This begs the question as to what factors determine whether diversifying selection or selective sweeps will dominate as the driving evolutionary force within a population, and how these forces alternate throughout infection as conditions evolve. Building upon previous work (51), conducting deep sampling of populations over long time scales may help us illuminate the full evolutionary dynamics of *P. aeruginosa* populations in the CF lung.

Previous reports on the *in vitro* long-term evolution of *P. aeruginosa* lab strains in conditions mimicking the CF lung have reported high prevalence of genetic variations in short time spans driven by hypermutation. One study evolving PA14 in biofilm-promoting arginine media found 145.7 ± 55 SNPs per population over 90 days (≈600 generations) of evolution (52), while another study evolving PAO1 found an average of 282 ± 13 SNPs after only 10 days (≈160 generations) of selection in an SCFM bead biofilm model (53). This suggests that rapid genomic diversification is not a unique signature of adaptation to a chronic infection environment. In contrast, an experimental evolution study of PA14 evolved in media of varying viscosities and nutritional complexities found on average between 5.48 to 10.29 SNPs per evolved population after ≈220 generations and detected no mutations in DNA mismatch repair mechanisms (54). This, in conjunction with our findings, suggests that while hypermutation is a major determinant of genetic diversity both *in vitro* and *in vivo*, it is not a necessary precursor to diversity, as evidenced by Patient 3 in our study. We cannot definitively say why some normomutator populations achieve substantial degrees of diversity while others do not, and it may ultimately prove challenging to precisely predict how various evolutionary forces will drive diversity in a given population, due to the stochastic processes involved.

We found that while genomic diversity was positively correlated with phenotypic diversity in AMR, even populations with lower genomic diversity displayed notable variation in resistance profiles. Previous reports of CF patients harboring infections with diverse resistance profiles are numerous (2). However, we have a poor understanding of the social interactions that may influence how microbes with varying resistance levels respond to antibiotic treatment and no reliable methodology for assessing population-level resistance. Our ability to assess resistance even at the isolate-level is inadequate. All individuals in this cohort were being treated with inhaled tobramycin, yet only one population displayed clinical resistance to tobramycin. An overwhelming number of isolates in our study were classified as clinically susceptible to the majority of antimicrobials we tested despite persisting within the lung. This is in accordance with the wide array of literature that has already called into question the utility of antimicrobial susceptibility testing in the clinic (55). We were particularly concerned to find isolates with increased growth in the presence of tobramycin, as inhaled tobramycin is one of the most commonly prescribed drugs for CF patients with *P. aeruginosa* infection.

Combining single-isolate whole genome sequencing and phenotypic characterization approaches further allowed us to understand how the evolution of various genotypes and epistatic interaction of alleles impact AMR within a population. Although we were able to identify a number of candidate genotypes responsible for these phenotypic variations, there were a number of unexplained phenotypic outliers, highlighting the presence of novel genetic signatures of AMR that have yet to be discovered. Previous reports have primarily focused on the role that hypermutation plays in evolving increased AMR in clinical populations (56-61). We found ample evidence that hypermutation can also lead to increased susceptibility, such as the hypermutator isolates in Patient 1 that were up to 36 times more susceptible to tobramycin, despite this patient undergoing treatment with inhaled tobramycin. This may be a function of either antimicrobial treatment regimens exerting uneven selection pressure on a population, or collateral sensitivity as a trade-off with resistance to other antimicrobials. The explanation to this finding may be even more deceptively simple. Mutations are more or less random, and their survival in a population is influenced by either natural selection or genetic drift. Our current understanding of microbes primarily derives from *in vitro* experiments, where it has been shown that antimicrobial treatment leads to increased resistance due to a selective advantage (i.e. populations are under strong selective pressure) (62-68). However, it may be that the evolution of genetic resistance for these populations is inconsequential *in vivo* because antimicrobials are failing to penetrate phenotypic barriers, such as biofilms, persister cells, and other mechanisms of antibiotic tolerance (69-73). In other words, the development of resistance or sensitivity *in vivo* may, in some ways, be random if antibiotic treatment regimens are only exerting weak selective pressure.

There are a few limitations of our study. Our construction of the aggregated genome of this cohort is restricted by inherent limitations of the pangenome software and short-read sequencing. As a result, we present the gene content variations as representative of broad patterns, but not exact in nature. We cannot know definitively whether variable genes have been lost or acquired, although we hypothesize that these variations overwhelmingly represent losses, as *P. aeruginosa* is known to undergo genome reduction as an adaptation to chronic infection (14, 74, 75), and we found no evidence of any co-infecting strains of *P. aeruginosa* from which DNA could be acquired. Another limitation lies in that our populations are isolated from expectorated sputum samples, which may not accurately recapitulate the whole populations. Despite these limitations, our study greatly adds to the breadth of knowledge on within-patient genomic diversity and AMR in *P. aeruginosa* CF lung infection. Based on our findings, we propose that clinicians and researchers re-consider single-isolate sampling and testing of CF sputum. We hope that future studies will expand upon these analyses with more patients to identify potential predictors of *P. aeruginosa* infection diversity, such as antimicrobial treatment history, length of chronic infection, or presence of co-infecting microbiota, and to better understand the role diversity plays in patient outcomes.

Acknowledgments
For funding, we thank the Georgia Institute of Technology, the Cystic Fibrosis Foundation for grants (DIGGLE18I0 and DIGGLE20G0) to S.P.D. and a fellowship to S.A. (AZIMI18F0), CF@LANTA for a fellowship to S.A. (3206AXB), the National Institutes of Health for a grant (R01AI153116) to S.P.D and T.R., and the National Science Foundation for a fellowship to J.V. (DGE-2039655). Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies. Access to the Cystic Fibrosis Biospecimen Registry at Emory Children's Center for CF and Airways Disease Research was provided through Children's Healthcare of Atlanta and the Emory University Pediatric CF Discovery Core. We thank Arlene Stecenko and Katy Clemmer for assistance in acquiring sputum samples.

Author Contributions
All authors contributed to research design; J.V. performed research and analyzed data; all authors contributed to writing the paper.

LEGENDS
Figure 1. Genomic diversity as measured by core genome SNPs varies greatly from one population to another, despite similar duration of chronic *P. aeruginosa* infection. Populations are presented in order of decreasing genomic diversity: Patient 1 (A), Patient 2 (B), Patient 3 (C), and Patient 4 (D). Isolates with one DNA mismatch repair mutation are highlighted in yellow on phylogenies. Isolates with two DNA mismatch repair mutations are highlighted in red.

Figure 2. Phylogeny and aggregated genomes of the 300-isolate dataset. Each tile on the x-axis indicates a gene family, where a blue tile indicates presence of that gene family at ≥90% BlastP sequence identity in a sample, and a white tile indicates absence.

Figure 3. Violin plot of the antimicrobial susceptibility profiles of all four populations against amikacin, meropenem, piperacillin-tazobactam, ciprofloxacin, tobramycin, and ceftazidime as measured by zone of inhibition in a standard disc diffusion assay.

Figure 4. Comparative antimicrobial susceptibility profiles of hypermutators and normomutators in Patient 1 (A) and Patient 2 (B) as measured by zone of inhibition in a standard disc diffusion assay. (*) indicates $p \leq .05$, (**) indicates $p \leq .01$, (***) indicates $p \leq .001$, and (****) indicates $p < .0001$ in a Mann-Whitney U test.

Figure 5. Resistomes of Patients 1 (A), 2 (B), 3 (C), and 4 (D) as predicted by the Comprehensive Antibiotic Resistance Database Resistance Gene Identifier (CARD RGI). Yellow indicates a perfect hit to the database, teal indicates a strict hit, and purple indicates no hit (or loose hit in some cases). Histogram indicates the number of isolates in the population that share a unique resistome profile. An asterisk (*) indicates a gene with resistance conferred by a mutation (i.e. CARD RGI protein variant model).

Table 1. Metadata on the four patients in our cohort: sex, cystic fibrosis transmembrane conductance regulator (CFTR) mutation status, length of *P. aeruginosa* infection, clinical status, forced expiratory volume (% FEV1), modulator therapy, antibiotic treatment, and dominant infection strain type.

Table 2. Genetic variations in each population: single nucleotide polymorphisms (SNPs), multiple nucleotide polymorphisms (MNPs), and insertions and deletions (indels).

Table 3. Proportion of genes in each population that are core, soft core, shell, or cloud genes.

Supplemental Table 1. Genome size, average sequencing coverage, and number of contigs of each assembly.

Supplemental Table 2. Genes that were impacted by non-synonymous mutations in at least one isolate in all four populations.

Supplemental Table 3. Full details of the chi-squared goodness of fit and Monte Carlo simulation exact multinomial tests, with all associated chi-squared and p-values.

Supplemental Table 4. Genes that were impacted by non-synonymous mutations in at least one isolate in three out of four populations.

Supplemental Table 5. Annotated genetic variants discovered in Patient 1.

Supplemental Table 6. Annotated genetic variants discovered in Patient 2.

Supplemental Table 7. Annotated genetic variants discovered in Patient 3.

Supplemental Table 8. Annotated genetic variants discovered in Patient 4.

Supplemental Table 9. Complete list of all genes in the aggregate genome, with information on gene annotation, frequency in the collection, average frequency per isolate, minimum nucleotide size, maximum nucleotide size, average nucleotide size, and presence in each isolate.

Supplemental Table 10. Antimicrobial susceptibility testing measurements for Patient 1 as measured by zone of inhibition (ZOI) in a standard disc diffusion assay for amikacin, meropenem, piperacilin-tazobactam, ciprofloxacin, tobramycin, and ceftazidime. Data in the left columns represent raw measurements of zone of inhibition radii (mm units). Data in the right columns represent calculated zone of inhibition values as diameters (mm units).

Supplemental Table 11. Antimicrobial susceptibility testing measurements for Patient 2 as measured by zone of inhibition (ZOI) in a standard disc diffusion assay for amikacin, meropenem, piperacilin-tazobactam, ciprofloxacin, tobramycin, and ceftazidime. Data in the left columns represent raw measurements of zone of inhibition radii (mm units). Data in the right columns represent calculated zone of inhibition values as diameters (mm units).

Supplemental Table 12. Antimicrobial susceptibility testing measurements for Patient 3 as measured by zone of inhibition (ZOI) in a standard disc diffusion assay for amikacin, meropenem, piperacilin-tazobactam, ciprofloxacin, tobramycin, and ceftazidime. Data in the left columns represent raw measurements of zone of inhibition radii (mm units). Data in the right columns represent calculated zone of inhibition values as diameters (mm units).

Supplemental Table 13. Antimicrobial susceptibility testing measurements for Patient 4 as measured by zone of inhibition (ZOI) in a standard disc diffusion assay for amikacin, meropenem, piperacilin-tazobactam, ciprofloxacin, tobramycin, and ceftazidime. Data in the left columns represent raw measurements of zone of inhibition radii (mm units). Data in the right columns represent calculated zone of inhibition values as diameters (mm units).

Supplemental Figure 1. Phylogeny of Patients 1-4 with PAO1 and PA14. Patients 1, 2, and 4 cluster with PAO1, while Patient 3 clusters with PA14.

Supplemental Figure 2. Donut plot of the relative frequencies of genes categorized within each of the 27 different PseudoCAP functional categories in the PAO1 genome (A). Donut plots of the relative frequencies of non-synonymous SNPs and indels located in each of the 27 different

PseudoCAP functional categories in Patient 1 (B), 2 (C), 3 (D), and 4 (E). Enrichment analysis of the frequency of functional categories in which genetic variants are found in each of the four populations relative to the proportions of these functional categories in the PAO1 genome shows that protein secretion/ export apparatuses and transcriptional regulators are enriched for such variants, while phage/ transposon/ plasmid and non-coding RNA are less likely to be impacted by such variants.

Supplemental Figure 3. Violin plot of the antimicrobial susceptibility profiles of all four populations against amikacin, meropenem, piperacillin-tazobactam, ciprofloxacin, tobramycin, and ceftazidime as measured by zone of inhibition in a standard disc diffusion assay. Black horizontal bars indicate the cut-off values for susceptibility (top bar) and resistance (bottom bar) for each antibiotic as determined by the Clinical and Laboratory Standards Institute (CLSI). This plot displays the same data as Fig. 3, displayed in clusters by antibiotic rather than by population.

Supplemental Figure 4. Genomic diversity (as measured by total SNPs in the population) is positively correlated with diversity in antimicrobial susceptibility (as measured by standard deviation in the disc diffusion assay) in our four-patient cohort for all six antibiotics: amikacin (A), meropenem (B), piperacillin-tazobactam (C), ciprofloxacin (D), tobramycin (E), and ceftazidime (F).

Bibliography

1. L. Jackson, V. Waters, Factors influencing the acquisition and eradication of early Pseudomonas aeruginosa infection in cystic fibrosis. *J Cyst Fibros* **20**, 8-16 (2021).
2. S. T. Clark, D. S. Guttman, D. M. Hwang, Diversification of Pseudomonas aeruginosa within the cystic fibrosis lung and its effects on antibiotic resistance. *FEMS Microbiol Lett* **365** (2018).
3. J. A. Bartell *et al.*, Omics-based tracking of Pseudomonas aeruginosa persistence in "eradicated" cystic fibrosis patients. *Eur Respir J* **57** (2021).
4. L. Camus, F. Vandenesch, K. Moreau, From genotype to phenotype: adaptations of Pseudomonas aeruginosa to the cystic fibrosis environment. *Microb Genom* **7** (2021).
5. H. H. Mehta *et al.*, The Essential Role of Hypermutation in Rapid Adaptation to Antibiotic Stress. *Antimicrob Agents Chemother* **63** (2019).
6. N. Cramer *et al.*, Microevolution of the major common Pseudomonas aeruginosa clones C and PA14 in cystic fibrosis lungs. *Environ Microbiol* **13**, 1690-1704 (2011).
7. R. L. Marvig, H. K. Johansen, S. Molin, L. Jelsbak, Genome analysis of a transmissible lineage of pseudomonas aeruginosa reveals pathoadaptive mutations and distinct evolutionary paths of hypermutators. *PLoS Genet* **9**, e1003741 (2013).
8. S. Feliziani *et al.*, Coexistence and within-host evolution of diversified lineages of hypermutable Pseudomonas aeruginosa in long-term cystic fibrosis infections. *PLoS Genet* **10**, e1004651 (2014).
9. T. Markussen *et al.*, Environmental heterogeneity drives within-host diversification and evolution of Pseudomonas aeruginosa. *mBio* **5**, e01592-01514 (2014).
10. R. L. Marvig *et al.*, Within-host microevolution of Pseudomonas aeruginosa in Italian cystic fibrosis patients. *BMC Microbiol* **15**, 218 (2015).
11. R. L. Marvig, L. M. Sommer, S. Molin, H. K. Johansen, Convergent evolution and adaptation of Pseudomonas aeruginosa within patients with cystic fibrosis. *Nat Genet* **47**, 57-64 (2015).
12. I. Bianconi *et al.*, Persistence and Microevolution of Pseudomonas aeruginosa in the Cystic Fibrosis Lung: A Single-Patient Longitudinal Genomic Study. *Front Microbiol* **9**, 3242 (2018).
13. J. Klockgether, N. Cramer, S. Fischer, L. Wiehlmann, B. Tummler, Long-Term Microevolution of Pseudomonas aeruginosa Differs between Mildly and Severely Affected Cystic Fibrosis Lungs. *Am J Respir Cell Mol Biol* **59**, 246-256 (2018).
14. M. Gabrielaite, H. K. Johansen, S. Molin, F. C. Nielsen, R. L. Marvig, Gene Loss and Acquisition in Lineages of Pseudomonas aeruginosa Evolving in Cystic Fibrosis Patient Airways. *mBio* **11** (2020).
15. R. Datar *et al.*, Phenotypic and Genomic Variability of Serial Peri-Lung Transplantation Pseudomonas aeruginosa Isolates From Cystic Fibrosis Patients. *Front Microbiol* **12**, 604555 (2021).
16. S. J. T. Wardell *et al.*, Genome evolution drives transcriptomic and phenotypic adaptation in Pseudomonas aeruginosa during 20 years of infection. *Microb Genom* **7** (2021).
17. D. I. Andersson, H. Nicoloff, K. Hjort, Mechanisms and clinical relevance of bacterial heteroresistance. *Nat Rev Microbiol* **17**, 479-496 (2019).
18. E. Reece, P. H. A. Bettio, J. Renwick, Polymicrobial Interactions in the Cystic Fibrosis Airway Microbiome Impact the Antimicrobial Susceptibility of Pseudomonas aeruginosa. *Antibiotics (Basel)* **10** (2021).
19. E. Mowat *et al.*, Pseudomonas aeruginosa population diversity and turnover in cystic fibrosis chronic infections. *Am J Respir Crit Care Med* **183**, 1674-1679 (2011).

20. A. Ashish *et al.*, Extensive diversification is a common feature of Pseudomonas aeruginosa populations during respiratory infections in cystic fibrosis. *J Cyst Fibros* **12**, 790-793 (2013).
21. M. L. Workentine *et al.*, Phenotypic heterogeneity of Pseudomonas aeruginosa populations in a cystic fibrosis patient. *PLoS One* **8**, e60225 (2013).
22. S. T. Clark *et al.*, Phenotypic diversity within a Pseudomonas aeruginosa population infecting an adult with cystic fibrosis. *Sci Rep* **5**, 10932 (2015).
23. D. Williams *et al.*, Divergent, coexisting Pseudomonas aeruginosa lineages in chronic cystic fibrosis lung infections. *Am J Respir Crit Care Med* **191**, 775-785 (2015).
24. D. Williams *et al.*, Transmission and lineage displacement drive rapid population genomic flux in cystic fibrosis airway infections of a Pseudomonas aeruginosa epidemic strain. *Microb Genom* **4** (2018).
25. P. Jorth *et al.*, Regional Isolation Drives Bacterial Diversification within Cystic Fibrosis Lungs. *Cell Host Microbe* **18**, 307-319 (2015).
26. K. L. Palmer, L. M. Aye, M. Whiteley, Nutritional cues control Pseudomonas aeruginosa multicellular behavior in cystic fibrosis sputum. *J Bacteriol* **189**, 8079-8087 (2007).
27. R. R. Wick, L. M. Judd, C. L. Gorrie, K. E. Holt, Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* **13**, e1005595 (2017).
28. O. N. T. Research (2018) Medaka. (Oxford Nanopore Technologies).
29. R. A. Petit, 3rd, T. D. Read, Bactopia: a Flexible Pipeline for Complete Analysis of Bacterial Genomes. *mSystems* **5** (2020).
30. R. R. Wick, K. E. Holt, Polypolish: Short-read polishing of long-read bacterial genome assemblies. *PLoS Comput Biol* **18**, e1009802 (2022).
31. B. J. Walker *et al.*, Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963 (2014).
32. T. Seemann (2020) Snippy.
33. T. Seemann, Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-2069 (2014).
34. T. D. Otto, G. P. Dillon, W. S. Degrave, M. Berriman, RATT: Rapid Annotation Transfer Tool. *Nucleic Acids Res* **39**, e57 (2011).
35. K. A. Jolley, J. E. Bray, M. C. J. Maiden, Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res* **3**, 124 (2018).
36. S. Guindon *et al.*, New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307-321 (2010).
37. P. Danecek *et al.*, The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158 (2011).
38. K. Brinda (2017) Disty McMatrixface.
39. P. Cingolani *et al.*, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80-92 (2012).
40. B. Engels (2015) XNomial.
41. N. J. Croucher *et al.*, Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* **43**, e15 (2015).
42. T. Seemann, Shovill. (2020).
43. A. J. Page *et al.*, Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691-3693 (2015).
44. L. T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**, 268-274 (2015).

45. B. P. Alcock *et al.*, CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res* **51**, D690-D699 (2023).
46. S. C. Bayliss, H. A. Thorpe, N. M. Coyle, S. K. Sheppard, E. J. Feil, PIRATE: A fast and scalable pangenomics toolbox for clustering diverged orthologues in bacteria. *Gigascience* **8** (2019).
47. A. Toledo-Arana, F. Repoila, P. Cossart, Small noncoding RNAs controlling pathogenesis. *Curr Opin Microbiol* **10**, 182-188 (2007).
48. J. C. Chung *et al.*, Genomic variation among contemporary Pseudomonas aeruginosa isolates from chronically infected cystic fibrosis patients. *J Bacteriol* **194**, 4857-4866 (2012).
49. S. E. Darch *et al.*, Recombination is a key driver of genomic and phenotypic diversity in a Pseudomonas aeruginosa population during cystic fibrosis infection. *Sci Rep* **5**, 7649 (2015).
50. A. Kordes *et al.*, Genetically diverse Pseudomonas aeruginosa populations display similar transcriptomic profiles in a cystic fibrosis explanted lung. *Nat Commun* **10**, 3397 (2019).
51. J. Diaz Caballero *et al.*, Selective Sweeps and Parallel Pathoadaptation Drive Pseudomonas aeruginosa Evolution in the Cystic Fibrosis Lung. *mBio* **6**, e00981-00915 (2015).
52. K. B. Harris, K. M. Flynn, V. S. Cooper, Polygenic Adaptation and Clonal Interference Enable Sustained Diversity in Experimental Pseudomonas aeruginosa Populations. *Mol Biol Evol* **38**, 5359-5375 (2021).
53. S. Azimi *et al.*, Allelic polymorphism shapes community function in evolving Pseudomonas aeruginosa populations. *ISME J* 10.1038/s41396-020-0652-0 (2020).
54. A. Schick, S. Shewaramani, R. Kassen, Genomics of Diversification of Pseudomonas aeruginosa in Cystic Fibrosis Lung-like Conditions. *Genome Biol Evol* **14** (2022).
55. R. Somayaji *et al.*, Antimicrobial susceptibility testing (AST) and associated clinical outcomes in individuals with cystic fibrosis: A systematic review. *J Cyst Fibros* **18**, 236-243 (2019).
56. M. D. Macia *et al.*, Hypermutation is a key factor in development of multiple-antimicrobial resistance in Pseudomonas aeruginosa strains causing chronic lung infections. *Antimicrob Agents Chemother* **49**, 3382-3386 (2005).
57. G. Cabot *et al.*, Evolution of Pseudomonas aeruginosa Antimicrobial Resistance and Fitness under Low and High Mutation Rates. *Antimicrob Agents Chemother* **60**, 1767-1778 (2016).
58. P. P. Khil *et al.*, Dynamic Emergence of Mismatch Repair Deficiency Facilitates Rapid Evolution of Ceftazidime-Avibactam Resistance in Pseudomonas aeruginosa Acute Infection. *mBio* **10** (2019).
59. V. E. Rees *et al.*, Characterization of Hypermutator Pseudomonas aeruginosa Isolates from Patients with Cystic Fibrosis in Australia. *Antimicrob Agents Chemother* **63** (2019).
60. C. A. Colque *et al.*, Hypermutator Pseudomonas aeruginosa Exploits Multiple Genetic Pathways To Develop Multidrug Resistance during Long-Term Infections in the Airways of Cystic Fibrosis Patients. *Antimicrob Agents Chemother* **64** (2020).
61. A. Dulanto Chiang *et al.*, Hypermutator strains of Pseudomonas aeruginosa reveal novel pathways of resistance to combinations of cephalosporin antibiotics and beta-lactamase inhibitors. *PLoS Biol* **20**, e3001878 (2022).
62. A. Wong, N. Rodrigue, R. Kassen, Genomics of adaptation during experimental evolution of the opportunistic pathogen Pseudomonas aeruginosa. *PLoS Genet* **8**, e1002928 (2012).
63. K. M. Jorgensen *et al.*, Sublethal ciprofloxacin treatment leads to rapid development of high-level ciprofloxacin resistance during long-term experimental evolution of Pseudomonas aeruginosa. *Antimicrob Agents Chemother* **57**, 4215-4221 (2013).

64. P. Jorth *et al.*, Evolved Aztreonam Resistance Is Multifactorial and Can Produce Hypervirulence in Pseudomonas aeruginosa. *mBio* **8** (2017).
65. M. N. Ahmed, A. Porse, M. O. A. Sommer, N. Hoiby, O. Ciofu, Evolution of Antibiotic Resistance in Biofilm and Planktonic Pseudomonas aeruginosa Populations Exposed to Subinhibitory Levels of Ciprofloxacin. *Antimicrob Agents Chemother* **62** (2018).
66. F. Sanz-Garcia, S. Hernando-Amado, J. L. Martinez, Mutation-Driven Evolution of Pseudomonas aeruginosa in the Presence of either Ceftazidime or Ceftazidime-Avibactam. *Antimicrob Agents Chemother* **62** (2018).
67. S. J. T. Wardell *et al.*, A large-scale whole-genome comparison shows that experimental evolution in response to antibiotics predicts changes in naturally evolved clinical Pseudomonas aeruginosa. *Antimicrob Agents Chemother* **63** (2019).
68. M. N. Ahmed *et al.*, The evolutionary trajectories of P. aeruginosa in biofilm and planktonic growth modes exposed to ciprofloxacin: beyond selection of antibiotic resistance. *NPJ Biofilms Microbiomes* **6**, 28 (2020).
69. T. F. Moriarty, J. C. McElnay, J. S. Elborn, M. M. Tunney, Sputum antibiotic concentrations: implications for treatment of cystic fibrosis lung infection. *Pediatr Pulmonol* **42**, 1008-1017 (2007).
70. O. Ciofu, T. Tolker-Nielsen, P. O. Jensen, H. Wang, N. Hoiby, Antimicrobial resistance, respiratory tract infections and role of biofilms in lung infections in cystic fibrosis patients. *Adv Drug Deliv Rev* **85**, 7-23 (2015).
71. I. Martin, V. Waters, H. Grasemann, Approaches to Targeting Bacterial Biofilms in Cystic Fibrosis Airways. *Int J Mol Sci* **22** (2021).
72. I. Santi, P. Manfredi, E. Maffei, A. Egli, U. Jenal, Evolution of Antibiotic Tolerance Shapes Resistance Development in Chronic Pseudomonas aeruginosa Infections. *mBio* **12** (2021).
73. C. Witzany, R. R. Regoes, C. Igler, Assessing the relative importance of bacterial resistance, persistence and hyper-mutation for antibiotic treatment failure. *Proc Biol Sci* **289**, 20221300 (2022).
74. M. H. Rau, R. L. Marvig, G. D. Ehrlich, S. Molin, L. Jelsbak, Deletion and acquisition of genomic content during early stage adaptation of Pseudomonas aeruginosa to a human host environment. *Environ Microbiol* **14**, 2200-2211 (2012).
75. C. R. Armbruster *et al.*, Adaptation and genomic erosion in fragmented Pseudomonas aeruginosa populations in the sinuses of people with cystic fibrosis. *Cell Rep* **37**, 109829 (2021).

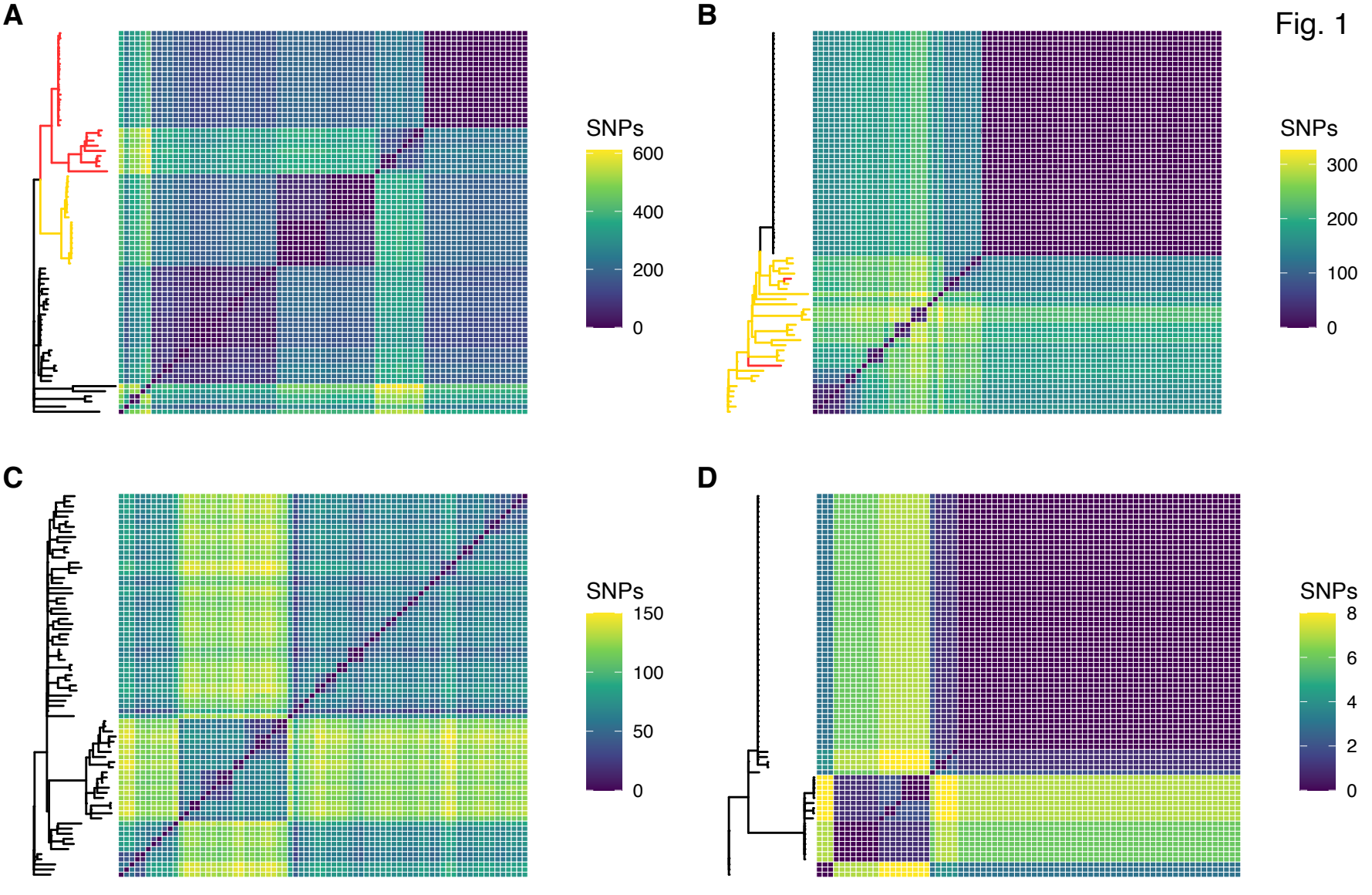

A
SNPs
600
400
200
0
B
SNPs
300
200
100
0
C
SNPs
150
100
50
0
D
SNPs
8
6
4
2
0

Fig. 1

Fig. 2

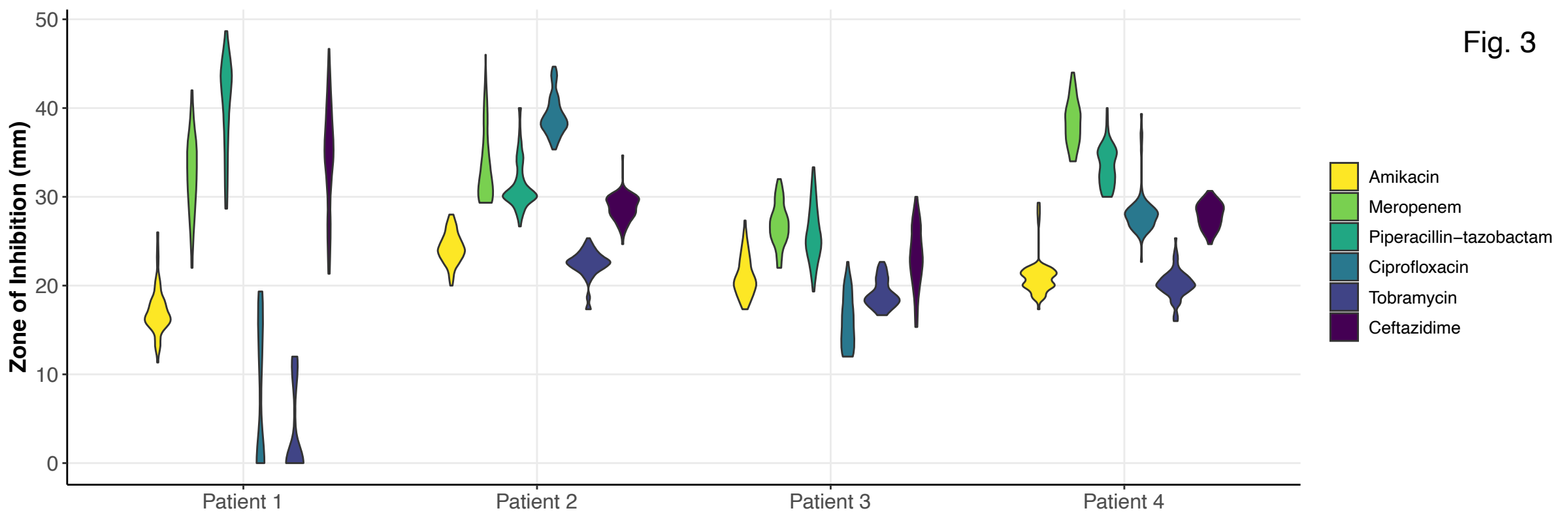


Fig. 3
Zone of Inhibition (mm)
50
40
30
20
10
0
Patient 1
Patient 2
Patient 3
Patient 4
Amikacin
Meropenem
Piperacillin–tazobactam
Ciprofloxacin
Tobramycin
Ceftazidime

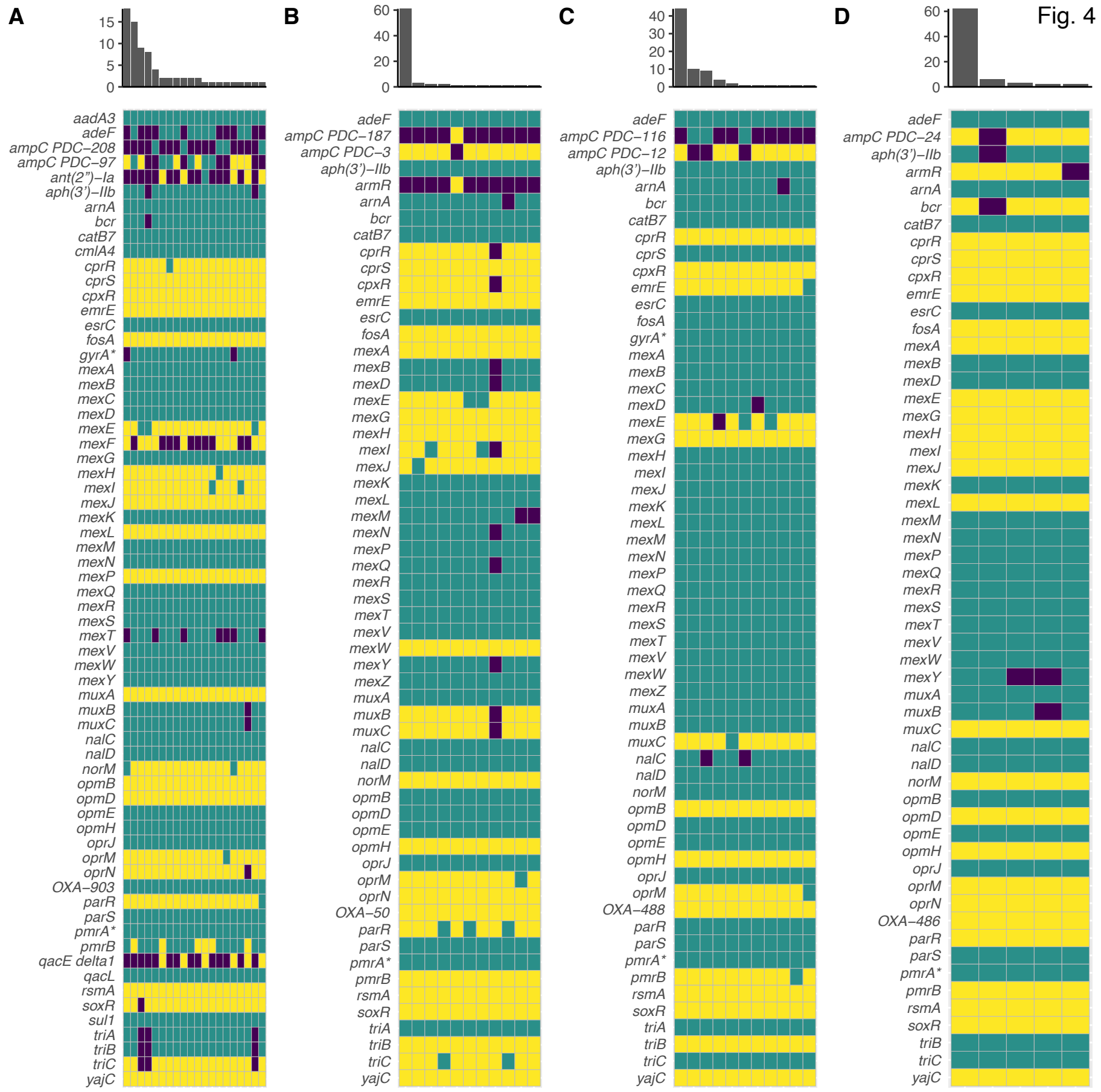


Fig. 4
A
B
C
D
aadA3
adeF
ampC PDC–208
ampC PDC–97
ant(2”)–Ia
aph(3’)–IIb
arnA
bcr
catB7
cmlA4
cprR
cprS
cpxR
emrE
esrC
fosA
gyrA*
mexA
mexB
mexC
mexD
mexE
mexF
mexG
mexH
mexI
mexJ
mexK
mexL
mexM
mexN
mexP
mexQ
mexR
mexS
mexT
mexV
mexW
mexY
muxA
muxB
muxC
nalC
nalD
norM
opmB
opmD
opmE
opmH
oprJ
oprM
oprN
OXA–903
parR
parS
pmrA*
pmrB
qacE delta1
qacL
rsmA
soxR
sul1
triA
triB
triC
yajC
adeF
ampC PDC–187
ampC PDC–3
aph(3’)–IIb
armR
arnA
bcr
catB7
cprR
cprS
cpxR
emrE
esrC
fosA
mexA
mexB
mexD
mexE
mexG
mexH
mexI
mexJ
mexK
mexL
mexM
mexN
mexP
mexQ
mexR
mexS
mexT
mexV
mexW
mexY
mexZ
muxA
muxB
muxC
nalC
nalD
norM
opmB
opmD
opmE
opmH
oprJ
oprM
oprN
OXA–50
parR
parS
pmrA*
pmrB
rsmA
soxR
triA
triB
triC
yajC
adeF
ampC PDC–116
ampC PDC–12
aph(3’)–IIb
arnA
bcr
catB7
cprR
cprS
cpxR
emrE
esrC
fosA
gyrA*
mexA
mexB
mexC
mexD
mexE
mexG
mexH
mexI
mexJ
mexK
mexL
mexM
mexN
mexP
mexQ
mexR
mexS
mexT
mexV
mexW
mexZ
muxA
muxB
muxC
nalC
nalD
norM
opmB
opmD
opmE
opmH
oprJ
oprM
OXA–488
parR
parS
pmrA*
pmrB
rsmA
soxR
triA
triB
triC
yajC
adeF
ampC PDC–24
aph(3’)–IIb
armR
arnA
bcr
catB7
cprR
cprS
cpxR
emrE
esrC
fosA
mexA
mexB
mexD
mexE
mexG
mexH
mexI
mexJ
mexK
mexL
mexM
mexN
mexP
mexQ
mexR
mexS
mexT
mexV
mexW
mexY
muxA
muxB
muxC
nalC
nalD
norM
opmB
opmD
opmE
opmH
oprJ
oprM
oprN
OXA–486
parR
parS
pmrA*
pmrB
rsmA
soxR
triB
triC
yajC

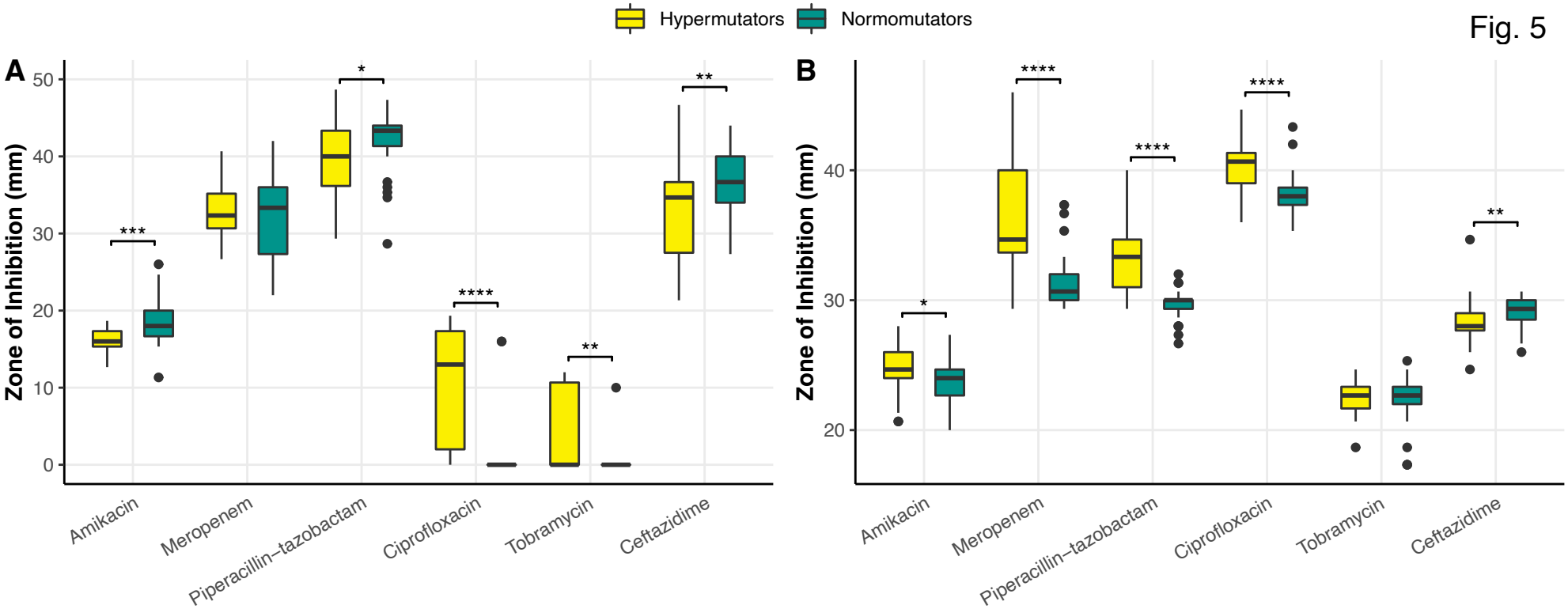

Hypermutators
Normomutators
A
B
Zone of Inhibition (mm)
0
10
20
30
40
50
***
*
****
**
Amikacin
Meropenem
Piperacillin–tazobactam
Ciprofloxacin
Tobramycin
Ceftazidime

Fig. 5

Table 1. Metadata on the four patients in our cohort: sex, cystic fibrosis transmembrane conductance regulator (CFTR) mutation status, length of *P. aeruginosa* infection, clinical status, forced expiratory volume (% FEV1), modulator therapy, antibiotic treatment, and dominant infection strain type.

	Patient 1	**Patient 2**	**Patient 3**	**Patient 4**
Patient Sex	F	F	F	M
CFTR Mutation	F508del/R1162X	F508del/F508del	F508del/L467P	F508del/ 621+1G->T
Length of *Pa* infection	15 years, 2 months	12 years, 5 months	10 years, 4 months	13 years
Clinical status	APE Outpatient	APE Outpatient	Stable	APE Outpatient
FEV1 (%)	67.96%	74.92%	67.83%	60.30%
Modulator	None	None	None	None
Antibiotic Treatment	Inhaled tobramycin, oral azithromycin	Inhaled tobramycin, oral Trimethoprim / Sulfamethoxazole	Inhaled tobramycin, inhaled aztreonam, oral azithromycin	Inhaled tobramycin, oral Trimethoprim / Sulfamethoxazole, oral levofloxavcin
Dominant ST	870	2999	1197	274

Table 2. Genetic variations in each population: single nucleotide polymorphisms (SNPs), multiple nucleotide polymorphisms (MNPs), and insertions and deletions (indels).

	Patient 1	Patient 2	Patient 3	Patient 4
Total # unique SNPs/ MNPs	4592	1972	1638	31
# SNPs/ MNPs separating most divergent isolates	611	326	150	8
Non-synonymous SNPs/ MNPs	2803	1294	1024	24
Synonymous SNPs/ MNPs	1248	484	425	5
SNPs in non-coding regions	541	194	189	2
Total # indels	498	307	330	14
Indels in non-coding regions	204	99	115	2

Table 3. Proportion of genes in each population that are core, soft core, shell, or cloud genes.

	1	2	3	4	All
Core (>98%)	4816 (84.8%)	5537 (96.7%)	5895 (97.7%)	5463 (97.0%)	4762 (69.2%)
Soft core (95%-98%)	120 (2.1%)	9 (0.16%)	42 (0.70%)	10 (0.18%)	202 (2.9%)
Shell (15%-95%)	655 (11.5%)	32 (0.56%)	38 (0.63%)	103 (1.8%)	1615 (23.5%)
Cloud (<15%)	90 (1.6%)	148 (2.6%)	57 (0.94%)	57 (1.0%)	301 (4.4%)
Total	**5681**	**5726**	**6032**	**5633**	**6880**

Fig. S1

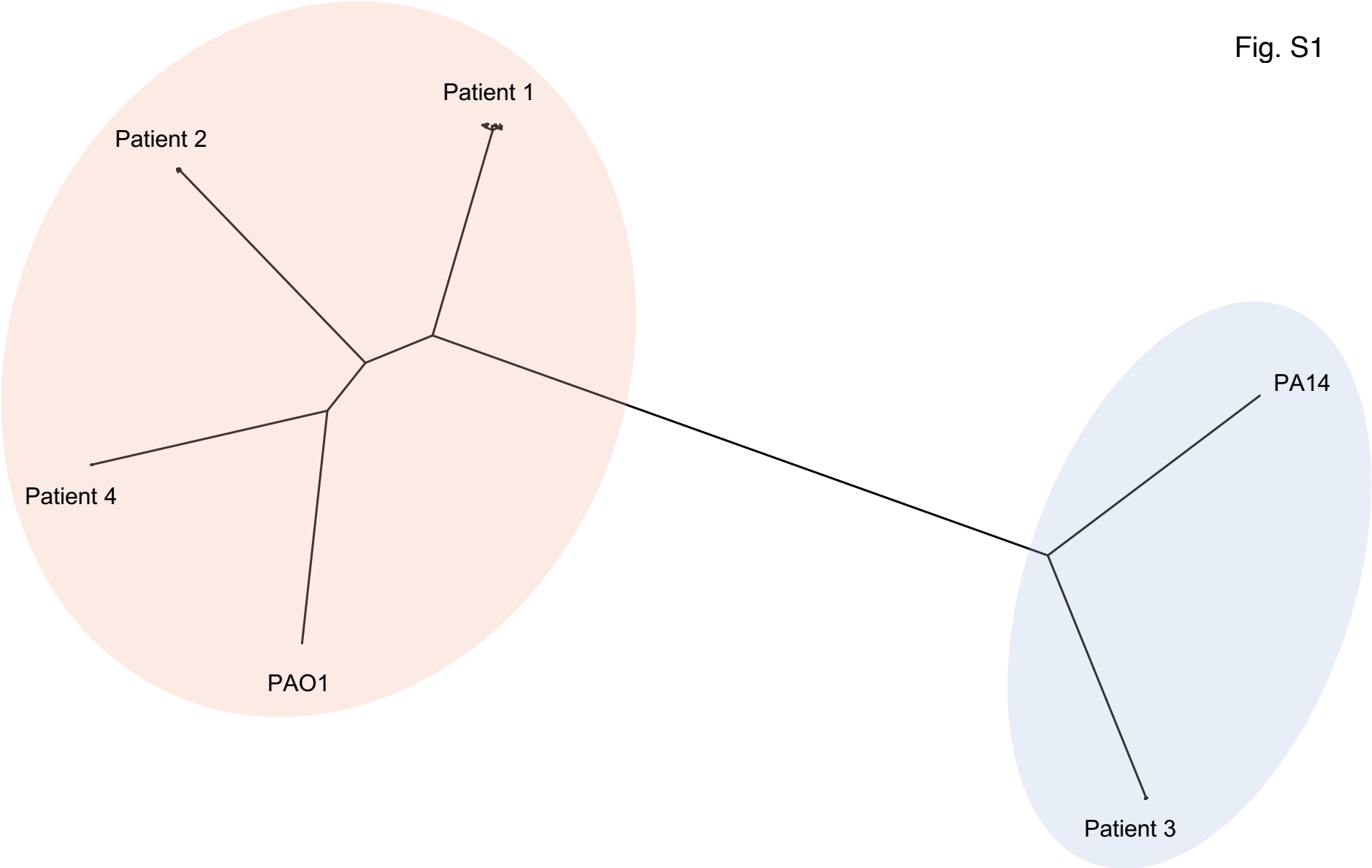
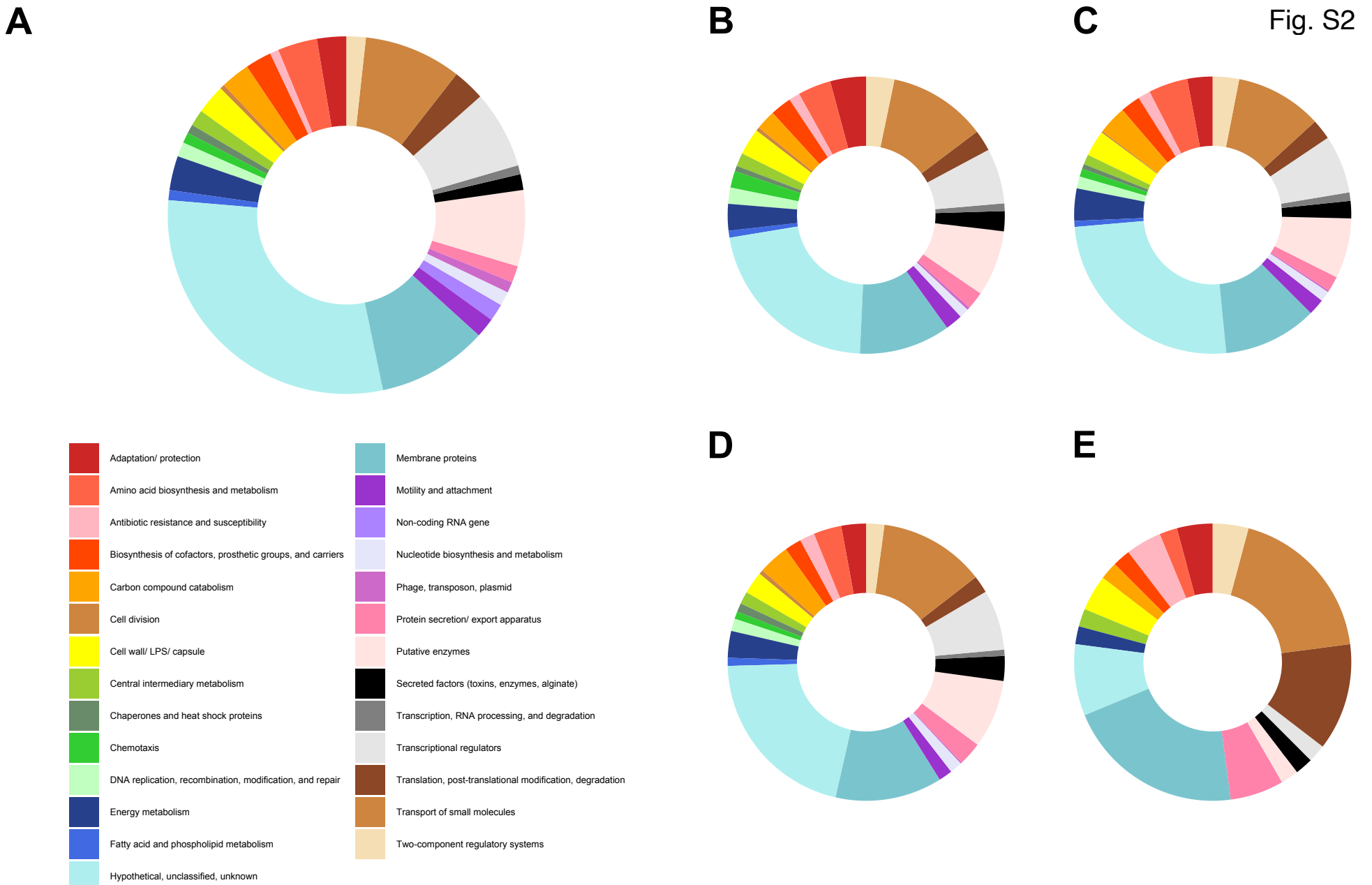


Fig. S2


A
B
C
D
E
Adaptation/ protection
Amino acid biosynthesis and metabolism
Antibiotic resistance and susceptibility
Biosynthesis of cofactors, prosthetic groups, and carriers
Carbon compound catabolism
Cell division
Cell wall/ LPS/ capsule
Central intermediary metabolism
Chaperones and heat shock proteins
Chemotaxis
DNA replication, recombination, modification, and repair
Energy metabolism
Fatty acid and phospholipid metabolism
Hypothetical, unclassified, unknown
Membrane proteins
Motility and attachment
Non-coding RNA gene
Nucleotide biosynthesis and metabolism
Phage, transposon, plasmid
Protein secretion/ export apparatus
Putative enzymes
Secreted factors (toxins, enzymes, alginate)
Transcription, RNA processing, and degradation
Transcriptional regulators
Translation, post-translational modification, degradation
Transport of small molecules
Two-component regulatory systems

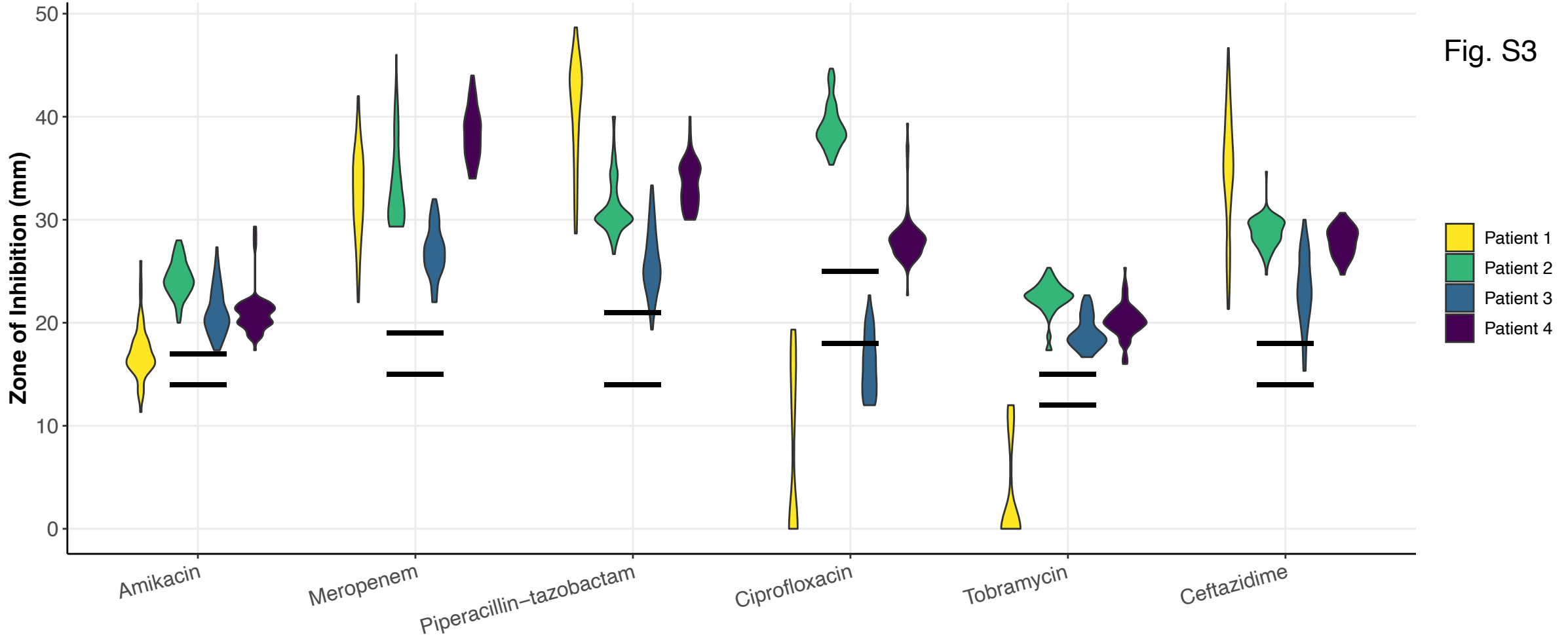
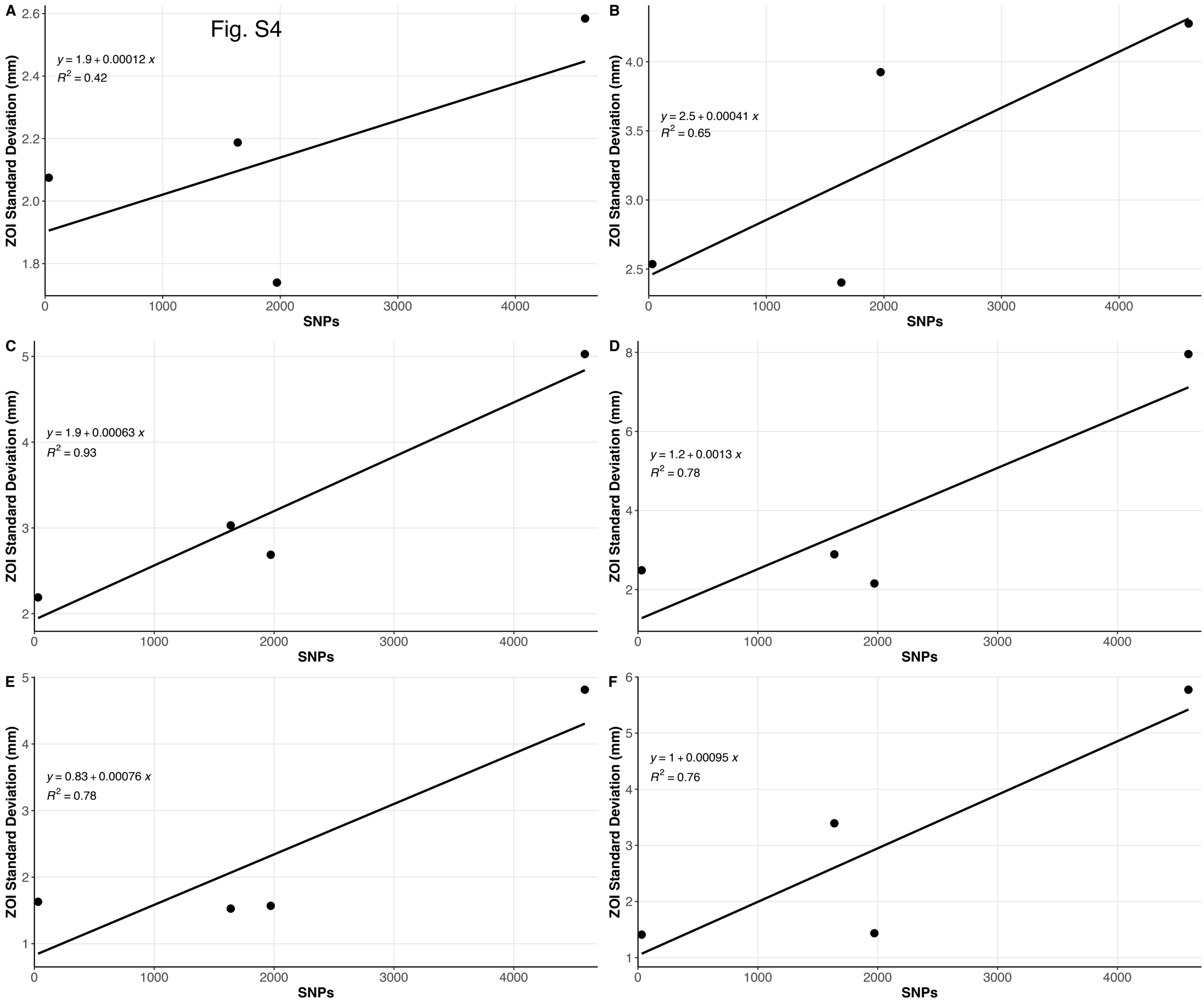


Fig. S3
Zone of Inhibition (mm)
50
40
30
20
10
0
Amikacin
Meropenem
Piperacillin–tazobactam
Ciprofloxacin
Tobramycin
Ceftazidime
Patient 1
Patient 2
Patient 3
Patient 4

Fig. S4


A
$y = 1.9 + 0.00012\,x$
$R^2 = 0.42$
B
$y = 2.5 + 0.00041\,x$
$R^2 = 0.65$
C
$y = 1.9 + 0.00063\,x$
$R^2 = 0.93$
D
$y = 1.2 + 0.0013\,x$
$R^2 = 0.78$
E
$y = 0.83 + 0.00076\,x$
$R^2 = 0.78$
F
$y = 1 + 0.00095\,x$
$R^2 = 0.76$
ZOI Standard Deviation (mm)
SNPs