Long-term Persistence of an Extensively Drug Resistant Subclade of Globally Distributed *Pseudomonas aeruginosa* Clonal Complex 446 in an Academic Medical Center

SUPPLEMENTARY MATERIALS

SUPPLEMENTARY METHODS

Antimicrobial Resistance Determination

Minimal inhibitory concentrations (MICs) were determined in triplicate using the broth microdilution protocol by Wiegand, et al. [1] and are reported in Supplementary Table 2 and Table 2. The following antibiotics were prepared from commercially available sources and were used to assess MICs: gentamicin, cefepime, ceftazidime, piperacillin-tazobactam, meropenem, aztreonam, ciprofloxacin, and colistin. Where discordant values were obtained, the median was used. Isolates were classified as susceptible or non-susceptible (intermediate and resistant) to each antibiotic based on 2018 Clinical and Laboratory Standards Institute (CLSI) breakpoints [2]. An isolate was classified as multidrug resistant (MDR) if it was non-susceptible to at least one antibiotic from ≥3 classes tested and classified as extensively drug resistant (XDR) if non-susceptible to at least one antibiotic from ≥5 classes tested (susceptible to antibiotics tested from ≤2 classes) [3].

Susceptibility to ceftazidime-avibactam and ceftolozane-tazobactam was assessed through Kirby-Bauer disk diffusion testing using HardyDisk AST disks (Hardy Diagnostics). Isolates were classified as susceptible or non-susceptible based on 2018 CLSI breakpoints [2].

BURST Analysis

To investigate the relationships between ST298, ST446, and other related sequence types, BURST analysis was performed using the goeBURST algorithm [4] as implemented in PHYLOViZ (v2.0) [5], and the resulting clonal complex containing these sequence types was

identified. All *Pseudomonas aeruginosa* sequence types listed in the PubMLST database (accessed August 12, 2019) were considered [6].

Whole Genome Sequencing

To construct a complete genome sequence for PABL048, long-read sequencing was performed on a PacBio RS II machine at the University of Maryland Institute for Genome Sciences. PacBio raw data were corrected and assembled using HGAP assembler (SMRT Analysis 2.3.0), Canu assembler v1.2, and Celera assembler v8.2. The assemblers were run using default settings. Resulting contigs were combined and circularized using Circlator v1.5.1. The final assembly was polished using Quiver (SMRT Analysis 2.3.0). Indel errors were corrected with Pilon v1.21 using 100-bp paired-end reads generated on an Illumina HiSeq 2000 system, with an average read coverage of 190-fold. The genome was annotated through the NCBI Prokaryotic Genome Annotation Pipeline [7] and has been deposited to GenBank with the accession numbers CP039293.1 (chromosome) and CP039294.1 (plasmid). Potential virulence factors present within the PABL048 chromosome and plasmid were identified using the VFanalyzer pipeline to screen against the virulence factor database (VFDB) [8].

CC446 isolates were whole genome sequenced using Illumina HiSeq and MiSeq platforms. Sequencing was performed at Northwestern University Feinberg School of Medicine and at the University of Maryland Institute for Genome Sciences. Reads were trimmed using Trimmomatic (v0.36) [9] to remove low-quality bases and assembled into contigs using SPAdes (v3.9.1) [10]. Contigs shorter than 200 bp were filtered out. Short-read sequencing was additionally performed to investigate heterogenous plasmid presence in PABL036 and PABL067 and to confirm plasmid curing in PABL048 using a MiSeq Reagent Kit v3 (600 cycle) on an Illumina MiSeq instrument to yield 2 x 300 bp paired-end reads.

Sequence Alignment

All CC446 genomes were aligned to the complete genome sequence of PABL048, with separate alignments to the PABL048 chromosome and pPABL048 plasmid. For isolates with

reads available (Supplementary Table 1), read-trimming was performed with Trimmomatic (v0.36) [9] to remove low quality bases, and alignment was then performed using BWA (v0.7.15) with the BWA-MEM algorithm [11]. Single nucleotide variants (SNVs) relative to the reference were called using the mpileup function of samtools (v0.1.19-44428cd) [12, 13] with the following settings: -E (recalculate extended BAQ), -M 0 (cap mapping quality at 0), -Q 25 (skip bases with BAQ less than 25), -q 30 (skip alignments with mapQ less than 30), -m 2 (minimum gapped reads for indel candidates of 2), -D (output per-sample DP in binary call format [BCF]), -S (output per-sample strand bias P-value in BCF), and -g (generate BCF output). SNVs were filtered if they failed to meet 1 or more of the following criteria: minimum SNV quality score of 200, minimum read consensus of 75%, minimum of 5 reads covering the SNV position, maximum of 3 times the median read depth of the total alignment, minimum of 1 read in either direction covering the SNV position, homozygous under the diploid model, and not within a repetitive region as determined by BLAST alignment of fragments of the reference sequence against itself. Any positions in the reference sequence with SNVs that passed the above filters were changed to the SNV base. Positions with SNVs that did not pass the above filters were changed to a missing base character. Non-SNV positions with coverage of fewer than 5 reads were also changed to a missing base character. For 13 NCBI genomes usable reads were not available, so draft genome contigs were aligned to PABL048 using NUCmer (v3.1) [14] with SNVs within repetitive regions masked (replaced with “N”).

For subsequent phylogenetic analysis, alignments of all CC446 isolates to the PABL048 chromosome were concatenated into a single fasta file. The core genome was defined as all non-missing and non-filtered positions present in 91 (98%) of the 92 genomes. Bases in all non-core positions were replaced by the corresponding base in the PABL048 reference.

Phylogenetic Analysis

A maximum likelihood phylogenetic tree was constructed based on core genome alignments to PABL048 using RAxML (v8.2.11) [15]. Tree construction was performed using a

gamma model of rate heterogeneity (-m GTRGAMMA) with 1000 rapid bootstraps (-f a -N 1000) to assess support. The phylogenetic tree was corrected for the impact of recombination using ClonalFrameML (v1.11-3-g4f12f23) with default settings [16]. The recombination-corrected phylogenetic tree was visualized and annotated using iTOL [17]. Predicted recombinant regions were then masked in the core genome alignment.

To model the evolution of the ST298* subclade, a time-scaled phylogenetic tree was constructed for bacteria from this subclade with known isolation dates (Supplementary Tables 1 and 2). Recombination-filtered core genome alignments of these isolates to the PABL048 chromosome were extracted, and the recombination-corrected maximum likelihood tree was pruned to contain only these isolates using the ape package in R [18, 19]. These, along with isolation dates, were used as input for Bayesian analysis. Bayesian analysis was performed using BEAST (v2.5.1) with a gamma site model, strict clock rate, Yule tree prior, and chain length of 100 million, sampling every 1000 states [20]. Sampled states were analyzed with Tracer (v1.7.1) to determine the clock rate and last common ancestor date for ST298*, considering the first 10% of states as burn-in [21]. SNVs/year were determined by multiplying estimated clock rate (SNVs per site per year) by PABL048 chromosome size. To construct the final time-scaled tree, sample trees were used to form a maximum clade credibility tree with node heights as common ancestor heights, using the first 10% of trees as burn-in. The time-scaled tree was visualized using FigTree (v1.4.4).

Determination of Heterogenous Plasmid Presence in ST298* and Curing Plasmid from PABL048

ST298 isolates from NMH were screened for heterogenous resistance to gentamicin by patching individual colonies onto LB agar supplemented with gentamicin (50 μg/mL). Gentamicin-resistant and -sensitive colonies of PABL036 and PABL067 were selected for further analysis. These underwent MIC testing and whole genome sequencing as described

above. Plasmid presence and chromosomal SNVs were determined by read alignment to the PABL048 complete genome.

To cure pPABL048 from PABL048, we used a combination of sodium dodecyl sulfate (SDS) and elevated temperature [22]. Colonies were inoculated into 5 mL LB with 2% SDS, cultured at 42ºC for approximately 24 hours, and plated on LB agar with irgasan (5 µg/mL). Colonies were then screened for loss of gentamicin resistance on LB agar supplemented with gentamicin (50 µg/mL) and irgasan (5 µg/mL) as a marker for pPABL048 presence. Loss of pPABL048 was confirmed through whole genome sequencing and alignment as described above.

Characterization of in1697, pPABL048, and Plasmid Comparative Genomics

The AMR integron in1697 was identified through detection of several AMR genes in multiple NMH ST298 isolates using the ResFinder database [23]. The sequence of this locus was referenced against the PABL048 complete genome to determine its genomic context. In1697 was further characterized through sequence alignment of translated coding sequences to the NCBI non-redundant protein database and through the INTEGRALL integron database [24], through which it was assigned the unique name in1697.

Plasmids similar to pPABL048 were identified using BLASTn, separately screening *P. aeruginosa*, non-*aeruginosa Pseudomonas*, and non-*Pseudomonas* Gammaproteobacteria sequences in the NCBI nucleotide database (nr/nt). This identified 16 plasmids with a minimum of query coverage of 70% (Supplementary Table 6). SPINE (v0.3) was used to determine the plasmid backbone of pPABL048 based on sequences conserved in 16/17 complete plasmids analyzed [25].

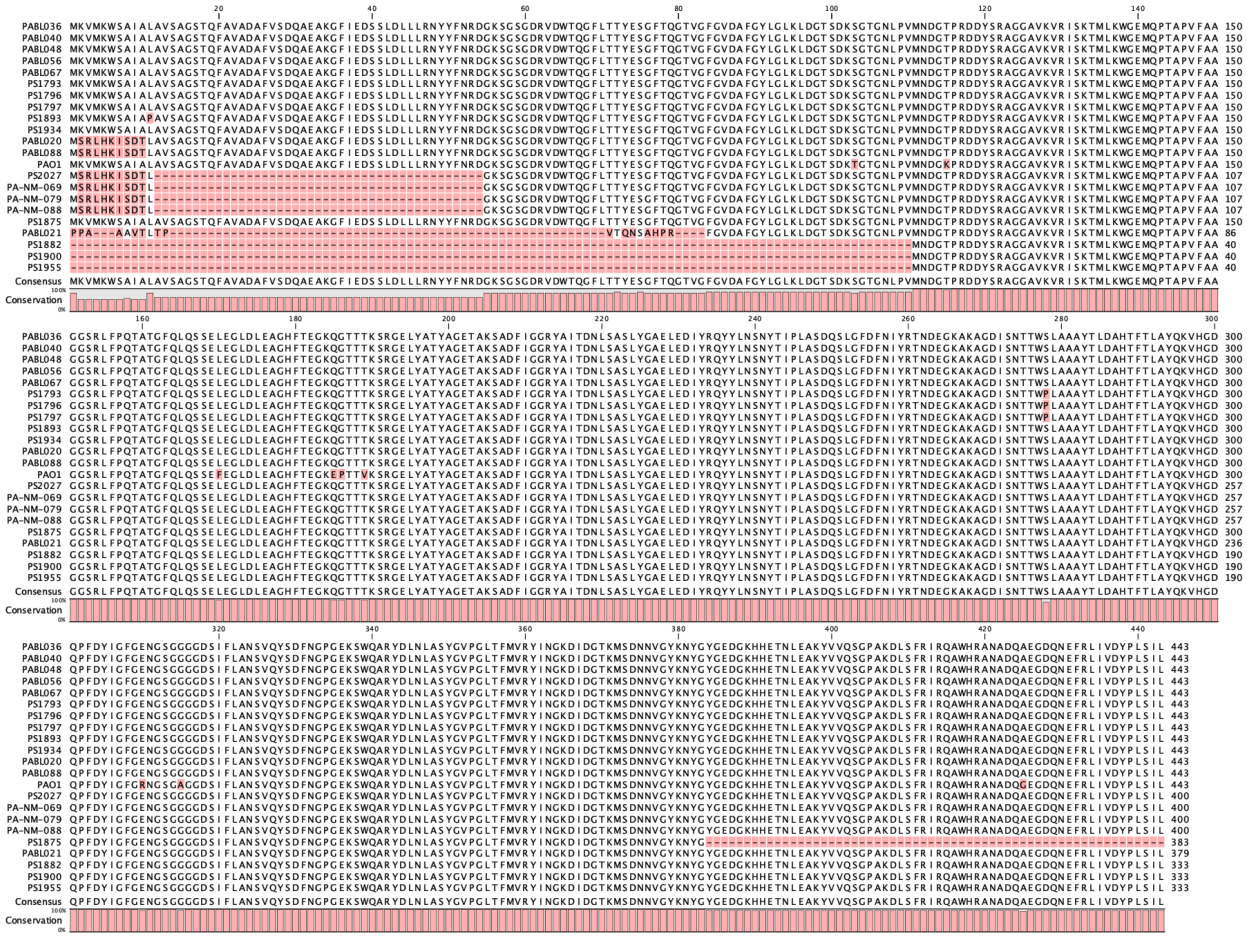
To identify other isolates which harbor plasmids similar to pPABL048, 3133 *Pseudomonas* genus draft genomes cataloged by the *Pseudomonas* Genome Database (accessed January 2019) [26] were aligned to pPABL048 using NUCmer as described above and screened for genome sequences with >70% alignment by length (Supplementary Table 7).

A 98% “core” sequence alignment to pPABL048 (considering all non-missing and non-filtered positions in 62/63 sequences) was determined. A maximum likelihood phylogenetic tree was constructed to show relationships between these plasmids using RAxML (GTRGamma model, 1000 rapid bootstraps) [15].

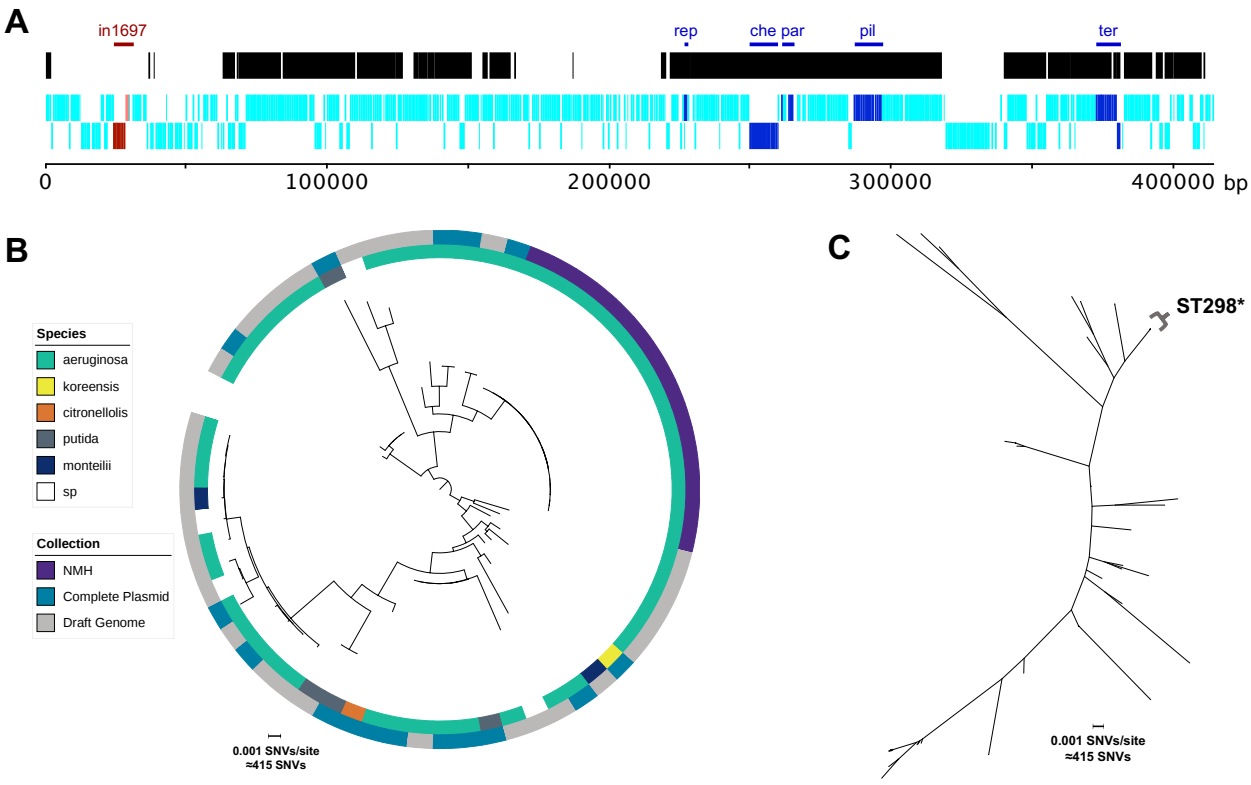
Mutational Resistance Analysis

To examine the role of mutational resistance in the observed AMR phenotype, a panel of PABL048 genes was screened for mutations known to confer resistance in *P. aeruginosa* [27]. In cases where resistance is imparted through specific gain-of-function mutations, translated coding sequences were screened for previously reported alleles known to be involved in resistance. In cases where resistance is conferred from loss-of-function mutations (e.g. gene disruption), translated coding sequences were compared to that of PAO1 as a reference to assess for gross changes in the amino acid sequence. The genomes of ceftazidime-resistant ST298* isolates PS1793, PS1796, and PS1797 were similarly screened to investigate mechanisms of ceftazidime resistance. OprD protein sequences for isolates in the ST298* subclade were extracted, and multiple sequence alignment was performed using CLC Sequence Viewer (v8.0) with default parameters.

SUPPLEMENTARY FIGURES



Supplementary Figure 1. Multiple alignment of OprD protein sequences from ST298* isolates. Deviations from the consensus sequence are highlighted in pink. The sequence for PAO1 OprD is included as a reference.



152

153 **Supplementary Figure 2.** Comparative genomic analysis of pPABL048. (A) Linear diagram of
154 pPABL084 showing coding sequences (light blue) and the plasmid backbone (black) defined as
155 positions present in at least 16 of 17 similar plasmids. In1697 is indicated in red. Plasmid
156 backbone features including putative replication (rep) and partitioning (par) genes, chemotaxis
157 locus (che), putative pili locus (pil), and tellurium resistance locus (ter) are indicated in dark
158 blue. (B) Midpoint-rooted circular and (C) unrooted radial maximum likelihood phylogenetic trees
159 based on alignment of 63 *Pseudomonas* genus sequences to pPABL048. Sequences (ST298*
160 read alignments, complete plasmids, and draft genomes) with >70% alignment to pPABL048 by
161 length were included, and SNVs in positions present in 62/63 alignments (plasmid backbone)
162 were considered. The circular tree is annotated with species (inner ring) and collection (outer
163 ring). On the radial tree, pPABL048 alignments from ST298* isolates are indicated.

REFERENCES

1. Wiegand I, Hilpert K, Hancock REW. Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. Nature Protocols **2008**; 3: 163.
2. Performance standards for antimicrobial susceptibility testing: 28th informational supplement, M100-S28. Wayne, PA: Clinical and Laboratory Standards Institute, **2018**.
3. Magiorakos AP, Srinivasan A, Carey RB, et al. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. Clinical Microbiology and Infection **2012**; 18(3): 268-81.
4. Francisco AP, Bugalho M, Ramirez M, Carriço JA. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. BMC Bioinformatics **2009**; 10(1): 152.
5. Nascimento M, Sousa A, Ramirez M, Francisco AP, Carriço JA, Vaz C. PHYLOViZ 2.0: providing scalable data integration and visualization for multiple phylogenetic inference methods. Bioinformatics **2016**; 33(1): 128-9.
6. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome open research **2018**; 3: 124-.
7. Tatusova T, DiCuccio M, Badretdin A, et al. NCBI prokaryotic genome annotation pipeline. Nucleic acids research **2016**; 44(14): 6614-24.
8. Liu B, Zheng D, Jin Q, Chen L, Yang J. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. Nucleic Acids Research **2018**; 47(D1): D687-D92.
9. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **2014**; 30(15): 2114-20.
10. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **2012**; 19(5): 455-77.
11. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:13033997 **2013**.
12. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics **2009**; 25(16): 2078-9.
13. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics **2011**; 27(21): 2987-93.
14. Kurtz S, Phillippy A, Delcher AL, et al. Versatile and open software for comparing large genomes. Genome Biology **2004**; 5(2): R12.
15. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **2014**; 30(9): 1312-3.
16. Didelot X, Wilson DJ. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. PLOS Computational Biology **2015**; 11(2): e1004041.
17. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. **2019**.
18. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics **2004**; 20(2): 289-90.
19. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, **2016**.
20. Bouckaert R, Heled J, Kühnert D, et al. BEAST 2: a software platform for bayesian evolutionary analysis. PLOS Computational Biology **2014**; 10(4): e1003537.

21. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in bayesian phylogenetics using Tracer 1.7. Systematic Biology **2018**; 67(5): 901-4.
22. Trevors JT. Plasmid curing in bacteria. FEMS Microbiology Letters **1986**; 32(3-4): 149-57.
23. Zankari E, Hasman H, Cosentino S, et al. Identification of acquired antimicrobial resistance genes. The Journal of antimicrobial chemotherapy **2012**; 67(11): 2640-4.
24. Moura A, Correia A, Pereira C, Henriques I, Soares M, Leitão N. INTEGRALL: a database and search engine for integrons, integrases and gene cassettes. Bioinformatics **2009**; 25(8): 1096-8.
25. Ozer EA, Allen JP, Hauser AR. Characterization of the core and accessory genomes of *Pseudomonas aeruginosa* using bioinformatic tools Spine and AGEnt. BMC Genomics **2014**; 15(1): 737.
26. Winsor GL, Griffiths EJ, Lo R, Dhillon BK, Shay JA, Brinkman FSL. Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. Nucleic acids research **2016**; 44(D1): D646-D53.
27. López-Causapé C, Cabot G, Del Barrio-Tofiño E, Oliver A. The versatile mutational resistome of *Pseudomonas aeruginosa*. Frontiers in microbiology **2018**; 9: 685-.