Cartography of opportunistic pathogens and antibiotic resistance genes in a tertiary hospital environment

Kern Rei Chng[1,*], Chenhao Li[1,*], Denis Bertrand[1,*], Amanda Hui Qi Ng[1], Junmei Samantha Kwah[1], Hwee Meng Low[1], Chengxuan Tong[1], Licheng Xu[4], Tamar V. Av-Shalom[1], Chiea Chuen Khor[1], MetaSUB Consortium, Swaine L. Chen[1], Christopher E. Mason[2], Oon Tek Ng[3], Kalisvar Marimuthu[3], Brenda Ang[3], Niranjan Nagarajan[1,5,#]

[1]Genome Institute of Singapore, 60 Biopolis Street, #02-01 Genome, Singapore 138672, Singapore

[2]Department of Physiology and Biophysics, Weill Cornell Medicine, New York 10065, USA

[3]Institute of Infectious Diseases and Epidemiology, Tan Tock Seng Hospital, 11 Jalan Tan Tock Seng, Singapore 304833, Singapore

[4]Singapore University of Technology and Design, 8 Somapah Rd, Singapore 487372, Singapore

[5]National University of Singapore, 21 Lower Kent Ridge Road, Singapore 119077, Singapore

*Contributed equally

[#]Corresponding author

Abstract

There is growing attention surrounding hospital acquired infections (HAIs) due to high associated healthcare costs, compounded by the scourge of widespread multi-antibiotic resistance. Although hospital environment disinfection is well acknowledged to be key for infection control, an understanding of colonization patterns and resistome profiles of environment-dwelling microbes is currently lacking. We report the first extensive genomic characterization of microbiomes (355), common HAI-associated microbes (891) and transmissible drug resistance cassettes (1435) in a tertiary hospital environment based on a 2-timepoint sampling of 179 sites from 45 beds. Deep shotgun metagenomic sequencing unveiled two distinct ecological niches of microbes and antibiotic resistance genes characterized by biofilm-forming and human microbiome influenced environments that display corresponding patterns of divergence over space and time. To study common nosocomial pathogens that were typically present at low abundances, a combination of culture enrichment and long-read nanopore sequencing was used to obtain thousands of high contiguity genomes (2347) and closed plasmids (5910), a significant fraction of which (>58%) are not represented in current sequence databases. These high-quality assemblies and metadata enabled a rich characterization of resistance gene combinations, plasmid architectures, and the dynamic nature of hospital environment resistomes and their reservoirs. Phylogenetic analysis identified multidrug resistant clonal strains as being more widely disseminated and stably colonizing across hospital sites. Further genomic comparisons with clinical isolates across multiple species supports the hypothesis that multidrug resistant strains can persist in the hospital environment for extended periods (>8 years) to opportunistically infect patients. These findings highlight the importance of characterizing antibiotic resistance reservoirs in the hospital environment and establishes the feasibility of systematic genomic surveys to help target resources more efficiently for preventing HAIs.

Introduction

The global epidemic of antibiotic resistance and associated infections has put hospitals at the epicenter of this challenge[1]. It is estimated that if the spread of antibiotic resistance grows unchecked, it will cause millions of deaths worldwide, with a concomitant economic impact of >100 trillion dollars by 2050[2]. In general, hospital acquired infections (HAIs) pose a high healthcare burden in both developed and developing countries[3]. In the U.S., approximately one in 25 acute-care hospital patients have active HAIs each day, translating to 721,800 HAIs each year[4]. It is estimated that 11.5% of such patients will die during hospitalisation[4]. The problem of HAIs is further compounded by the global spread of multidrug resistant organisms (MDROs) that complicate infection management, limit therapy options and result in poorer outcomes[5]. The risk of HAIs can be mitigated through good infection prevention practice, with hand hygiene advocated as one of the key strategies to limit spread of micro-organisms between patients and medical staff[6].

In addition to human-to-human transfer, the inanimate hospital environment is another key node in the transmission network, with mounting evidence that it harbors opportunistic antibiotic resistant pathogens that contribute to HAIs[7,8]. Reinforced environmental cleaning measures have also been shown to be effective in reducing HAIs[9-11]. As a built-environment with unique clinical significance, the microbial ecology and uncharacterized genetic reservoirs of hospital environments are of special interest for both infection epidemiology and microbiology. For example, the transmission and recombination profiles of antibiotic resistant genes in hospital environments remains largely unknown, and could be valuable in gauging the potential risk for emergence of novel resistance combinations in specific species. Similarly, comparative genomics of hospital adapted strains and epidemic strains can help in identifying the source of HAI outbreaks and inform infection control. While large-scale surveillance holds the promise to reveal clinical and biological insights pertaining to the hospital environment microbiome as a functional reservoir of pathogens and antibiotic resistance genes, significant technological challenges still remain. Traditionally, efforts to survey the hospital environment have been focused on culture-based isolation of specific pathogens. Each isolate is then individually characterized via functional profiling, genotyping and/or whole genome sequencing[12-17]. This is a laborious process, prone to isolation bias and precludes insights into microbial community dynamics and its relationship with the hospital built-environment for the transmission of HAIs[18].

The development of metagenomics has made it possible to profile the entire community structure, characterizing individual microbes without the need for isolation, and represents a high-throughput and scalable method for surveying the hospital environment microbiome[19]. This capability has been leveraged in the form of 16S rRNA sequencing to study bacterial diversity, particularly in intensive care units (ICU), in several early studies[20-23]. In a landmark study in 2017, Lax et al used this approach to extensively characterize microbial ecology, colonization and succession patterns in a newly built hospital environment (a subset with shotgun metagenomics)[24]. Using bioinformatics approaches, they were also able to identify

ecological signatures of exchange of bacteria between the hospital environment, patients and healthcare workers. In general, however, 16S rRNA sequencing precludes the detailed analysis of nosocomial strains, resistomes, metabolic pathways, and transmission of pathogenomes[25]. The use of Illumina shotgun metagenomics by Brooks et al was key to characterizing strain polymorphisms and relatedness of pathogens in low diversity neonatal ICU environments[26]. Several limitations remain for the use of shotgun metagenomics in general, including low microbial biomass, the presence of nosocomial strains at low relative abundances, the presence of multiple strains, computational constraints in strain-level analysis[27], and the shortcomings of short-read metagenomics for assembling high-contiguity, strain-resolved genomes for detailed genetic analysis[24,26,28,29].

The availability of portable, real-time, long-read sequencing platforms presents new opportunities and challenges for pathogenome and resistome monitoring[30,31]. In this study, we combined extensive short-read shotgun metagenomic sequencing of multiple sites, wards and timepoints (355 samples), with enrichment and nanopore sequencing of antibiotic-resistant mixed cultures (1661 samples), to provide the most extensive genetic characterization of hospital environments to date. The combination of metagenomic surveys (short-read based) with detailed genomic analysis of nosocomial strains (long-read based) is ideal for studying the interplay between dissemination, abundance and turnover patterns of pathogens and antibiotic resistance genes with the abiotic components of the hospital environment, informing the development of targeted cleaning protocols and prioritization of high-risk areas for cleaning. Nanopore metagenomics enabled us to generate thousands of high-contiguity genomes (2347) and closed plasmids (5910), unveiling the significant under-characterized genetic diversity (>58% novel) for potentially pathogenic microbes. These were harbored in two distinct ecological niches characterized by biofilm-forming and human microbiome associated bacteria, with divergent patterns of temporal and spatial variation. The availability of a large collection of high-quality genomes helped establish through phylogenetic analysis the notable observation that multi-drug resistant clonal strains are more likely to be widely disseminated and stably colonizing across hospital sites. In addition, the genomes delineated common resistance gene combinations and plasmid architectures, revealing the dynamic nature of hospital environment resistomes and their understudied reservoirs. Further genomic comparisons with clinical isolates across multiple species supports the hypothesis that multidrug resistant strains can persist in the hospital environment for extended periods of time (>8 years) to opportunistically infect patients. Together these findings highlight the importance of characterizing hospital environment microbiomes for understanding its niches and genetic reservoirs, the need for improved methods of disinfection, as well as the feasibility and cost-effectiveness of large-scale genomic surveys for informing infection control measures.

Results

Hospital environment microbiomes offer distinct ecological niches for opportunistic pathogens and antibiotic resistance genes

A diverse set of sites (n=7) of concern for infection control[32-34], and different room types distributed around the building (5 single-bed isolation rooms, 4 MDRO and 4 standard 5-bedded wards) were selected for two time-point sampling (1 week apart) of a tertiary care hospital in Singapore (45 beds, 4% of total; 179 sites, 358 samples; **Suppl. Fig. S1** and **Suppl. File 1**). Illumina shotgun metagenomic sequencing (2×101bp) was used to deeply characterize each sample (30 million reads on average; 3/358 libraries failed due to low biomass) and obtain taxonomic and functional profiles, as well as resistomes (**Suppl. File 2**; **Methods**). Taxonomic profiles were visualized using a principal coordinates analysis (PCoA) plot to identify two distinct microbial community types in the hospital environment (CTA and CTB; **Fig. 1a**). While community type A (CTA) was largely composed of more taxonomically diverse, high-touch surfaces with frequent contact from patients and health-care workers[35], community type B (CTB) represents sites that are increasingly of concern for infection control for their propensity to harbor MDROs[16,33,36] (Wilcoxon p-value<10^{-4}; **Suppl. Fig. S2**). Joint analysis of these community types helped identify key taxonomic features that differentiate them, including several human microbiome-associated (e.g. *Cutibacterium*) and aquatic/terrestrial environment associated (e.g. *Achromobacter*) genera in community types A and B respectively (**Fig. 1bi**). At the species level, key differences observed include the enrichment of common skin bacteria (e.g. *Cutibacterium acnes* and *Staphylococcus epidermidis*) and biofilm forming organisms (e.g. *Elizabethkingia anophelis* and *Serratia marcescens*) in CTA and CTB respectively, though their occurrences were not mutually exclusive indicating shared influences (**Fig. 1bii**).

Microbiomes associated with the two community types exhibited varying stability across the two sampled timepoints, with CTA sites demonstrating larger fluctuations (except door handles; Wilcoxon p-value<10^{-5}; **Fig. 1c**). In general, microbial profiles diverged with physical distance (within bed, within wards and across wards) and time (1 week apart), with temporal variability within a week being lower than spatial variability within a ward (Wilcoxon p-value<10^{-3}; **Suppl. Fig. S3a**). The microbial composition of any particular site is however expected to be influenced by a range of factors including abiotic conditions (humidity, temperature, surface type), seeding from microbial reservoirs (human or environmental) and exchange across sites. Based on the sequencing data, we computed various scores to quantify these factors including a microbiome turnover index (fraction of taxa that are gained or lost across timepoints), human influence index (fraction of human reads) and site specificity index (uniqueness of site-specific taxonomic composition in relation to physically proximal sites). This analysis reinforces the notion that CTB sites tend to have stable compositions (low turnover) based on site-specific biofilm configurations with limited human microbiome seeding (**Fig. 1d**). Sites associated with community type A tend to have higher human influence (Wilcoxon p-value<10^{-15}) and turnover indices (Wilcoxon p-value<10^{-3}), though

they are not directly correlated, and tend to have weaker site-specificity (Wilcoxon p-value<10^{-12}), concordant with a model where human activities (patient discharge/admittance events) have a systemic role in shaping their compositions (**Fig. 1d**). Species that were enriched in CTA sites were also observed in CTB sites (and *vice versa*), but tended to have higher turnover in these sites (**Suppl. Fig. S3b**), with a few exceptions such as *Klebsiella pneumoniae* having low turnover in both CTA and CTB sites.

We observed that overall patterns of microbiome variability were consistent across ward types, though isolation rooms exhibited lower variability across two timepoints (**Suppl. Fig. S4**). In line with MDRO management guidelines in Singapore[37], patients colonized with Carbapenem-resistant Enterobacteriaceae (CRE; e.g. *K. pneumoniae*) were typically warded in single-bed isolation rooms, while methicillin resistant *Staphylococcus aureus* (MRSA) carrying patients were largely assigned to MDRO wards. Analysis of differentially abundant common nosocomial pathogens (curated from https://www.cdc.gov/hai/organisms/organisms.html and published literature[4,38]) across ward types identified *K. pneumoniae* and *S. aureus* as being enriched in CTA sites for isolation rooms and MDRO wards respectively, providing further evidence for the influence of patient microbiomes on CTA sites (**Fig. 1e**). Consistent with the observed taxonomic differences, community types A and B also harbored distinct complements of antibiotic resistance genes in their resistomes (**Fig. 1f** and **Suppl. Fig. S5**). While some resistance genes were frequently detected in CTB sites (e.g. ges, oxa-7 in **Fig. 1f**), CTA sites carried a wider diversity of genes at lower frequencies. Despite the growing recent focus on CTB sites as drug resistance reservoirs[16,36], clinically important resistance genes such as oxa-23 (carbapenamase) and mecA (methicillin resistance) were more frequently found in CTA sites. Different sites also exhibited distinct resistome patterns, e.g. specific tetracycline (tetC) and macrolide (mphE) resistance genes were highly enriched only in aerators, while vancomycin resistance genes were only seen in bedside lockers and bedrails, highlighting the importance of considering site and ward-specific patterns when defining infection control and drug resistance mitigation strategies (**Suppl. Fig. S5**).

Nanopore sequencing reveals distribution of pathogenomes in the hospital environment with preferential persistence and spread of multidrug resistant opportunistic pathogens

Based on Illumina shotgun metagenomic profiles we noted that common nosocomial pathogens were generally present at very low relative abundances (median relative abundance <0.5%; **Fig. 2a**), precluding detailed genomic characterization and analysis of transmission patterns, resistance gene combinations and plasmids. The distribution of common nosocomial pathogens exhibited site-specific patterns (PERMANOVA p-value <0.001; **Fig. 2a**), in agreement with the distinct ecological niches observed in the hospital environment (**Fig. 1a**, **b**), and indicated that enrichment cultures could capture a diverse set of species. We exploited this observation to use a culturing, antibiotic selection (5 antibiotics) and metagenomic nanopore sequencing approach to obtain a large database of high-contiguity genome assemblies (n=2347) from the hospital environment (median N50 >1Mbp; **Fig. 2b**, **Suppl. File 3**; **Methods**), expanding substantially on genomic resources established by previous hospital environment

surveillance studies[16,17]. Overall, a large percentage of sites led to viable initial cultures (>95%), with antibiotic selection still resulting in growth in >80% of plates (1495/1790), and >42% of sites resulting in cultures for all 5 antibiotics. Control swabs led to no cultures (0/10) confirming that cultures were not likely due to contamination (**Methods**), and further testing of isolates for antibiotic resistance confirmed that the vast majority of strains in the cultures were likely antibiotic resistant (99%; **Suppl. Note 1**).

DNA was extracted from 1661 plates and sequenced on the GridION system to provide 535Mbp of nanopore sequencing on average per sample (median read length >2.5kbp). Long read metagenomic assembly enabled the frequent reconstruction of megabase-pair sized contigs (versus average N50 <5kbp for Illumina metagenomic assemblies) as the communities were largely simple (**Suppl. Fig. S6**, **Fig. 2b**; **Methods**). An evaluation of these draft genomes based on evolutionarily conserved single-copy genes confirmed that they were on average of high quality (completeness >99%, contamination <0.5%; **Methods**). In total, we obtained genomes for 69 different known species from the hospital environment, 40% of which belonged to common nosocomial pathogens (**Methods**). Our results confirm the viability of these species in different hospital environments and the ability to substantially enrich them for sequencing and genome reconstruction compared to their acutely low abundances in the hospital environment microbiome (median relative abundance = 0.68%, averaged across species; **Fig. 2c**). Large-scale homology analysis of these genomes with public databases[39] also helped identify 42 species-level clusters which do not have representatives from known species, highlighting the potential to obtain high-contiguity and high-quality genomes for novel species in the hospital environment using this approach (**Methods**). Rarefaction analysis of our data showed that >90% of the species and resistance gene diversity that could be sampled for sites in this study was captured by our sample size (**Suppl. Fig. S7**), while significant additional diversity remains to be captured for plasmids and HAI-associated strains (~50%; **Suppl. Note 2**). In addition, this analysis revealed that future surveys of resistance gene families in the hospital environment could be done with much fewer samples (~50), making more regular surveys feasible, affordable and potentially actionable.

As plasmids serve as an important medium for the evolution and spread of antibiotic resistance genes and emergence of multidrug resistant bacteria[40-42], we identified sequences belonging to plasmids in our genomic database and further characterized them (**Methods**). In total, we recovered 696Mbp of putative plasmid sequences (5910 closed/circular sequences and 493Mbp of linear fragments), with most not present in existing databases for complete plasmids[43] (>90%, 1505/1588 plasmid clusters; **Methods**), highlighting the unexplored genetic diversity in the hospital environment. Many closed plasmid sequences were >100kbp long (>9%, n=536), repeat rich and present at low abundances in the microbiome, impeding their characterization using Illumina shotgun metagenomics. In particular, we noted the presence of a group of large mecA carrying plasmids that also contained antiseptic and disinfectant resistance genes (qacA or qacC), a combination that is not represented in existing databases[43], but is in agreement with high biocide-resistant rates for MRSA that has been reported in clinical settings[44,45].

Remarkably, one of these plasmids has genes belonging to several additional antibiotic resistance classes that have not been described before in combination (e.g. dfrC, lnuA and aac6-Aph2), highlighting the value of closed plasmid sequences for detailed characterization of novel resistance gene combinations.

The availability of a large collection of highly contiguous plasmid (closed) and chromosomal (megabase-pair contigs) genome assemblies from the hospital environment allowed us to do genomic relatedness (with environmental and patient strains) and structural analysis (common gene cassettes and exchange across cassettes), as we discuss in the following sections. We first analyzed the evolutionary relationships between genomes for hospital environment strains to understand their spatio-temporal distribution. For many species, a diverse set of strains were observed to be disseminated across the hospital environment (>99.9% average nucleotide identity; from 7 strains for *Pseudomonas aeruginosa* to 39 strains for *S. epidermidis*; **Fig. 2d** and **Suppl. Fig. S8**). A subset of strains were frequently detected at multiple sites and ward types in the hospital environment, and these were also significantly enriched for strains detected at both timepoints (Fisher's exact test p-value=3.1×10^{-11}). Overlaying antibiotic resistance information with these patterns, we noted an enrichment of multi-antibiotic resistance among strains that are widely disseminated across the hospital environment through space and time (Fisher's exact test p-value=4.2×10^{-13}; **Fig. 2d** and **Suppl. Fig. S8**). This was also consistently observed across several common nosocomial pathogens in the hospital environment (Fisher's exact test p-value: 3.7×10^{-2} for *S. aureus*; 3.7×10^{-5} for *S. epidermidis*, 1.1×10^{-4} for *Enterococcus faecalis*, 3.6×10^{-4} for *Enterococcus faecium*, 6.1×10^{-2} for *K. pneumoniae*) highlighting the presence of stable, viable environmental reservoirs for common HAIs and the need to further understand the mechanisms that contribute to enrichment of multidrug resistant strains[46,47].

Diversity and dynamics of antibiotic resistance gene cassettes in the hospital environment

With growing multidrug resistance, the specific combination of antibiotic resistance genes (ARGs) harbored in a strain is important to know from a clinical perspective. In the context of the hospital environment, little is known about the diversity of gene combinations and how often genes are exchanged across genomic cassettes and plasmids. Analyzing our database of 2347 high-contiguity genomes and 5910 closed plasmids together with existing genomic databases, we found that 34% of the ARG combinations observed in the hospital environment were novel (255/752) (**Suppl. File 4**). Certain ARG combinations have obvious clinical significance e.g. we noted the co-occurrence of the mecA gene (methicillin resistance) with fosB gene (fosfomycin resistance) in several environmental *S. aureus* strains, an observation that is concerning given the potential utility of fosfomycin for treating methicillin resistant *S. aureus* infections[48]. Notably, we detected Enterobacteriaceae-associated genes that confer resistance to gentamycin (e.g. aac3-IIa), fosfomycin (e.g. fosA and fosA2) and colistin (e.g. mcr1), all last resort antibiotics for treating CRE infections. Two Enterobacteriaceae-associated plasmids, one carrying a fosA gene and the other carrying a mcr1 gene were obtained from the same bedside locker sample,

highlighting the potential reservoir in the hospital environment for emergence of co-resistance to colistin and fosfomycin. Another Enterobacteriaceae-associated plasmid carried a rifampicin resistance gene (arr), a telling observation given the growing interest to use rifampicin in combination treatments for a variety of gram negative (e.g. *Acinetobacter baumannii*[49,50]) infections in hospitals.

We next identified common resistance gene pairs in close genomic proximity (<10kbp apart) to identify chromosomal resistance gene cassettes that may serve as the unit of evolution, co-regulation and gene exchange (**Methods**). Chromosomal cassettes for resistance genes were generally small (2-6 genes, Average=3) and specific to a species, though two large cassettes carrying extended-spectrum beta-lactamases were found to overlap very well for *K. pneumonia* and *Enterobacter cloacae* (KpnC1, KpnC2 and EclC1, EclC2; KpnC3 and EclC3; **Fig. 3a; Suppl. File S5**). Selective pressure from the rampant use of beta-lactams and plasmid-mediated transmission is likely to have contributed to the sharing of these large cassettes across species. In general, cassettes for gram-negative species were larger and more stable (solid lines to genes) while those for gram-positives were smaller with many variably present members (dashed lines to genes). The largest shared cassette among gram positives (aminoglycoside-streptothricin resistance; ant6-Ia, sat4A and aph3-III) was seen in *Enterococcus* and *Staphylococcus* species, but with no discernible signals of mobile elements[51]. While most genes were observed to be stably present in cassettes (solid lines; except for a few e.g. tetK, far1, catA), exchange of genes across cassettes were rarely observed (e.g. blaZ), indicating that chromosomal cassettes tend to be relatively fixed.

Performing similar analysis for the closed plasmid sequences identified in this study, we first clustered them into shared plasmid backbones and annotated them for their known hosts (identity ≥95%; **Methods**). Analyzing ARGs in the context of these backbones, we found that many resistance genes are variably present in backbones (93/143), and for genes that are stably found in one backbone, a high frequency are also variably present in another backbone (19/31), highlighting the dynamic nature of resistance genes combinations in the plasmids that we recovered in the hospital environment (**Fig. 3b**). In this background, we observed a few gene combinations that were stably present in multiple plasmid backbones, pointing to strong selection for co-existence. For example, the genes strA, strB (streptomycin resistance) and sulII (sulfonamide resistance) co-occur in two distinct backbones (Sen1 and Kpn2, sequence overlap <54%) as a signature from past co-administration of streptomycin and sulphonamides for clinical use[52,53]. Similarly, while aminoglycoside resistance genes such as aadD and aac6-Aph2 are widely and variably disseminated across plasmid backbones, the genes ant6-Ia and aph3-III are stably shared by two distinct backbones (Efa4 and Efs1, sequence overlap <11%) indicating that they may provide synergistic resistance to aminoglycosides by catalysing different modifications. Interestingly, we noted that genes that are widely disseminated across plasmids (e.g. tetK, far1 and blaZ) can come together in a novel, clinically relevant plasmid backbone (**Fig. 3b**; Slu3, with 38 sequences in our database) as described for a cytotoxin producing MRSA strain[54]. While the previously isolated strain was

resistant to fusidic acid and tertracycline, but susceptible to erythromycin and clindamycin, we noted the presence of a common plasmid backbone in our database (Sha2 with 88 sequences) that carries a novel combination of resistance genes for all 4 antibiotics (**Fig. 3b** and **Suppl. File 4**). In general, we observed that antibiotic resistance genes found in plasmids tend to have more genetic linkages in chromosomes than chromosome-exclusive genes (Wilcoxon test p-value=6×10^{-7}), characteristic of higher gene mobility and shuffling for plasmid-associated genes. Thus the plasmid backbones seen in the hospital environment likely represent a more plastic framework to generate diverse ARG combinations, many of which are not seen in genomic cassettes (25%), despite the strong overlap in the complement of resistance genes that they harbour (84% of plasmid genes).

Hospital environment strains overlapping with patient isolates are preferentially multidrug resistant and globally disseminated

The availability of a large database of genomes from many species in the hospital environment, an obvious hub for patient colonization, prompted us to ask the question: "How are environmental strains related to patient colonizing strains?". To examine this relationship, we jointly analyzed environmental strains and patient isolates for different species to construct phylogenetic trees (**Fig. 4**). We started by analyzing Singaporean *E. anopheles* strains from a 2012 HAI outbreak[55]. Strikingly, despite sampling from two different hospitals in Singapore and after a span of 5 years, the outbreak associated strain matched environmental strains from this study with as low as 111 SNPs (average nucleotide identity, ANI 99.997%) in the genome. Both strains were obtained from hospital sinks, which as we noted earlier tends to have a stable community, indicating that the strains may have originated from a common reservoir upstream of water piping systems[56]. In addition, we noted that the shared strain was resistant to all 5 antibiotics used in this study (**Fig. 4a**).

Encouraged by the high-similarity match observed for *E. anopheles*, we analyzed *S. aureus* genomes obtained from a surveillance study of hospitalized patients in the same hospital almost a decade ago[57]. These strains were found to match 7 out of 16 strains that were found in the current study, with some genomes having as few as 65 SNPs and average nucleotide identity as high 99.998% (**Fig. 4b**). The overlapping strains were also found to be significantly enriched in multidrug resistant strains in the environment (Fisher's exact test p-value=4.8×10^{-3}, >2 antibiotics), highlighting the marked stability of antibiotic resistance reservoirs in the hospital environment and persistence of strains.

Extending our analysis to a regional context, we analyzed genomes from infectious *A. baumannii* strains isolated from patients in two major Kuwaiti hospitals, along with our environmental genomes (**Fig. 4c**), and found surprisingly that one of the environmental strains matched with as few as 233 SNPs and high identity (ANI 99.994%). The shared strain was also observed to be resistant to all 5 antibiotics used in this study. Our data thus points to rapid global dissemination of multidrug resistant *A. baumannii* strains in hospitals.

Similar patterns were observed recently for *S. epidermidis* lineages (ST2/ST2-mixed) which seem to have disseminated across the globe within a short period[38]. We confirmed the detection of these rifampicin and fusidic acid resistant[38] lineages in our data (**Fig. 4d**) with as few as 71 SNPs (ANI 99.997%) to our hospital environment genomes. In addition, two other lineages (ST16 and ST59) not known to be globally disseminated (but isolated from patient samples in USA and Australia respectively[38]) were also represented by highly similar genomes in our database (ST16: 288 SNPs, ANI 99.988%; ST59: 299 SNPs, ANI 99.988%). In addition, we again found that the overlap between globally disseminated *S. epidermidis* patient isolates and environmental genomes from this study is enriched for multidrug resistant strains (Fisher's exact test p-value=0.038, >2 antibiotics). Together with the observation that multi-antibiotic resistance strains are persistent and widely disseminated across the hospital (**Fig. 2d**), this data points to selective advantages for multi-drug resistant organisms to persist and spread in hospital environments and patients.

Discussion

While the importance of a well-designed built-environment for preventing infections in hospitals is well appreciated[62], the use of shotgun metagenomic approaches for surveying microbial communities established in functioning hospitals or medical environments remains underexplored[18]. A large baseline survey such as the one conducted here can help provide a reference map that can then be updated based on periodic scans whose frequency and locations can be informed by the initial survey. For example, the turnover score and specificity of a site can help determine whether and how frequently it should be sampled. Variations in human influence scores could also help fine-tune cleaning practices, and the distribution of specific opportunistic pathogens of concern could be informative in an outbreak setting for infection control. As genomics-guided infection control is further refined, the knowledge gained could then feed back into better hospital designs. With further improvements in cost and ease of short-read sequencing, we anticipate that hospital-wide surveys will be increasingly feasible, will provide valuable information for infection control, and eventually be part of routine practice.

In recent years, infection control studies have primarily focused on sites associated with wash areas (sinks, showers etc.; CTB sites), as in many cases the presumptive pathogens for an outbreak have been isolated from such sites[56]. This focus on community type B sites is in line with them being composed of biofilm-forming bacteria and having the ability to harbor viable reservoirs for extended periods of time (e.g. in the plumbing system). Our data however suggests that many medically-relevant species (e.g. *K. pneumoniae*) and resistance plasmids (present in >85% of CTA sites) are frequently harbored in community type A sites. While CTA sites tend to be more dynamic, the ability to detect very similar strains over long durations of time indicates that they may also have distinct reservoirs (e.g. in ventilation or air-conditioning ducts), and that culture-based screening protocols may bias against sites that have lower biomass or are variably colonized. Combining the strengths of metagenomic sequencing and culturing may therefore be needed to more systematically explore the source of outbreaks.

Large-scale genomic studies of nosocomial pathogens through strain isolation can be a laborious and time-consuming process, while as shown here, direct shotgun metagenomics will often not provide detailed genetic information for species present at low relative abundances. The intermediate approach proposed here attempts to address both issues. The use of culture-based enrichment allows us to effectively shift the distribution away from abundant species such as *C. acnes*, and towards nosocomial pathogens that are typically at low relative abundances (e.g. *K. pneumoniae*, *S. aureus* and *A. baumannii*), while also allowing for functional selection such as for antibiotic resistance as we did here. Its combination with long-read metagenomic sequencing is then powerful as it allows us to directly recover high-contiguity genomes without a tedious isolation step. With further automation (e.g. in library preparation) this workflow is suitable for high-throughput analysis and amenable for wider surveillance, in line with the vision to achieve precision epidemiology for infectious diseases[63]. Future improvements in nanopore sequencing throughput, the ability to work with lower DNA input, and use in remote settings, could help accelerate time-to-answer by reducing the culturing period, or eliminating it altogether.

The availability of a database of high-contiguity assemblies, with >8Gbp of sequence, 2347 genomes and 5910 closed plasmids, provides a unique resource for studying transmission patterns of strains and diversity of resistance gene cassettes across species in the hospital environment. Based on this resource, we observed that multidrug resistant strains are preferentially disseminated and persistent in hospital environments across a range of species. This represents a worrisome pattern, with several possible explanations that deserve further investigation. One hypothesis is that hospital cleaning measures may also be selecting for more antibiotic-resistant organisms[64], a model that is also supported by the presence of multiple copies of disinfectant resistance genes in the widely disseminated, multidrug resistant *S. aureus* strains in our study. Another plausible scenario is that hospital environments are preferentially seeded with multidrug resistant strains by the patients they house, though this would not necessarily explain the wide spatial distribution of a few strains in the hospital unless these are also the common strains in patients or in the general population. Comparisons with surveys from other built-environments such as those that are intensively cleaned but do not house patients (e.g. operating rooms) or are not intensively cleaned but have high patient traffic (e.g. waiting areas in clinics) might help explore these questions further. Studies across wards or hospitals with different cleaning protocols could also be illuminating for understanding how antibiotic resistance reservoirs in hospitals can be shaped by infection control practices[65,66].

Despite its importance as the epicenter for the battle against growing antibiotic resistance[1], hospital environments have received relatively little attention compared to studies on agricultural and animal farms[18]. Our analysis highlights that hospital environments harbor significant uncharacterized genetic diversity in terms of microbial species (42 novel species) and antibiotic resistance gene combinations (255). This reservoir can be the origin of new opportunistic infections and serve as fertile ground for evolution of antibiotic resistance combinations that are of clinical concern (e.g. colistin and fosfomycin

resistance). This can be further facilitated by the presence of resistance gene carrying plasmids that serve as a vehicle for gene transfer across species[67], and were commonly found in the hospital environment (n=1400). The development and use of anti-plasmid agents[68,69] could thus be a complimentary strategy to reduce the spread of antibiotic resistance through hospital environments.

The genetic relatedness between environmental and patient-colonizing strains in hospitals is important to study for understanding the potential risk that environmental strains carry for causing infections[24,26]. For strains that are contemporary and co-located, high genetic relatedness between a subset of environmental and patient strains is expected. The observation that highly similar genomes can be obtained despite being temporally separated by >8 years suggests that large reservoirs of nosocomial multidrug resistant strains are being maintained with limited genetic drift. The identification and elimination of these reservoirs may help reduce the incidence of corresponding infections and the risk from maintenance of antibiotic resistance properties. Another distinct observation is the high genetic similarity between MDRO strains observed in Singaporean hospitals and those obtained from patients around the world. The consistency of these patterns across species emphasizes the rapid global dissemination of newly emerged MDRO lineages and the role of hospital environments in this deserves investigation, in conjunction with global environmental metagenomic datasets[70,71].

Taken together, our data points to selective advantages for multidrug resistant organisms to persist and spread in hospital environments (**Fig. 2d**) and be shared with patients (**Fig. 4**). The enrichment of virulence factors (one-sided Wilcoxon p-value=0.027 for *S. aureus*) and the presence of multiple copies of disinfectant resistance genes[44,45] (as many as 3 copies of qacA in s2 and s3; **Fig. 2d**) in the genomes of shared multidrug resistant strains may enable them to successfully colonize both hospital environments and patients, and thus be readily transferred between them. This points to a vicious cycle where disinfectant resistance, antibiotic resistance and virulence may in turn be selected for, enriching for strains that are adept at colonizing both niches in the presence of depleted microbial competition, and offers an explanation for the high incidence of multidrug resistant HAIs worldwide despite increased surveillance and aggressive cleaning measures in hospitals[58-61].

Methods

Sample collection and storage

Environmental swabs were collected from Tan Tock Seng Hospital (TTSH), a major tertiary care hospital with >2000 patient visits every day that serves as the national referral center for communicable diseases in Singapore. Sampling was done in November 2017, with samples for the first timepoint being collected in 2 days and the second timepoint in 3 days, with 1 week separating the timepoints. Samples were collected from isolation rooms (1 bed, typically warding CRE colonized patients), MDRO wards (5 beds, typically warding MRSA colonized patients) and standard wards (5 beds), at 7 different sites, including aerator, sink trap, bed rail, bedside locker, cardiac table, pulse oxymeter and door handle (**Suppl. Fig. S1** and **Suppl. File 1**). Standard cleaning protocols at TTSH require that high-touch areas and sinks are cleaned daily (chlorine 5000ppm and cleaning detergent, respectively; excluding beds that are cleaned upon patient discharge). Isohelix DNA Buccal Swabs (SK-4S) were used for sampling carried out based on MetaSUB protocols[71]. Briefly, a total of 4 swabs were collected with 1 swab (for culturing) moistened with 1X phosphate buffer saline (PBS, pH 7.2) and 3 swabs (2 for metagenomic DNA isolation, 1 for storage) moistened with DNA/RNA shield (Zymo Research, Cat. No. ZYR.R1100-250). Swabbing was performed for 2 min in each site and stored in respective storage liquids (i.e. 1X PBS, pH 7.2 or Zymo DNA/RNA shield). Swabs in PBS were placed on ice and sent for culturing while the other swabs were transported back at room temperature to the laboratory and stored at -80˚C. In total, 1432 swabs were collected from 179 sites in the hospital at 2 timepoints, representing 358 samples.

DNA extraction from swabs

DNA was extracted from swabs using a bead-beating and automated DNA purification system. Briefly, 300 µL of lysis buffer was added to lysing matrix E tubes (MP Biomedicals, Cat. No. 116914500). Samples were homogenized using the FastPrep-24 instrument at 6 m/s for 40s prior to centrifugation at maximum speed for 5 min. The supernatant was treated with Proteinase K (Qiagen Singapore Pte. Ltd, Cat. No. 19133) for 20 min at 56˚C before DNA was purified with Maxwell RSC Blood DNA Kit (Promega Pte. Ltd., Cat. No. AS1400). DNA concentration was quantified using Qubit® 2.0 fluorometer, prepared with Qubit dsDNA HS Assay Kit (Life Technologies Holdings Pte. Ltd., Cat. No. Q32854). DNA extraction from backup swabs was carried out for samples with less than 0.2 ng of DNA. Samples that still had undetectable DNA concentrations were excluded for library preparation (3/358).

Illumina library preparation

Extracted DNA was sheared using Adaptive Focused Acoustics™ (Covaris) with the following parameters: 240s, Duty Factor: 30, PIP: 450, 200 cycles/burst. Metagenomic libraries were prepared with NEBNext Ultra DNA Kit (New England Biolabs, Cat. No. E7370) according to manufacturer's instructions. Paired-end sequencing (2×101bp reads) was performed on the Illumina HiSeq2500 platform.

Culture enrichment

Following MetaSUB protocols, swabs were directly incubated with 7 mL of Brain Heart Infusion (BHI) broth (Thermo Scientific Microbiology, Cat. No. CM1135B) at 37˚C till turbidity was observed (14-16 hours for >95% of samples) or up to a maximum of 48 h. Culture tubes were centrifuged at 3200 g for 12 min and cell pellets were re-suspended with 550 µL of 1X PBS. Fifty microliters of re-suspended cultures were then plated on each of the 6 agar plates (BHI without antibiotics, Ampicillin 100 µg/mL, Streptomycin sulfate 100 µg/mL, Tetracycline 10 µg/mL, Kanamycin 50 µg/mL and Chloramphenicol 35 µg/mL). Plates were incubated overnight at 37˚C. Cells were harvested and were pelleted down by centrifugation at 8000 g for 15 min at 4 ˚C. Cells were stored at -80 ˚C until further processing.

DNA extraction from enrichment cultures

Frozen cells were thawed on ice and manually mixed with a wide bore tip. A volume of 30-50 µL of cells was re-suspended in 100 µL of 1X PBS, pH 7.4. Twenty microliters of suspended cells were added to 20 µL of metapolyzyme (6.7µg/µL) (Sigma Aldrich, Cat. No. MAC4L). The mixture was incubated at 35˚C for 4 hours. RNase treatment was carried out by adding 350 µL of 1X TE buffer and 10 µL RNase A (4mg/µL) and incubated on a rotator for 10 min at room temperature. DNA was extracted with Maxwell RSC cultured cells kit (Promega Pte. Ltd., Cat. No. AS1620). DNA was cleaned up and concentrated with 0.4X Agencourt AMPure XP beads. DNA purity and concentration was measured with Nanodrop and Qubit fluorometer. DNA integrity was assessed on a 0.5% agarose gel. DNA samples with the following quality measurements were selected for nanopore sequencing (DNA amount >400ng, A260/280: 1.8-2.0, A260/230: 1.7-3.0, Qubit:Nanodrop: 0.7-1.3, DNA integrity on 0.5% agarose gel: >1kb).

Nanopore library preparation

DNA was prepared with either 1D[2] sequencing kit (SQK-LSK308) or 1D sequencing kit (SQK-LSK108) together with native barcoding kit (EXP-NBD103) according to the 1D native barcoding genomic DNA protocol. Samples were multiplexed (9-12 samples per pool) and sequenced with either MIN106 or MIN107 flowcells on a GridION machine.

Taxonomic and resistome profiling with Illumina shotgun metagenomic data

Illumina shotgun metagenomic sequencing reads were processed using a Snakemake[72] pipeline (https://github.com/gis-rpd/pipelines/tree/master/metagenomics/shotgun-metagenomics). Briefly, raw reads were filtered to remove low quality bases using skewer[73] (v0.2.2; -q 3 -l 30 -n) and human reads were removed by mapping to the hg19 reference using BWA-MEM[74] (v0.7.10-r789; requiring 50% of the read to be covered). The remaining microbial reads were profiled with MetaPhlAn2[75] (v2.6.0) and SRST2[76] (v0.1.4; --min_coverage 90) for taxonomic and antibiotic resistance gene abundances, respectively. Microbial reads were also assembled using MEGAHIT[75] (v1.0.4-beta; default parameters) for comparison to nanopore assemblies. Site specificity score was computed as the z-score for the closest taxonomic profile for a sample (Bray-Curtis dissimilarity) among physically proximal sites (in the same

room/cubicle and same timepoint), compared to the distribution of Bray-Curtis dissimilarities across all samples of a site (e.g. all bed rails).

Preprocessing of nanopore sequencing data

Raw nanopore reads were basecalled with the latest version of the basecaller available at the point of sequencing (Guppy v0.5.1 to v2.0.5 or Albacore v2.3.1 to v2.3.3 for libraries that failed live basecalling). Basecalled nanopore reads were demultiplexed and filtered for adapters with Porechop (v0.2.3; https://github.com/rrwick/Porechop). Sequencing statistics were summarized using SeqKit[77] (v0.10.1). Reads were taxonomically classified with Kraken[78] (v0.10.5-beta) against the miniKraken database to assess the diversity of cultures on the plates.

Genome assembly and species assignment

Reads for each plate were assembled using Canu[79] (v1.3/v1.7; genomeSize=8m). Assembled contigs were mapped to the NCBI nt database with BLAST (v2.2.28), to identify the microbial species, plasmid or phage assignment according to the best BLAST hit (highest total reference coverage). Circular sequences were identified using MUMmer[80] (v3.23; -maxmatch -nosimplify, alignments <1kbp long or with identity <95% were filtered out) as recommended in the documentation for the Canu assembler (https://canu.readthedocs.io/en/latest/faq.html#my-circular-element-is-duplicated-has-overlap). Contigs assigned to the same species were binned into genomic bins. Metagenomic Illumina reads were used to polish Canu assemblies where feasible using Pilon[81] (v1.22, --fix indel). We noted that annotation errors were significantly reduced after polishing, and that genomic bins whose length was within 10% of the expected length met the criteria for high-quality genomes (completeness >90% and contamination <5% using CheckM[82]; v1.0.7 --reduced_tree). Genomic bins that met this criteria were therefore designated as high-quality and incomplete bins (<50% of expected length) were removed from further analysis. Genomes corresponding to novel species were identified as those with identity <95% or coverage <80% compared to known genomes (BLAST with nt) and a recent catalog of microbiome assembled genomes[39] (with Mash[83]). The genomes were hierarchically clustered (single-linkage with Mash distance[83]) to identify species-level clusters at 95% identity. Similarly, novel circular plasmids were identified by comparing to the PLSDB[43] database with Mash distance and identifying clusters at 99% identity (single-linkage) with no known sequence.

Analysis of antibiotic resistance gene combinations

Antibiotic resistance genes (ARGs) were annotated to contigs by mapping them to the SRST2 database (v3) with BLAST (best hit with >90% identity and >90% reference coverage). ARG combinations present in plasmid sequences were considered novel when they were not found in the reference databases (nt or PLSDB[43]). Assembled circular plasmids were clustered and annotated based on their best BLAST hit with identity >95% and >60% query coverage. A bipartite graph was constructed by connecting each plasmid cluster to ARGs found in it, with edge weights representing the frequency of occurrence (clusters with <5

representatives were excluded). For each species, an ARG co-occurrence graph was created for ARGs found in the assembled genomes by connecting the ARG pairs that were found within 10kbp on the same contig (discarding ARG pairs occurring less than 5 times). Each edge was weighted by the frequency of ARG pairs divided by the minimal frequency of the two ARGs. All ARG co-occurrence graphs were merged into a final co-occurrence multigraph. The graphs were visualized using Cytoscape (v3.7.1)[84].

Analysis of virulence factor and biocide resistance genes

Nanopore assemblies were aligned to virulence factors in the PATRIC database[85] (2018/12/20) with DIAMOND (v0.9.24.125; blastx --long-reads) and alignments with E-value >0.001 were filtered out. To identify biocide resistance genes, the assemblies were aligned to nucleotide sequences for the genes qacA (NC_014369.1) and qacC (NC_013339.1) with BLAST (>90% identity and >90% reference coverage).

Analysis of patient isolates and strain relationships

Raw reads corresponding to genomes for outbreak isolates[38,55,57,86] were downloaded and assembled using the Velvet assembler[87] (version 1.2.10) with parameters optimized by Velvet Optimiser (k-mer length ranging from 81 to 127), scaffolded with Opera[88] (version 1.4.1), and gapfilled with FinIS[89] (version 0.3). Outbreak genomes were jointly analyzed with high-quality genomes from the hospital environment. For each species of interest, a distance matrix (1-average nucleotide identity) was computed by first aligning each pair of genomes using nucmer (--maxmatch --nosimplify) and then estimating their identity based on SNPs using MUMmer's 'dna-diff' utility. The distance matrix was clustered hierarchically (single linkage) and strain groups were obtained at 99.9% identity. The cladogram of strains was visualized using the 'ggtree' R package[90]. Strain distributions across sites were visualized with the 'HiveR' R package (academic.depauw.edu/~hanson/HiveR/HiveR.html). The method in Brooks et al[26] was adapted to identify high confidence SNPs. Specifically, SNPs between patient isolates and environmental genomes were called using MUMmer's 'show-snps' function and SNPs within 40bp of each other were filtered out to avoid repeat associated artefacts. Rarefaction analysis for species, plasmids, strains and resistance genes was performed using the iNEXT R package[91].

Data and source code availability

All sequencing reads are available from the European Nucleotide Archive (ENA) under project PRJEB31632 (https://www.ebi.ac.uk/ena/data/view/PRJEB31632). Source code and associated data for reproducing the figures in this manuscript are available on GitHub under an MIT license (https://github.com/csb5/hospital_microbiome).

Acknowledgements

Funding for this work was provided by A*STAR, and we are grateful for support from NMRC (NMRC CGAug16C005). C.E.M. acknowledges support from the WorldQuant Foundation, the Bill and Melinda Gates Foundation (OPP1151054), and the Alfred P. Sloan Foundation (G-2015-13964). The funders had

no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We would like to thank Prof Jack Gilbert for insightful comments and feedback on this work.

Figure legends

Figure 1: Distinct ecological niches in the hospital environment for microbes and antibiotic resistance genes. (a) Principle coordinates analysis plot based on genus-level Bray-Curtis dissimilarity of taxonomic profiles indicating two distinct community types (denoted as CTA and CTB) for microbiomes from the hospital environment. (bi) Heatmap showing relative abundances (log-scale, Log_2RA) for differentially abundant genera between community types CTA and CTB (FDR-adjusted Wilcoxon p-value <0.01) (bii) Boxplots showing relative abundances for differentially abundant species between community types CTA and CTB (FDR-adjusted Wilcoxon p-value <0.01). (c) Principle coordinates analysis plot (genus-level Bray-Curtis dissimilarity) showing how much environmental microbiomes vary over time (lines connect two timepoints, 1 week apart) for different sites. (d) Radar plot showing the microbiome turnover index (fraction of taxa that are gained or lost across timepoints), human influence index (fraction of human reads) and site specificity index (uniqueness of site-specific taxonomic composition in relation to physically proximal sites). A positive site specificity index indicates a stronger site-specific microbiome composition signature. (e) Boxplots showing relative abundances of common nosocomial pathogens that were differentially abundant across ward types in high human-contact sites (FDR-adjusted Wilcoxon p-value <0.01). (f) Heatmap depicting the frequency of detection for beta-lactamases at different sites in hospital wards. Multiple carbapenemases and the mecA gene were detected as part of the resistomes that were primarily defined by the community types (CTA and CTB).

Figure 2: Genome-resolved characterization of nosocomial multidrug resistant strains that spread and persist at low relative abundances in the hospital environment. (a) Heatmap displaying the distinct median relative abundances (RA) of common nosocomial pathogens at different sites in hospital environments (PERMANOVA p-value <0.001). (b) Distribution of assembly contiguity statistics (N50=fragment size s.t. more than 50% of the genome is in longer sequences) for common nosocomial pathogens, highlighting the high genomic contiguity that was obtained (median N50 >1Mbp). (c) Dotplots highlighting that genomes can be rapidly obtained for several nosocomial pathogens despite their low relative abundances in corresponding environmental microbiomes (y-axis), through an enrichment and long-read metagenomic sequencing based protocol. Represented species are associated with more than 20 genome drafts in the overall database of 2347 genomes. (d) (Left panel) Phylogenetic relationships of *S. aureus* clonal strains detected in the hospital environment together with their antibiotic resistance profiles. (Right panel) Hive-map representation showing localization of *S. aureus* strains that spread (detected at 2 or more locations) and/or persist (detected at 2 timepoints) in the hospital environment. Orange lines represent occurrence at time point 1 while blue lines represent occurrence at timepoint 2. Line thickness represents the number of instances of such occurrences. Note that multidrug resistant strains such as s3, s2 and s1 tend to be more widely disseminated and persistent in the hospital environment.

Figure 3: Species distribution and genetic linkage of drug resistance genes in the hospital environment microbiome. Genetic-linkage network and clustering of antibiotic resistance genes based on 2347 microbial genomes and 5910 closed plasmids obtained from the hospital environment. (a) Multigraph of genetic linkage between antibiotic resistance genes, where edges indicate gene pairs found <10kbp apart in the genomes for a species (not including plasmids). Edges are colored according to species and line widths indicate frequency of occurrence of gene pairs (normalized by count for rarer gene) with frequencies >80% marked with solid lines. Solid line cliques in each species were used to define cassettes and assign names (**Suppl. File 5**) and the number after the colon sign indicates clique size. Genes are colored according to their respective antibiotic classes. (b) Circles represent different plasmid clusters (95% identity) and corresponding antibiotic resistance genes belonging to them are connected by edges and indicated by diamonds. Plasmid nodes are labelled based on a three letter short form for the host species and a number (e.g. Kpn1 for a *K. pneumoniae* plasmid) and the number after the colon sign indicates the number of representative of the plasmid family that were observed in the database. Edges are weighted by the frequency at which a gene is present in a plasmid and frequencies >80% are indicated with red solid lines. Genes and backbones are color-coded according to their respective antibiotic resistance classes and inferred host species for ease of reference.

Figure 4: Multi-species analysis of phylogenetic relationships between environmental and patient genomes. Phylogenies depicting the evolutionary relationships between clonal strains for (a) *Elizabethkingia anophelis* from a nosocomial outbreak in Singapore in 2012, (b) patient colonizing *Staphylococcus aureus* from a 2009-2011 surveillance study in Singaporean hospitals, (c) infectious *Acinetobacter baumannii* strains isolated from patients in two major Kuwaiti hospitals, and (d) recent globally disseminated multi-drug resistant *Staphylococcus epidermidis* lineages, together with environmental strains for corresponding species from this study. While subfigures a and b highlight the close relationships between strains circulating in Singaporean hospitals that are up to 8 years apart, subfigures c and d reveal the global dissemination of several lineages. Note that the matrices next to the trees indicate the antibiotic resistance profiles for corresponding clonal strains. For all species, multidrug resistant strains were more often observed to be shared between the environment and patients (Fisher's exact test p-value <0.01).

References

1 Fischer, M. M. & Bild, M. Hospital use of antibiotics as the main driver of infections with antibiotic-resistant bacteria - a reanalysis of recent data from the European Union. *bioRxiv* (2019).

2 O'Neill, J. *Tackling drug-resistant infections globally: final report and recommendations*, <https://amr-review.org/Publications.html> (2016).

3 Allegranzi, B. *et al.* Burden of endemic health-care-associated infection in developing countries: systematic review and meta-analysis. *Lancet* **377**, 228-241, doi:10.1016/S0140-6736(10)61458-4 (2011).

4 Magill, S. S. *et al.* Multistate point-prevalence survey of health care-associated infections. *N Engl J Med* **370**, 1198-1208, doi:10.1056/NEJMoa1306801 (2014).

5 Laxminarayan, R. *et al.* Antibiotic resistance-the need for global solutions. *Lancet Infect Dis* **13**, 1057-1098, doi:10.1016/S1473-3099(13)70318-9 (2013).

6 Longtin, Y., Sax, H., Allegranzi, B., Schneider, F. & Pittet, D. Videos in clinical medicine. Hand hygiene. *N Engl J Med* **364**, e24, doi:10.1056/NEJMvcm0903599 (2011).

7 Otter, J. A., Yezli, S., Salkeld, J. A. & French, G. L. Evidence that contaminated surfaces contribute to the transmission of hospital pathogens and an overview of strategies to address contaminated surfaces in hospital settings. *Am J Infect Control* **41**, S6-11, doi:10.1016/j.ajic.2012.12.004 (2013).

8 Weber, D. J., Rutala, W. A., Miller, M. B., Huslage, K. & Sickbert-Bennett, E. Role of hospital surfaces in the transmission of emerging health care-associated pathogens: norovirus, Clostridium difficile, and Acinetobacter species. *Am J Infect Control* **38**, S25-33, doi:10.1016/j.ajic.2010.04.196 (2010).

9 Hayden, M. K. *et al.* Reduction in acquisition of vancomycin-resistant enterococcus after enforcement of routine environmental cleaning measures. *Clin Infect Dis* **42**, 1552-1560, doi:10.1086/503845 (2006).

10 Datta, R., Platt, R., Yokoe, D. S. & Huang, S. S. Environmental cleaning intervention and risk of acquiring multidrug-resistant organisms from prior room occupants. *Arch Intern Med* **171**, 491-494, doi:10.1001/archinternmed.2011.64 (2011).

11 Mitchell, B. G. *et al.* An environmental cleaning bundle and health-care-associated infections in hospitals (REACH): a multicentre, randomised trial. *Lancet Infect Dis* **19**, 410-418, doi:10.1016/S1473-3099(18)30714-X (2019).

12 Sabino, R. *et al.* Analysis of clinical and environmental Candida parapsilosis isolates by microsatellite genotyping--a tool for hospital infection surveillance. *Clin Microbiol Infect* **21**, 954 e951-958, doi:10.1016/j.cmi.2015.06.001 (2015).

13 Obeidat, N., Jawdat, F., Al-Bakri, A. G. & Shehabi, A. A. Major biologic characteristics of Acinetobacter baumannii isolates from hospital environmental and patients' respiratory tract sources. *Am J Infect Control* **42**, 401-404, doi:10.1016/j.ajic.2013.10.010 (2014).

14 Price, J. R. *et al.* Transmission of Staphylococcus aureus between health-care workers, the environment, and patients in an intensive care unit: a longitudinal cohort study based on whole-genome sequencing. *Lancet Infect Dis* **17**, 207-214, doi:10.1016/S1473-3099(16)30413-3 (2017).

15 Faires, M. C. *et al.* A prospective study to examine the epidemiology of methicillin-resistant Staphylococcus aureus and Clostridium difficile contamination in the general environment of three community hospitals in southern Ontario, Canada. *BMC Infect Dis* **12**, 290, doi:10.1186/1471-2334-12-290 (2012).

16 Weingarten, R. A. *et al.* Genomic Analysis of Hospital Plumbing Reveals Diverse Reservoir of Bacterial Plasmids Conferring Carbapenem Resistance. *MBio* **9**, doi:10.1128/mBio.02011-17 (2018).
17 Conlan, S. *et al.* Single-molecule sequencing to track plasmid diversity of hospital-associated carbapenemase-producing Enterobacteriaceae. *Sci Transl Med* **6**, 254ra126, doi:10.1126/scitranslmed.3009845 (2014).
18 Lax, S. & Gilbert, J. A. Hospital-associated microbiota and implications for nosocomial infections. *Trends Mol Med* **21**, 427-432, doi:10.1016/j.molmed.2015.03.005 (2015).
19 Miller, R. R., Montoya, V., Gardy, J. L., Patrick, D. M. & Tang, P. Metagenomics for pathogen detection in public health. *Genome Med* **5**, 81, doi:10.1186/gm485 (2013).
20 Chen, C. H. *et al.* Bacterial diversity among four healthcare-associated institutes in Taiwan. *Sci Rep* **7**, 8230, doi:10.1038/s41598-017-08679-3 (2017).
21 Hewitt, K. M. *et al.* Bacterial diversity in two Neonatal Intensive Care Units (NICUs). *PLoS One* **8**, e54703, doi:10.1371/journal.pone.0054703 (2013).
22 Oberauner, L. *et al.* The ignored diversity: complex bacterial communities in intensive care units revealed by 16S pyrosequencing. *Sci Rep* **3**, 1413, doi:10.1038/srep01413 (2013).
23 Poza, M. *et al.* Exploring bacterial diversity in hospital environments by GS-FLX Titanium pyrosequencing. *PLoS One* **7**, e44105, doi:10.1371/journal.pone.0044105 (2012).
24 Lax, S. *et al.* Bacterial colonization and succession in a newly opened hospital. *Sci Transl Med* **9**, doi:10.1126/scitranslmed.aah6500 (2017).
25 Janda, J. M. & Abbott, S. L. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J Clin Microbiol* **45**, 2761-2764, doi:10.1128/JCM.01228-07 (2007).
26 Brooks, B. *et al.* Strain-resolved analysis of hospital rooms and infants reveals overlap between the human and room microbiome. *Nat Commun* **8**, 1814, doi:10.1038/s41467-017-02018-w (2017).
27 McIntyre, A. B. R. *et al.* Comprehensive benchmarking and ensemble approaches for metagenomic classifiers. *Genome Biol* **18**, 182, doi:10.1186/s13059-017-1299-7 (2017).
28 Minich, J. J. *et al.* KatharoSeq Enables High-Throughput Microbiome Analysis from Low-Biomass Samples. *mSystems* **3**, doi:10.1128/mSystems.00218-17 (2018).
29 Bertrand, D. *et al.* Nanopore sequencing enables high-resolution analysis of resistance determinants and mobile elements in the human gut microbiome. *bioRxiv* (2018).
30 Che, Y. *et al.* Mobile antibiotic resistome in wastewater treatment plants revealed by Nanopore metagenomic sequencing. *Microbiome* **7**, 44, doi:10.1186/s40168-019-0663-0 (2019).
31 van der Helm, E. *et al.* Rapid resistome mapping using nanopore sequencing. *Nucleic Acids Res* **45**, e61, doi:10.1093/nar/gkw1328 (2017).
32 Dancer, S. J. Controlling hospital-acquired infection: focus on the role of the environment and new technologies for decontamination. *Clin Microbiol Rev* **27**, 665-690, doi:10.1128/CMR.00020-14 (2014).
33 Kanamori, H., Weber, D. J. & Rutala, W. A. Healthcare Outbreaks Associated With a Water Reservoir and Infection Prevention Strategies. *Clin Infect Dis* **62**, 1423-1435, doi:10.1093/cid/ciw122 (2016).
34 Dancer, S. J. Importance of the environment in meticillin-resistant Staphylococcus aureus acquisition: the case for hospital cleaning. *Lancet Infect Dis* **8**, 101-113, doi:10.1016/S1473-3099(07)70241-4 (2008).
35 Sehulster, L., Chinn, R. Y., Cdc & Hicpac. Guidelines for environmental infection control in health-care facilities. Recommendations of CDC and the Healthcare Infection Control Practices Advisory Committee (HICPAC). *MMWR Recomm Rep* **52**, 1-42 (2003).

36 Kizny Gordon, A. E. *et al.* The Hospital Water Environment as a Reservoir for Carbapenem-Resistant Organisms Causing Hospital-Acquired Infections-A Systematic Review of the Literature. *Clin Infect Dis* **64**, 1435-1444, doi:10.1093/cid/cix132 (2017).
37 Ministry of Health Singapore. Guidelines for control and prevention of multi-drug resistant organisms (MDROs) in healthcare facilities. (2013).
38 Lee, J. Y. H. *et al.* Global spread of three multidrug-resistant lineages of Staphylococcus epidermidis. *Nat Microbiol* **3**, 1175-1185, doi:10.1038/s41564-018-0230-7 (2018).
39 Pasolli, E. *et al.* Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell* **176**, 649-662 e620, doi:10.1016/j.cell.2019.01.001 (2019).
40 Carattoli, A. Plasmids and the spread of resistance. *Int J Med Microbiol* **303**, 298-304, doi:10.1016/j.ijmm.2013.02.001 (2013).
41 Alekshun, M. N. & Levy, S. B. Molecular mechanisms of antibacterial multidrug resistance. *Cell* **128**, 1037-1050, doi:10.1016/j.cell.2007.03.004 (2007).
42 San Millan, A. Evolution of Plasmid-Mediated Antibiotic Resistance in the Clinical Context. *Trends Microbiol* **26**, 978-985, doi:10.1016/j.tim.2018.06.007 (2018).
43 Galata, V., Fehlmann, T., Backes, C. & Keller, A. PLSDB: a resource of complete bacterial plasmids. *Nucleic Acids Res* **47**, D195-D202, doi:10.1093/nar/gky1050 (2019).
44 Ho, C. M. *et al.* High rate of qacA- and qacB-positive methicillin-resistant Staphylococcus aureus isolates from chlorhexidine-impregnated catheter-related bloodstream infections. *Antimicrob Agents Chemother* **56**, 5693-5697, doi:10.1128/AAC.00761-12 (2012).
45 Duran, N., Temiz, M., Duran, G. G., Eryilmaz, N. & Jenedi, K. Relationship between the resistance genes to quaternary ammonium compounds and antibiotic resistance in staphylococci isolated from surgical site infections. *Med Sci Monit* **20**, 544-550, doi:10.12659/MSM.890177 (2014).
46 Fahimipour, A. K. *et al.* Antimicrobial Chemicals Associate with Microbial Function and Antibiotic Resistance Indoors. *mSystems* **3**, doi:10.1128/mSystems.00200-18 (2018).
47 Hartmann, E. M. *et al.* Antimicrobial Chemicals Are Associated with Elevated Antibiotic Resistance Genes in the Indoor Dust Microbiome. *Environ Sci Technol* **50**, 9807-9815, doi:10.1021/acs.est.6b00262 (2016).
48 Falagas, M. E., Roussos, N., Gkegkes, I. D., Rafailidis, P. I. & Karageorgopoulos, D. E. Fosfomycin for the treatment of infections caused by Gram-positive cocci with advanced antimicrobial drug resistance: a review of microbiological, animal and clinical studies. *Expert Opin Investig Drugs* **18**, 921-944, doi:10.1517/13543780902967624 (2009).
49 Urban, C., Mariano, N. & Rahal, J. J. In vitro double and triple bactericidal activities of doripenem, polymyxin B, and rifampin against multidrug-resistant Acinetobacter baumannii, Pseudomonas aeruginosa, Klebsiella pneumoniae, and Escherichia coli. *Antimicrob Agents Chemother* **54**, 2732-2734, doi:10.1128/AAC.01768-09 (2010).
50 Izadpanah, M. & Khalili, H. Antibiotic regimens for treatment of infections due to multidrug-resistant Gram-negative pathogens: An evidence-based literature review. *J Res Pharm Pract* **4**, 105-114, doi:10.4103/2279-042X.162360 (2015).
51 Werner, G., Hildebrandt, B. & Witte, W. Aminoglycoside-streptothricin resistance gene cluster aadE-sat4-aphA-3 disseminated among multiresistant isolates of Enterococcus faecium. *Antimicrob Agents Chemother* **45**, 3267-3269, doi:10.1128/AAC.45.11.3267-3269.2001 (2001).
52 Weinstein, L., Chang, T. W., Hudson, J. B. & Hartl, W. The concurrent use of sulfonamides and antibiotics in the treatment of infections: in vivo and in vitro studies of the effect of sulfonamide-antibiotic combinations on the emergence of drug resistance. *Ann N Y Acad Sci* **69**, 408-416 (1957).

53 Sundin, G. W. & Bender, C. L. Dissemination of the strA-strB streptomycin-resistance genes among commensal and pathogenic bacteria from humans, animals, and plants. *Mol Ecol* **5**, 133-143 (1996).

54 Monecke, S. *et al.* Microarray-based characterisation of a Panton-Valentine leukocidin-positive community-acquired strain of methicillin-resistant Staphylococcus aureus. *Clin Microbiol Infect* **12**, 718-728, doi:10.1111/j.1469-0691.2006.01420.x (2006).

55 Teo, J. *et al.* First case of E anophelis outbreak in an intensive-care unit. *Lancet* **382**, 855-856, doi:10.1016/S0140-6736(13)61858-9 (2013).

56 Johnson, R. C. *et al.* Investigation of a Cluster of Sphingomonas koreensis Infections. *N Engl J Med* **379**, 2529-2539, doi:10.1056/NEJMoa1803238 (2018).

57 Htun, H. L. *et al.* Methicillin-resistant Staphylococcus aureus colonisation: epidemiological and molecular characteristics in an acute-care tertiary hospital in Singapore. *Epidemiol Infect* **146**, 1785-1792, doi:10.1017/S0950268818001966 (2018).

58 Sievert, D. M. *et al.* Antimicrobial-resistant pathogens associated with healthcare-associated infections: summary of data reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2009-2010. *Infect Control Hosp Epidemiol* **34**, 1-14, doi:10.1086/668770 (2013).

59 Weiner, L. M. *et al.* Antimicrobial-Resistant Pathogens Associated With Healthcare-Associated Infections: Summary of Data Reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2011-2014. *Infect Control Hosp Epidemiol* **37**, 1288-1301, doi:10.1017/ice.2016.174 (2016).

60 Cornejo-Juarez, P. *et al.* The impact of hospital-acquired infections with multidrug-resistant bacteria in an oncology intensive care unit. *Int J Infect Dis* **31**, 31-34, doi:10.1016/j.ijid.2014.12.022 (2015).

61 Teerawattanapong, N. *et al.* A Systematic Review of the Burden of Multidrug-Resistant Healthcare-Associated Infections Among Intensive Care Unit Patients in Southeast Asia: The Rise of Multidrug-Resistant Acinetobacter baumannii. *Infect Control Hosp Epidemiol* **39**, 525-533, doi:10.1017/ice.2018.58 (2018).

62 Walker, J. T. *et al.* Hospital and community acquired infection and the built environment--design and testing of infection control rooms. *J Hosp Infect* **65 Suppl 2**, 43-49, doi:10.1016/S0195-6701(07)60014-0 (2007).

63 Ladner, J. T., Grubaugh, N. D., Pybus, O. G. & Andersen, K. G. Precision epidemiology for infectious disease control. *Nat Med* **25**, 206-211, doi:10.1038/s41591-019-0345-2 (2019).

64 Mahnert, A. *et al.* Man-made microbial resistances in built environments. *Nat Commun* **10**, 968, doi:10.1038/s41467-019-08864-0 (2019).

65 Caselli, E. *et al.* Impact of a probiotic-based hospital sanitation on antimicrobial resistance and HAI-associated antimicrobial consumption and costs: a multicenter study. *Infect Drug Resist* **12**, 501-510, doi:10.2147/IDR.S194670 (2019).

66 Goodman, E. R. *et al.* Impact of an environmental cleaning intervention on the presence of methicillin-resistant Staphylococcus aureus and vancomycin-resistant enterococci on surfaces in intensive care unit rooms. *Infect Control Hosp Epidemiol* **29**, 593-599, doi:10.1086/588566 (2008).

67 Lerminiaux, N. A. & Cameron, A. D. S. Horizontal transfer of antibiotic resistance genes in clinical environments. *Can J Microbiol* **65**, 34-44, doi:10.1139/cjm-2018-0275 (2019).

68 Buckner, M. M. C., Ciusa, M. L. & Piddock, L. J. V. Strategies to combat antimicrobial resistance: anti-plasmid and plasmid curing. *FEMS Microbiol Rev* **42**, 781-804, doi:10.1093/femsre/fuy031 (2018).

69 Chen, S., Larsson, M., Robinson, R. C. & Chen, S. L. Direct and convenient measurement of plasmid stability in lab and clinical isolates of E. coli. *Sci Rep* **7**, 4788, doi:10.1038/s41598-017-05219-x (2017).
70 McIntyre, A. B. R. *et al.* Single-molecule sequencing detection of N6-methyladenine in microbial reference materials. *Nat Commun* **10**, 579, doi:10.1038/s41467-019-08289-9 (2019).
71 The MetaSUB International Consortium. The Metagenomics and Metadesign of the Subways and Urban Biomes (MetaSUB) International Consortium inaugural meeting report. *Microbiome* **4**, 24, doi:10.1186/s40168-016-0168-z (2016).
72 Koster, J. & Rahmann, S. Snakemake--a scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520-2522, doi:10.1093/bioinformatics/bts480 (2012).
73 Jiang, H., Lei, R., Ding, S. W. & Zhu, S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* **15**, 182, doi:10.1186/1471-2105-15-182 (2014).
74 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
75 Truong, D. T. *et al.* MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat Methods* **12**, 902-903, doi:10.1038/nmeth.3589 (2015).
76 Inouye, M. *et al.* SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med* **6**, 90, doi:10.1186/s13073-014-0090-6 (2014).
77 Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS One* **11**, e0163962, doi:10.1371/journal.pone.0163962 (2016).
78 Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* **15**, R46, doi:10.1186/gb-2014-15-3-r46 (2014).
79 Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **27**, 722-736, doi:10.1101/gr.215087.116 (2017).
80 Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol* **5**, R12, doi:10.1186/gb-2004-5-2-r12 (2004).
81 Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963, doi:10.1371/journal.pone.0112963 (2014).
82 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**, 1043-1055, doi:10.1101/gr.186072.114 (2015).
83 Ondov, B. D. *et al.* Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol* **17**, 132, doi:10.1186/s13059-016-0997-x (2016).
84 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).
85 Wattam, A. R. *et al.* PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res* **42**, D581-591, doi:10.1093/nar/gkt1099 (2014).
86 Nasser, K. *et al.* Draft Genome Sequences of Six Multidrug-Resistant Clinical Strains of Acinetobacter baumannii, Isolated at Two Major Hospitals in Kuwait. *Genome Announc* **6**, doi:10.1128/genomeA.00264-18 (2018).
87 Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**, 821-829, doi:10.1101/gr.074492.107 (2008).
88 Gao, S., Bertrand, D., Chia, B. K. & Nagarajan, N. OPERA-LG: efficient and exact scaffolding of large, repeat-rich eukaryotic genomes with performance guarantees. *Genome Biol* **17**, 102, doi:10.1186/s13059-016-0951-y (2016).

89 Gao, S., Bertrand, D., Nagarajan, N. in *Algorithms in Bioinformatics* (ed Tang J. Raphael B.) 314-325 (Springer, Berlin, Heidelberg, WABI, 2012).

90 Yu, G., Lam, T. T., Zhu, H. & Guan, Y. Two Methods for Mapping and Visualizing Associated Data on Phylogeny Using Ggtree. *Mol Biol Evol* **35**, 3041-3043, doi:10.1093/molbev/msy194 (2018).

91 Hsieh, T. C., Ma, K. H. & Chao, A. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution* **7**, 1451-1456, doi:10.1111/2041-210x.12613 (2016).

Figure 1

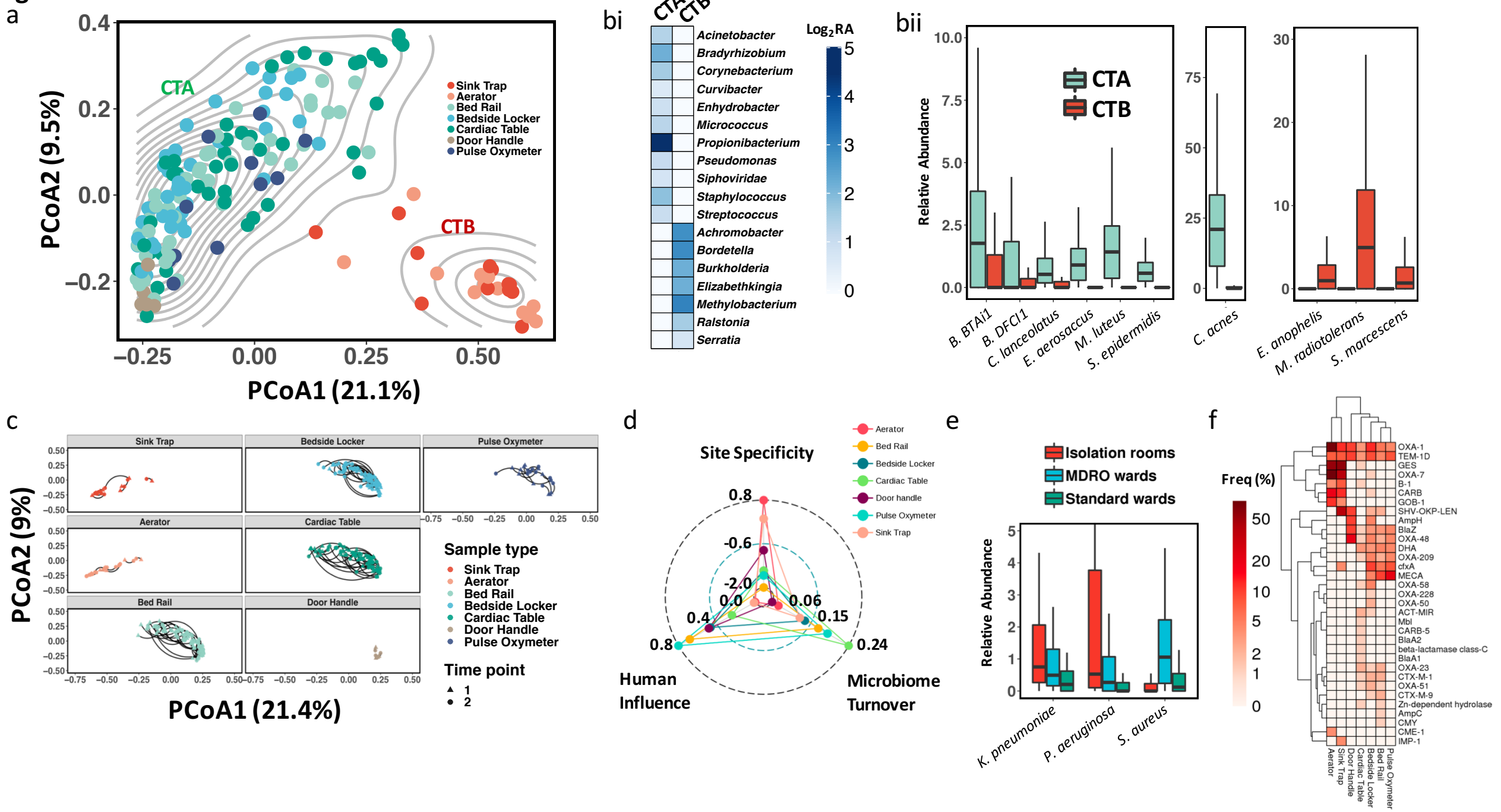
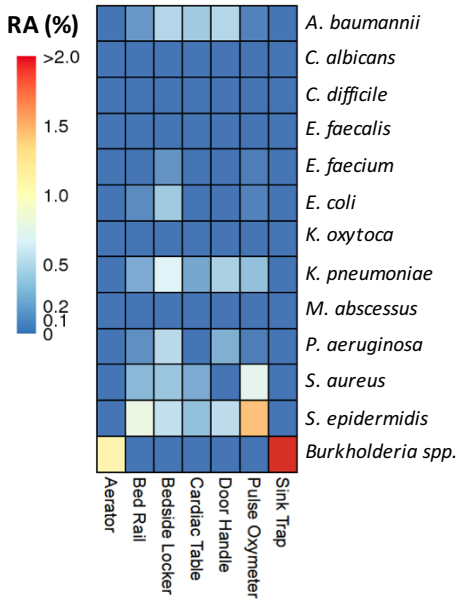
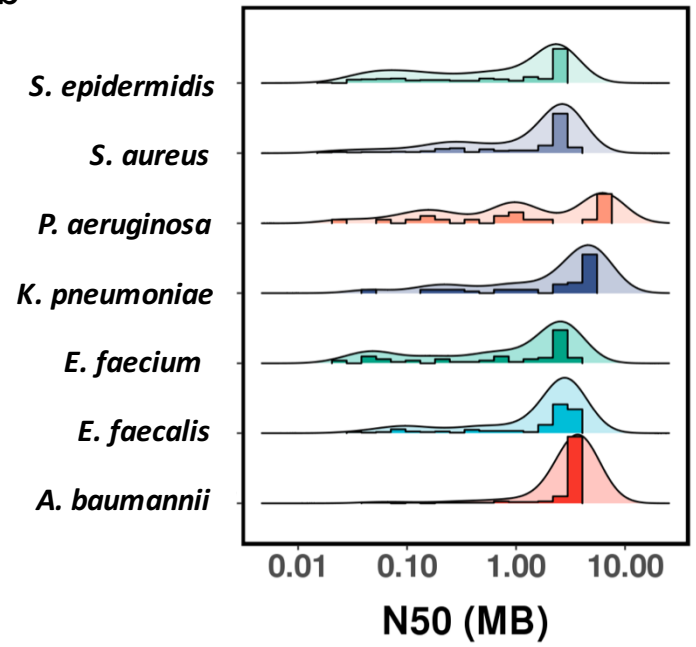

a
PCoA2 (9.5%)
PCoA1 (21.1%)
CTA
CTB
Sink Trap
Aerator
Bed Rail
Bedside Locker
Cardiac Table
Door Handle
Pulse Oxymeter
bi
CTA
CTB
Log2RA
Acinetobacter
Bradyrhizobium
Corynebacterium
Curvibacter
Enhydrobacter
Micrococcus
Propionibacterium
Pseudomonas
Siphoviridae
Staphylococcus
Streptococcus
Achromobacter
Bordetella
Burkholderia
Elizabethkingia
Methylobacterium
Ralstonia
Serratia
bii
Relative Abundance
CTA
CTB
B. BTAi1
B. DFCI1
C. lanceolatus
E. aerosaccus
M. luteus
S. epidermidis
C. acnes
E. anophelis
M. radiotolerans
S. marcescens
c
PCoA2 (9%)
PCoA1 (21.4%)
Sink Trap
Bedside Locker
Pulse Oxymeter
Aerator
Cardiac Table
Bed Rail
Door Handle
Sample type
Sink Trap
Aerator
Bed Rail
Bedside Locker
Cardiac Table
Door Handle
Pulse Oxymeter
Time point
1
2
d
Site Specificity
Aerator
Bed Rail
Bedside Locker
Cardiac Table
Door handle
Pulse Oxymeter
Sink Trap
Human Influence
Microbiome Turnover
e
Isolation rooms
MDRO wards
Standard wards
Relative Abundance
K. pneumoniae
P. aeruginosa
S. aureus
f
Freq (%)
OXA-1
TEM-1D
GES
OXA-7
B-1
CARB
GOB-1
SHV-OKP-LEN
AmpH
BlaZ
OXA-48
DHA
OXA-209
cfxA
MECA
OXA-58
OXA-228
OXA-50
ACT-MIR
Mbl
CARB-5
BlaA2
beta-lactamase class-C
BlaA1
OXA-23
CTX-M-1
OXA-51
CTX-M-9
Zn-dependent hydrolase
AmpC
CMY
CME-1
IMP-1
Aerator
Sink Trap
Door Handle
Cardiac Table
Bedside Locker
Bed Rail
Pulse Oxymeter

Figure 2

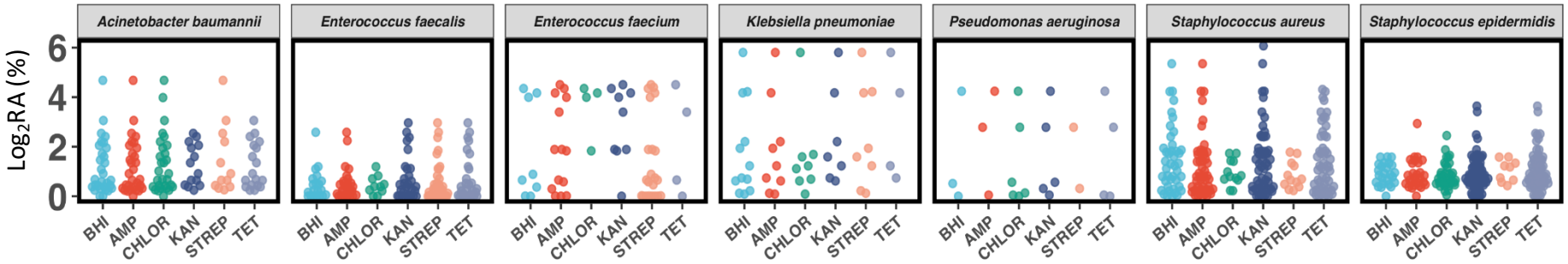
a



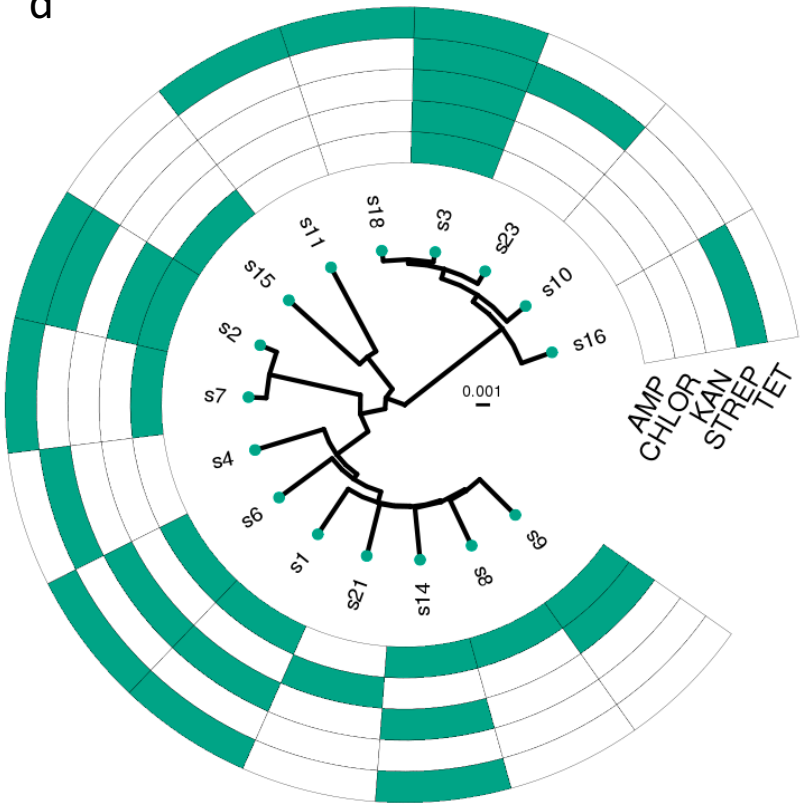
b



c



d



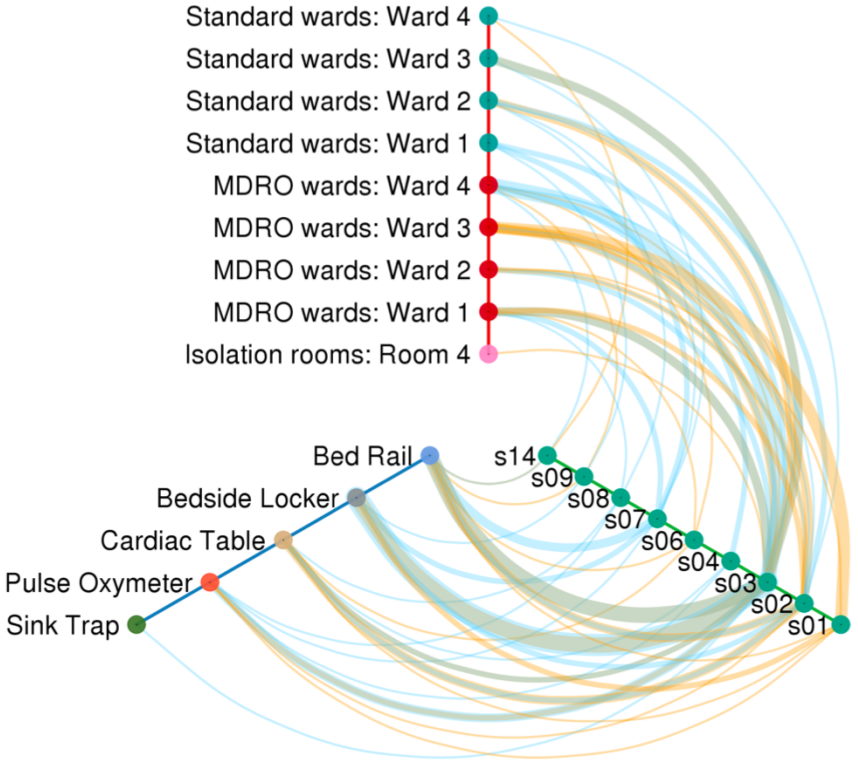


Figure 3

Microbial Species

- *A. baumannii*
- *E. cloacae*
- *E. faecalis*
- *E. faecium*
- *Enterococcus spp.*
- *K. pneumoniae*
- *S. aureus*
- *S. capitis*
- *S. epidermidis*
- *S. haemolyticus*
- *S. hominis*
- *S. warneri*

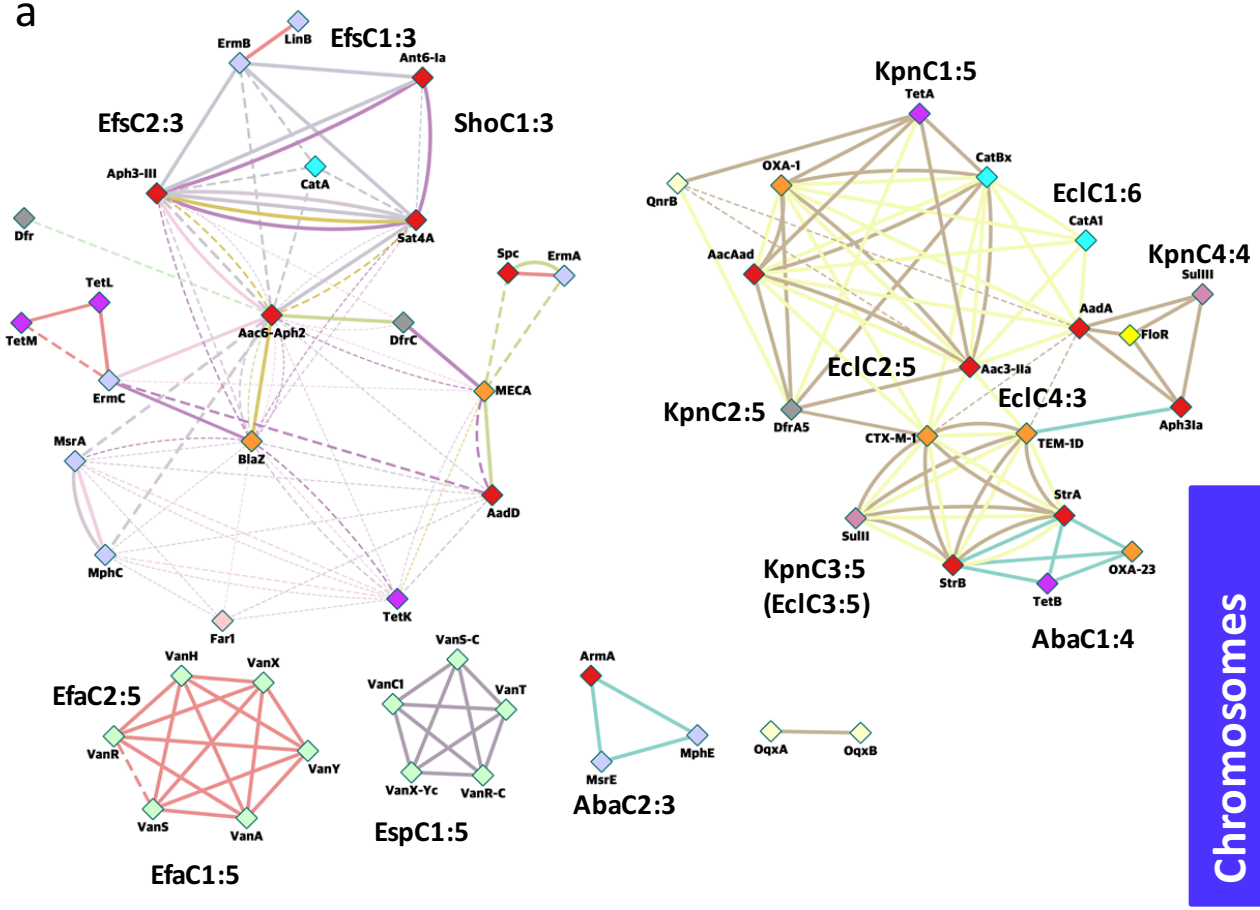
Plasmid Host Species

- Aba *A. baumannii*
- Asp *Acinetobacter spp.*
- Bth *B. thuringiensis*
- Cfr *C. freundii*
- Ecl *E. cloacae*
- Efa *E. faecium*
- Efs *E. faecalis*
- Kpn *K. pneumoniae*
- Sen *S. enterica*
- Sau *S. aureus*
- Sca *S. capitis*
- Sch *S. chromogenes*
- Sep *S. epidermidis*
- Sha *S. haemolyticus*
- Sho *S. hominis*
- Slu *S. lugdunensis*
- Spa *S. pasteuri*
- Ssa *S. saprophyticus*
- Ssi *S. simulans*

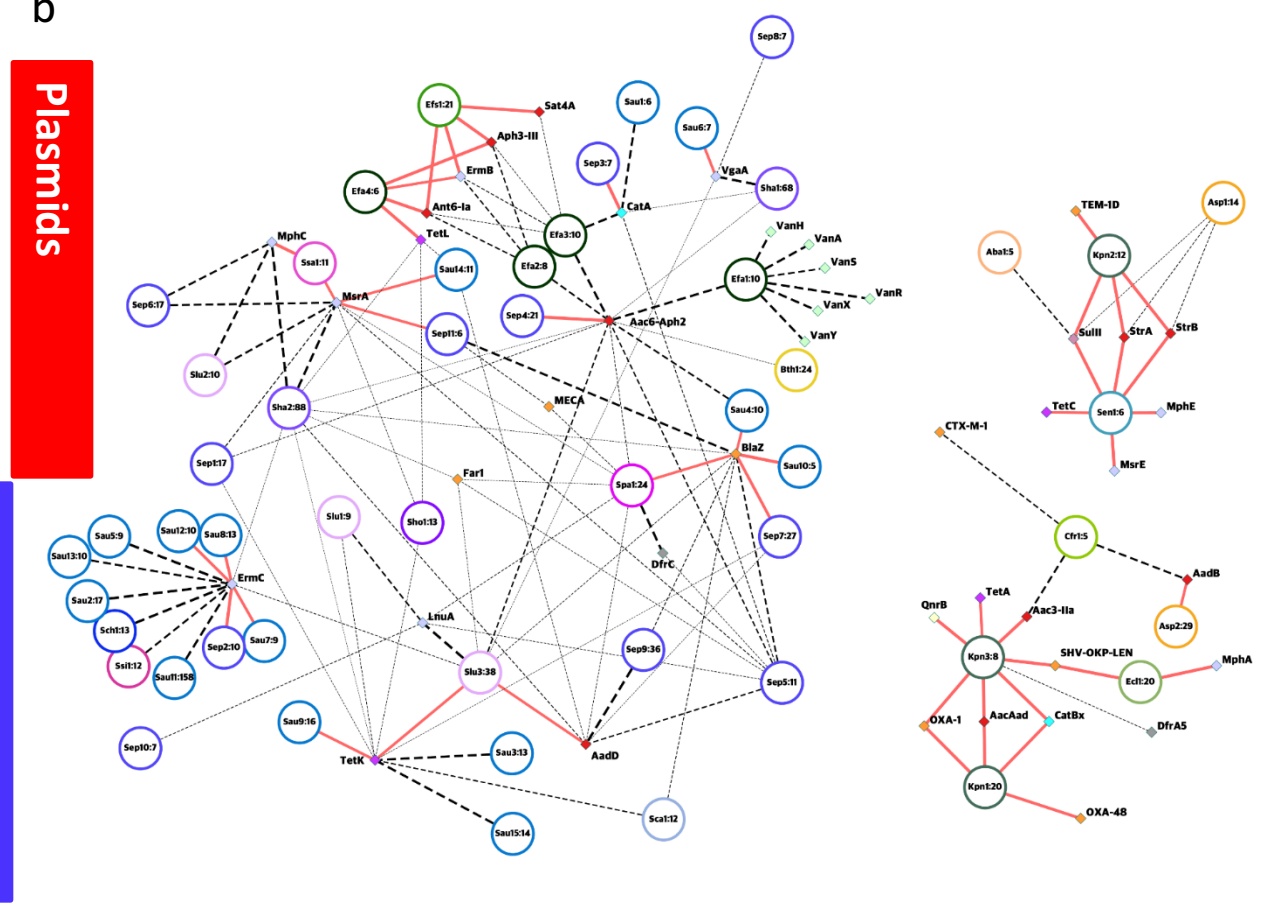
Antibiotic Resistance Class

- Aminoglycosides
- Beta-lactamases
- Fosfomycin
- Fluoroquinolones
- Glycopeptides
- Macrolide-Lincosamide-Streptogramin
- Phenicols
- Sulfonamides
- Tetracyclines
- Trimethoprim

a



b



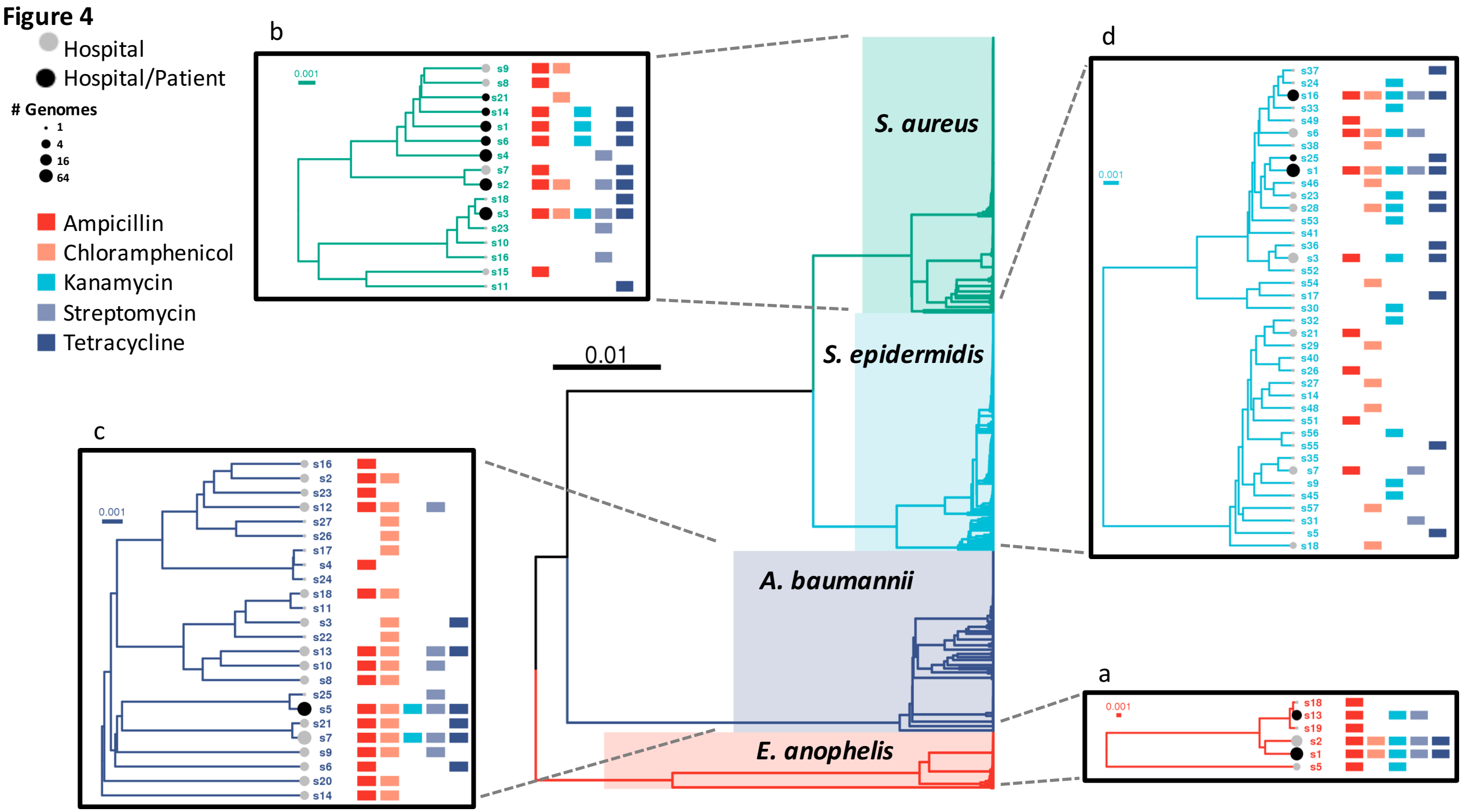


Figure 4
Hospital
Hospital/Patient
Genomes
1
4
16
64
Ampicillin
Chloramphenicol
Kanamycin
Streptomycin
Tetracycline
b
0.001
s9
s8
s21
s14
s1
s6
s4
s7
s2
s18
s3
s23
s10
s16
s15
s11
S. aureus
S. epidermidis
A. baumannii
E. anophelis
0.01
d
0.001
s37
s24
s16
s33
s49
s6
s38
s25
s1
s46
s23
s28
s53
s41
s36
s3
s52
s54
s17
s30
s32
s21
s29
s40
s26
s27
s14
s48
s51
s56
s55
s35
s7
s9
s45
s57
s31
s5
s18
c
0.001
s16
s2
s23
s12
s27
s26
s17
s4
s24
s18
s11
s3
s22
s13
s10
s8
s25
s5
s21
s7
s9
s6
s20
s14
a
0.001
s18
s13
s19
s2
s1
s5