Title: Detecting, mapping, and suppressing the spread of a decade-long *Pseudomonas aeruginosa* nosocomial outbreak with genomics

Authors: William Stribling[1], Lindsey R. Hall[1], Aubrey Powell[1,2], Casey Harless[1], Melissa J. Martin[1], Brendan W. Corey[1], Erik Snesrud[1], Ana Ong[1], Rosslyn Maybank[1], Jason Stam[1], Katie Bartlett[1], Brendan T. Jones[1], Lan N. Preston[1], Katherine F. Lane[1], Bernadette Thompson[3], Lynn M. Young[3], Yoon I. Kwak[1], Alice E. Barsoumian[4], Ana-Elizabeth Markelz[4], John L. Kiley[4], Robert J. Cybulski[1,5], Jason W. Bennett[1], Patrick T. Mc Gann[1,*] and Francois Lebreton[1,*]

Affiliations:

[1] Multidrug-Resistant Organism Repository and Surveillance Network (MRSN), Walter Reed Army Institute of Research, Silver Spring, Maryland, USA.

[2] Department of Pathology, Brooke Army Medical Center, Joint Base San Antonio-Fort Sam Houston, Texas, USA.

[3] Infection Prevention & Control, Brooke Army Medical Center, Joint Base San Antonio-Fort Sam Houston, Texas, USA.

[4] Infectious Disease Service, Department of Medicine, Brooke Army Medical Center, Joint Base San Antonio Fort Sam Houston, Texas, USA.

[5] Bacterial Disease Branch, Walter Reed Army Institute of Research, Silver Spring, Maryland, USA.

*Address correspondence to: francois.lebreton.ctr@health.mil and patrick.t.mcgann4.civ@health.mil

Keywords: nosocomial transmission, MDR, ESKAPE, antibiotic resistance, infection control, microbial genomics, evolution.

ABSTRACT:

Whole-genome sequencing is revolutionizing bacterial outbreak investigation but its application to the clinic remains limited. In 2020, prospective and retrospective surveillance detected a *Pseudomonas aeruginosa* outbreak with 253 isolates collected from 82 patients in 26 wards of a hospital. Its origin was dated to the late 90s, just after the facility opened, and patient-to-patient and environment-to-patient cases of transmission were inferred. Over time, two epidemic subclones evolved in separate hosts and hospital areas, including newly opened wards, and hospital-wide sampling confirmed reservoirs persisted in the plumbing. Pathoadaptive mutations in genes associated with virulence, cell wall biogenesis, and antibiotic resistance were identified. While the latter correlated with the acquisition of phenotypic resistances to 1st (cephalosporin), 2nd (carbapenems) and 3rd (colistin) lines of treatment, maximum parsimony suggested that a truncation in a lipopolysaccharide component coincided with the emergence of a subclone prevalent in long-term infections. Since initial identification, extensive infection control efforts guided by routine, near real-time surveillance have proved successful at slowing transmission.

MAIN TEXT:

INTRODUCTION

Pseudomonas aeruginosa is a versatile, opportunistic human pathogen causing a wide range of infections of both community and hospital origin[1–3]. In the clinic, patients with cystic fibrosis (CF), severe burns or neutropenia are of highest risk and cases are associated with high morbidity and mortality[3]. Furthermore, the management of these infections is increasingly difficult due to the global emergence of multidrug resistant (MDR) strains, especially carbapenem-resistant *P. aeruginosa* (CRPA), which now account for 16-30% of cases in the US[4]. In response, global and national agencies have categorized CRPA as a "Critical" pathogen posing a serious threat on public health[4].

While high-risk MDR lineages, exemplified by the most globally prevalent sequence type (ST) 235, are enriched in isolates carrying acquired carbapenemases and/or extended-spectrum beta lactamases (ESBLs)[5,6], the hallmark of *P. aeruginosa* lies in its remarkable propensity for developing resistances through chromosomal mutations[7]. Within the hundreds of intrinsic genes comprising this mutational resistome, several harbor well-characterized modifications that are the most common cause of resistance for nearly all classes of drugs[7]. These include gain-of-function mutations in gyrase subunits GyrAB (fluoroquinolones), penicillin-binding protein PBP3 (cephalosporins), various Mex efflux systems (multiple antimicrobials) as well as inactivation of the porin OprD (carbapenems) or the sensor component of the PhoPQ system, the latter associated with resistance to last-line polymyxin antibiotics[8–10].

Being capable of both acute and chronic infection, *P. aeruginosa* possesses a large arsenal of virulence factors such as a type 4 pili (T4P), O-antigens and lipopolysaccharides (LPS), extracellular enzymes, exotoxins (*e.g.* ExoS/U), a flagellum and five types of secretion systems (TSSs)[11]. These are generally part of complex regulatory pathways in

which two-component systems (TCSs) often play a major role[11,12]. The ability of *P. aeruginosa* to form biofilm is a key aspect of its persistence in patients with chronic infections as well as in environmental niches[13]. In the hospital, biofilms are often found in water-related environments and eradication is particularly difficult due to increased disinfectant resistance[1,14]. This in turn poses a challenge for infection prevention and control (IPAC) as illustrated by the increasing reports of *P. aeruginosa* outbreaks and hospital-acquired infections (HAI) linked to contaminated sinks, drains or taps[15–18].

Here, routine genome-based surveillance of MDR clinical isolates across the US Military Health System (MHS) identified an epidemic cluster of *P. aeruginosa* in a single hospital in 2020. Retrospective analysis of isolates from the preceding ten years, together with Bayesian phylogenetics, revealed a protracted outbreak clone with cases from all hospital floors and an origin dated to the late 1990s. Dissection of this unique dataset sheds new light on the evolution, persistence, and routes of transmission of *P. aeruginosa* in the clinic.

RESULTS

P. aeruginosa ST-621 is endemic in a large US tertiary hospital.

The Multidrug-Resistant Organism Repository and Surveillance Network (MRSN) performed whole genome sequencing (WGS) of 5,129 *P. aeruginosa* isolates received between 2011 and 2020 from 71 facilities around the globe[19]. While a total of 547 STs were represented, high-risk epidemic lineages ST-235/*exoU+* (10%) and ST-244/*exoS+* (5%) were the most prevalent and widespread with isolates recovered from 37 and 28 facilities, respectively (**Fig. S1A**). By contrast, isolates from lineage ST-621/*exoS+* also represented 5% of the total STs, but originated from just 6 facilities, with 89% from a single US tertiary hospital (Facility A) and sharing high-level genetic relatedness (0-23 allelic differences) by core genome MLST (cgMLST) (**Fig. S1B**). Furthermore, of the 5 additional facilities from which ST-621 was detected, patient transfer from facility A was observed in three instances and isolates from the remaining two were genetically distinct strains (88-89 allelic differences) (**Fig. S1B**).

Retrospective analysis revealed that a total of 253 ST-621 isolates were collected from facility A (a 425-bed medical center and the only level 1 trauma center in the Military Health System) between May 2011 and December 2020. The majority were isolated from respiratory (42%) and urine (39%) samples, followed by wound (7%), surveillance (6%), tissue (3%), and blood cultures (2%) (**Table S1**).

A protracted outbreak with cases dispersed throughout floors and wards.

Phylogenetic analysis confirmed that all ST-621 isolates from facility A belonged to the same outbreak clone with an average distance (all vs all pairwise comparison) of just 38 single nucleotide polymorphisms (SNPs) and an interquartile range of 19 (**Fig. 1A, Fig. S2 and Table S2**). A total of 82 patients (average age of. 57.2 +/- 21.0 and male/female ratio of

1.7) were represented, with the majority (56%) contributing serial isolates. In particular, six patients (ID 10, 27, 34, 35, 49 and 63) showed prolonged carriage, best exemplified by patient 35 from whom 21 urine isolates were collected over 4.5 years (**Fig. 1A and B**).

Two outbreak subclones, designated SC1 and SC2, were identified from the tree topology and showed increased genetic relatedness within each (average distance of 24 and 31 SNPs, respectively) (**Fig. 1A**). Overall ST-621 isolates were observed in 26 wards/units from all 7 floors of two buildings (Main and Tower Buildings) in facility A (**Fig. 1C**), yet subclones SC1 and SC2 were spatially and temporally distinct. 78% of SC1 isolates were collected from patients in the Main Building and 60% were detected before 2014. In contrast, 52% of SC2 isolates originated from patients in the Tower Building and 77% were collected after 2014 (**Fig. 1A**). Specifically, the early spread of SC1 was largely connected to patients in adjacent ICUs and step-down units on the 2nd and 3rd floors (wards 1-10) of the Main Building. The subsequent emergence of SC2 largely centered in patients receiving care in adjacent ICUs (wards 20-21) on the 3rd floor within the Tower Building (**Fig. 1A and C**). Throughout the outbreak, new cases due to either SC1 or SC2 were also detected from patients in the emergency room (ER). Unlike the isolates from other locations, these ER isolates were nearly all (32 vs 92%, $p < .0001$) from urine cultures and, as expected, the majority (72%) were serial isolates from returning patients at Facility A (**Fig. 1C, Table S1**). The remaining were collected from 9 patients for whom a previous visit at Facility A is suspected but could not be established as no previous isolates were collected from them and no historical patient data was available.

The epidemic curve revealed that between 2011 and 2016 an average of 11.5 new cases were detected every year. From 2017 to 2021, only 12 new patients were identified (3 cases a year) (**Fig. 1D)**. Although the rate of outbreak cases appeared to decrease, the overall number of *P. aeruginosa* cases in facility A (irrespective of lineages) remained stable

throughout the studied period. However, due to staffing issues, the number of isolates received/sequenced by the MRSN every year was not consistent (**Fig. 1D** and **Fig. S3**).

An origin dated soon after the opening of facility A.

Though the collection of clinical isolates at facility A only started in 2011, genetic diversity was already evident (16.3 SNPs on average), suggesting an earlier origin (**Table S2**). Molecular clock phylogenetic analyses of all studied isolates inferred the time of the most recent common ancestor (MRCA) to be in 1999 (± 3 years), soon after the opening of the Main Building of facility A in 1996 (**Fig. 2**). Furthermore, this analysis predicted that, despite the observed temporal succession in sequenced isolates, SC2 was not a descendant of SC1. Instead, the two subclones emerged concurrently with respective MRCAs dated in 2004 (± 2 years) (**Fig. 2**).

Despite this predicted early origin, it is only after facility expansion and the opening of the Tower Building in 2012 that the first SC2 isolates were ever sampled (**Table S1**). Specifically, the earliest SC2 isolate collected was from a wound infection in patient 27. Remarkably, the 15 serial isolates collected over 3 years (2012 to 2015) from this patient were not monophyletic and their MRCA coincided with the emergence of this subclone (**Fig. 2**).

Environmental reservoirs sustain nosocomial transmission

The presumed origin of an infection, either directly from patient-to-patient or from environment-to-patient, was designated based on SNP thresholds (starting from pairwise comparisons for all inter-patient isolates, **Table S2**) and known epidemiological links, or lack thereof. Indeed, while a unimodal distribution of SNPs (peak at 40.5) was observed for the whole cohort, a bimodal profile appeared when inter-patient isolates from the same month, the same ward, or both were selected (local maxima of the lower peak at 12.4, 15.3, and 11.6

SNPs respectively) (**Fig. 1E**). This pool of virtually identical isolates from patients with spatial and temporal overlap represented the most likely cases of direct, patient-to-patient transmission. Pairs of isolates from only 10 patients satisfied these conditions (**Table S1**): All carried subclone SC1, occurred in 2011 and 2012 (with two exceptions), and six were hospitalized in ICU #1 in the Main Building of facility A (**Fig. 1C**).

In contrast, a presumed environmental origin was designated when a new patient was infected with a virtually identical strain from a preceding patient in the same ward, but no known temporal overlap was established for at least 1 year (based on isolates collection dates). Twelve cases met these criteria (**Fig. 1C and Table S2**). Among them, 3 patients (ID 57, 61, and 74) in ICU #20 on the Tower Building had primary isolates (first isolates from a patient) of subclone SC2 that were distinct by only 0-4 SNPs despite being collected years apart (in 2015, 2016, and 2018, respectively) (**Fig. 1C**). Furthermore, an origin from contaminated surfaces was also predicted for 7 of the 9 most recent patients (2018-2020). This included 5 cases caused by the apparent resurgence of SC1 in ICUs #1-2 of the Main Building (**Fig. 1C)**.

Genomic inferences were supported by contemporary environmental sampling, conducted in 2021 and 2022, during which a total of 159 swabs were taken from 55 distinct locations in 8 wards/units including the ER in Facility A. This confirmed the existence of reservoirs (*i.e*. sink drains located in patient rooms) of both SC1 and SC2 isolates (n = 17) in 7 distinct patient rooms in wards #1, 2, 6 and 20 throughout the Main and Tower Buildings of facility A (**Fig. 3, Table S3**). In that timeframe, an additional 56 clinical isolates (25 SC1 and 31 SC2) from 13 patients (including 11 new cases) were collected, indicating transmission events were still occurring (**Fig. 3**).

Selection on resistance genes and rising prevalence of CRPA due to porin defects.

Phenotypically, a prototypical outbreak isolate showed non-susceptibility to aminoglycosides (except for amikacin), 3rd and 4th generation cephalosporins, carbapenems, and fluoroquinolones (**Table S4**). However, resistance profiles were not homogenous and temporal patterns were observed (**Fig. 4A**). In particular, while 63% of the isolates collected in 2011 were non-susceptible to imipenem, this fraction gradually increased to >95% by 2014 and beyond. In the same timeframe, the prevalence of isolates with non-susceptibility to cefepime decreased from 81% to 37% (**Fig. 4A**). Both phenomena were linked to the decline of SC1 and rise of SC2, with the latter displaying significantly higher (98% to 76% in SC1, chi-square = 53.5 and $p<0.001$) and lower (38% to 66%, chi-square = 25.3 and $p<0.001$) rates of resistance to imipenem and cefepime, respectively (**Fig. 2**). In contrast, resistance to tobramycin and ciprofloxacin was highly prevalent and remained relatively stable over the sampling period (**Fig. 4A**).

Genotypically, outbreak isolates contained no plasmids (based on the finished genomes of a diverse subset of isolates, **Table S1**) and, besides an intrinsic AmpC-type β-lactamase, only carried two acquired resistance genes, known to confer resistance to ciprofloxacin (*crpP*) and tobramycin (*aac(6')-Ib4*) (**Table S1**). To explain other phenotypic resistances, and in the absence of an acquired ESBL or carbapenemase, a genome-wide analysis of non-synonymous (NSY) and loss of function (LOF) mutations was performed for all outbreak isolates (**Table S5**). From a list of 164 genes (**Table S6**) for which known chromosomal mutations have been linked to antibiotic resistance[11-13], 68 distinct mutations (including 20 predicted LOF) in 37/164 genes were identified in at least one outbreak isolate (**Table S7**). Of particular interest, 8 genes (or functional clusters) were each mutated independently three or more times (**Fig. 4B**).

The most frequently mutated gene was *oprD*, encoding a porin that facilitates the diffusion of amino acids, small peptides, and carbapenem antibiotics into the cell[8]. Five of the

six independent mutations identified in OprD were predicted LOF, including an R6* stop codon and a R318fs frameshift observed in 115 SC2 and 92 SC1 isolates, respectively (**Fig. 4B, Table S7**). Phylogenetic Analysis Using Parsimony (PAUP) predicted that both the R318fs and R6* mutations occurred early (between 2006-2008) within the evolution of their respective subclones (**Fig. 2**). Despite this early origin, highly carbapenem resistant SC2-OprD-R6* isolates were first sampled from patients in mid-2012 and only became predominant from 2014 onward (**Fig. 2**).

Within all other mutations in antibiotic resistance associated genes (**Table S7**), only two more variants were found in > 50 outbreak isolates: a W97* stop codon in β-lactamase expression regulator AmpD and a R504C substitution in penicillin-binding protein PBP3 (**Fig. 4B**). The LOF in AmpD, predicted to have emerged in 2012, was observed in a subset of 51 monophyletic SC2-OprD-R6* background isolates that were not consistently resistant to cefepime (**Fig. 2**). In contrast, the R504C substitution in PBP3 was acquired independently twice (within SC1-OprD-R318fs in 2007 and within SC2-OprD-R6* in 2014) and was strongly associated with a resistance to 4th-generation cephalosporins (**Fig. 2**). Overall, *ampD* and *pbp3* were among the genes with the most independently acquired NSY or LOF mutations, with six and four distinct sites, respectively (**Fig. 4B**). Accumulation of mutations, albeit in small numbers of isolates, was also observed in other genes involved in β-lactam resistance (*e.g.* catabolite repression control *crc*, and *mpl* UDP-muramic acid/peptide ligase) as well as antibiotic efflux systems (*mexXY* and *mexCD-oprJ*) (**Fig. 4B**).

Sporadic emergence of colistin resistance.

Though rare, the presence and impact of mutations in genes associated with resistance to colistin, a last-line antibiotic, was examined. NSY mutations were detected in the TCSs ColRS (three distinct sites in four isolates) (**Fig. 4B**) and PmrAB (two distinct sites in three

isolates) (**Table S7**) but colistin MIC's were the same as wild-type isolates (4 mg/L). By contrast, patient 2 (who had a history of colistin therapy) acquired a predicted LOF mutation (E77fs) in sensor protein PhoQ in two serial isolates resulting in an 8-fold increase in MIC (32 mg/L) (**Fig. 4C**). Notably, the emergence of this colistin resistant strain, which occurred in early 2011 (**Fig. 2**), remained constrained to a single patient and no further spread was detected.

Mutational convergence in virulence, cell wall biogenesis and signaling pathways.

Genome-wide, and excluding synonymous mutations, 64 genes were each mutated independently two or more times over the course of this outbreak (**Table S5**). Compared to the distribution observed for the whole genome, these 64 genes were functionally enriched for signal transduction (28% vs 6% in the whole genome, chi-square $p<0.001$) and cell wall/membrane biogenesis (17% vs 7%, $p= 0.0017$) (**Fig. 5A**). For the latter category, besides genes already discussed for their direct role in antibiotic resistance (i.e *pbp3, oprD, mpl, and mex* loci*)*, six other genes were recurrently mutated in outbreak isolates including LPS biosynthesis *rmlA* and choline transporter *betT2* (**Fig. 5A**).

For genes associated with signaling functions, various TCS involved in pathogenesis were represented, including *narXL*, *fleRS*, and the catabolic regulation *cbrAB* (**Fig. 5A**). Notably, *cbrAB* was the most independently mutated loci in this dataset with 16 distinct NSY mutations in 80 isolates from both subclones. Beyond TCS, sigma factor *pvdS*, required for toxin expression (6 NSY and 1 LOF), motility regulator *morA* (11 NSY and 2 LOF) and mucoidy regulatory system *mucAB* (4 sites including three predicted LOF) were also recurrently mutated (though rarely fixed) in outbreak isolates (**Fig. 5A).**

Finally, at the level of functional pathways, and compared to the theoretical frequency for an even distribution of mutated sites across the whole genome (0.11 sites per kb of coding

sequence), NSY and LOF mutations were significantly enriched in genes involved in the biosynthesis of the Type IV secretion system (0.59 sites/kb, chi-square $p<0.001$), flagellar biosynthesis and chemotaxis (0.21 sites/kb, $p= 0.0264$), alginate production (0.31 sites/kb, $p= 0.0088$) and O-antigen production (0.36 sites/kb, $p=0.0126$) (**Fig. 5B, Table S8**). Importantly, the WbpX D-rhamnosyltransferase, part of the O-antigen biosynthesis pathway, was disrupted (S19fs) in 127 SC2 isolates (**Table S5**). PAUP predicted this frameshift occurred in 2005 (± 24 months), concomitant with the predicted emergence of SC2 (**Fig. 2**).

DISCUSSION

WGS technology is revolutionizing infection prevention and control. Here, routine surveillance of MDR pathogens across the US Military Health System was critical in uncovering a decades-long *P. aeruginosa* epidemic cluster. With traditional approaches alone[20], comparable outbreaks may avoid detection due to the sporadic nature of patient infections, scattering of cases throughout the hospital, and changing antibiotic susceptibility profiles. Furthermore, because of budget constraints, surveillance programs focusing on "high-risk" isolates carrying select resistance genes (*e.g.* ESBLs, carbapenemases)[4–6], would also fail to detect this ST-621 outbreak clone. Indeed, while the ST-621 lineage was the focus of several studies in the late 2000's, these examined a $bla_{\text{IMP-13}}$ carrying epidemic clone which spread throughout Europe and has since been sporadically detected in South America and Southeast Asia[21–24]. To our knowledge this IMP-carrying strain has not been reported in the USA to date, and the outbreak in Facility A is the result of a distinct MDR clone lacking the carbapenemase.

Similar to the US Military Health System, healthcare networks worldwide that already benefit from prospective WGS surveillance programs are reporting large numbers of unrecognized outbreaks[18]. However, two resources rarely available in other settings were key to this investigation: a genome repository of isolates from the preceding decade and the ability to conduct prospective environmental sampling. The former was integral to date the presumed origin of the outbreak to the late 90s, soon after the hospital opened, and to trace the evolution and separate spread of two subclones throughout floors and wards. The latter revealed that, more than 20 years later, reservoirs of both subclones persist in sink drains from patient rooms throughout the facility. This supports previous evidence implicating these sites as major reservoirs for *P. aeruginosa* in the clinic[1,25]. Decontamination of drains can be extremely challenging due to the limited penetration of disinfectants, the lack of access (*e.g.*

to perform scrubbing), recolonization (*e.g.* disposal of contaminated patient specimens) or retrograde growth from p-traps[1,14,25]. Furthermore, sinks with shallow basins and gooseneck faucets directing water straight into the drain, similar to those in Facility A (**Fig. 3A**), have been linked to increased back splash onto nearby surfaces and medical equipment[26].

Besides environmental contamination, reservoirs within patients likely contributed to the spread and longevity of the outbreak. In particular, long-term infections, documented for a significant fraction of patients, provided a recurring source of the epidemic clone. It is noteworthy that 5 of the 6 patients with long-term (> 1 year) infections (and 9 out of 11 if reduced to 6 months) were carriers of subclone SC2. Compared to SC1, all but one SC2 isolates carried a truncated WbpX glycosyltransferase, an essential component of the common polysaccharide antigen (CPA) biosynthesis[27]. While this mutation could result in increased immune evasion, the rough LPS phenotype recurrently observed *in vivo* (in particular in CF patients) usually results from a functional CPA but the lack of O-specific antigen[11,28]. In further evidence for a progression toward a host-adapted lifestyle, a convergence of mutations in genes involved in alginate production, quorum-sensing deficiency, loss of motility, and decreased protease secretion, all phenotypes associated with chronic infections[29], was observed throughout the 10 years of sampling.

Considering its role in long-term infections and the distinctive association with patients and sinks in the Tower Building of Facility A, the origin and emergence of subclone SC2 is of particular importance. Although obfuscated by the limited sampling, a plausible scenario (supported by BEAST2 inferences) would be that a patient with an initial infection acquired during a stay in the Main Building before the SC1-2 split (ca. 2004) proceeded to shed the epidemic strain in the newly opened (2012) Tower Building during successive visits. Importantly, the Main and Tower Buildings of facility A have distinct plumbing systems, providing the necessary ecological separation (minus inter-hospital patient transfers) for the

further spread and divergence of the two sub-clones. Patient 27 is just such an example, but the lack of early sampling (2000-2010) and incomplete collection thereafter precludes a definitive answer. Indeed, other scenarios cannot be discounted, with a recent study showing sinks in a newly built ICU were already contaminated with an outbreak clone of *P. aeruginosa* before the arrival of the first patients[17].

One of the most notable features of this outbreak was the evolution of antibiotic resistance over time, with the emergence of resistance to 1st (cephalosporin), 2nd (carbapenems) and 3rd (colistin) line treatments. The potent R504C substitution in PBP3 was selected in each subclone, and in addition to facilitating cephalosporin resistance, has also been linked to ceftazidime/avibactam resistance[30]. While no patient prescription data is available from Facility A, it can be speculated that the high prevalence of cephalosporin resistance in early outbreak isolates led to increased carbapenem use, which in turn selected for the many independently evolved OprD mutants, one of the most common mechanisms for carbapenem resistance in *P. aeruginosa*[11]. Finally, albeit limited to a single patient, the emergence of colistin resistance, through a well characterized mechanism, is a reminder that the threat of extensively drug resistant *P. aeruginosa* is only a prescription and a few mutations away[31].

Some limitations were noted for this study and complicated the genomic inferences. First, although criteria remained the same, the partial and inconsistent sampling of clinical *P. aeruginosa* isolates prevented from capturing the exact magnitude of the outbreak through the years and sampling biased cannot be ruled out. Second, detailed epidemiological data for the patients (date of admission(s), bed location, reason for sampling, medications prescribed, etc.) were unavailable. As a result, possible patient overlaps were only inferred from the location and culture date of the isolates, and many were likely not detected. Third, fixed thresholds (*i.e.* genetic relatedness, time between isolates collection, and shared hospital

location) were applied to predict a possible origin of infection. Acknowledging the other limitations, these were designed to predict the most likely cases of direct patient-to-patient transmission, when an overlap was identified, or environment-to-patient transmission, when a patient overlap was ruled out. Because of conservative criteria, the origins of most cases remained undetermined and an alternative to fixed threshold, using TransPhylo[32], is being explored in a follow-up study. This approach uses a probabilistic model to predict whom infected whom in outbreak scenarios and reconstruct transmission trees[32]. Fitting our dataset, TransPhylo was recently extended to allow the use of multiple genomes per host[33] and to incorporate epidemiological data into the analysis[34].

The data generated during this study has resulted in various ongoing interventions (*e.g.* closing sinks, replacing tubing, using foaming detergents[35]) at Facility A. Sampling and sequencing, in near real-time, of all *P. aeruginosa* clinical isolates (and not just MDR) also commenced in 2021. Notably, as of May 2025, no new ST-621 patient cases have been reported in over 15 months, and the unbiased contemporary data using all *P. aeruginosa* from patients at this facility confirms the spread has slowed, with just 4 cases identified in 2022-2024. Similar to Facility A, the roll-out of routine WGS surveillance in the clinic has the potential to improve patient care and prevent some of the estimated 136 million hospital-associated drug-resistant infections per year globally[36].

MATERIALS AND METHODS

Data Collection & Antibiotic Susceptibility Testing

As mandated by the US National Action Plan for Combating Antibiotic-Resistant Bacteria (CARB), the Multidrug Resistant Organism Repository and Surveillance Network (MRSN) collects and analyses clinically relevant MDR organisms across the U.S. Military Health System (MHS) and around the world in collaboration with the Global Emerging Infections Surveillance (GEIS) Branch. All isolates received by the MRSN underwent whole genome sequencing (single colony from a pure culture) and criteria for isolate collection were as follow. Between 2011-2020, all MDR isolates (resistance to ≥ 3 antibiotic families) were requested from all US facilities, but compliance varied between years and facilities. For Facility A, starting in 2021, all clinical *P. aeruginosa* isolates (MDR or not) were requested.

As a result of this sampling, from 2011 to 2020, 5,129 MDR *P. aeruginosa* isolates were collected from 71 facilities and sent to the MRSN. Of these, 1,511 were collected from Facility A, a large tertiary U.S. military hospital located in the continental United States. In total, 253 isolates belonged to the ST-621 outbreak clone.

Antibiotic susceptibility testing (AST) was performed in a College of American Pathologists (CAP)-accredited lab at Facility A using a Vitek 2 (card GN AST 71 and GN ID; bioMérieux, NC, USA). For isolates without MICs, AST was performed in the MRSN CAP-accredited clinical lab as previously described[37].

Whole Genome Sequencing & Bioinformatic Analysis

Genomic DNA was extracted and sequenced via Illumina MiSeq or NextSeq benchtop sequencer (Illumina, Inc., San Diego, CA) as previously described[37]. Long read sequencing was performed using PacBio RS II (Pacific Biosciences of California, Inc., Menlo Park, CA). Kraken 2 was used to identify isolate species and check for contamination[38]. Short-read

sequencing data were trimmed for adapter sequence content and quality using bbduk[39]. *De novo* assembly was performed using Newbler v2.9[40]. Minimum thresholds for contig size and coverage were set at 200 bp and 49.5×, respectively. Long-read sequencing data were assembled using HGAP 3.0 in the SMRT Analysis portal. *In silico* multilocus sequence typing (MLST) was performed using the scheme developed by Curran and Dawson[41]. Antimicrobial resistance genes were annotated using a combination of AMRFinderPlus and ARIBA[42,43]. The genomes of all 253 ST-621 isolates have been deposited in the National Center for Biotechnology Information under BioProject PRJNA852179.

Core Genome MLST, Gene Annotation, SNP Calling, and Phylogenetic Analysis

Core genome MLST were determined in SeqSphere + (Ridom, Germany) using the cgMLST scheme developed by Curran *et al*[41] and a 90% cut-off (3,960 of 4,400 genes). SNP calling was performed with Snippy v.4.4.5 (https://github.com/tseemann/snippy) using error corrected [Pilon v1.23][44] and annotated [Prokka v1.14.6][45] draft assembly, with isolate MRSN 4605 as the reference (chosen because it was both collected early in the outbreak and had low root to tip divergence). Type strain PAO1 was used as the basis for annotation[46]. The core SNP alignment was filtered for recombination using Gubbins v2.3.1[47]. A SNP-based phylogeny was created by inferring a maximum-likelihood tree with RaxML-NG v0.9.0[48] using GTR+G (50 parsimony, 50 random). The phylogeny was visualized in iTOL[49].

Gene annotation and clustering showed 8,754 total genes in the pangenome of the outbreak. The core genome (present in ≥ 99% of isolates) was comprised of 5,674 genes, with a shell genome (defined as genes found in ≥ 95% of isolates) of 297 genes, and a cloud genome (≥ 15% of isolates) of 530 genes.

Curation and analysis of recurring mutations in outbreak isolates

Genome-wide analysis of non-synonymous and loss-of-function mutations in all outbreak isolates showed a total of 629 mutated sites (257 singletons and 372 sites shared by ≥ 2 isolates) in 488 distinct genes (**Table S5**). For variant sites identified in >99% of outbreak isolates, individual nucleotide-based sequence alignments were performed using a wild-type allele from the PAO1 genome (https://www.pseudomonas.com/). This allowed the identification of sites for which the MRSN 4605 reference genome was carrying the mutation and not the remaining outbreak isolates. These were curated (e.g. entry was modified to reflect the true mutation and its presence/absence in all outbreak isolates) manually and indicated as such in **Table S5**.

SNP distances between isolates from different patients

SNP distances were obtained from the alignment by using snp-dist (https://github.com/treemann/snp-dists). To create violin plots, the complete all-vs-all dataset was limited, based on 5 sets of criteria, to pairs of isolates that (1) originated from different patients; (2) originated from different patients, and were collected in the same month; (3) originated from different patients and were collected in the same ward; (4) originated from different patients, were collected in the same month, and the same ward; or (5) originated from within the same patient (Fig. 1E).

Primary isolates were designated as originating from a Patient-to-Patient transmission if they were part of a pair from different patients, were collected in the same ward, were separated by ≤ 10 SNPs (to account for the observed within-host variability, Fig. 1E), and were collected ≤ 31 days apart (as a proxy for a possible patient overlap). Primary isolates were conservatively designated as environmental if they were part of a pair from different patients, were collected in the same ward, were collected > 365 days apart (as a proxy for unlikely patient overlap), and were separated by a number of SNPs ≤ 3x the number of years they

were separated to account for the 2.987E-7 substitutions/site/year from the BEAST analysis. This conservative threshold was chosen to favor high specificity as this inference relied solely on clinical isolates (i.e. the identical environmental strain in the patient-environment-patient chain was not sampled). For these clinical isolates to have acquired no/very little mutation in that much time, no/low replication is expected and, although unsampled, we propose this most likely happened on hospital surfaces.

In each of these cases, direct transmission or transmission mediated by a healthcare worker were not distinguished as no sampling was available from the later. All other primary isolates were designated as Undetermined.

Bayesian evolutionary phylogenetic analysis

To evaluate the strength of the temporal signal, TempEst v1.5.3 was utilized to visualize the relationship between root-to-tip genetic distances for samples with known collection dates[50]. To date internal nodes of interest on the phylogeny, bayesian phylogenetic inference was performed using BEAST2 v2.6.5 on a recombination free alignment, removing samples with uncertain collection dates, and accounting for constant sites with beast2_constsites (https://github.com/andersgs/beast2_constsites)[51]. The HKY substitution model was selected based on evaluation of all possible substitution models in bModelTest v1.2.1[52]. The random clock model was selected based on support by the marginal likelihood value using the Nested Sampling package v1.1.0[53]. BEAST2 was run under a coalescent constant population model with a Markov chain Monte Carlo length of 1×10^8 sampling every 5×10^3 steps. Analyses were repeated five times to confirm consistency between the obtained posterior distributions. Parameter estimates were computed using Tracer v1.7.1. Posterior trees were combined with LogCombiner and summarized in TreeAnnotator after a 50% burn-in. The final MCC target tree was visualized and annotated using iTOL (https://itol.embl.de/)[49].

List of Supplementary Materials

Table S1 - Isolate Metadata

Table S2 - All vs All SNP Comparison

Table S3 - Environmental Isolates

Table S4 - Antibiotic Susceptibility Testing

Table S5 - Prediction of variants in outbreak isolates

Table S6 - Resistome reference intrinsic alleles

Table S7 - Mutations in intrinsic genes associated with antibiotic resistance

Table S8 - Mutations in functional groups or pathways

References and Notes

1. Kerr, K. G. & Snelling, A. M. Pseudomonas aeruginosa: a formidable and ever-present adversary. *Journal of Hospital Infection* **73**, 338–344 (2009).
2. September, S. M., Els, F. A., Venter, S. N. & Brö, V. S. Prevalence of bacterial pathogens in biofilms of drinking water distribution systems. doi:10.2166/wh.2007.004
3. Rodríguez-Rojas, A., Oliver, A. & Blázquez, J. Intrinsic and Environmental Mutagenesis Drive Diversification and Persistence of Pseudomonas aeruginosa in Chronic Lung Infections. *J. Infect. Dis*. **205**, 121–127 (2012).
4. World Health Organization. Prioritization of pathogens to guide discovery, research and development of new antibiotics for drug resistant bacterial infections, including tuberculosis. *Essential medicines and health products* 1–88 (2017). doi:WHO reference number: WHO/EMP/IAU/2017.12
5. Hong, D. J. *et al.* Epidemiology and characteristics of metallo-ß-lactamase-producing Pseudomonas aeruginosa. *Infect. Chemother*. **47**, 81–97 (2015).
6. del Barrio-Tofiño, E., López-Causapé, C. & Oliver, A. Pseudomonas aeruginosa epidemic high-risk clones and their association with horizontally-acquired β-lactamases: 2020 update. *Int. J. Antimicrob. Agents* **56**, 106196 (2020).
7. Blair, J. M. A., Webber, M. A., Baylay, A. J., Ogbolu, D. O. & Piddock, L. J. V. Molecular mechanisms of antibiotic resistance. *Nat. Rev. Microbiol. 2014 131* **13**, 42–51 (2014).
8. Li, H., Luo, Y. F., Williams, B. J., Blackwell, T. S. & Xie, C. M. Structure and function of OprD protein in Pseudomonas aeruginosa: From antibiotic resistance to novel therapies. *International Journal of Medical Microbiology* **302**, 63–68 (2012).
9. Bruchmann, S., Dötsch, A., Nouri, B., Chaberny, I. F. & Häussler, S. Quantitative

contributions of target alteration and decreased drug accumulation to pseudomonas aeruginosa fluoroquinolone resistance. *Antimicrob. Agents Chemother*. **57**, 1361–1368 (2013).

10. Clark, S. T. *et al.* Penicillin-binding protein 3 is a common adaptive target among Pseudomonas aeruginosa isolates from adult cystic fibrosis patients treated with β-lactams. *Int. J. Antimicrob. Agents* **53**, 620–628 (2019).

11. Jurado-Martín, I., Sainz-Mejías, M. & McClean, S. Pseudomonas aeruginosa: An Audacious Pathogen with an Adaptable Arsenal of Virulence Factors. *Int. J. Mol. Sci.* **22**, 1–37 (2021).

12. Juan, C., Peña, C. & Oliver, A. Host and pathogen biomarkers for severe Pseudomonas aeruginosa infections. *J. Infect. Dis.* **215**, S44–S51 (2017).

13. Thi, M. T. T., Wibowo, D. & Rehm, B. H. A. Pseudomonas aeruginosa Biofilms. *Int. J. Mol. Sci.* **21**, 1–25 (2020).

14. Smith, K. & Hunter, I. S. Efficacy of common hospital biocides with biofilms of multi-drug resistant clinical isolates. *J. Med. Microbiol.* **57**, 966–973 (2008).

15. Naze, F. *et al.* Pseudomonas aeruginosa outbreak linked to mineral water bottles in a neonatal intensive care unit: Fast typing by use of high-resolution melting analysis of a variable-number tandem-repeat locus. *J. Clin. Microbiol.* **48**, 3146–3152 (2010).

16. Quick, J. *et al.* Seeking the source of Pseudomonas aeruginosa infections in a recently opened hospital: An observational study using whole-genome sequencing. *BMJ Open* **4**, e006278 (2014).

17. Sukhum, K. V. *et al.* Antibiotic-resistant organisms establish reservoirs in new hospital built environments and are related to patient blood infection isolates. *Commun. Med.* **2**, 62 (2022).

18. Parcell, B. J. *et al.* Pseudomonas aeruginosa intensive care unit outbreak: winnowing

of transmissions with molecular and genomic typing. *J. Hosp. Infect.* **98**, 282 (2018).

19. Lebreton, F. *et al.* A panel of diverse Pseudomonas aeruginosa clinical isolates for research and development . *JAC-Antimicrobial Resist.* **3**, (2021).

20. Giani, T. *et al.* Italian nationwide survey on pseudomonas aeruginosa from invasive infections: Activity of ceftolozane/tazobactam and comparators, and molecular epidemiology of carbapenemase producers. *J. Antimicrob. Chemother.* **73**, 664–671 (2018).

21. Mereuță, A. I., Docquier, J. D., Rossolini, G. M. & Buiuc, D. [Detection of metallo-beta-lactamases in gram-negative bacilli isolated in hospitals from Romania--research fellowship report]. *Bacteriol. Virusol. Parazitol. Epidemiol.* **52**, 45–49 (2007).

22. Arhin, F. F., Draghi, D. C., Pillar, C. M., Moeck, G. & Sahm, D. F. Spread of the blaIMP-13 gene in French Pseudomonas aeruginosa through sequence types ST621, ST308 and ST111. *Int. J. Antimicrob. Agents* **40**, 562–563 (2012).

23. Santella, G. *et al.* Intercontinental dissemination of IMP-13-producing Pseudomonas aeruginosa belonging in sequence type 621. *J. Clin. Microbiol.* **48**, 4342–4343 (2010).

24. Teo, J. Q. M. *et al.* Genomic characterization of carbapenem-non-susceptible Pseudomonas aeruginosa in Singapore. *Emerg. Microbes Infect.* **10**, 1706–1716 (2021).

25. Shireen Kotay, a Weidong Chai, a William Guilford, b Katie Barry, a Amy J. Mathersa, C. & Division. Spread from the Sink to the Patient□: In Situ Study Using Green Fluorescent. **83**, 1–12 (2017).

26. Hota, S. *et al.* Outbreak of Multidrug-Resistant Pseudomonas aeruginosa Colonization and Infection Secondary to Imperfect Intensive Care Unit Room Design . *Infect. Control Hosp. Epidemiol.* **30**, 25–33 (2009).

27. Lam, J. S., Taylor, V. L., Islam, S. T., Hao, Y. & Kocíncová, D. Genetic and

Functional Diversity of Pseudomonas aeruginosa Lipopolysaccharide. *Front. Microbiol.* **2**, (2011).

28. Hu, P. *et al.* Molecular epidemiology, resistance, and virulence properties of Pseudomonas aeruginosa cross-colonization clonal isolates in the non-outbreak setting. *Infect. Genet. Evol.* **55**, 288–296 (2017).

29. 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**, 57–64 (2015).

30. López-Causapé, C. *et al.* Evolution of the Pseudomonas aeruginosa mutational resistome in an international Cystic Fibrosis clone. *Sci. Rep.* **7**, 1–15 (2017).

31. Chambers, J. R. & Sauer, K. The MerR-like regulator BrlR impairs Pseudomonas aeruginosa biofilm tolerance to colistin by repressing PhoPQ. *J. Bacteriol.* **195**, 4678–4688 (2013).

32. Didelot X, Kendall M, Xu Y, White PJ, McCarthy N. Genomic Epidemiology Analysis of Infectious Disease Outbreaks Using TransPhylo. *Curr Protoc*. **2**:e60. (2019). *Erratum* in: Curr Protoc. 2022 Aug;2(8):e552.

33. Carson J, Keeling M, Wyllie D, Ribeca P, Didelot X. Inference of Infectious Disease Transmission through a Relaxed Bottleneck Using Multiple Genomes Per Host. *Mol Biol Evol*. **41**(1):msad288. (2024).

34. Carson J, Keeling M, Ribeca P, Didelot X. Incorporating Epidemiological Data into the Genomic Analysis of Partially Sampled Infectious Disease Outbreaks. *Mol Biol Evol*. **42**(4):msaf083. (2025).

35. Jones, L. D. *et al.* Effectiveness of foam disinfectants in reducing sink-drain gram-negative bacterial colonization. *Infect. Control Hosp. Epidemiol.* **41**, 280–285 (2020).

36. Balasubramanian, R., Van Boeckel, T. P., Carmeli, Y., Cosgrove, S. & Laxminarayan,

R. Global incidence in hospital-associated infections resistant to antibiotics: An analysis of point prevalence surveys from 99 countries. *PLoS Med.* **20**, e1004178 (2023).

37. Galac, M. R. *et al.* A Diverse Panel of Clinical Acinetobacter baumannii for Research and Development . *Antimicrob. Agents Chemother.* **64**, 1–11 (2020).

38. DE, W., J, L. & B, L. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, (2019).

39. Bushnell, B. BBMap. Available at: https://sourceforge.net/projects/bbmap/. (Accessed: 5th October 2021)

40. Silva, G. G. *et al.* Combining de novo and reference-guided assembly with scaffold_builder. *Source Code Biol. Med.* **8**, 23 (2013).

41. Curran, B., Jonas, D., Grundmann, H., Pitt, T. & Dowson, C. G. Development of a Multilocus Sequence Typing Scheme for the Opportunistic Pathogen Pseudomonas aeruginosa. *J. Clin. Microbiol.* **42**, 5644 (2004).

42. Feldgarden, M. *et al.* Validating the AMRFINder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrob. Agents Chemother.* **63**, (2019).

43. Hunt, M. *et al.* ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb. Genomics* **3**, e000131 (2017).

44. Walker, B. J. *et al.* Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS One* **9**, (2014).

45. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

46. Stover, C. K. *et al.* Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen. *Nat. 2000 4066799* **406**, 959–964 (2000).

47. Croucher, N. J. *et al.* Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43**, e15 (2015).

48. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 1–3 (2019). doi:10.1093/bioinformatics/btz305

49. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. **49**, W293–W296 (2021).

50. Rambaut, A., Lam, T. T., Carvalho, L. M. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, (2016).

51. Bouckaert, R. *et al.* BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput. Biol.* **10**, (2014).

52. Bouckaert, R. R. & Drummond, A. J. bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evol. Biol.* **17**, 1–11 (2017).

53. Russel, P. M., Brewer, B. J., Klaere, S. & Bouckaert, R. R. Model Selection and Parameter Inference in Phylogenetics Using Nested Sampling. *Syst. Biol.* **68**, 219–233 (2019).

Acknowledgments: The authors are thankful to all the staff of the MRSN and the clinical microbiology laboratory of the MRSN MTF network. Additionally, we would like to thank Dr. Madeline Galac for sharing insight on phylogenetic analysis using parsimony and Dr. Xavier Didelot for helpful discussions in predicting patient transmission events during bacterial outbreaks. The manuscript has been reviewed by the Walter Reed Army Institute of Research and there is no objection to its presentation. The views expressed herein are those of the author(s) and do not necessarily reflect the official policy or position of the Defense Health Agency, Brooke Army Medical Center, the Department of Defense, nor any agencies under the U.S. Government.

Funding: This study was funded by the U.S. Army Medical command and the Defense Medical Research and Development Program. In addition, this study was partly funded by the Armed Forces Health Surveillance Division (AFHSD), Global Emerging Infections Surveillance (GEIS) Branch ProMIS P0079_22_WR. The funders have no role in the decision to publish or the preparation of this article.

Author Contributions: P.T.M and F.L. designed research; W.S., L.R.H, A.P., C.H., M.J.M, B.W.C., E.S., A.O., R.M., J.S., K.B., B.T.J, L.P., K.L., B.T., L.M.Y., Y.I.K., P.T.M. and F.L. performed research. W.S., L.R.H., C.H., E.S., J.W.B., P.T.M. and F.L. analyzed research; B.T., A.B., A.E.M, J.L.K, R.J.C., J.W.B. contributed clinical metadata and isolates. W.S., M.J.M, P.T.M and F.L. wrote the paper with input from all authors.

Data and materials availability: Both genomic assemblies and raw sequencing data of all isolates analyzed in this study are publicly available in the NCBI database under the BioProject number PRJNA8852179

Competing interests: The authors declare that they have no competing interests.

Ethics approval and consent to participate: The isolates and clinical information were collected as part of the public health surveillance activities of the MRSN, as determined by the WRAIR branch director and Human Subjects Protection Branch (HSPB) which granted ethical approval. Informed patient consent was waived as samples were taken under a hospital surveillance framework for routine sampling. The research conformed to the principles of the Helsinki Declaration.

Figure 1. Phylogenetic and spatio-temporal analysis of 253 clinical ST-621 *P. aeruginosa* isolates. (A) SNP-based, core genome phylogeny. Outbreak subclones are colored with red (SC1) or blue (SC2) branches. From inner- to outermost, datasets show: the year, the floor, the building, and the source of isolation. Patients with prolonged carriage (>1 year) are indicated. **(B)** Timeline of isolate collection. Each line indicates a single patient. Isolates are shown as dots colored by the floor the patient was on. Patients with prolonged carriage are highlighted. **(C)** Simplified floor plan of the Main (light blue) and Tower (dark blue) Buildings of Facility A. Wards and floors are numbered. For each floor, a pie chart indicates the proportion of isolates collected from urine, respiratory, or other cultures. Within each ward, isolates are colored by year of collection with a red outline indicating primary (first) isolates. Symbols indicate the presumed origin of infection. **(D)** Bar charts showing for each year the number of: all *P. aeruginosa* primary clinical isolates collected at Facility A (bottom), the subset of genome-sequenced primary isolates (middle), and the subset of identified ST-621 isolates (top). Bars are color-coded by the isolation source. **(E)** Distribution of SNP distances for all inter-patient isolates or subsets collected in the same month, the same ward, or both. SNP distances within patient isolates are also indicated. Dotted lines indicate the local maxima.

Figure 2. Dated phylogeny of 253 ST-621 outbreak isolates. Time-stamped phylogeny. Error bars for annotated nodes are shown as gray boxes. Years are indicated on the left, along with the opening of the Main Building (T1), and the Tower Building extension (T2). The MRCA of SC1 and SC2 are indicated in black text above the node. The MRCA of isolates from patients with prolonged carriage are shown in red, indicated by their patient ID. The predicted emergence of select variants are shown in blue. Data showing the floor and source isolates were collected from is shown, as in Figure 1. Select datasets are shown including the antibiotic susceptibility testing results for cefepime (cephalosporin) and imipenem (carbapenem).

Figure 3. Identification of environmental reservoirs of the ST-621 outbreak clone in sink drains. (A) Photos depict a swab from a sink drain in a patient room and the typical sink design (shallow basins and gooseneck faucets) in facility A. **(B)** Minimum Spanning Tree of 326 Facility A ST-621 *P. aeruginosa* isolates, including contemporary isolates collected in 2021 and 2022 from clinical (n= 56) and environmental (n= 17) sources. Isolates from wards #1, 6 and 20 where environmental contamination was identified are colored. Environmental isolates are indicated with the letter E.

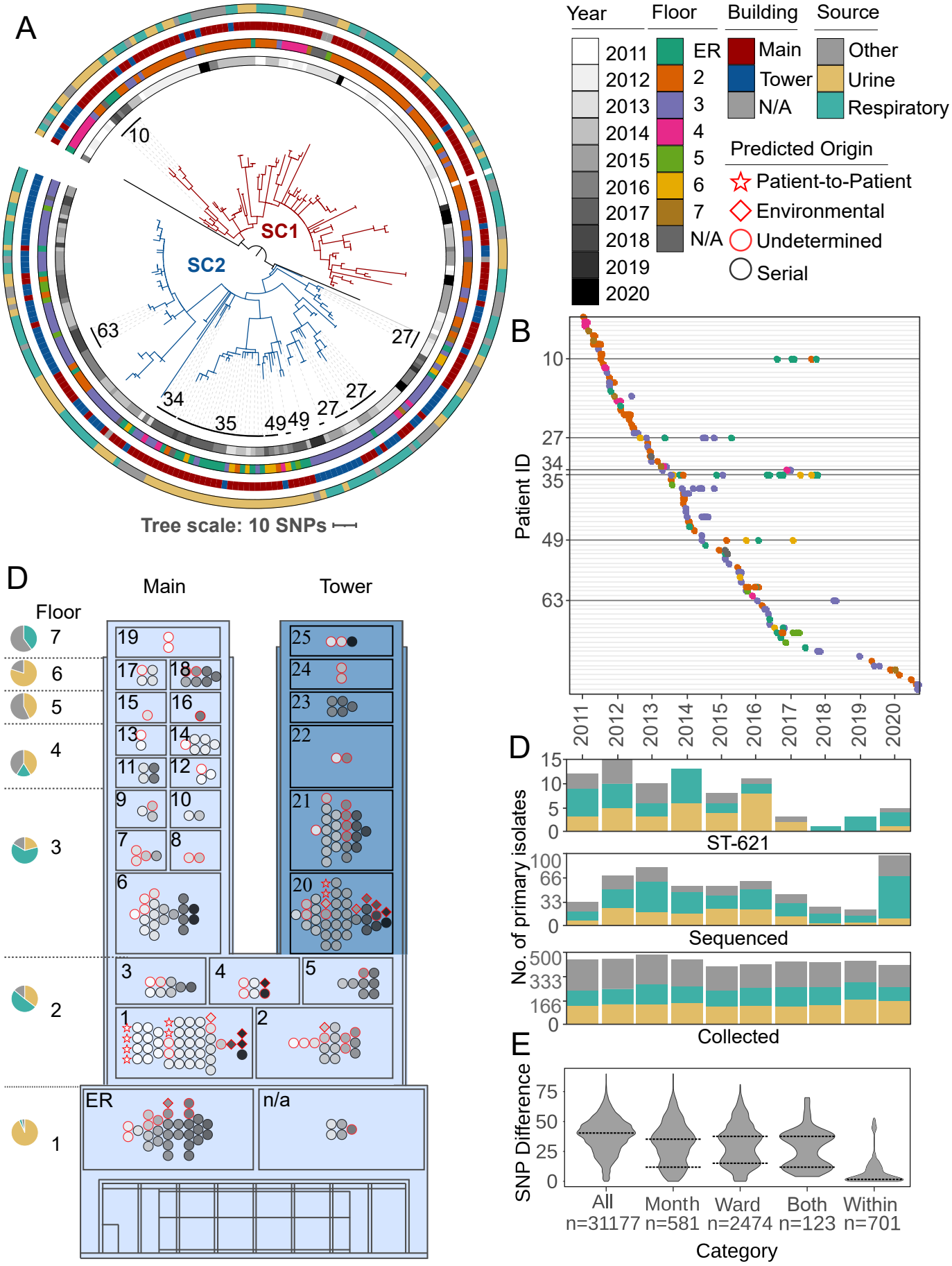
Figure 4. Antibiotic resistance and poly-variant sites of 253 outbreak isolates. (A) Bar charts showing the proportion of clinical isolates resistant to tobramycin, cefepime, imipenem, and ciprofloxacin each year from 2011-2020. **(B)** Chart of mutations (excluding synonymous) in resistance-associated genes. Each block indicates one distinct mutation observed within the corresponding gene. Blocks are colored by the number of isolates sharing this exact mutation. Blocks are outlined to indicate whether the mutation is found within SC1, SC2, or isolates from both subclones. Mutations causing a predicted loss-of-function (LOF) are indicated with a star. **(C)** Pruned SNP-based phylogeny showing the acquisition of colistin resistance in Patient 2. Patient ID, day of collection (since the first outbreak isolate), and available colistin MIC are shown on the right. The MRCA of the PhoQ E77fs mutation is indicated with a triangle.

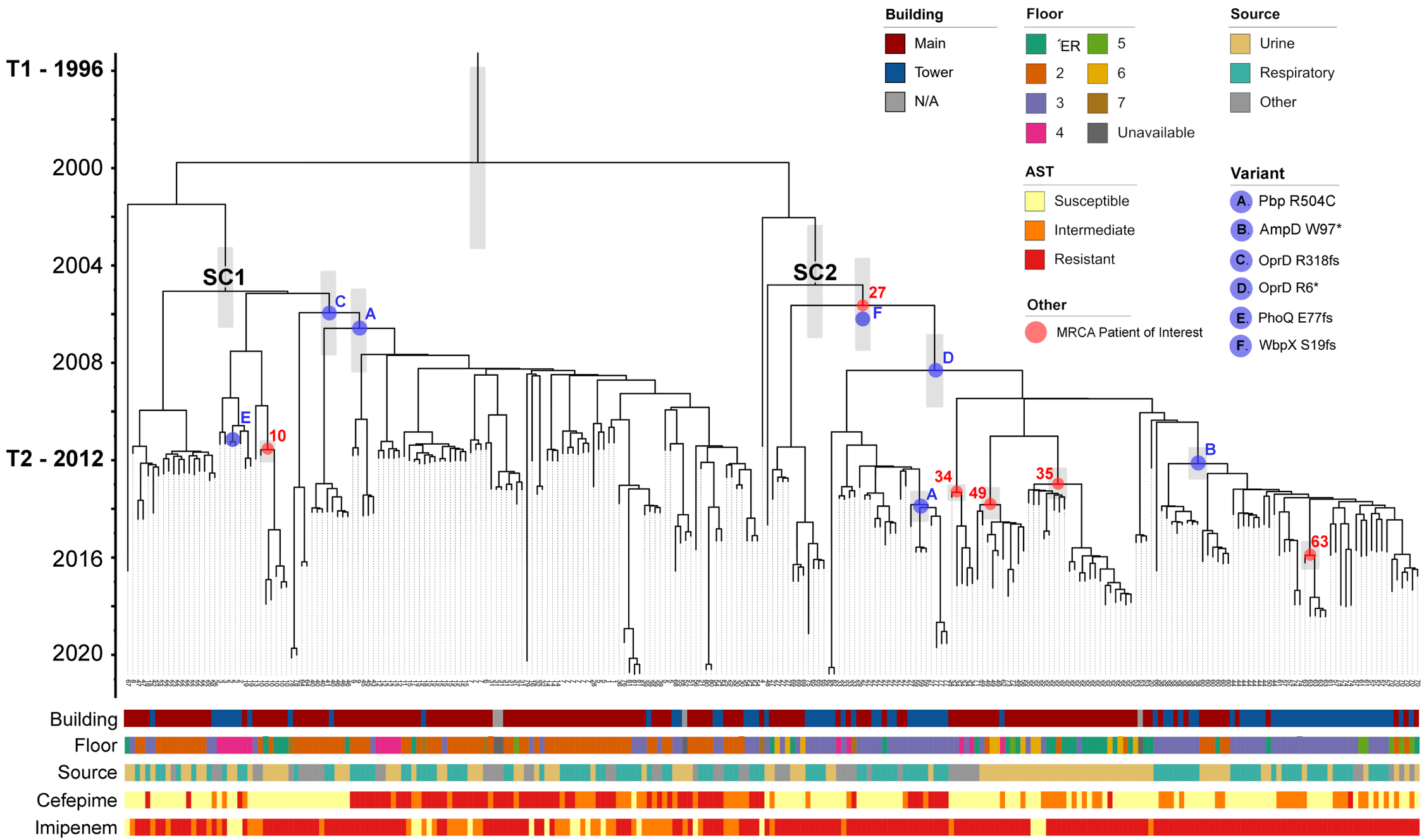
Figure 5. Identification of pathoadaptive mutations. (A) Distribution of COG functional categories for all genes (Ref) compared to the subset of genes (Sel) with 2 or more distinct mutated sites (excluding synonymous) in outbreak isolates. Significant enrichment in cell wall biogenesis and signal transduction categories are indicated. A positional chart of mutations across the entire chromosome of closed reference genome for isolate 4605 is provided. The number of unique mutation sites within each gene is indicated (top) as well as the number of outbreak isolates which carried each mutation (bottom). Genes with 3 or more distinct mutated sites are labeled and genes involved in cell wall biogenesis and signal transduction genes are color coded. **(B)** Accumulation of mutations (excluding synonymous) in genes and pathways associated with *P. aeruginosa* pathogenesis. Each block indicates one distinct mutation observed within a gene part of the indicated functional group (data available in Table S8). Blocks are colored by the number of isolates sharing this exact mutation. Blocks are outlined to indicate whether the mutation is found within SC1, SC2, or isolates from both subclones. Mutations causing a predicted loss-of-function (LOF) are indicated with a star.

Figure S1. Minimum Spanning Trees of *Pseudomonas aeruginosa* cgMLST. Isolates are color coded by facility, showing the 10 most prominent. **(A)** Includes all 5,129 *Pseudomonas aeruginosa* in the MRSN's collection from 2011-2020. Clusters of isolates belonging to ST-235, -244, and -621 are indicated. **(B)** Includes all ST-621 isolates from 2011-2020. Genetic distances higher than 23 allelic differences are indicated.

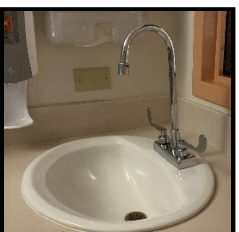
Figure S2. Bar chart showing the distribution of SNP distances for all inter-patient isolates. For a given SNP distance (x-axis) the number of observed isolate pairs is plotted (y-axis). Raw data available in Table S2.

Figure S3. Bar charts showing the percentage of sequenced *P. aeruginosa* isolates compared to the total collected by Hospital A stratified by year (2011-2020) and isolate types (respiratory, urine, or other).

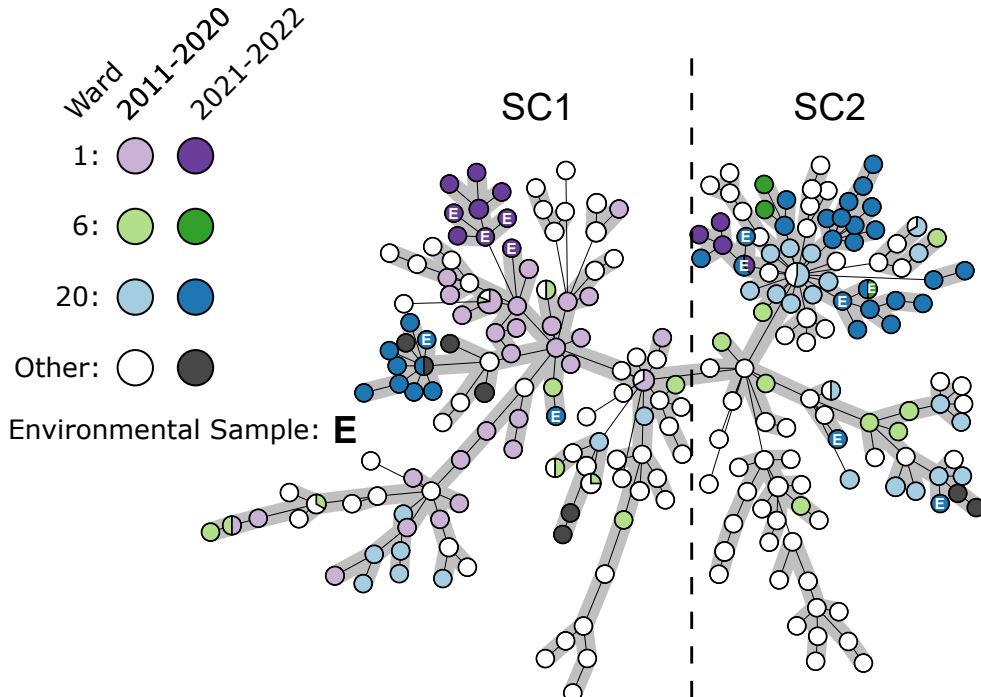

A
SC1
SC2
10
63
27
34
35
49
49
27
27
Tree scale: 10 SNPs
Year
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Floor
ER
2
3
4
5
6
7
N/A
Building
Main
Tower
N/A
Source
Other
Urine
Respiratory
Predicted Origin
Patient-to-Patient
Environmental
Undetermined
Serial
B
Patient ID
10
27
34
35
49
63
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
D
Main
Tower
Floor
7
6
5
4
3
2
1
19
17
18
15
16
13
14
11
12
9
10
7
8
6
25
24
23
22
21
20
3
4
5
1
2
ER
n/a
D
No. of primary isolates
ST-621
Sequenced
Collected
E
SNP Difference
All
n=31177
Month
n=581
Ward
n=2474
Both
n=123
Within
n=701
Category

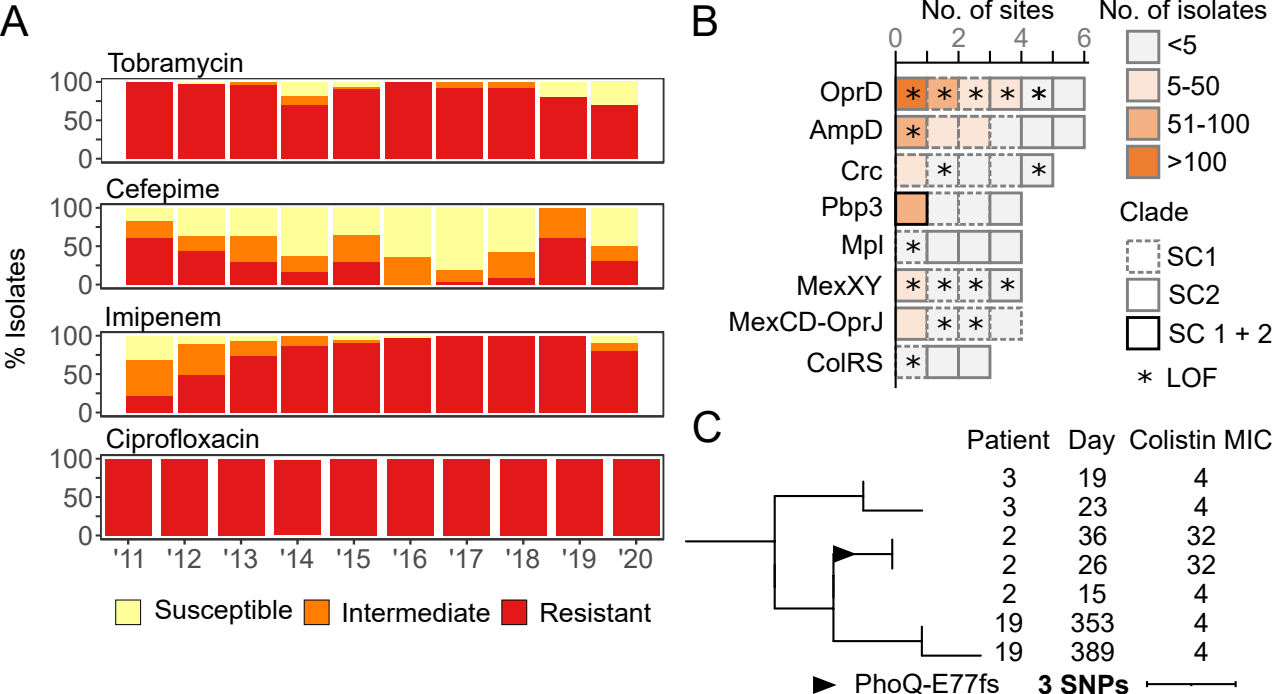

Building
Main
Tower
N/A
Floor
ER
2
3
4
5
6
7
Unavailable
Source
Urine
Respiratory
Other
AST
Susceptible
Intermediate
Resistant
Other
MRCA Patient of Interest
Variant
A. Pbp R504C
B. AmpD W97*
C. OprD R318fs
D. OprD R6*
E. PhoQ E77fs
F. WbpX S19fs
T1 - 1996
2000
2004
2008
T2 - 2012
2016
2020
SC1
SC2
C
A
E
10
27
F
D
34
49
35
B
63
Building
Floor
Source
Cefepime
Imipenem

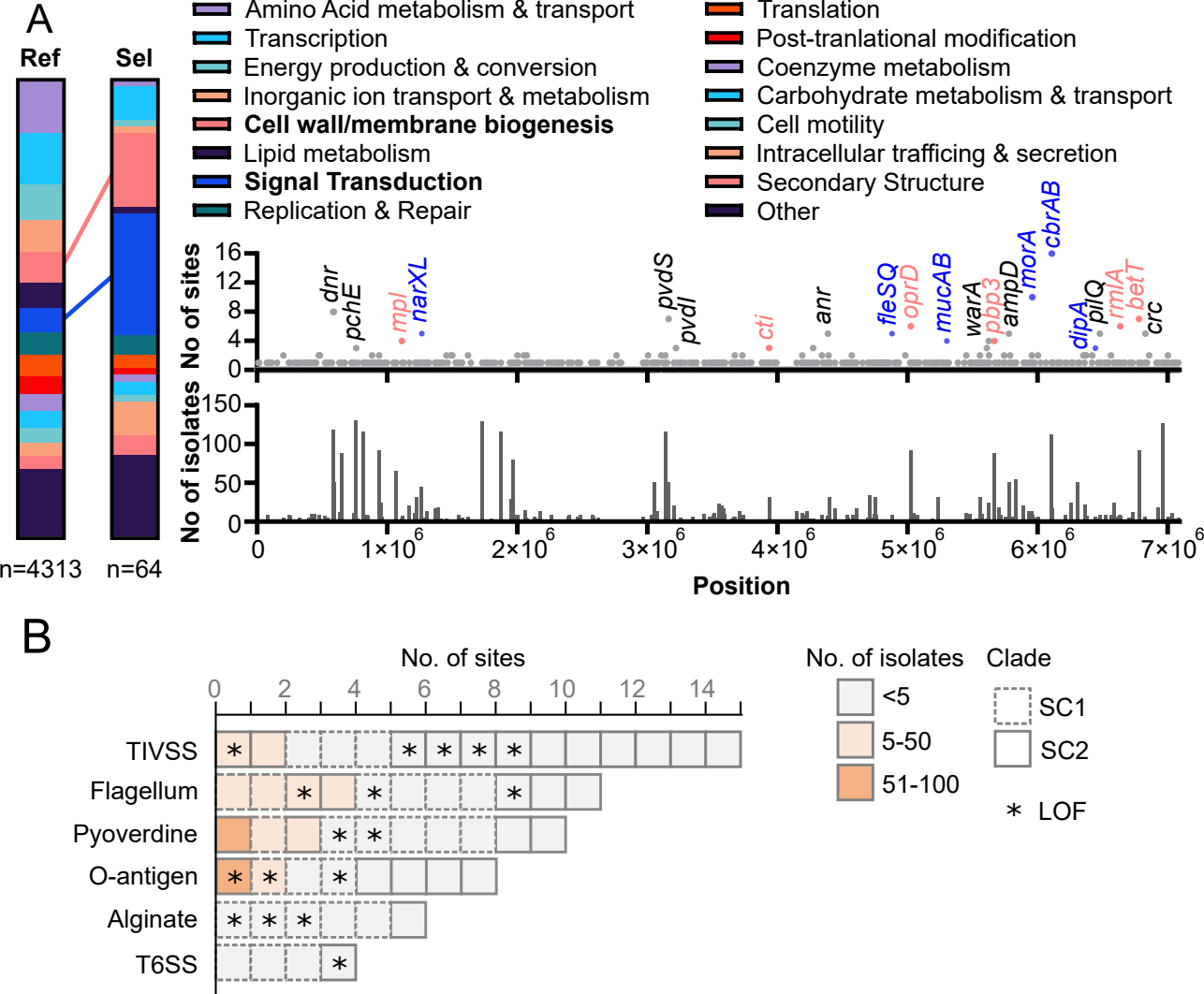
A

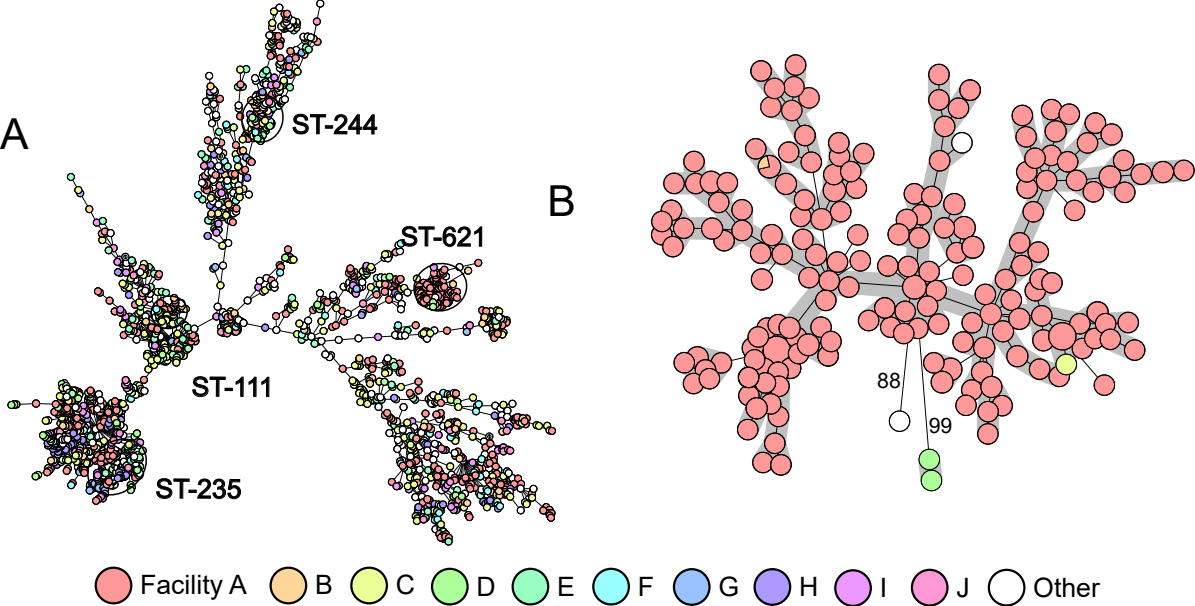


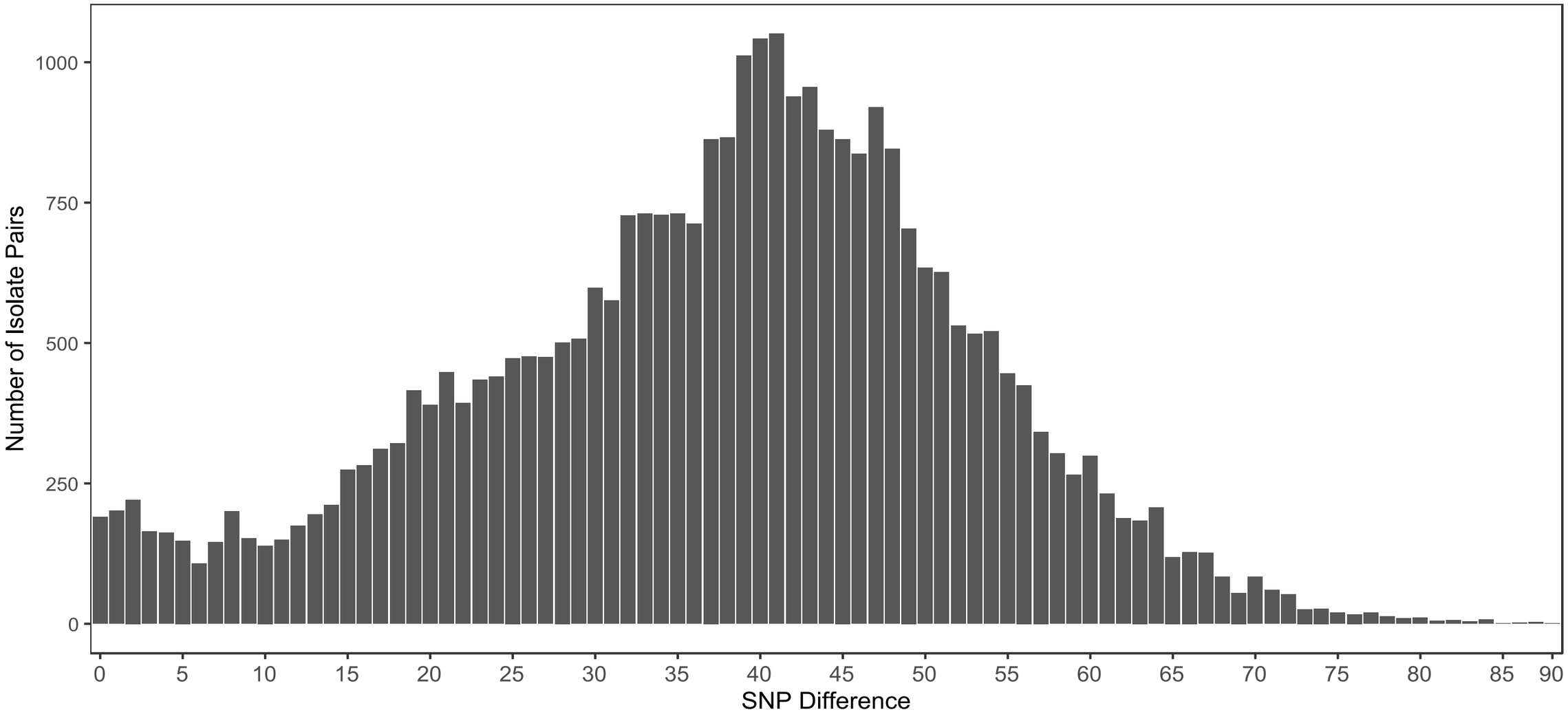
B

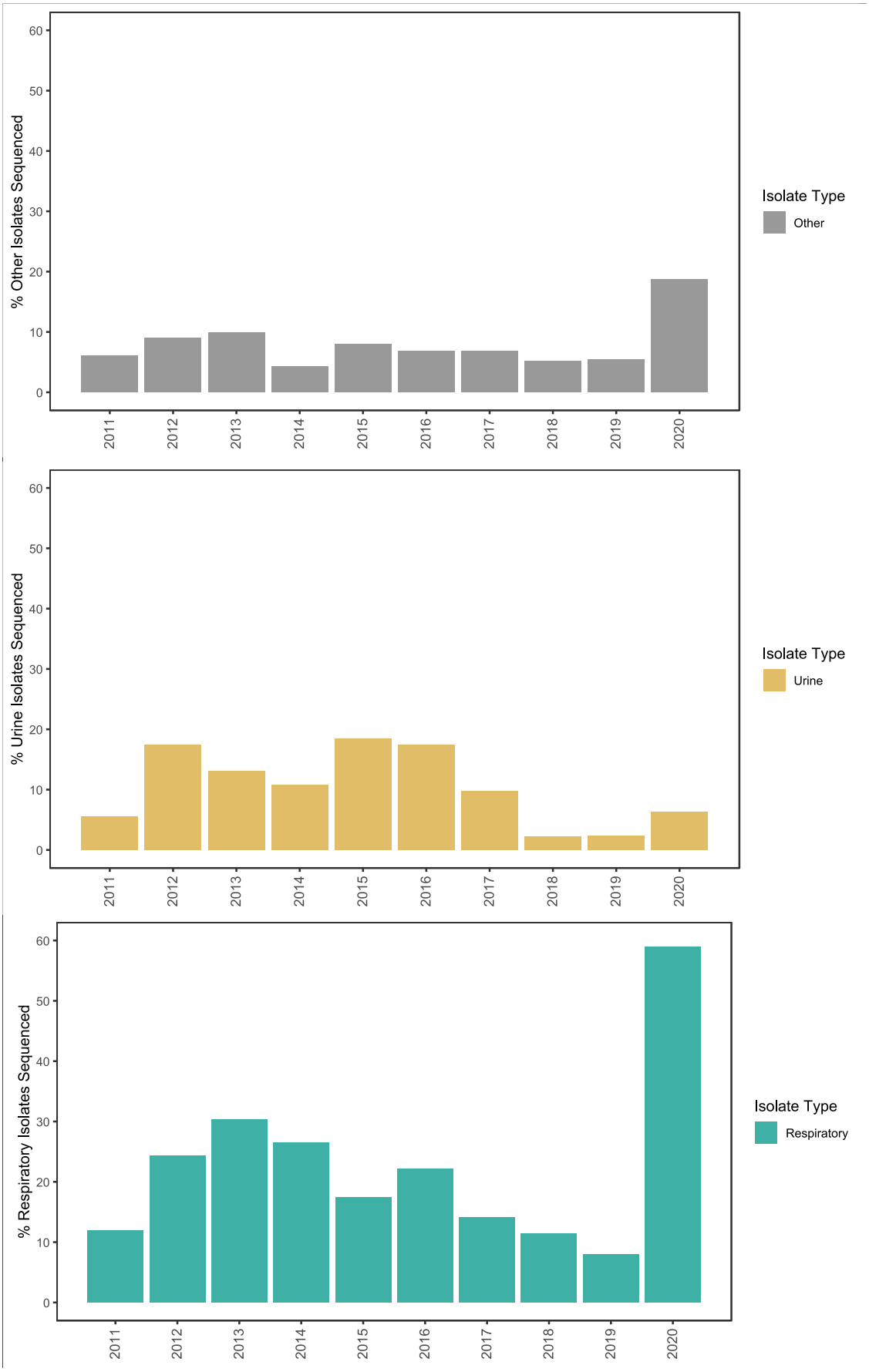

Ward
2011-2020
2021-2022
1:
6:
20:
Other:
Environmental Sample: E
SC1
SC2


A
Tobramycin
Cefepime
Imipenem
Ciprofloxacin
% Isolates
100
50
0
'11
'12
'13
'14
'15
'16
'17
'18
'19
'20
Susceptible
Intermediate
Resistant
B
No. of sites
0
2
4
6
No. of isolates
<5
5-50
51-100
>100
OprD
AmpD
Crc
Pbp3
Mpl
MexXY
MexCD-OprJ
ColRS
Clade
SC1
SC2
SC 1 + 2
* LOF
C
Patient
Day
Colistin MIC
3 19 4
3 23 4
2 36 32
2 26 32
2 15 4
19 353 4
19 389 4
PhoQ-E77fs
3 SNPs


A
Ref
Sel
n=4313
n=64
Amino Acid metabolism & transport
Transcription
Energy production & conversion
Inorganic ion transport & metabolism
Cell wall/membrane biogenesis
Lipid metabolism
Signal Transduction
Replication & Repair
Translation
Post-tranlational modification
Coenzyme metabolism
Carbohydrate metabolism & transport
Cell motility
Intracellular trafficing & secretion
Secondary Structure
Other
No of sites
No of isolates
dnr
pchE
mpl
narXL
pvdS
pvdI
cti
anr
fleSQ
oprD
mucAB
warA
pbp3
ampD
morA
cbrAB
dipA
pilQ
rmlA
betT
crc
Position
B
No. of sites
No. of isolates
Clade
<5
5-50
51-100
SC1
SC2
* LOF
TIVSS
Flagellum
Pyoverdine
O-antigen
Alginate
T6SS


A
ST-244
ST-621
ST-111
ST-235
B
88
99
Facility A
B
C
D
E
F
G
H
I
J
Other


Number of Isolate Pairs
1000
750
500
250
0
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
SNP Difference


% Other Isolates Sequenced
0
10
20
30
40
50
60
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Isolate Type
Other
% Urine Isolates Sequenced
0
10
20
30
40
50
60
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Isolate Type
Urine
% Respiratory Isolates Sequenced
0
10
20
30
40
50
60
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Isolate Type
Respiratory