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], Maanasa Natrajan[1], Michael Hongjie Zhang[1], Licheng Xu[2], Karrie Kwan Ki Ko[3,4,5], Eliza Xin Pei Ho[1], Tamar V. Av-Shalom[1], Jeanette Woon Pei Teo[6], Chiea Chuen Khor[1], MetaSUB Consortium, Swaine L. Chen[1], Christopher E. Mason[7], Oon Tek N[8], Kalisvar Marimuthu[8], Brenda Ang[8], Niranjan Nagarajan[1,9,#]

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

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

[3]Department of Microbiology, Singapore General Hospital, Singapore 169856, Singapore

[4]Department of Molecular Pathology, Singapore General Hospital, Singapore 169856, Singapore

[5]Duke-NUS Graduate Medical School, Singapore 169857, Singapore

[6]Department of Laboratory Medicine, National University Hospital, 5 Lower Kent Ridge Road, Main Building I, Singapore 119074, Singapore

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

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

[9]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 (428), common HAI-associated microbes (891) and transmissible drug resistance cassettes (1435) in a tertiary hospital environment based on a 3-timepoint sampling (1 week and >1 year apart) 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), phage sequences (1693) and closed plasmids (5910), a significant fraction of which (>60%) are not represented in current sequence databases. These high-quality assemblies and metadata enabled a rich characterization of resistance gene combinations, phage diversity, plasmid architectures, and the dynamic nature of hospital environment resistomes and their reservoirs. Phylogenetic analysis identified multidrug resistant strains as being more widely distributed 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 has refocused attention on infection prevention and control measures in hospitals[1]. It is estimated that if the spread of antibiotic resistance grows unchecked, it will cause millions of deaths worldwide, with 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 resistance genes (ARGs) 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 ARGs, 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 overall microbial community structure and interactions with the hospital built-environment to impact 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[28].

The availability of portable, real-time, long-read sequencing platforms presents new opportunities and challenges for pathogenome and resistome monitoring[29,30]. In this study, we combined extensive short-read shotgun metagenomic sequencing of multiple sites, wards and timepoints (428 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 ARGs 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), phage sequences (1693) and closed plasmids (5910), unveiling the significant under-characterized genetic diversity (>60% 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 strains are more likely to be widely distributed and stably colonizing across hospital sites. In addition, the genomes delineated common resistance gene combinations, phage diversity 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 ARGs

A diverse set of sites (n=7) of concern for infection control[31-33], and different room types distributed around the building (5 single-bed isolation rooms, 4 MDRO and 4 standard 5-bedded wards) were selected for initial 2-timepoint sampling (1 week apart) of a tertiary care hospital in Singapore (45 beds, 4% of total; 179 sites, 358 samples; **Fig. 1a** 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 excluded due to low biomass) and obtain taxonomic and functional profiles, as well as resistomes (**Fig. 1b**, **Suppl. File 2**; **Methods**). Controls, spike-ins and validation experiments were used to assess and account for the impact of kit contaminants on low biomass samples[34], with likely contaminant species being identified using batch and correlation analysis[34], and filtered from profiles (**Suppl. Note 1, Suppl. File 2**; **Methods**). Taxonomic profiles were visualized using a principal coordinates analysis (PCoA) plot to identify two distinct microbial community configurations in the hospital environment (CTA and CTB; **Fig. 2a**). While community type A (CTA) sites were more taxonomically diverse (Wilcoxon p-value<10^{-3}; **Suppl. Fig. S1**), and largely 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[16,32,36]. 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, though not all genera (e.g. *Pseudomona*, *Acinetobacte*r and *Ralstonia*) could be defined in these terms (**Fig. 2bi**). At the species level, key differences observed include the enrichment of common skin bacteria (e.g. *Cutibacterium acnes* and *Staphylococcus epidermidis*) and common biofilm-associated organisms in hospitals (e.g. *Elizabethkingia anophelis* and *Serratia marcescens*) in CTA and CTB sites respectively, though their occurrences were not mutually exclusive, indicating shared influences (**Fig. 2bii**). Comparison of hospital microbiome CTA and CTB sites, with similar sites in an indoor office environment (n=30, Office; **Suppl. File 1**, **Methods**) and other high-touch environmental microbiomes[37] (n=99, MetaSUB Singapore; **Suppl. File 1**), further highlighted the distinctness of hospital environments and their community types (**Suppl. Fig. S2**), and corresponding utility as an organizing principle for studying clinical impact[22,38].

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^{-3}; **Fig. 2c**). 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**). Analysis of a subset of sites (n=80) resampled at a third timepoint >1 year after the first two timepoints, confirmed the long-term

152 stability of community types for various sites (**Suppl. Fig. S3b**). The microbial composition of any
153 particular site is however expected to be influenced by a range of factors including abiotic conditions
154 (humidity, temperature, surface type), seeding from microbial reservoirs (human or environmental) and
155 exchange across sites. Based on the sequencing data, we computed various scores to quantify these
156 factors including a *microbiome turnover* index (fraction of taxa that are gained or lost across timepoints),
157 *human influence* index (fraction of human reads) and *site specificity* index (uniqueness of site-specific
158 taxonomic composition in relation to physically proximal sites), each of which exhibits significantly
159 correlated trends across the sampled timepoints (**Suppl. Fig. S4a**). The computed indices reinforce the
160 notion that CTB sites (primarily sink traps/aerators) tend to have stable compositions (low turnover) based
161 on site-specific biofilm configurations with limited human microbiome seeding (**Fig. 2d**). Sites associated
162 with community type A tend to have higher human influence (Wilcoxon p-value<10^{-15}) and microbiome
163 turnover indices (Wilcoxon p-value<10^{-4}), though they are not directly correlated, and tend to have weaker
164 site-specificity (Wilcoxon p-value<10^{-12}), concordant with a model where human activities (patient
165 discharge/admittance events) have a systemic role in shaping their compositions (**Fig. 2d**). Species that
166 were enriched in CTA sites were also observed in CTB sites (and *vice versa*), but tended to have higher
167 turnover in these cases (**Suppl. Fig. S4b**), with a few exceptions such as *Siphoviridae* having high
168 turnover in both CTA and CTB sites.

169 We observed that overall patterns of microbiome variability were consistent across ward types,
170 though isolation rooms exhibited lower variability across two timepoints (**Suppl. Fig. S5**). In line with
171 MDRO management guidelines in Singapore[39], patients colonized with Carbapenem-resistant
172 Enterobacteriaceae (CRE; e.g. *K. pneumoniae*) were typically warded in single-bed isolation rooms, while
173 methicillin resistant *Staphylococcus aureus* (MRSA) carrying patients were largely assigned to MDRO
174 wards. Analysis of differentially abundant common nosocomial pathogens (curated from
175 https://www.cdc.gov/hai/organisms/organisms.html and published literature[4,40]) across ward types
176 identified *K. pneumoniae* and *S. aureus* as being enriched in CTA sites for isolation rooms and MDRO
177 wards respectively, providing further evidence for the influence of patient microbiomes on CTA sites (**Fig.
178 2e**). Consistent with the observed taxonomic differences, community types A and B also harbored distinct
179 complements of ARGs in their resistomes (**Fig. 2f**, **Suppl. Fig. S6**). While some resistance genes were
180 frequently detected in CTB sites (e.g. ges, oxa-7 in **Fig. 2f**), CTA sites carried a wider diversity of genes
181 at lower frequencies. Despite the growing recent focus on CTB sites as drug resistance gene
182 reservoirs[16,36], some clinically important resistance genes such as oxa-23 (carbapenemase) and mecA
183 (methicillin resistance) were more frequently found in CTA sites, while others such as imp-1
184 (carbapenemase) and cme-1 (extended-spectrum beta-lactamase) were more common in CTB sites.
185 Different sites also exhibited distinct resistome patterns, e.g. specific tetracycline (tetC) and macrolide
186 (mphE) resistance genes were highly enriched only in aerators, while vancomycin resistance genes were
187 only seen in bedside lockers and bedrails, highlighting the importance of considering site and ward-
188 specific patterns when defining infection control and drug resistance mitigation strategies (**Suppl. Fig.**

189 **S6**). While CTB sites revealed a higher proportion of ARGs as being consistently detected across
190 timepoints compared to CTA sites (**Suppl. Fig. S7a**), overall hospital environment microbiomes exhibit
191 significantly higher abundance (>3× compared to MetaSUB Singapore and >12× compared to office sites,
192 Wilcoxon *p*-value<10^{-15} for both comparisons, **Suppl. Fig. S7b**) and diversity (Wilcoxon *p*-value<10^{-15};
193 **Suppl. Fig. S7c**) of ARGs compared to other high-touch urban environmental microbiomes. Even though
194 the presence of ARGs does not always translate to resistance phenotypes, these results further
195 underscore the distinctness of hospital environment microbiomes as ARG reservoirs[42].

196 **Quasi-metagenomics with nanopore sequencing reveals distribution of multidrug resistant**
197 **opportunistic pathogens in the hospital environment**

198 Based on Illumina shotgun metagenomic profiles we noted that common nosocomial pathogens were
199 generally present at very low relative abundances (median relative abundance <0.5%; **Fig. 3a**) in hospital
200 environments (even though this was higher than in other urban sites; **Suppl. Fig. S7d**), precluding
201 detailed genomic characterization and analysis of transmission patterns, resistance gene combinations
202 and plasmids. The distribution of common nosocomial pathogens exhibited site-specific patterns
203 (PERMANOVA p-value<0.001; **Fig. 3a**), in agreement with the distinct ecological niches observed in the
204 hospital environment (**Fig. 2a**, **b**), and indicated that enrichment cultures could capture a diverse set of
205 species. We exploited this observation to use a culturing, antibiotic selection (5 antibiotics) and
206 metagenomic nanopore sequencing approach (**Fig. 1b**) to obtain a large database of high-contiguity
207 genome assemblies (n=2347) from the hospital environment (median N50 >1Mbp; **Fig. 3b**, **Suppl. File 3**;
208 **Methods**), expanding substantially on genomic resources established by previous hospital environment
209 surveillance studies[16,17]. Overall, a large percentage of sites led to viable initial cultures (>95%), with
210 antibiotic selection still resulting in growth in >80% of plates (1495/1790), and >42% of sites resulting in
211 cultures for all 5 antibiotics. Control swabs led to no cultures (0/10) confirming that cultures were not likely
212 due to contamination (**Methods**), and further testing of isolates for antibiotic resistance confirmed that the
213 vast majority of strains in the cultures were likely antibiotic resistant (99%; **Suppl. Note 2**).

214 DNA was extracted from 1661 plates and sequenced on the GridION system to provide 535Mbp
215 of nanopore sequencing on average per sample (median read length >2.5kbp). Long read metagenomic
216 assembly enabled the frequent reconstruction of megabase-pair sized contigs (versus average N50
217 <5kbp for Illumina metagenomic assemblies) as the communities were largely simple (**Suppl. Fig. S8**,
218 **Fig. 3b**; **Methods**). An evaluation of these draft genomes based on evolutionarily conserved single-copy
219 genes confirmed that they were on average of high quality (completeness >99%, contamination <0.5%;
220 **Methods**). In total, we obtained genomes for 69 different known species from the hospital environment,
221 40% of which belonged to common nosocomial pathogens (**Methods**). Our results confirm the viability of
222 these species in different hospital environments and the ability to substantially enrich them for sequencing
223 and genome reconstruction compared to their acutely low abundances in the hospital environment
224 microbiome (median relative abundance = 0.68%, averaged across species; **Fig. 3c**). Large-scale

225 homology analysis of these genomes with public databases[43-45] also helped identify 13 (out of >80)
226 species-level clusters (from 11 different genera including *Bacillus*, *Pseudomonas* and *Staphylococcus*)
227 which do not have representatives from known species, highlighting the potential to obtain high-contiguity
228 and high-quality genomes for novel species in the hospital environment using this approach (**Methods**).
229 Rarefaction analysis of our data showed that >90% of the species and resistance gene diversity (>50% of
230 richness) that could be sampled for sites in this study was captured by our sample size (**Suppl. Fig. S9**),
231 while significant additional diversity remains to be captured for plasmids and HAI-associated strains
232 (**Suppl. Note 3**). In addition, this analysis revealed that future surveys of resistance gene families in the
233 hospital environment could be done with much fewer samples (~50), making more regular surveys
234 feasible, affordable and potentially actionable.

235 As plasmids and phages serve as an important medium for the evolution and spread of ARGs
236 and emergence of multidrug resistant bacteria[46-49], we identified sequences belonging to them in our
237 genomic database and further characterized them (**Methods**). In total, we recovered 696Mbp of putative
238 plasmid sequences (5910 closed/circular sequences and 493Mbp of linear fragments) and 63Mbp of
239 phage sequences (1693 sequences, of which 277 are circular), with most not present in existing
240 databases for complete plasmids[50] or phages[51] (>90%, 1505/1588 plasmid clusters and 501/557 phage
241 clusters; **Methods**) despite being commonly distributed in the hospital (**Suppl. Fig. S10**), highlighting the
242 unexplored genetic diversity in the hospital environment. Many closed plasmid sequences were >100kbp
243 long (>9%, n=536), repeat rich and present at low abundances in the microbiome, impeding their
244 characterization using Illumina shotgun metagenomics. In particular, we noted the presence of a group of
245 large mecA carrying plasmids that also contained antiseptic and disinfectant resistance genes (qacA or
246 qacC), a combination that is not represented in existing databases[50], but is in agreement with high
247 biocide-resistant rates for MRSA that has been reported in clinical settings[52,53]. Remarkably, one of these
248 plasmids has genes belonging to several additional ARG classes that have not been described before in
249 combination (e.g. dfrC, lnuA and aac6-Aph2), highlighting the value of closed plasmid sequences for
250 detailed characterization of novel resistance gene combinations.

251 The availability of a large collection of highly contiguous plasmid (closed) and chromosomal
252 (megabase-pair contigs) genome assemblies from the hospital environment allowed us to do genomic
253 relatedness (with environmental and patient strains) and structural analysis (common gene cassettes and
254 exchange across cassettes), as we discuss in the following sections. We first analyzed the evolutionary
255 relationships between genomes from the hospital environment, with previously used thresholds of
256 average nucleotide identity (ANI) to define strain-level[54] (>99.9% ANI), derivative[16] (>99.99% ANI) and
257 direct-transfer[26] (>99.999% ANI) genome clusters, and understand their spatio-temporal distribution. For
258 many species, a diverse set of genome clusters were observed to be distributed across the hospital
259 environment (from 6 clusters for *Pseudomonas aeruginosa* to 46 clusters for *S. epidermidis*; **Fig. 3d** and
260 **Suppl. Fig. S11**). A subset of them were frequently detected at multiple sites and ward types in the

261 hospital environment, and these were also significantly enriched for those detected in the 1st and 2nd
262 timepoints (Fisher's exact test p-value<1.5×10^{-9}). Even at the most stringent threshold (direct-transfer), a
263 substantial fraction of genomes observed in the 3rd timepoint (1.5 years later) clustered with genomes
264 from the first two timepoints for various species (92% for *E. anopheles* and as few as 5 SNPs; 20% for *S.*
265 *marcescens* and as few as 16 SNPs; 21% for *S. haemolyticus* and as few as 8 SNPs), emphasizing the
266 stability of environmental pathogenomes.

267 Overlaying antibiotic resistance information with these patterns, we noted an enrichment of multi-
268 antibiotic resistance among strains that are widely distributed across the hospital environment through
269 space and time (>2 antibiotics; Fisher's exact test p-value<3×10^{-8}; **Fig. 3d** and **Suppl. Fig. S12**). This
270 was also consistently observed across several common nosocomial pathogens in the hospital
271 environment (Fisher's exact test p-value: 1.6×10^{-2} for *S. aureus*, 2.3×10^{-3} for *S. epidermidis*, 3.7×10^{-3} for
272 *Enterococcus faecalis*, and 5.0×10^{-2} for *Acinetobacter baumannii*). For a subset of species (*S. aureus, S.*
273 *epidermidis* and *A. baumannii*) we additionally did Illumina sequencing to generate hybrid assemblies and
274 reliably detect derivative and direct-transfer relationships between genomes (**Methods**). Genomes that
275 were related across the first 2 timepoints based on these stringent criteria continued to be significantly
276 enriched for multidrug resistance (Binomial test p-value<10^{-5} for all species and both thresholds), and
277 were enriched in the third sampling timepoint (derivative clusters, Binomial test p-value: 0.028 for *S.*
278 *epidermidis* and 5.0×10^{-5} for *S. aureus*), highlighting the presence of stable, viable environmental
279 reservoirs for common HAIs and the need to further understand the mechanisms that contribute to
280 enrichment of multidrug resistant strains[55,56].

281 **Diversity and dynamics of ARG cassettes in the hospital environment**

282 With growing multidrug resistance, the specific combination of ARGs harbored in a strain is important to
283 know from a clinical perspective. In the context of the hospital environment, little is known about the
284 diversity of gene combinations and how often genes are exchanged across genomic cassettes and
285 plasmids. Analyzing our database of 2347 high-contiguity genomes and 5910 closed plasmids together
286 with existing genomic databases, we found that 34% of the ARG combinations observed in the hospital
287 environment were novel (255/752) (**Suppl. File 4**). Certain ARG combinations have obvious clinical
288 significance e.g. we noted the co-occurrence of the mecA gene (methicillin resistance) with fosB gene
289 (fosfomycin resistance) in several environmental *S. aureus* strains, an observation that is concerning
290 given the potential utility of fosfomycin for treating methicillin resistant *S. aureus* infections[57]. Notably, we
291 detected Enterobacteriaceae-associated genes that can confer resistance to gentamycin (e.g. aac3-IIa),
292 fosfomycin (e.g. fosA and fosA2) and colistin (e.g. mcr1), all last resort antibiotics for treating CRE
293 infections. Two Enterobacteriaceae-associated plasmids, one carrying a fosA gene and the other carrying
294 a mcr1 gene were obtained from the same bedside locker sample, highlighting the potential reservoir in
295 the hospital environment for emergence of co-resistance to colistin and fosfomycin. Another
296 Enterobacteriaceae-associated plasmid carried a rifampicin resistance gene (arr), a telling observation

297 given the growing interest to use rifampicin in combination treatments for a variety of gram negative (e.g.
298 *Acinetobacter baumannii*[58,59]) infections in hospitals.

299 We next identified common resistance gene pairs in close genomic proximity (<10kbp apart) to
300 identify chromosomal resistance gene cassettes that may serve as the unit of evolution, co-regulation and
301 gene exchange (**Methods**). Chromosomal cassettes for resistance genes were generally small (2-6
302 genes, average=3) and specific to a species, though two large cassettes carrying extended-spectrum
303 beta-lactamases were found to overlap very well for *K. pneumonia* and *Enterobacter cloacae* (KpnC1,
304 KpnC2 and EclC1, EclC2; KpnC3 and EclC3; **Fig. 4a** and **Suppl. File 5**). Selective pressure from the
305 rampant use of beta-lactams and plasmid-mediated transmission is likely to have contributed to the
306 sharing of these large cassettes across species. In general, cassettes for gram-negative species were
307 larger and more stable (solid lines to genes) while those for gram-positives were smaller with many
308 variably present members (dashed lines to genes). The largest shared cassette among gram-positives
309 (aminoglycoside-streptothricin resistance; ant6-Ia, sat4A and aph3-III) was seen in *Enterococcus* and
310 *Staphylococcus* species, but with no discernible signals of mobile elements[60]. While most genes were
311 observed to be stably present in cassettes (solid lines; except for a few e.g. tetK, far1, catA), exchange of
312 genes across cassettes were rarely observed (e.g. blaZ), indicating that chromosomal cassettes tend to
313 be relatively fixed.

314 Performing similar analysis for the closed plasmid sequences identified in this study, we first
315 clustered them into shared plasmid backbones and annotated them for their known hosts (identity ≥95%;
316 **Methods**). Analyzing ARGs in the context of these backbones, we found that many resistance genes are
317 variably present in backbones (93/143 genes), and for genes that are stably found in one backbone, a
318 high frequency are also variably present in another backbone (19/31 genes), highlighting the dynamic
319 nature of resistance genes combinations in the plasmids that we recovered in the hospital environment
320 (**Fig. 4b**). In this background, we observed a few gene combinations that were stably present in multiple
321 plasmid backbones, pointing to strong selection for co-existence. For example, the genes strA, strB
322 (streptomycin resistance) and sulII (sulfonamide resistance) co-occur in two distinct backbones (Sen1 and
323 Kpn2, sequence overlap <54%) as a signature from past co-administration of streptomycin and
324 sulphonamides for clinical use[61,62]. Similarly, while aminoglycoside resistance genes such as aadD and
325 aac6-Aph2 are widely and variably distributed across plasmid backbones, the genes ant6-Ia and aph3-III
326 are stably shared by two distinct backbones (Efa4 and Efs1, sequence overlap <11%) indicating that they
327 may provide synergistic resistance to aminoglycosides by catalysing different modifications. Interestingly,
328 we noted that genes that are widely distributed across plasmids (e.g. tetK, far1 and blaZ) can come
329 together in a novel, clinically relevant plasmid backbone (**Fig. 4b**; Slu3, with 38 sequences in our
330 database) as described for a cytotoxin producing MRSA strain[63]. While the previously isolated strain was
331 resistant to fusidic acid and tetracycline, but susceptible to erythromycin and clindamycin, we noted the
332 presence of a common plasmid backbone in our database (Sha2 with 88 sequences) that carries a novel

333 combination of resistance genes for all 4 antibiotics (**Fig. 4b** and **Suppl. File 4**). Similarly, we found that
334 resistance genes found in phages, such as aac6-Aph2 and far1, tend to be more widely present (**Fig. 4**
335 and **Suppl. Fig. S13a**), with evidence for recent phage-mediated dissemination of far1 across
336 *Staphylococcus* species (**Suppl. Fig. S13b**). In general, we observed that ARGs found in plasmids tend
337 to have more genes in close proximity (<10kbp apart) in chromosomes than chromosome-exclusive
338 genes (Wilcoxon test p-value=6×10^{-7}), characteristic of higher gene mobility and shuffling for plasmid-
339 associated genes. Thus the plasmid backbones seen in the hospital environment likely represent a more
340 plastic framework to generate diverse ARG combinations, many of which are not seen in genomic
341 cassettes (25%), despite the strong overlap in the complement of resistance genes that they harbour
342 (84% of plasmid genes).

343 **Hospital environment strains overlapping with patient isolates are globally disseminated and**
344 **enriched for multidrug resistance**

345 The availability of a large database of genomes from many species in the hospital environment, an
346 obvious hub for patient colonization, prompted us to ask the question: "How are environmental strains
347 related to patient colonizing strains?". To examine this relationship, we jointly analyzed environmental
348 strains and patient isolates for different species to construct phylogenetic trees (**Fig. 5**). We started by
349 analyzing Singaporean *E. anophelis* isolates from a 2012 HAI outbreak[64] (n=10) and an additional set of
350 patients isolates collected from 2009-2012 (n=52; **Fig. 5a**; **Methods)**. Strikingly, despite sampling from
351 different hospitals in Singapore and after a span of 5-8 years, patient associated genomes matched
352 environmental genomes with as few as 16 SNPs (s1; average nucleotide identity, ANI 99.9996%). The
353 environmental *E. anophelis* genomes in our studies primarily originated from hospital sinks, which as we
354 noted earlier tends to have a stable community, indicating that these strains may have originated from a
355 common reservoir upstream of water piping systems[54]. In addition, we noted that *E. anophelis* clusters
356 shared between patients and the environment were also detected at the 3rd timepoint sampling 1.5 years
357 later (ANI >99.999%, corresponding to direct-transfer), and exhibited resistance to more antibiotics (fold-
358 change=1.25×, one-sided Wilcoxon p-value=0.059).

359 Encouraged by the high-similarity match observed for *E. anophelis*, we analyzed *S. aureus*
360 genomes (n=221) obtained from a surveillance study of hospitalized patients in the same facility almost a
361 decade ago[65]. These strains were found to match 5 out of 17 strains that were obtained in the current
362 study, with genomes having as few as 39 SNPs and average nucleotide identity as high as 99.9985%
363 (**Fig. 5b**). The co-occurrence of patient and environmental genomes was found to be significantly
364 enriched in multidrug resistant clusters at the derivative genomes threshold (e.g. s1, s2 and s3; Binomial
365 test p-value<10^{-15}). These clusters were also enriched for genomes detected in the 3rd timepoint sampling
366 (Binomial test p-value<2×10^{-7}) with <60 SNPs (ANI 99.998%) distinguishing them from genomes in the 1st
367 and 2nd timepoints, highlighting the stability of antibiotic resistant derivative clusters in the hospital
368 environment.

369 To further extend our observations, we sequenced *A. baumannii* patient isolates (n=108) from a
370 hospital surveillance cohort in Singapore that was established >8 years ago. Patient isolate genomes
371 from this cohort were found to have high identity to some of our environmental genomes (99.995% ANI,
372 cluster s6), while being temporally separated by almost a decade. In addition, patient isolates that
373 overlapped with environmental genomes were enriched for multidrug resistance status (derivative
374 clusters; Binomial test p-value<4×10^{-3}). Extending our analysis to a regional context, we analyzed
375 genomes (n=36) for *A. baumannii* patient isolates from two major Kuwaiti hospitals[66] along with our
376 environmental genomes (**Fig. 5c**), and found that a derivative cluster shared between environmental and
377 Singaporean patient genomes exhibited resistance to all 5 antibiotics in this study and also included
378 Kuwaiti isolates at high identity (s6, ANI >99.99%). Our data thus highlights the presence of multidrug
379 resistant *A. baumannii* derivative clusters in hospital environments that are persistent, enriched in the
380 overlap with patient isolates and globally disseminated.

381 Similar patterns were observed recently for *S. epidermidis* lineages (ST2/ST2-mixed) which seem
382 to have disseminated across the globe within a short period (n=229)[40]. We confirmed the detection of
383 these rifampicin resistant[40] lineages in our data (**Fig. 5d**) with as few as 80 SNPs (ANI 99.997%) to our
384 hospital environment genomes. In addition, one other lineage (ST16) not known to be globally
385 disseminated (isolated from a patient sample in USA[40]) was also represented by a genome in our
386 database that is similar at the derivative threshold (ANI 99.991%). Finally, we again found that the overlap
387 between *S. epidermidis* patient isolates (surveillance samples from Austin Health in Australia[40]) and
388 environmental genomes from this study was enriched for multidrug resistance in derivative clusters
389 (Binomial test p-value<4×10^{-12}). Together with the observation that multi-antibiotic resistant strains are
390 persistent and widely distributed across the hospital (**Fig. 3d** and **Suppl. Fig. 11**), this data points to
391 selective advantages for multi-drug resistant organisms to persist and spread in hospital environments
392 and patients.

393 Discussion

394 While the importance of a well-designed built-environment for preventing infections in hospitals is well
395 appreciated[67], the use of shotgun metagenomic approaches for surveying microbial communities
396 established in functioning hospitals or medical environments remains underexplored[18]. A large baseline
397 survey such as the one conducted here can help provide a reference map (potentially visualized using 3D
398 tools, e.g. https://github.com/csb5/hospital_microbiome_explorer) that can then be updated based on
399 periodic scans whose frequency and locations can be informed by the initial survey. For example, the
400 turnover score and specificity of a site can help determine whether and how frequently it should be
401 sampled. Variations in human influence scores could also help fine-tune cleaning practices, and the
402 distribution of specific opportunistic pathogens of concern could be informative in an outbreak setting for
403 infection control. As genomics-guided infection control is further refined, the knowledge gained could then
404 feed back into better hospital designs. With further improvements in cost and ease of short-read

405 sequencing, we anticipate that hospital-wide surveys will be increasingly feasible, will provide valuable
406 information for infection control, and eventually be part of routine practice.

407 The two microbial community types observed in this study, highlight the distinct niches found in
408 hospital environments compared to other urban environments, and provide an organizing principle for
409 further study. For example, while many pathogens are substantially enriched in the hospital environment,
410 this appears to be more a feature of CTA sites which also have a greater diversity of ARGs (**Suppl. Fig.**
411 **S7c**, **d**). From a clinical perspective, recent infection control studies have primarily focused on sites
412 associated with wash areas (sinks, showers etc.; CTB sites), as in many cases the presumptive
413 pathogens for an outbreak have been isolated from such sites[54]. This focus on community type B sites is
414 in line with them being composed of biofilm-forming bacteria and having the ability to harbor viable
415 reservoirs for extended periods of time (e.g. in the plumbing system). Our data however suggests that
416 many medically-relevant species (e.g. *K. pneumoniae*), ARGs (e.g. carbapenemases such as oxa-23)
417 and resistance plasmids (present in >85% of CTA sites) are more frequently harbored in community type
418 A sites. While CTA sites have higher microbiome turnover, the ability to detect very similar strains over
419 long durations of time indicates that they may also have distinct reservoirs (e.g. in ventilation or air-
420 conditioning ducts), and that culture-based screening protocols may bias against sites that have lower
421 biomass or are variably colonized. Combining the strengths of metagenomic sequencing and culturing
422 may therefore be needed to more systematically explore the source of outbreaks.

423 Large-scale genomic studies of nosocomial pathogens through strain isolation can be a laborious
424 and time-consuming process, while as shown here, direct shotgun metagenomics will often not provide
425 detailed genetic information for species present at low relative abundances. The intermediate approach
426 proposed here attempts to address both issues. The use of culture-based enrichment allows us to
427 effectively shift the distribution away from abundant species such as *C. acnes*, and towards nosocomial
428 pathogens that are typically at low relative abundances (e.g. *K. pneumoniae*, *S. aureus* and *A.*
429 *baumannii*), while also allowing for functional selection such as for antibiotic resistance as we did here. Its
430 combination with long-read metagenomic sequencing is then powerful as it allows us to directly recover
431 high-contiguity genomes (chromosomal, plasmid and phage) without a tedious isolation step. With further
432 automation (e.g. in library preparation) this workflow is suitable for high-throughput analysis and
433 amenable for wider surveillance, in line with the vision to achieve precision epidemiology for infectious
434 diseases[68]. Future improvements in nanopore sequencing throughput, the ability to work with lower DNA
435 input, and use in remote settings, could help accelerate time-to-answer by reducing the culturing period,
436 or eliminating it altogether.

437 The availability of a database of high-contiguity assemblies, with >8Gbp of sequence, 2347
438 genomes, 1693 phage sequences and 5910 closed plasmids, provides a unique resource for studying
439 transmission patterns of strains and diversity of resistance gene cassettes across species in the hospital
440 environment. Based on this resource, we observed that multidrug resistant strains are preferentially

441 distributed and persistent in hospital environments across a range of species. This represents a
442 worrisome pattern, with several possible explanations that deserve further investigation. One plausible
443 scenario is that hospital environments are repeatedly seeded by multidrug resistant strains that are
444 preferentially persistent in the community (humans or environment). This scenario seems less likely to be
445 generally true given that some of the species where we observe this pattern are rarely found in humans
446 (e.g. *E. anophelis*, *A. baumannii*), and based on the observations here that other urban environmental
447 microbiomes are distinct from hospital environments in terms of composition, frequency at which they
448 harbor species associated with nosocomial infections, and diversity and abundance of ARGs.
449 Nevertheless, the data in this study does not rule out the possibility that urban environments, (i) harbor
450 pathogens and resistant strains at lower abundances compared to hospitals, and (ii) that antibiotic
451 resistant strains are also widespread and persistent in these environments. Another potential hypothesis
452 is that hospital cleaning measures may be selecting for more antibiotic-resistant organisms[69,70], a model
453 that is also supported by the presence of multiple copies of disinfectant resistance genes in the widely
454 distributed, multidrug resistant *S. aureus* strains in our study. Comparisons with surveys from built-
455 environments that are intensively cleaned but do not house patients (e.g. operating rooms) or are not
456 intensively cleaned but have high patient traffic (e.g. waiting areas in clinics) might help explore this
457 model further. Studies across wards or hospitals with different cleaning protocols could also be
458 illuminating for understanding how ARG reservoirs in hospitals can be shaped by infection control
459 practices[71,72].

460 Despite its importance as the epicenter for the battle against growing antibiotic resistance[1],
461 hospital environments have received relatively little attention compared to studies on agricultural and
462 animal farms[18]. Our analysis highlights that hospital environments harbor significant uncharacterized
463 genetic diversity in terms of microbial species (13 novel species) and ARG combinations (255). This
464 reservoir can be the origin of new opportunistic infections and serve as fertile ground for evolution of ARG
465 combinations that are of clinical concern (e.g. colistin and fosfomycin resistance). This can be further
466 facilitated by the presence of resistance gene carrying plasmids that serve as a vehicle for gene transfer
467 across species[73], and were commonly found in the hospital environment (n=1400). The development and
468 use of anti-plasmid agents[74,75] could thus be a complimentary strategy to reduce the spread of ARGs
469 through hospital environments.

470 While other studies have focused on patient isolates[76], the genetic relatedness between
471 environmental and patient-colonizing strains in hospitals is important to study for understanding the
472 potential risk that environmental strains carry for causing infections[24,26]. For strains that are contemporary
473 and co-located, high genetic relatedness between a subset of environmental and patient strains is
474 expected. The observation that highly similar genomes can be obtained despite being temporally
475 separated by >8 years suggests that large reservoirs of nosocomial multidrug resistant strains are being
476 maintained with limited diversification. The identification and elimination of these reservoirs may help

477 reduce the incidence of corresponding infections and the risk from maintenance of antibiotic resistance
478 properties. Another distinct observation is the high genetic similarity between MDRO strains observed in
479 Singaporean hospitals and those obtained from patients around the world. The consistency of these
480 patterns across species emphasizes the global dissemination of newly emerged MDRO lineages and the
481 role of hospital environments in this deserves investigation, in conjunction with global environmental
482 metagenomic datasets[37].

483 Taken together, our data points to selective advantages for multidrug resistant organisms to
484 persist and spread in hospital environments (**Fig. 3d** and **Suppl. Fig. S11**) and be shared with patients
485 (**Fig. 5**). The *S. aureus* derivative clusters that persisted in the hospital environment across time are
486 enriched in virulence factors (1.5×, one-sided Wilcoxon p-value=0.015) and have as many as three copies
487 of disinfectant resistance genes[52,53] (**Fig. 3d**). This may enable them to successfully colonize both
488 hospital environments and patients, and thus be readily transferred between them. This points to a
489 vicious cycle where disinfectant resistance, antibiotic resistance and virulence may in turn be selected for,
490 enriching for strains that are adept at colonizing both niches in the presence of depleted microbial
491 competition, and offers an explanation for the high incidence of multidrug resistant HAIs worldwide
492 despite increased surveillance and aggressive cleaning measures in hospitals[77-80].

493 # Methods

494 **Sample collection and storage**

495 Environmental swabs were collected from Tan Tock Seng Hospital (TTSH), a major tertiary care hospital
496 with >2000 patient visits every day that serves as the national referral center for communicable diseases
497 in Singapore. Sampling was carried out in November 2017 and in May 2019. Samples for the first
498 timepoint being collected in 2 days and the second timepoint in 3 days, with 1 week separating the
499 timepoints. The third timepoint was 1.5 years later, collected in 4 days across 2 weeks. Samples were
500 collected from isolation rooms (1 bed, typically warding CRE colonized patients), MDRO wards (5 beds,
501 typically warding MRSA colonized patients) and standard wards (5 beds), at 7 different sites, including
502 aerator, sink trap, bed rail, bedside locker, cardiac table, pulse oxymeter and door handle (**Fig. 1** and
503 **Suppl. File 1**). Standard cleaning protocols at TTSH require that high-touch areas and sinks are cleaned
504 daily (chlorine 5000ppm and cleaning detergent, respectively; excluding beds that are cleaned upon
505 patient discharge). Isohelix DNA Buccal Swabs (SK-4S) were used for sampling carried out based on
506 MetaSUB protocols[37]. Briefly, a total of 4 swabs were collected with 1 swab (for culturing) moistened with
507 1X phosphate buffer saline (PBS, pH 7.2) and 3 swabs (2 for metagenomic DNA isolation, 1 for storage)
508 moistened with DNA/RNA shield (Zymo Research, Cat. No. ZYR.R1100-250). Swabbing was performed
509 for 2 min in each site and stored in respective storage liquids (i.e. 1X PBS, pH 7.2 or Zymo DNA/RNA
510 shield). Swabs in PBS were placed on ice and sent for culturing while the other swabs were transported
511 back at room temperature to the laboratory and stored at -80˚C. In total, 1752 swabs were collected from
512 179 sites in the hospital at 3 timepoints, representing 438 unique samples. Swabs were also collected
513 from an office environment (Genome Institute of Singapore) with sites selected to approximately match
514 those collected in the hospital (aerator, sink trap, chair handle, office desk, keyboard and door handle;
515 n=30; **Suppl. File 1**). MetaSUB Singapore samples were collected from high-touch surfaces in different
516 parts of the city and analyzed based on MetaSUB protocols as described in Danko et al[37] (n=99; **Suppl.**
517 **File 1**).

518 **DNA extraction from swabs**

519 DNA was extracted from swabs using a bead-beating and automated DNA purification system. Briefly,
520 300 µL of lysis buffer was added to lysing matrix E tubes (MP Biomedicals, Cat. No. 116914500).
521 Samples were homogenized using the FastPrep-24 instrument at 6 m/s for 40s prior to centrifugation at
522 maximum speed for 5 min. The supernatant was treated with Proteinase K (Qiagen Singapore Pte. Ltd,
523 Cat. No. 19133) for 20 min at 56˚C before DNA was purified with Maxwell® RSC Blood DNA Kit
524 (Promega Pte. Ltd., Cat. No. AS1400). DNA concentration was quantified using Qubit® 2.0 fluorometer,
525 prepared with Qubit® dsDNA HS Assay Kit (Life Technologies Holdings Pte. Ltd., Cat. No. Q32854). DNA
526 extraction from backup swabs was carried out for samples with insufficient amounts of DNA. Samples that
527 still had less than 0.5ng of DNA were excluded for library preparation (10/438).

528 **Illumina library preparation**

529 Extracted DNA was sheared using Adaptive Focused Acoustics™ (Covaris) with the following
530 parameters: 240s, Duty Factor: 30, PIP: 450, 200 cycles/burst. Metagenomic libraries for the first two
531 timepoints were prepared with NEBNext® Ultra™ DNA Kit (New England Biolabs, Cat. No. E7370)
532 according to manufacturer's instructions. Paired-end sequencing (2×101bp reads) was performed on the
533 Illumina HiSeq2500 platform. For the third timepoint, metagenomic libraries were prepared using
534 NEBNext® Ultra™ II DNA Library Prep Kit (New England Biolabs, Cat. No. E7645) according to
535 manufacturer's instructions. Paired-end sequencing (2×151bp reads) was performed on the Illumina
536 HiSeq4K platform.

537 **Culture enrichment**

538 Following MetaSUB protocols, swabs were directly incubated with 7 mL of Brain Heart Infusion (BHI)
539 broth (Thermo Scientific Microbiology, Cat. No. CM1135B) at 37˚C till turbidity was observed (14-16 h for
540 >95% of samples) or up to a maximum of 48 h. Culture tubes were centrifuged at 3200 g for 12 min. For
541 the first two time points, cell pellets were re-suspended with 550 µL of 1× PBS while the cell pellets for the
542 third time point were re-suspended with 1 mL of 1× PBS. Fifty microliters of re-suspended cultures were
543 then plated on each of the 6 agar plates (BHI without antibiotics, Ampicillin 100 µg/mL - AMP,
544 Streptomycin sulfate 100 µg/mL - STREP, Tetracycline 10 µg/mL - TET, Kanamycin 50 µg/mL – KAN and
545 Chloramphenicol 35 µg/mL - CHLOR) and incubated overnight at 37˚C. Cells were harvested by a plate
546 sweep and were pelleted down by centrifugation at 8000 *g* for 15 min at 4 ˚C for the first two time points.
547 For the third time point, a loopful of harvested cells was streaked out on an antibiotic-free BHI plate to
548 obtain single colonies for whole genome sequencing. Plates were only excluded if no cells were growing
549 on the plates or when the growth was insufficient to generate enough DNA for sequencing.

550 **DNA extraction from enrichment cultures**

551 Frozen cells were thawed on ice and manually mixed with a wide bore tip. A volume of 30-50 µL of cells
552 was re-suspended in 100 µL of 1× PBS, pH 7.4. Twenty microliters of suspended cells were added to 20
553 µL of metapolyzyme (6.7 µg/µL) (Sigma Aldrich, Cat. No. MAC4L). The mixture was incubated at 35˚C for
554 4 h. RNase treatment was carried out by adding 350 µL of 1× TE buffer and 10 µL RNase A (4 mg/µL)
555 and incubated on a rotator for 10 min at room temperature. DNA was extracted with Maxwell® RSC
556 cultured cells kit (Promega Pte. Ltd., Cat. No. AS1620). DNA was cleaned up and concentrated with 0.4×
557 Agencourt AMPure XP beads (Beckman Coulter, Cat. No. A63882). DNA purity and concentration was
558 measured with Nanodrop and Qubit® fluorometer. DNA integrity was assessed on a 0.5% agarose gel.
559 DNA samples with the following quality measurements were selected for nanopore sequencing (DNA
560 amount >400 ng, A260/280: 1.8-2.0, A260/230: 1.7-3.0, Qubit:Nanodrop: 0.7-1.3, DNA integrity on 0.5%
561 agarose gel: >1 kb). The Qubit:Nanodrop ratio was used to estimate and control the amount of single-
562 stranded DNA in the sample, and ensure successful nanopore sequencing.

563 **Collection and testing of bacterial isolates from patients**

564 *Elizabethkingia anophelis* isolates (n=52) were obtained from consecutive positive blood cultures and
565 respiratory samples collected in a 3-year period (2009-2012) at the National University Hospital in
566 Singapore (DSRB reference number 2017/00879). *Acinetobacter baumannii* complex isolates (n=108)
567 were consecutively obtained from all clinical specimens (including blood, tissue, respiratory and urine
568 samples) sent for routine bacterial culture between February 2009 and May 2009 at the Singapore
569 General Hospital Diagnostic Bacteriology Laboratory (de-identified and archived and hence no IRB
570 approval was required). Antibiotic susceptibility testing for *E. anopheles* isolates was done with 13
571 antimicrobial agents (cefotaxime, ceftazadime, cefepime, imipenem, meropenem, ampicillin-sulbactam,
572 piperacillin/tazobactam, tigecycline, gentamicin, nalidixic acid, ciprofloxacin, levofloxacin and
573 trimethoprim/sulfamethoxazole) using Etest strips (bioMérieux). Minimum inhibitory concentrations (MICs)
574 were interpreted according to the Clinical and Laboratory Standards Institute (CLSI) guidelines for non-
575 Enterobacteriaceae Gram-negative bacilli (Performance standards for antimicrobial susceptibility testing,
576 M100-S22, CLSI 2012; **Suppl. File 6**). Antibiotics that all strains were resistant to were excluded from
577 statistical analysis. Antibiotic susceptibility testing for *A. baumannii* complex isolates was done with 11
578 antimicrobial agents (ampicillin/sulbactam, piperacillin/tazobactam, cefepime, imipenem, gentamicin,
579 amikacin, ciprofloxacin, levofloxacin, trimethoprim/sulfamethoxazole, minocycline and polymixin B).
580 Polymixin B susceptibility testing was performed using Etest strips (bioMérieux) and disk diffusion was
581 performed for all other antimicrobial agents. Polymixin B minimum inhibitory concentrations (MICs) and
582 zone diameters for all other tested agents were interpreted in accordance to Clinical and Laboratory
583 Standards Institute (CLSI) breakpoints for *Acinetobacter* spp. (Performance standards for antimicrobial
584 susceptibility testing, M100-S19, CLSI 2009; **Suppl. File 6**). Multidrug resistant status for patient isolates
585 was defined according to CDC guidelines (ttps://www.cdc.gov/nhsn/pdfs/ps-analysis-
586 resources/phenotype_definitions.pdf).

587 **DNA extraction for bacterial isolates**

588 Cell pellets were allowed to thaw slowly on ice and re-suspended in 400 µL of ATL buffer (Qiagen
589 Singapore Pte. Ltd, Cat. No. 19076). Cells were lysed in Lysing Matrix E tubes (MP Biomedicals, Cat. No.
590 SKU 116914500) on a vortex adapter at maximum speed for 10 min. Cell lysates were centrifuged at
591 16,000 *g* for 5 min and supernatant was treated with 4 µL of RNase A (100 mg/mL) (Qiagen Singapore
592 Pte. Ltd, Cat. No. 19101), gently mixed by flicking of tube and incubated at room temperature for 2 min.
593 The cell lysate was further treated with 25 µL of proteinase K (20 mg/mL) (Qiagen Singapore Pte. Ltd,
594 Cat. No. 19133), gently mixed by flicking of tube and incubated at 56 ˚C for 20 min. DNA was purified
595 twice using 1× volume of AMPure XP beads (Beckman Coulter, Cat. No. A63882) with slight
596 modifications to the manufacturer's protocol. All mixing steps were replaced with gentle flicking of tube
597 and incubation on the hula rotor for gentle mixing. Fresh 70% ethanol was prepared for washing and
598 magnetic beads were incubated on a 37 ˚C heat block for 3–5 min to dry off residual ethanol. The quality

599 and quantity of DNA was assessed using Nanodrop, Qubit® fluorometer and 0.5% agarose gel. Samples
600 that were unable to pass the following criteria were omitted from sequencing: DNA amount measured by
601 qubit >510 ng, DNA concentration measured by qubit >11 ng/μL, A260/280 ratio between 1.7-2.0,
602 A260/230 ratio between 1.5-3.3 and DNA length >1 kb. Purified DNA was stored at 4 ˚C.

603 **Nanopore library preparation**

604 DNA was prepared with either 1D[2] sequencing kit (SQK-LSK308) or 1D sequencing kit (SQK-LSK108 or
605 SQK-LSK109) together with native barcoding kit (EXP-NBD103 or EXP-NBD104 and EXP-NBD114)
606 according to the native barcoding genomic DNA protocol. DNA was not sheared and was used directly for
607 DNA repair and end-prep. Both native barcode ligation and adapter ligation steps were extended to 30
608 min instead of 10 min. In addition, to maximize library yields, more than 700 ng of pooled sample (where
609 possible) was used for adapter ligation. Samples were multiplexed (9-12 samples per pool for culture
610 enriched samples and 24 samples per pool for isolates) and sequenced with either MIN106, MIN106D or
611 MIN107 flowcells on a GridION machine.

612 **Taxonomic and resistome profiling with Illumina shotgun metagenomic data**

613 Illumina shotgun metagenomic sequencing reads were processed using a Snakemake[81] pipeline
614 (https://github.com/gis-rpd/pipelines/tree/master/metagenomics/shotgun-metagenomics). Briefly, raw
615 reads were filtered to remove low quality bases using skewer[82] (v0.2.2; -q 3 -l 30 -n) and human reads
616 were removed by mapping to the hg19 reference using BWA-MEM[83] (v0.7.10-r789). The remaining
617 microbial reads were profiled with MetaPhlAn2[84] (v2.6.0) and SRST2[85] (v0.1.4; --min_coverage 100, hits
618 with identity <99% were filtered out) for taxonomic and ARG abundances, respectively. Microbial reads
619 were also assembled using MEGAHIT[84] (v1.0.4-beta; default parameters) for comparison to nanopore
620 assemblies. Site specificity score was computed as the z-score for the closest taxonomic profile for a
621 sample (Bray-Curtis dissimilarity) among physically proximal sites (in the same room/cubicle and same
622 timepoint), compared to the distribution of Bray-Curtis dissimilarities across all samples of a site (e.g. all
623 bed rails). Results based on analysis of taxonomic and resistome profiles were obtained for each
624 timepoint independently and compared across timepoints to check for consistency and filter out potential
625 sequencing artefacts[34].

626 **Removal of likely contaminant species**

627 Likely contaminant species were identified based on batch and correlation analysis[34] (**Suppl. Note 2**) and
628 removed from species-level abundance profiles. For genus-level profiles, relative abundances of the
629 filtered species were subtracted from the abundance of the corresponding genera for each sample.
630 Filtered profiles were then renormalized to sum to 100% and used for all downstream analyses.

631 **Preprocessing of nanopore sequencing data**

632 Raw nanopore reads were basecalled with the latest version of the basecaller available at the point of
633 sequencing (Guppy v0.5.1 to v3.0.6 or Albacore v2.3.1 to v2.3.3 for libraries that failed live basecalling).
634 Basecalled nanopore reads were demultiplexed and filtered for adapters with Porechop (v0.2.3;
635 https://github.com/rrwick/Porechop) or qcat (v.1.1.0 https://github.com/nanoporetech/qcat). Sequencing
636 statistics were summarized using SeqKit[86] (v0.10.1). Reads were taxonomically classified with Kraken[87]
637 (v0.10.5-beta) against the miniKraken database (minikraken_201711_01_8GB_dustmasked) to assess
638 the diversity of cultures on the plates.

639 **Genome assembly and species assignment**

640 Nanopore reads for each plate were assembled using Canu[88] (v1.3/v1.7; genomeSize=8m). For samples
641 where both Illumina and Nanopore reads were available, a higher quality hybrid assembly was obtained
642 using OPERA-MS[28] (v0.8.3; --polish --no-gap-filling --short-read-assembler spades). Assembled contigs
643 were mapped to the NCBI nt database with BLAST (v2.2.28), to identify microbial species or plasmid
644 assignments according to the best BLAST hit (highest total reference coverage). Circular sequences were
645 identified using MUMmer[89] (v3.23; -maxmatch -nosimplify, alignments <1kbp long or with identity <95%
646 were filtered out) as recommended in the documentation for the Canu assembler
647 (https://canu.readthedocs.io/en/latest/faq.html#my-circular-element-is-duplicated-has-overlap). Contigs
648 assigned to the same species were binned into genomic bins. Metagenomic Illumina reads were used to
649 polish Canu assemblies where feasible using Pilon[90] (v1.22, --fix indel). We noted that annotation errors
650 were significantly reduced after polishing, and that genomic bins whose length was within 10% of the
651 expected length met the criteria for high-quality genomes (completeness >90% and contamination <5%
652 using CheckM[91]; v1.0.7 --reduced_tree). Genomic bins that met these criteria were therefore designated
653 as high-quality and incomplete bins (<50% of expected length) were removed from further analysis.
654 Genomes corresponding to novel species were identified as those with identity <95% or coverage <80%
655 compared to known genomes (BLAST with nt) and three recent catalogs that include environmental and
656 human microbiome assembled genomes[43-45] (with Mash[92]). The genomes were hierarchically clustered
657 (single-linkage with Mash distance[92]) to identify species-level clusters at 95% identity and genus level
658 taxonomic classification was obtained using sourmash[93]. Similarly, novel circular plasmids were identified
659 by comparing to the PLSDB[50] database with Mash distance and identifying clusters at 99% identity
660 (single-linkage) with no known sequence.

661 **Analysis of ARG combinations**

662 ARGs were annotated to contigs by mapping them to the ARG-ANNOT[94] database provided in SRST2
663 (v3) with BLAST (best hit with >90% identity and >90% reference coverage). ARG combinations present
664 in chromosomes and plasmid sequences were considered novel when they were not found in the
665 reference databases (nt or PLSDB[50]). Assembled circular plasmids were clustered and annotated based
666 on their best BLAST hit with identity >95% and >60% query coverage. A bipartite graph was constructed

667 by connecting each plasmid cluster to ARGs found in it, with edge weights representing the frequency of
668 occurrence (clusters with <5 representatives were excluded). For each species, an ARG co-occurrence
669 graph was created for ARGs found in the assembled genomes by connecting the ARG pairs that were
670 found within 10kbp on the same contig (discarding ARG pairs occurring less than 5 times). Each edge
671 was weighted by the frequency of ARG pairs divided by the minimal frequency of the two ARGs. All ARG
672 co-occurrence graphs were merged into a final co-occurrence multigraph. The graphs were visualized
673 using Cytoscape (v3.7.1)[95].

674 **Analysis of virulence factor and biocide resistance genes**

675 Nanopore assemblies were aligned to virulence factors in the PATRIC database[96] (2018/12/20) with
676 DIAMOND (v0.9.24.125; blastx --long-reads) and alignments with E-value >0.001 were filtered out. To
677 identify biocide resistance genes, the assemblies were aligned to nucleotide sequences for the genes
678 qacA (NC_014369.1) and qacC (NC_013339.1) with BLAST (>90% identity and >90% reference
679 coverage).

680 **Analysis of phages and prophages**

681 Phage-like-elements (phage and prophages) were identified using VirSorter[97] (v1.0.5, phages and
682 prophages in category 3 or with length <10kbp were filtered out). The assembled phages and reference
683 phages from the MVP database[51] were hierarchically clustered (single-linkage with Mash distance[92]) to
684 identify phage clusters at 95% identity. Clusters without any phages from the reference database were
685 considered novel. For each cluster, sub-clusters were defined at 99.9% ANI by single-linkage clustering
686 with nucleotide identities from nucmer (--maxmatch –nosimplify, followed by dnadiff and minimum
687 sequence overlap of 80%). Phage-like-elements were annotated using RAST[98] (virus domain, fix frame-
688 shifts parameters).

689 **Analysis of patient isolates and strain relationships**

690 Raw reads corresponding to genomes for outbreak isolates[40,64-66] were downloaded and assembled using
691 the Velvet assembler[99] (v1.2.10) with parameters optimized by Velvet Optimiser (k-mer length ranging
692 from 81 to 127), scaffolded with OPERA-LG[100] (v1.4.1), and gapfilled with FinIS[101] (v0.3). Outbreak
693 genomes from the same species were jointly analyzed with high-quality genomes from the hospital
694 environment. To identify high confidence SNPs we adapted the method from Brooks et al[26]. Specifically,
695 we performed pairwise alignments between genomes using nucmer, and considered genome pairs with
696 alignment coverage >80% for ANI computation. SNPs between genome pairs were called
697 using MUMmer's 'show-snps' function and regions containing more than 1 SNP within 20bp were filtered
698 out to mask potential artefacts from horizontal gene transfer, recombination or repeats. Finally, the
699 genomic distance matrix (based on ANI = #SNPs/alignment size) was clustered hierarchically (single
700 linkage) and clusters obtained at 99.99% identity for species with hybrid (Nanopore and Illumina)
701 assemblies (or 99.9% for species with Nanopore-only assemblies). Single-linkage clustering was used to
702 avoid having highly similar genomes being assigned to separate clusters and we confirmed that despite

703 this, most members (99%) had an average distance to other members of the cluster below the clustering
704 thresholds used. Antibiotic resistance profiles and multidrug resistance status (>2 antibiotic types) for
705 each cluster were derived from the union of resistance profiles for each genome obtained in various
706 selection plates.

707 For phylogenetic analysis, a consensus genome was derived for each cluster based on reference-guided
708 alignment with nucmer (*S. aureus*: NC_020529, *S. epidermidis*: NC_004461, *E. anophelleis*:
709 NZ_CP007547, *E. faecalis*: NC_017312, *E. faecium*: NC_017960, *P. aeruginosa*: NC_018080, *K.*
710 *pneumoniae*: NC_018522, *A. baumannii*: NC_009085) and the cons utility in the EMBOSS suite[102].
711 Maximum likelihood phylogenetic trees were constructed for each species with Parsnp[103] (v1.2; -c –x;
712 accounting for recombination events using PhiPack[104]) based on consensus genomes for each cluster,
713 where multiple sequence alignments for each species varied in length from 0.6Mbp (*S. epidermidis*) to
714 5.1Mbp (*P. aeruginosa*). For the species-level tree, full-length 16S rRNA sequences (*S.*
715 *epidermidis*: L37605.1, *S. aureus*: NR_118997.2, *E. anophelis*: NR_116021.1 and *A. baumannii*:
716 NR_026206.1) were aligned with MAFFT[106] (v7.154b, default parameters) and the phylogeny was
717 determined using FastTree2[107] (v2.1.8, default parameters). The trees were visualized using the 'ggtree'
718 R package[105]. Strain distributions across sites were visualized with the 'HiveR' R package
719 (academic.depauw.edu/~hanson/HiveR/HiveR.html). Rarefaction analysis for species, plasmids, strains
720 and resistance genes was performed using the iNEXT R package[108].

721 **Statistical Analysis**

722 Statistical tests were carried out using R and are two-sided unless otherwise specified. For enrichment
723 analysis at the cluster level (overlap across time, cohorts or resistance status) Fisher's exact test was
724 used, and for analysis at the genome level (fraction of genomes with a specific property) the binomial test
725 was used.

726 **Data and source code availability**

727 All sequencing reads and assemblies are available from the European Nucleotide Archive under project
728 PRJEB31632 (https://www.ebi.ac.uk/ena/data/view/PRJEB31632). Source code and associated data for
729 reproducing the figures are available on GitHub under an MIT license
730 (https://github.com/csb5/hospital_microbiome). Assemblies and annotations for genomes, plasmids and
731 phages from this study are also available at https://ndownloader.figshare.com/files/17290511.

732 **Acknowledgements**

733 Funding for this work was provided by A*STAR, and we are grateful for support from NMRC (NMRC
734 CGAug16C005). C.E.M. acknowledges support from the WorldQuant Foundation, the Bill and Melinda
735 Gates Foundation (OPP1151054), and the Alfred P. Sloan Foundation (G-2015-13964). The funders had
736 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. We would also like to acknowledge constructive feedback and valuable guidance from the reviewers, particularly in accounting for laboratory contaminants and artefacts.

740 # Figure legends

741 **Figure 1: Overview of sampling sites and analysis workflow.** (a) Diagram showing the various sites
742 that were sampled, including cardiac tables (A), bedrails (B), pulse oxymeters (C), bedside lockers (D),
743 sink aerators (E), sink traps (F) and door handles (G). Each ward (MDRO and Standard) had five beds
744 (sites A-D individually sampled) and one sink (E, F) and no doors, while isolation rooms had one bed (A-
745 D), one sink (E, F) and a door (G). (b) Diagram showing the analysis workflow for the four swabs that
746 were collected from each site in terms of culturing, DNA extraction and sequencing. Samples from each
747 site were analyzed with shotgun metagenomic sequencing on the Illumina platform and multiple (n=6)
748 culture-enriched quasi-metagenomics on a GridION system.

749 **Figure 2: Distinct ecological niches in the hospital environment for microbes and ARGs.** (a)
750 Principle coordinates analysis plot based on genus-level Bray-Curtis dissimilarity of taxonomic profiles
751 indicating two distinct community types (denoted as CTA and CTB) for microbiomes from the hospital
752 environment. (bi) Heatmap showing relative abundances (log-scale, Log_2RA) for differentially abundant
753 genera between community types CTA and CTB (FDR-adjusted Wilcoxon p-value<0.01) (bii) Boxplots
754 showing relative abundances for differentially abundant species between community types CTA and CTB
755 (FDR-adjusted Wilcoxon p-value<0.01). (c) Principle coordinates analysis plot (genus-level Bray-Curtis
756 dissimilarity) showing how much environmental microbiomes vary over time (lines connect two timepoints,
757 1 week apart) for different sites. (d) Radar plot showing the microbiome turnover index (fraction of taxa
758 that are gained or lost across timepoints), human influence index (fraction of human reads) and site
759 specificity index (uniqueness of site-specific taxonomic composition in relation to physically proximal
760 sites). A positive site specificity index indicates a stronger site-specific microbiome composition signature.
761 (e) Boxplots showing relative abundances of common nosocomial pathogens that were differentially
762 abundant across ward types in high human-contact sites (FDR-adjusted Kruskal-Wallis p-value<0.01). (f)
763 Heatmap depicting the frequency of detection for beta-lactamases at different sites in hospital wards.
764 Multiple carbapenemases and the mecA gene were detected as part of the resistomes that were primarily
765 defined by the community types (CTA and CTB). Boxplots are represented with center line: median; box
766 limits: upper and lower quartiles; whiskers: 1.5× interquartile range; outlier points not included in
767 visualization.

768 **Figure 3: Genome-resolved characterization of nosocomial multidrug resistant strains that spread**
769 **and persist at low relative abundances in the hospital environment.** (a) Heatmap displaying the
770 distinct median relative abundances (RA) of common nosocomial pathogens at different sites in
771 hospital environments (PERMANOVA p-value<0.001). (b) Distribution of assembly contiguity statistics
772 (N50=fragment size s.t. more than 50% of the genome is in longer sequences) for common
773 nosocomial pathogens, highlighting the high genomic contiguity that was obtained (median N50
774 >1Mbp). (c) Dotplots highlighting that genomes can be rapidly obtained for several nosocomial
775 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* derivative clusters (>99.99% ANI; each node represents the consensus genome for the cluster) detected in the hospital environment together with their antibiotic resistance profiles. The scale bar represents the number of substitutions per site in the core alignment. (Right panel) Hive-map representation showing localization of *S. aureus* clusters that spread (detected at 2 or more locations) and/or persist (detected in timepoints 1 and 2) 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 distributed and persistent in the hospital environment.

Figure 4: Species distribution and genomic proximity of drug resistance genes in the hospital environment microbiome. Genomic proximity network and clustering of ARGs based on 2347 microbial genomes and 5910 closed plasmids obtained from the hospital environment. (a) Multigraph of genomic proximity between ARGs, 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 ARGs 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 ARG classes and inferred host species for ease of reference.

Figure 5: Multi-species analysis of phylogenetic relationships between environmental and patient genomes. Phylogenies depicting the evolutionary relationships between derivative clusters (>99.99% ANI), with each node representing the consensus genome for a cluster, for (a) *Elizabethkingia anophelis* from a nosocomial outbreak in Singapore in 2012 and other patient isolates, (b) patient colonizing *Staphylococcus aureus* from a 2009-2011 surveillance study in Singaporean hospitals, (c) infectious *Acinetobacter baumannii* isolates from patients in two major Kuwaiti hospitals and Singaporean patient isolates, and (d) recent globally disseminated multi-drug resistant *Staphylococcus epidermidis* lineages, together with environmental genomes 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

812 years apart, subfigures c and d reveal the global dissemination of several lineages. Note that the matrices
813 next to the trees indicate the antibiotic resistance profiles for corresponding derivative clusters. Scale bars
814 depict the number of substitutions per site in the core alignment. For all species tested, clusters shared
815 between environmental and patient genomes were enriched for multidrug resistance.

816 # References

817 1 Fischer, M. M. & Bild, M. Hospital use of antibiotics as the main driver of infections with
818 antibiotic-resistant bacteria - a reanalysis of recent data from the European Union. *bioRxiv*
819 (2019).
820 2 O'Neill, J. *Tackling drug-resistant infections globally: final report and recommendations*,
821 <https://amr-review.org/Publications.html> (2016).
822 3 Allegranzi, B. *et al.* Burden of endemic health-care-associated infection in developing countries:
823 systematic review and meta-analysis. *Lancet* **377**, 228-241, doi:10.1016/S0140-6736(10)61458-4
824 (2011).
825 4 Magill, S. S. *et al.* Multistate point-prevalence survey of health care-associated infections. *N Engl*
826 *J Med* **370**, 1198-1208, doi:10.1056/NEJMoa1306801 (2014).
827 5 Laxminarayan, R. *et al.* Antibiotic resistance-the need for global solutions. *Lancet Infect Dis* **13**,
828 1057-1098, doi:10.1016/S1473-3099(13)70318-9 (2013).
829 6 Longtin, Y., Sax, H., Allegranzi, B., Schneider, F. & Pittet, D. Videos in clinical medicine. Hand
830 hygiene. *N Engl J Med* **364**, e24, doi:10.1056/NEJMvcm0903599 (2011).
831 7 Otter, J. A., Yezli, S., Salkeld, J. A. & French, G. L. Evidence that contaminated surfaces contribute
832 to the transmission of hospital pathogens and an overview of strategies to address
833 contaminated surfaces in hospital settings. *Am J Infect Control* **41**, S6-11,
834 doi:10.1016/j.ajic.2012.12.004 (2013).
835 8 Weber, D. J., Rutala, W. A., Miller, M. B., Huslage, K. & Sickbert-Bennett, E. Role of hospital
836 surfaces in the transmission of emerging health care-associated pathogens: norovirus,
837 Clostridium difficile, and Acinetobacter species. *Am J Infect Control* **38**, S25-33,
838 doi:10.1016/j.ajic.2010.04.196 (2010).
839 9 Hayden, M. K. *et al.* Reduction in acquisition of vancomycin-resistant enterococcus after
840 enforcement of routine environmental cleaning measures. *Clin Infect Dis* **42**, 1552-1560,
841 doi:10.1086/503845 (2006).
842 10 Datta, R., Platt, R., Yokoe, D. S. & Huang, S. S. Environmental cleaning intervention and risk of
843 acquiring multidrug-resistant organisms from prior room occupants. *Arch Intern Med* **171**, 491-
844 494, doi:10.1001/archinternmed.2011.64 (2011).
845 11 Mitchell, B. G. *et al.* An environmental cleaning bundle and health-care-associated infections in
846 hospitals (REACH): a multicentre, randomised trial. *Lancet Infect Dis* **19**, 410-418,
847 doi:10.1016/S1473-3099(18)30714-X (2019).
848 12 Sabino, R. *et al.* Analysis of clinical and environmental Candida parapsilosis isolates by
849 microsatellite genotyping--a tool for hospital infection surveillance. *Clin Microbiol Infect* **21**, 954
850 e951-958, doi:10.1016/j.cmi.2015.06.001 (2015).
851 13 Obeidat, N., Jawdat, F., Al-Bakri, A. G. & Shehabi, A. A. Major biologic characteristics of
852 Acinetobacter baumannii isolates from hospital environmental and patients' respiratory tract
853 sources. *Am J Infect Control* **42**, 401-404, doi:10.1016/j.ajic.2013.10.010 (2014).
854 14 Price, J. R. *et al.* Transmission of Staphylococcus aureus between health-care workers, the
855 environment, and patients in an intensive care unit: a longitudinal cohort study based on whole-
856 genome sequencing. *Lancet Infect Dis* **17**, 207-214, doi:10.1016/S1473-3099(16)30413-3 (2017).
857 15 Faires, M. C. *et al.* A prospective study to examine the epidemiology of methicillin-resistant
858 Staphylococcus aureus and Clostridium difficile contamination in the general environment of
859 three community hospitals in southern Ontario, Canada. *BMC Infect Dis* **12**, 290,
860 doi:10.1186/1471-2334-12-290 (2012).

861 16 Weingarten, R. A. *et al.* Genomic Analysis of Hospital Plumbing Reveals Diverse Reservoir of
862 Bacterial Plasmids Conferring Carbapenem Resistance. *MBio* **9**, doi:10.1128/mBio.02011-17
863 (2018).
864 17 Conlan, S. *et al.* Single-molecule sequencing to track plasmid diversity of hospital-associated
865 carbapenemase-producing Enterobacteriaceae. *Sci Transl Med* **6**, 254ra126,
866 doi:10.1126/scitranslmed.3009845 (2014).
867 18 Lax, S. & Gilbert, J. A. Hospital-associated microbiota and implications for nosocomial infections.
868 *Trends Mol Med* **21**, 427-432, doi:10.1016/j.molmed.2015.03.005 (2015).
869 19 Miller, R. R., Montoya, V., Gardy, J. L., Patrick, D. M. & Tang, P. Metagenomics for pathogen
870 detection in public health. *Genome Med* **5**, 81, doi:10.1186/gm485 (2013).
871 20 Chen, C. H. *et al.* Bacterial diversity among four healthcare-associated institutes in Taiwan. *Sci*
872 *Rep* **7**, 8230, doi:10.1038/s41598-017-08679-3 (2017).
873 21 Hewitt, K. M. *et al.* Bacterial diversity in two Neonatal Intensive Care Units (NICUs). *PLoS One* **8**,
874 e54703, doi:10.1371/journal.pone.0054703 (2013).
875 22 Oberauner, L. *et al.* The ignored diversity: complex bacterial communities in intensive care units
876 revealed by 16S pyrosequencing. *Sci Rep* **3**, 1413, doi:10.1038/srep01413 (2013).
877 23 Poza, M. *et al.* Exploring bacterial diversity in hospital environments by GS-FLX Titanium
878 pyrosequencing. *PLoS One* **7**, e44105, doi:10.1371/journal.pone.0044105 (2012).
879 24 Lax, S. *et al.* Bacterial colonization and succession in a newly opened hospital. *Sci Transl Med* **9**,
880 doi:10.1126/scitranslmed.aah6500 (2017).
881 25 Janda, J. M. & Abbott, S. L. 16S rRNA gene sequencing for bacterial identification in the
882 diagnostic laboratory: pluses, perils, and pitfalls. *J Clin Microbiol* **45**, 2761-2764,
883 doi:10.1128/JCM.01228-07 (2007).
884 26 Brooks, B. *et al.* Strain-resolved analysis of hospital rooms and infants reveals overlap between
885 the human and room microbiome. *Nat Commun* **8**, 1814, doi:10.1038/s41467-017-02018-w
886 (2017).
887 27 McIntyre, A. B. R. *et al.* Comprehensive benchmarking and ensemble approaches for
888 metagenomic classifiers. *Genome Biol* **18**, 182, doi:10.1186/s13059-017-1299-7 (2017).
889 28 Bertrand, D. *et al.* Hybrid metagenomic assembly enables high-resolution analysis of resistance
890 determinants and mobile elements in human microbiomes. *Nat Biotechnol* **37**, 937-944,
891 doi:10.1038/s41587-019-0191-2 (2019).
892 29 Che, Y. *et al.* Mobile antibiotic resistome in wastewater treatment plants revealed by Nanopore
893 metagenomic sequencing. *Microbiome* **7**, 44, doi:10.1186/s40168-019-0663-0 (2019).
894 30 van der Helm, E. *et al.* Rapid resistome mapping using nanopore sequencing. *Nucleic Acids Res*
895 **45**, e61, doi:10.1093/nar/gkw1328 (2017).
896 31 Dancer, S. J. Controlling hospital-acquired infection: focus on the role of the environment and
897 new technologies for decontamination. *Clin Microbiol Rev* **27**, 665-690, doi:10.1128/CMR.00020-
898 14 (2014).
899 32 Kanamori, H., Weber, D. J. & Rutala, W. A. Healthcare Outbreaks Associated With a Water
900 Reservoir and Infection Prevention Strategies. *Clin Infect Dis* **62**, 1423-1435,
901 doi:10.1093/cid/ciw122 (2016).
902 33 Dancer, S. J. Importance of the environment in meticillin-resistant Staphylococcus aureus
903 acquisition: the case for hospital cleaning. *Lancet Infect Dis* **8**, 101-113, doi:10.1016/S1473-
904 3099(07)70241-4 (2008).
905 34 de Goffau, M. C. *et al.* Recognizing the reagent microbiome. *Nat Microbiol* **3**, 851-853,
906 doi:10.1038/s41564-018-0202-y (2018).

907 35 Sehulster, L., Chinn, R. Y., Cdc & Hicpac. Guidelines for environmental infection control in health-
908 care facilities. Recommendations of CDC and the Healthcare Infection Control Practices Advisory
909 Committee (HICPAC). *MMWR Recomm Rep* **52**, 1-42 (2003).
910 36 Kizny Gordon, A. E. *et al.* The Hospital Water Environment as a Reservoir for Carbapenem-
911 Resistant Organisms Causing Hospital-Acquired Infections-A Systematic Review of the Literature.
912 *Clin Infect Dis* **64**, 1435-1444, doi:10.1093/cid/cix132 (2017).
913 37 Danko, D. C. *et al.* Global Genetic Cartography of Urban Metagenomes and Anti-Microbial
914 Resistance. *bioRxiv*, 724526, doi:10.1101/724526 (2019).
915 38 Beggs, C., Knibbs, L. D., Johnson, G. R. & Morawska, L. Environmental contamination and
916 hospital-acquired infection: factors that are easily overlooked. *Indoor Air* **25**, 462-474,
917 doi:10.1111/ina.12170 (2015).
918 39 Ministry of Health Singapore. Guidelines for control and prevention of multi-drug resistant
919 organisms (MDROs) in healthcare facilities. (2013).
920 40 Lee, J. Y. H. *et al.* Global spread of three multidrug-resistant lineages of Staphylococcus
921 epidermidis. *Nat Microbiol* **3**, 1175-1185, doi:10.1038/s41564-018-0230-7 (2018).
922 41 Bradford, P. A. Extended-spectrum beta-lactamases in the 21st century: characterization,
923 epidemiology, and detection of this important resistance threat. *Clin Microbiol Rev* **14**, 933-951,
924 table of contents, doi:10.1128/CMR.14.4.933-951.2001 (2001).
925 42 Wright, G. D. Environmental and clinical antibiotic resistomes, same only different. *Curr Opin*
926 *Microbiol* **51**, 57-63, doi:10.1016/j.mib.2019.06.005 (2019).
927 43 Pasolli, E. *et al.* Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000
928 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell* **176**, 649-662 e620,
929 doi:10.1016/j.cell.2019.01.001 (2019).
930 44 Almeida, A. *et al.* A new genomic blueprint of the human gut microbiota. *Nature* **568**, 499-504,
931 doi:10.1038/s41586-019-0965-1 (2019).
932 45 Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny substantially
933 revises the tree of life. *Nat Biotechnol* **36**, 996-1004, doi:10.1038/nbt.4229 (2018).
934 46 Carattoli, A. Plasmids and the spread of resistance. *Int J Med Microbiol* **303**, 298-304,
935 doi:10.1016/j.ijmm.2013.02.001 (2013).
936 47 Alekshun, M. N. & Levy, S. B. Molecular mechanisms of antibacterial multidrug resistance. *Cell*
937 **128**, 1037-1050, doi:10.1016/j.cell.2007.03.004 (2007).
938 48 San Millan, A. Evolution of Plasmid-Mediated Antibiotic Resistance in the Clinical Context.
939 *Trends Microbiol* **26**, 978-985, doi:10.1016/j.tim.2018.06.007 (2018).
940 49 Calero-Caceres, W., Ye, M. & Balcazar, J. L. Bacteriophages as Environmental Reservoirs of
941 Antibiotic Resistance. *Trends Microbiol* **27**, 570-577, doi:10.1016/j.tim.2019.02.008 (2019).
942 50 Galata, V., Fehlmann, T., Backes, C. & Keller, A. PLSDB: a resource of complete bacterial
943 plasmids. *Nucleic Acids Res* **47**, D195-D202, doi:10.1093/nar/gky1050 (2019).
944 51 Gao, N. L. *et al.* MVP: a microbe-phage interaction database. *Nucleic Acids Res* **46**, D700-D707,
945 doi:10.1093/nar/gkx1124 (2018).
946 52 Ho, C. M. *et al.* High rate of qacA- and qacB-positive methicillin-resistant Staphylococcus aureus
947 isolates from chlorhexidine-impregnated catheter-related bloodstream infections. *Antimicrob*
948 *Agents Chemother* **56**, 5693-5697, doi:10.1128/AAC.00761-12 (2012).
949 53 Duran, N., Temiz, M., Duran, G. G., Eryilmaz, N. & Jenedi, K. Relationship between the resistance
950 genes to quaternary ammonium compounds and antibiotic resistance in staphylococci isolated
951 from surgical site infections. *Med Sci Monit* **20**, 544-550, doi:10.12659/MSM.890177 (2014).
952 54 Johnson, R. C. *et al.* Investigation of a Cluster of Sphingomonas koreensis Infections. *N Engl J*
953 *Med* **379**, 2529-2539, doi:10.1056/NEJMoa1803238 (2018).

954 55 Fahimipour, A. K. *et al.* Antimicrobial Chemicals Associate with Microbial Function and Antibiotic
955 Resistance Indoors. *mSystems* **3**, doi:10.1128/mSystems.00200-18 (2018).
956 56 Hartmann, E. M. *et al.* Antimicrobial Chemicals Are Associated with Elevated Antibiotic
957 Resistance Genes in the Indoor Dust Microbiome. *Environ Sci Technol* **50**, 9807-9815,
958 doi:10.1021/acs.est.6b00262 (2016).
959 57 Falagas, M. E., Roussos, N., Gkegkes, I. D., Rafailidis, P. I. & Karageorgopoulos, D. E. Fosfomycin
960 for the treatment of infections caused by Gram-positive cocci with advanced antimicrobial drug
961 resistance: a review of microbiological, animal and clinical studies. *Expert Opin Investig Drugs*
962 **18**, 921-944, doi:10.1517/13543780902967624 (2009).
963 58 Urban, C., Mariano, N. & Rahal, J. J. In vitro double and triple bactericidal activities of
964 doripenem, polymyxin B, and rifampin against multidrug-resistant Acinetobacter baumannii,
965 Pseudomonas aeruginosa, Klebsiella pneumoniae, and Escherichia coli. *Antimicrob Agents*
966 *Chemother* **54**, 2732-2734, doi:10.1128/AAC.01768-09 (2010).
967 59 Izadpanah, M. & Khalili, H. Antibiotic regimens for treatment of infections due to multidrug-
968 resistant Gram-negative pathogens: An evidence-based literature review. *J Res Pharm Pract* **4**,
969 105-114, doi:10.4103/2279-042X.162360 (2015).
970 60 Werner, G., Hildebrandt, B. & Witte, W. Aminoglycoside-streptothricin resistance gene cluster
971 aadE-sat4-aphA-3 disseminated among multiresistant isolates of Enterococcus faecium.
972 *Antimicrob Agents Chemother* **45**, 3267-3269, doi:10.1128/AAC.45.11.3267-3269.2001 (2001).
973 61 Weinstein, L., Chang, T. W., Hudson, J. B. & Hartl, W. The concurrent use of sulfonamides and
974 antibiotics in the treatment of infections: in vivo and in vitro studies of the effect of
975 sulfonamide-antibiotic combinations on the emergence of drug resistance. *Ann N Y Acad Sci* **69**,
976 408-416 (1957).
977 62 Sundin, G. W. & Bender, C. L. Dissemination of the strA-strB streptomycin-resistance genes
978 among commensal and pathogenic bacteria from humans, animals, and plants. *Mol Ecol* **5**, 133-
979 143 (1996).
980 63 Monecke, S. *et al.* Microarray-based characterisation of a Panton-Valentine leukocidin-positive
981 community-acquired strain of methicillin-resistant Staphylococcus aureus. *Clin Microbiol Infect*
982 **12**, 718-728, doi:10.1111/j.1469-0691.2006.01420.x (2006).
983 64 Teo, J. *et al.* First case of E anophelis outbreak in an intensive-care unit. *Lancet* **382**, 855-856,
984 doi:10.1016/S0140-6736(13)61858-9 (2013).
985 65 Htun, H. L. *et al.* Methicillin-resistant Staphylococcus aureus colonisation: epidemiological and
986 molecular characteristics in an acute-care tertiary hospital in Singapore. *Epidemiol Infect* **146**,
987 1785-1792, doi:10.1017/S0950268818001966 (2018).
988 66 Nasser, K. *et al.* Draft Genome Sequences of Six Multidrug-Resistant Clinical Strains of
989 Acinetobacter baumannii, Isolated at Two Major Hospitals in Kuwait. *Genome Announc* **6**,
990 doi:10.1128/genomeA.00264-18 (2018).
991 67 Walker, J. T. *et al.* Hospital and community acquired infection and the built environment--design
992 and testing of infection control rooms. *J Hosp Infect* **65 Suppl 2**, 43-49, doi:10.1016/S0195-
993 6701(07)60014-0 (2007).
994 68 Ladner, J. T., Grubaugh, N. D., Pybus, O. G. & Andersen, K. G. Precision epidemiology for
995 infectious disease control. *Nat Med* **25**, 206-211, doi:10.1038/s41591-019-0345-2 (2019).
996 69 Mahnert, A. *et al.* Man-made microbial resistances in built environments. *Nat Commun* **10**, 968,
997 doi:10.1038/s41467-019-08864-0 (2019).
998 70 Westfall, C. *et al.* The Widely Used Antimicrobial Triclosan Induces High Levels of Antibiotic
999 Tolerance In Vitro and Reduces Antibiotic Efficacy up to 100-Fold In Vivo. *Antimicrob Agents*
1000 *Chemother* **63**, doi:10.1128/AAC.02312-18 (2019).

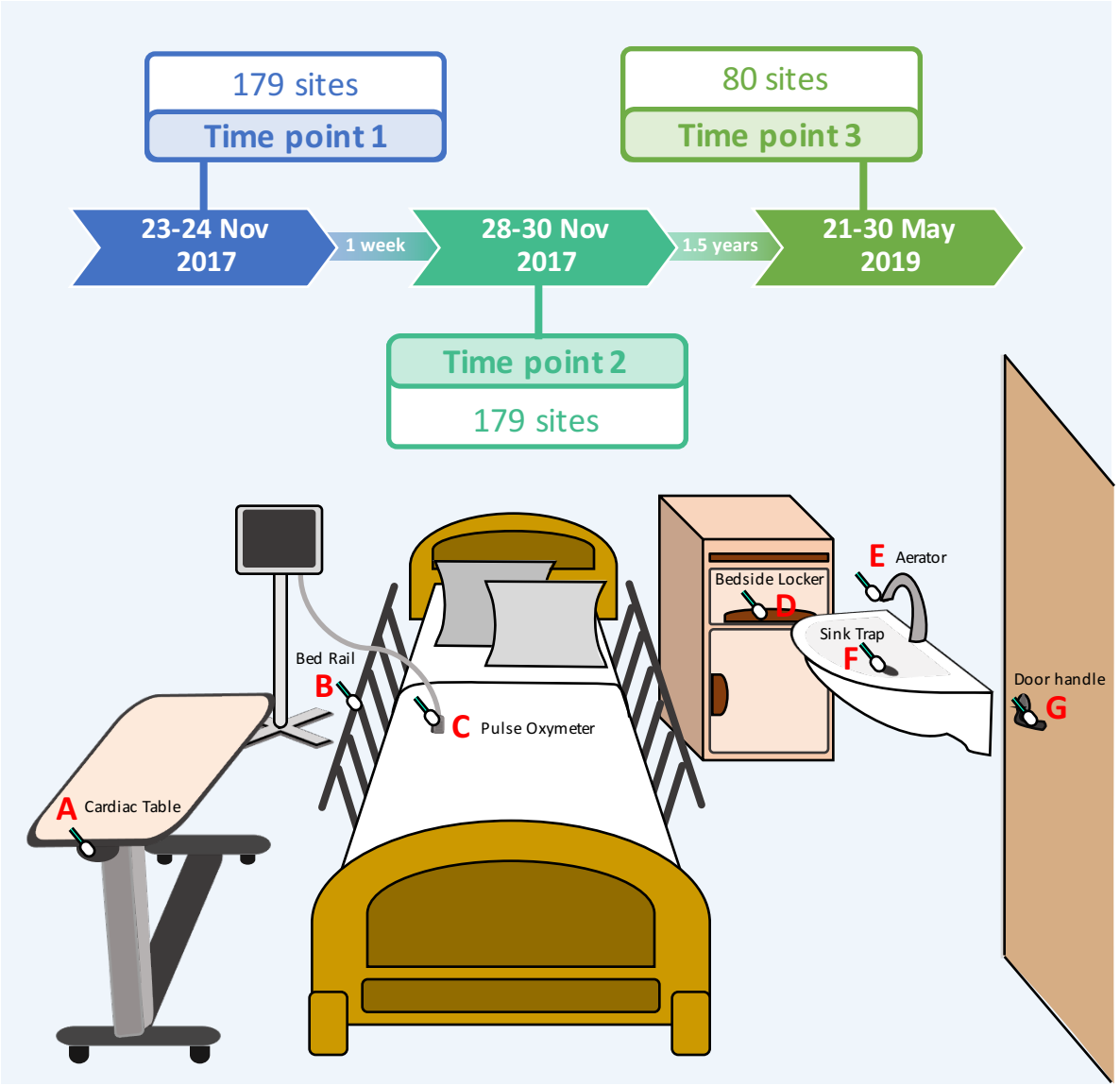
1001 71 Caselli, E. *et al.* Impact of a probiotic-based hospital sanitation on antimicrobial resistance and
1002 HAI-associated antimicrobial consumption and costs: a multicenter study. *Infect Drug Resist* **12**,
1003 501-510, doi:10.2147/IDR.S194670 (2019).
1004 72 Goodman, E. R. *et al.* Impact of an environmental cleaning intervention on the presence of
1005 methicillin-resistant Staphylococcus aureus and vancomycin-resistant enterococci on surfaces in
1006 intensive care unit rooms. *Infect Control Hosp Epidemiol* **29**, 593-599, doi:10.1086/588566
1007 (2008).
1008 73 Lerminiaux, N. A. & Cameron, A. D. S. Horizontal transfer of antibiotic resistance genes in clinical
1009 environments. *Can J Microbiol* **65**, 34-44, doi:10.1139/cjm-2018-0275 (2019).
1010 74 Buckner, M. M. C., Ciusa, M. L. & Piddock, L. J. V. Strategies to combat antimicrobial resistance:
1011 anti-plasmid and plasmid curing. *FEMS Microbiol Rev* **42**, 781-804, doi:10.1093/femsre/fuy031
1012 (2018).
1013 75 Chen, S., Larsson, M., Robinson, R. C. & Chen, S. L. Direct and convenient measurement of
1014 plasmid stability in lab and clinical isolates of E. coli. *Sci Rep* **7**, 4788, doi:10.1038/s41598-017-
1015 05219-x (2017).
1016 76 David, S. *et al.* Epidemic of carbapenem-resistant Klebsiella pneumoniae in Europe is driven by
1017 nosocomial spread. *Nat Microbiol* **4**, 1919-1929, doi:10.1038/s41564-019-0492-8 (2019).
1018 77 Sievert, D. M. *et al.* Antimicrobial-resistant pathogens associated with healthcare-associated
1019 infections: summary of data reported to the National Healthcare Safety Network at the Centers
1020 for Disease Control and Prevention, 2009-2010. *Infect Control Hosp Epidemiol* **34**, 1-14,
1021 doi:10.1086/668770 (2013).
1022 78 Weiner, L. M. *et al.* Antimicrobial-Resistant Pathogens Associated With Healthcare-Associated
1023 Infections: Summary of Data Reported to the National Healthcare Safety Network at the Centers
1024 for Disease Control and Prevention, 2011-2014. *Infect Control Hosp Epidemiol* **37**, 1288-1301,
1025 doi:10.1017/ice.2016.174 (2016).
1026 79 Cornejo-Juarez, P. *et al.* The impact of hospital-acquired infections with multidrug-resistant
1027 bacteria in an oncology intensive care unit. *Int J Infect Dis* **31**, 31-34,
1028 doi:10.1016/j.ijid.2014.12.022 (2015).
1029 80 Teerawattanapong, N. *et al.* A Systematic Review of the Burden of Multidrug-Resistant
1030 Healthcare-Associated Infections Among Intensive Care Unit Patients in Southeast Asia: The Rise
1031 of Multidrug-Resistant Acinetobacter baumannii. *Infect Control Hosp Epidemiol* **39**, 525-533,
1032 doi:10.1017/ice.2018.58 (2018).
1033 81 Koster, J. & Rahmann, S. Snakemake--a scalable bioinformatics workflow engine. *Bioinformatics*
1034 **28**, 2520-2522, doi:10.1093/bioinformatics/bts480 (2012).
1035 82 Jiang, H., Lei, R., Ding, S. W. & Zhu, S. Skewer: a fast and accurate adapter trimmer for next-
1036 generation sequencing paired-end reads. *BMC Bioinformatics* **15**, 182, doi:10.1186/1471-2105-
1037 15-182 (2014).
1038 83 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform.
1039 *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
1040 84 Truong, D. T. *et al.* MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat Methods*
1041 **12**, 902-903, doi:10.1038/nmeth.3589 (2015).
1042 85 Inouye, M. *et al.* SRST2: Rapid genomic surveillance for public health and hospital microbiology
1043 labs. *Genome Med* **6**, 90, doi:10.1186/s13073-014-0090-6 (2014).
1044 86 Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File
1045 Manipulation. *PLoS One* **11**, e0163962, doi:10.1371/journal.pone.0163962 (2016).
1046 87 Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact
1047 alignments. *Genome Biol* **15**, R46, doi:10.1186/gb-2014-15-3-r46 (2014).

1048 88 Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and
1049 repeat separation. *Genome Res* **27**, 722-736, doi:10.1101/gr.215087.116 (2017).
1050 89 Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol* **5**, R12,
1051 doi:10.1186/gb-2004-5-2-r12 (2004).
1052 90 Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and
1053 genome assembly improvement. *PLoS One* **9**, e112963, doi:10.1371/journal.pone.0112963
1054 (2014).
1055 91 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing
1056 the quality of microbial genomes recovered from isolates, single cells, and metagenomes.
1057 *Genome Res* **25**, 1043-1055, doi:10.1101/gr.186072.114 (2015).
1058 92 Ondov, B. D. *et al.* Mash: fast genome and metagenome distance estimation using MinHash.
1059 *Genome Biol* **17**, 132, doi:10.1186/s13059-016-0997-x (2016).
1060 93 Pierce, N. T., Irber, L., Reiter, T., Brooks, P. & Brown, C. T. Large-scale sequence comparisons
1061 with sourmash. *F1000Res* **8**, 1006, doi:10.12688/f1000research.19675.1 (2019).
1062 94 Gupta, S. K. *et al.* ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes
1063 in bacterial genomes. *Antimicrob Agents Chemother* **58**, 212-220, doi:10.1128/AAC.01310-13
1064 (2014).
1065 95 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular
1066 interaction networks. *Genome Res* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).
1067 96 Wattam, A. R. *et al.* PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic*
1068 *Acids Res* **42**, D581-591, doi:10.1093/nar/gkt1099 (2014).
1069 97 Roux, S., Enault, F., Hurwitz, B. L. & Sullivan, M. B. VirSorter: mining viral signal from microbial
1070 genomic data. *PeerJ* **3**, e985, doi:10.7717/peerj.985 (2015).
1071 98 Aziz, R. K. *et al.* The RAST Server: rapid annotations using subsystems technology. *BMC*
1072 *Genomics* **9**, 75, doi:10.1186/1471-2164-9-75 (2008).
1073 99 Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn
1074 graphs. *Genome Res* **18**, 821-829, doi:10.1101/gr.074492.107 (2008).
1075 100 Gao, S., Bertrand, D., Chia, B. K. & Nagarajan, N. OPERA-LG: efficient and exact scaffolding of
1076 large, repeat-rich eukaryotic genomes with performance guarantees. *Genome Biol* **17**, 102,
1077 doi:10.1186/s13059-016-0951-y (2016).
1078 101 Gao, S., Bertrand, D., Nagarajan, N. in *Algorithms in Bioinformatics* (ed Tang J. Raphael B.) 314-
1079 325 (Springer, Berlin, Heidelberg, WABI, 2012).
1080 102 Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software
1081 Suite. *Trends Genet* **16**, 276-277, doi:10.1016/s0168-9525(00)02024-2 (2000).
1082 103 Treangen, T. J., Ondov, B. D., Koren, S. & Phillippy, A. M. The Harvest suite for rapid core-
1083 genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome*
1084 *Biol* **15**, 524, doi:10.1186/s13059-014-0524-x (2014).
1085 104 Bruen, T. C., Philippe, H. & Bryant, D. A simple and robust statistical test for detecting the
1086 presence of recombination. *Genetics* **172**, 2665-2681, doi:10.1534/genetics.105.048975 (2006).
1087 105 Yu, G., Lam, T. T., Zhu, H. & Guan, Y. Two Methods for Mapping and Visualizing Associated Data
1088 on Phylogeny Using Ggtree. *Mol Biol Evol* **35**, 3041-3043, doi:10.1093/molbev/msy194 (2018).
1089 106 Katoh, K., Asimenos, G. & Toh, H. Multiple alignment of DNA sequences with MAFFT. *Methods*
1090 *Mol Biol* **537**, 39-64, doi:10.1007/978-1-59745-251-9_3 (2009).
1091 107 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood trees for
1092 large alignments. *PLoS One* **5**, e9490, doi:10.1371/journal.pone.0009490 (2010).
1093 108 Hsieh, T. C., Ma, K. H. & Chao, A. iNEXT: an R package for rarefaction and extrapolation of
1094 species diversity (Hill numbers). *Methods in Ecology and Evolution* **7**, 1451-1456,
1095 doi:10.1111/2041-210x.12613 (2016).

1096

Figure 1

a


179 sites
Time point 1
80 sites
Time point 3
23-24 Nov 2017
1 week
28-30 Nov 2017
1.5 years
21-30 May 2019
Time point 2
179 sites
Bed Rail
B
C Pulse Oxymeter
A Cardiac Table
Bedside Locker
D
E Aerator
Sink Trap
F
Door handle
G

b

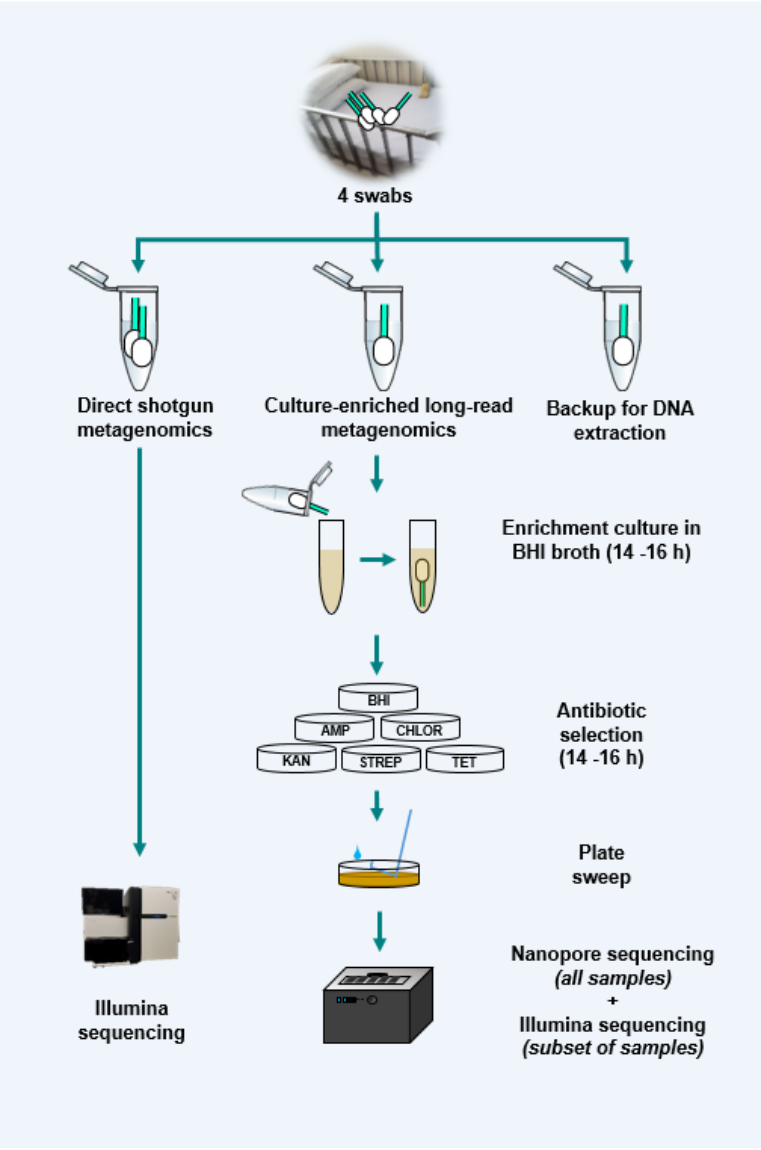

4 swabs
Direct shotgun metagenomics
Culture-enriched long-read metagenomics
Backup for DNA extraction
Enrichment culture in BHI broth (14 -16 h)
BHI
AMP
CHLOR
KAN
STREP
TET
Antibiotic selection (14 -16 h)
Plate sweep
Illumina sequencing
Nanopore sequencing (all samples) + Illumina sequencing (subset of samples)

Figure 2

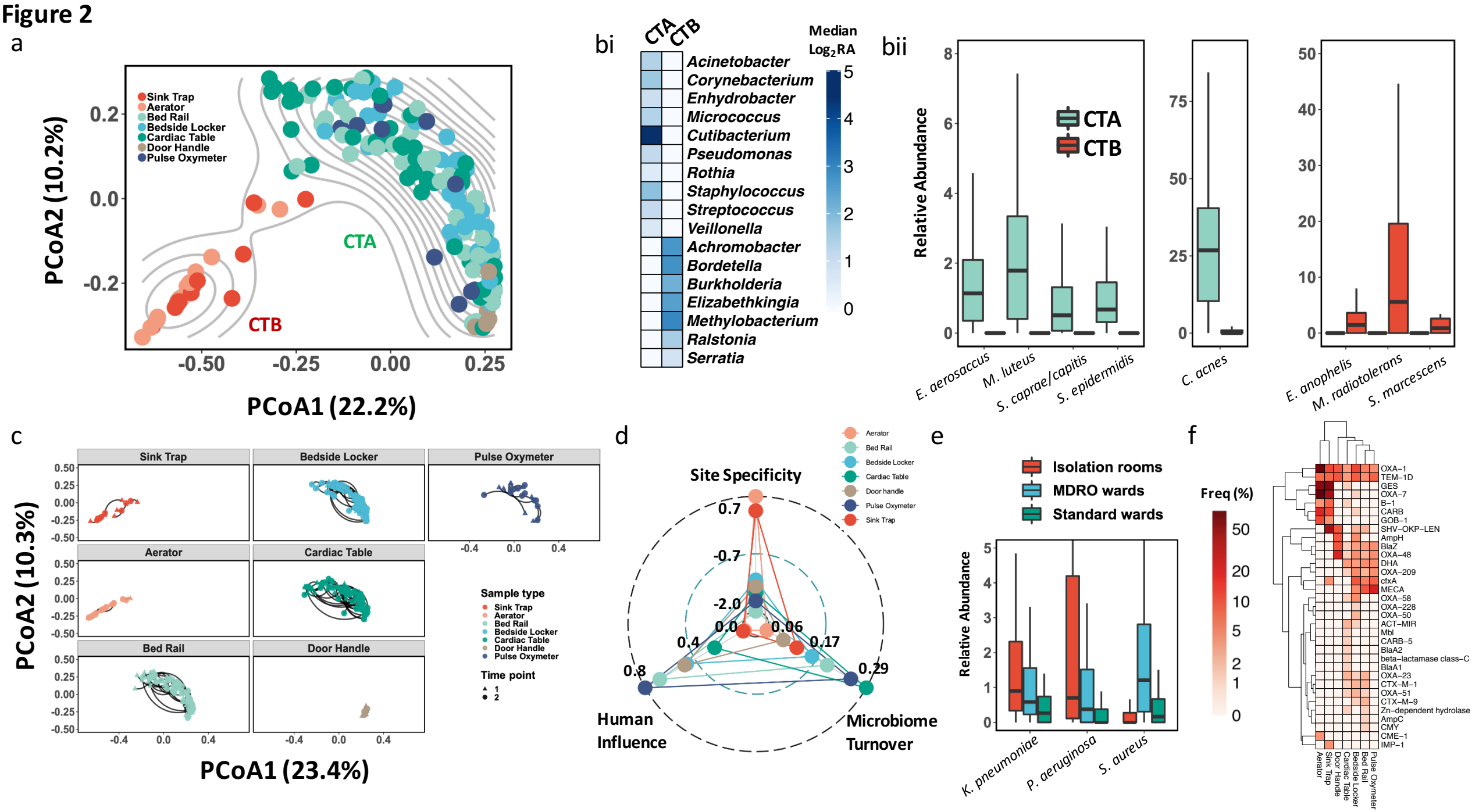

a
Sink Trap
Aerator
Bed Rail
Bedside Locker
Cardiac Table
Door Handle
Pulse Oxymeter
CTA
CTB
PCoA2 (10.2%)
PCoA1 (22.2%)
bi
CTA
CTB
Median Log2RA
Acinetobacter
Corynebacterium
Enhydrobacter
Micrococcus
Cutibacterium
Pseudomonas
Rothia
Staphylococcus
Streptococcus
Veillonella
Achromobacter
Bordetella
Burkholderia
Elizabethkingia
Methylobacterium
Ralstonia
Serratia
bii
Relative Abundance
CTA
CTB
E. aerosaccus
M. luteus
S. caprae/capitis
S. epidermidis
C. acnes
E. anophelis
M. radiotolerans
S. marcescens
c
Sink Trap
Bedside Locker
Pulse Oxymeter
Aerator
Cardiac Table
Bed Rail
Door Handle
PCoA2 (10.3%)
PCoA1 (23.4%)
Sample type
Sink Trap
Aerator
Bed Rail
Bedside Locker
Cardiac Table
Door Handle
Pulse Oxymeter
Time point
1
2
d
Site Specificity
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 3

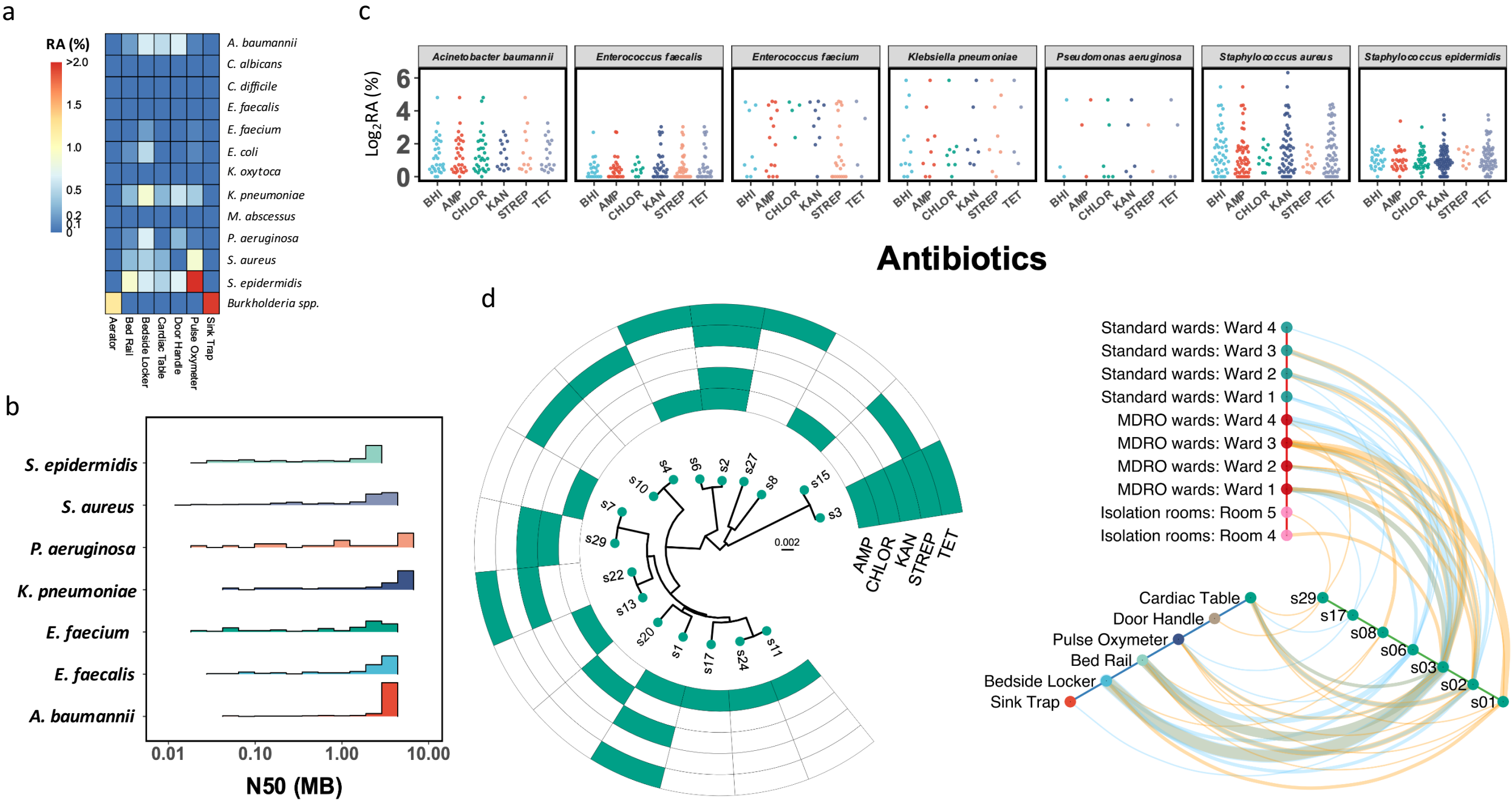

a
RA (%)
>2.0
1.5
1.0
0.5
0.2
0.1
0
A. baumannii
C. albicans
C. difficile
E. faecalis
E. faecium
E. coli
K. oxytoca
K. pneumoniae
M. abscessus
P. aeruginosa
S. aureus
S. epidermidis
Burkholderia spp.
Aerator
Bed Rail
Bedside Locker
Cardiac Table
Door Handle
Pulse Oxymeter
Sink Trap
b
S. epidermidis
S. aureus
P. aeruginosa
K. pneumoniae
E. faecium
E. faecalis
A. baumannii
0.01
0.10
1.00
10.00
N50 (MB)
c
Acinetobacter baumannii
Enterococcus faecalis
Enterococcus faecium
Klebsiella pneumoniae
Pseudomonas aeruginosa
Staphylococcus aureus
Staphylococcus epidermidis
Log2RA (%)
6
4
2
0
BHI
AMP
CHLOR
KAN
STREP
TET
Antibiotics
d
s4
s6
s2
s27
s8
s15
s3
s10
s7
s29
s22
s13
s20
s1
s17
s24
s11
0.002
AMP
CHLOR
KAN
STREP
TET
Standard wards: Ward 4
Standard wards: Ward 3
Standard wards: Ward 2
Standard wards: Ward 1
MDRO wards: Ward 4
MDRO wards: Ward 3
MDRO wards: Ward 2
MDRO wards: Ward 1
Isolation rooms: Room 5
Isolation rooms: Room 4
Cardiac Table
Door Handle
Pulse Oxymeter
Bed Rail
Bedside Locker
Sink Trap
s29
s17
s08
s06
s03
s02
s01

Figure 4

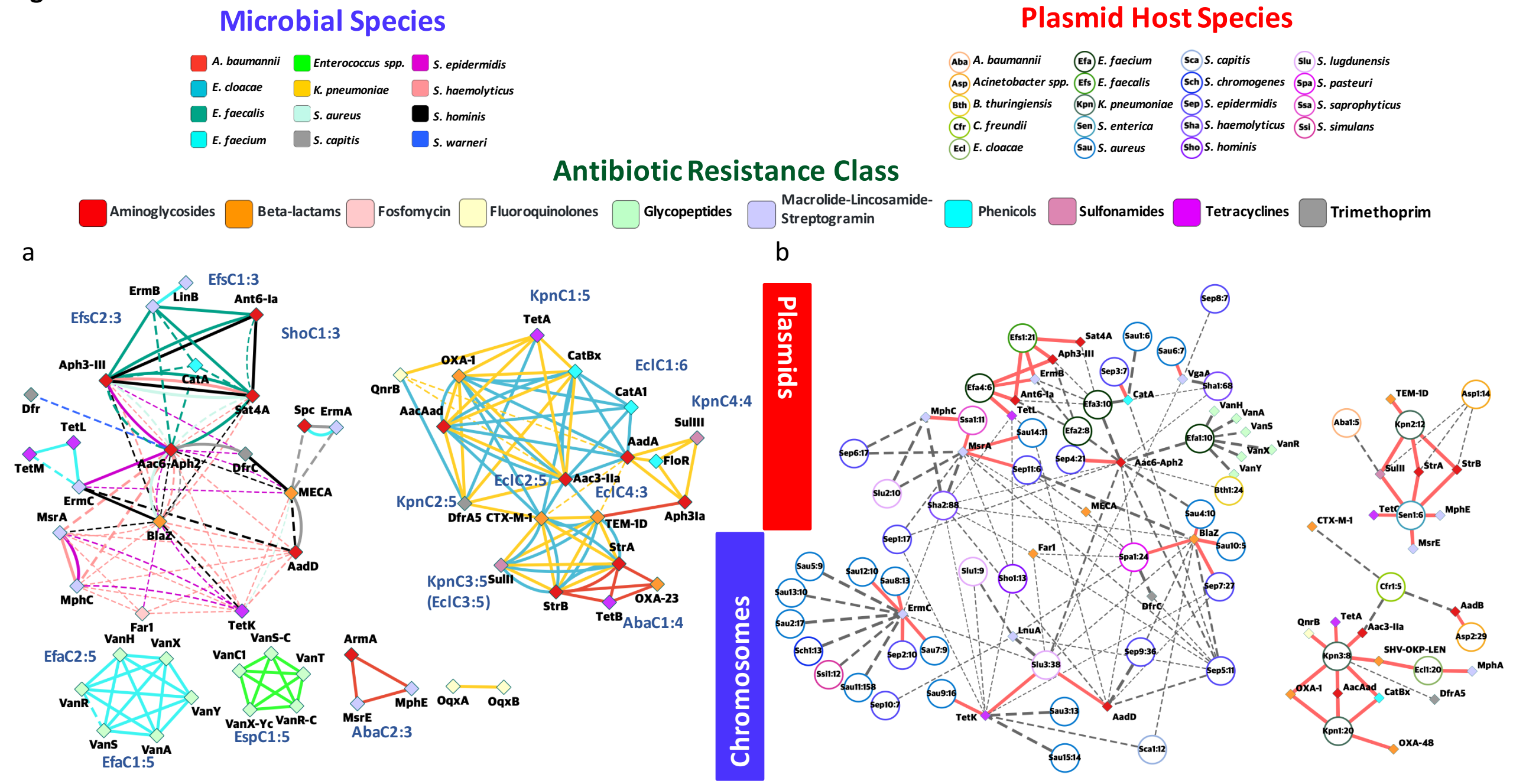
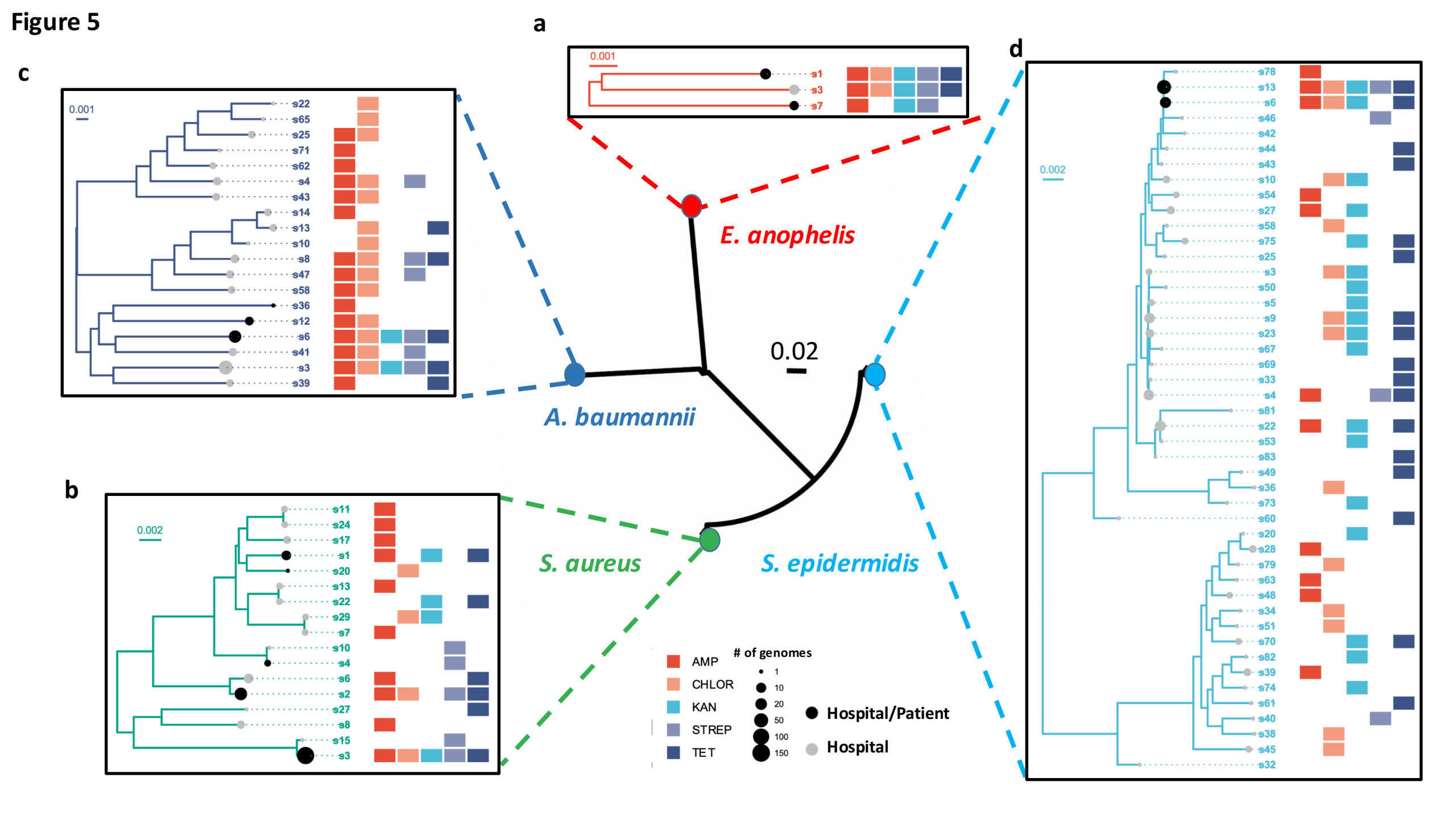

Microbial Species
A. baumannii
E. cloacae
E. faecalis
E. faecium
Enterococcus spp.
K. pneumoniae
S. aureus
S. capitis
S. epidermidis
S. haemolyticus
S. hominis
S. warneri
Plasmid Host Species
Aba A. baumannii
Asp Acinetobacter spp.
Bth B. thuringiensis
Cfr C. freundii
Ecl E. cloacae
Efa E. faecium
Efs 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-lactams
Fosfomycin
Fluoroquinolones
Glycopeptides
Macrolide-Lincosamide-Streptogramin
Phenicols
Sulfonamides
Tetracyclines
Trimethoprim
a
EfsC1:3
ErmB
LinB
Ant6-Ia
EfsC2:3
ShoC1:3
Aph3-III
CatA
Sat4A
Dfr
Spc
ErmA
TetL
TetM
Aac6-Aph2
DfrC
MECA
ErmC
MsrA
BlaZ
AadD
MphC
Far1
TetK
VanH
VanX
EfaC2:5
VanR
VanY
VanS
VanA
EfaC1:5
VanS-C
VanC1
VanT
VanX-Yc
VanR-C
EspC1:5
ArmA
MphE
MsrE
AbaC2:3
OqxA
OqxB
KpnC1:5
TetA
OXA-1
CatBx
EclC1:6
QnrB
CatA1
AacAad
KpnC4:4
SulIII
AadA
FloR
EclC2:5
Aac3-IIa
EclC4:3
KpnC2:5
DfrA5
CTX-M-1
TEM-1D
Aph3Ia
StrA
KpnC3:5
(EclC3:5)
SulII
StrB
TetB
OXA-23
AbaC1:4
b
Plasmids
Chromosomes
Efs1:21
Sat4A
Aph3-III
Sau1:6
Sau6:7
Sep8:7
Sep3:7
ErmB
VgaA
Efa4:6
Ant6-Ia
CatA
Sha1:68
Efa3:10
TetL
VanH
VanA
MphC
Ssa1:11
VanS
Sau14:1
Efa2:8
Efa1:10
VanR
Sep6:17
MsrA
VanX
Sep4:21
Aac6-Aph2
Sep11:6
VanY
Slu2:10
Bth1:24
Sha2:88
MECA
Sau4:10
Sep1:17
BlaZ
Sau10:5
Far1
Spa1:24
Sau5:9
Sau12:10
Sau8:13
Slu1:9
Sho1:13
Sep7:27
Sau13:10
ErmC
DfrC
Sau2:17
LnuA
Sch1:13
Sep2:10
Sau7:9
Sep9:36
Ssi1:12
Slu3:38
Sep5:11
Sau11:158
Sau9:16
Sep10:7
TetK
Sau3:13
AadD
Sca1:12
Sau15:14
Aba1:5
TEM-1D
Asp1:14
Kpn2:12
SulII
StrA
StrB
TetC
Sen1:6
MphE
CTX-M-1
MsrE
Cfr1:5
AadB
QnrB
TetA
Aac3-IIa
Asp2:29
Kpn3:8
SHV-OKP-LEN
Ecl1:20
MphA
OXA-1
AacAad
CatBx
DfrA5
Kpn1:20
OXA-48

Figure 5


a
0.001
s1
s3
s7
E. anophelis
0.02
A. baumannii
S. aureus
S. epidermidis
c
0.001
s22
s65
s25
s71
s62
s4
s43
s14
s13
s10
s8
s47
s58
s36
s12
s6
s41
s3
s39
b
0.002
s11
s24
s17
s1
s20
s13
s22
s29
s7
s10
s4
s6
s2
s27
s8
s15
s3
d
0.002
s78
s13
s6
s46
s42
s44
s43
s10
s54
s27
s58
s75
s25
s3
s50
s5
s9
s23
s67
s69
s33
s4
s81
s22
s53
s83
s49
s36
s73
s60
s20
s28
s79
s63
s48
s34
s51
s70
s82
s39
s74
s61
s40
s38
s45
s32
AMP
CHLOR
KAN
STREP
TET
# of genomes
1
10
20
50
100
150
Hospital/Patient
Hospital