A 21-year survey of *Escherichia coli* from bloodstream infections (BSIs) in a tertiary hospital reveals how community-hospital dynamics, influence local BSI rates, the trends of the B2 phylogroup and the STc131 pandemic clone

Irene Rodríguez,[1] Ana Sofia Figueiredo,[1,2] Melissa Sousa,[1,2] Val F. Lanza, [1,3,4] Concepción Rodríguez,[1] Patricia Mingo,[1] Javier Zamora,[4,5] Elena Loza,[1] Claire J. Brooks,[1] Rafael Cantón,[1,6] Fernando Baquero,[1,4] Teresa M Coque[1,4 *]

[1]Department of Microbiology, Ramón y Cajal University Hospital and Ramón y Cajal Health Research Institute (IRYCIS), Madrid, Spain; [2] Department of Biology, University of Aveiro, 3810-193 Aveiro, Portugal; [3] Bioinformatics Unit, Ramón y Cajal University Hospital and Ramón y Cajal Health Research Institute (IRYCIS), Madrid, Spain; [4] Biomedical Research Center Network of Epidemiology and Public Health (CIBERESP), Madrid, Spain; [5] Biostatistics Unit, Ramón y Cajal University Hospital and Ramón y Cajal Health Research Institute (IRYCIS), Madrid, Spain; [4]Spanish Network for Research in Infectious Diseases (REIPI), Madrid, Spain

Running title: Dynamics of Escherichia coli causing bloodstream infections

Keywords: Escherichia coli, ExPEC, bacteremia, STc131, pandemic clones, long-term dynamics

**Corresponding author.* Teresa M. Coque Mailing address: Servicio de Microbiología. Hospital Universitario Ramón y Cajal. Carretera Colmenar Viejo, Km 9.1. Madrid. Phone: +34913368152. E-mail: teresacoque@gmail.com; mariateresa.coque@salud.madrid.org

ABSTRACT

Escherichia coli is overrepresented in all bloodstream infections (BSIs) series, mostly associated with a few clonal lineages. Its population structure has been analyzed but the dynamics remains to be fully understood. We analyze the dynamics of E. coli-BSIs in a sample of 649 isolates, representing all 7165 E. coli BSI episodes recorded in a tertiary hospital (1996-2016) according to clonal identification (phylogenetic groups/subgroups, STc131 subclades), antibiotic susceptibility (13 antibiotics), and virulence-associated genes (VAGs, 29 genes). Patient data were obtained from the laboratory system and clinical charts. The incidence of BSI-EC doubled from 1996 to 2016 (5.5 to 10.8 BSI episodes/1000 hospitalizations). Intertwined waves of community-acquired (CA) and hospital-acquired isolates (HA) episodes of both B2 and non-B2 phylogroups, occurred until B2 overtook non-B2 BSI episodes. ST131 contributed to increasing the B2 rates, but only transiently altered the population structure. B2 isolates predominates (53%), overrepresented by subgroups B2-I (STc131), B2-II, B2-IX, and B2-VI (25%, 25%, 14%, and 9%). We observed a remarkable increase only for B2-I-STc131 (C1/C2 subclades), a decreasing trend for phylogroup D, and oscillations for other B2 subgroups throughout the years. According to VAG patterns, B2 strains exhibit a population structure compatible with the niche specialization theory. A reservoir of B2 and non-B2 strains represented in human microbiota, flows from the community to the hospital and vice-versa, where they can either be selected or coexist. The increase of BSI is determined by waves of CA that predate the amplification of HA episodes of both B2 and non-B2 phylogroups in various time periods, influenced by FQR and microbiota composition.

46 **IMPORTANCE**

47 The gut microbiota is an important reservoir for bacteria that cause extraintestinal infections
48 including sepsis, which is the third cause of mortality in Western countries and one of the Global
49 Health threads recognized by the WHO since 2017. Most of the bloodstream infections (BSI)
50 and UTIs due to *Escherichia coli* strains originate in the gut microbiota and belong to the
51 phylogenetic group B2. Most B2 strains recovered from BSI infections are clonal lineages
52 predominant in fecal isolates, often associated with outbreaks. Our study analyzes the long-
53 term dynamics of B2 *E. coli* subtypes and reveals waves of different *E.coli* lineages including
54 pandemic clones that emerge periodically and are established in the intestinal microbiota
55 afterward. It also reflects that clonal amplifications in the community predates the clonal
56 increases in hospitals which may favor the acquisition of antibiotic resistance in health centers
57 and further dissemination of well adapted clones that become multidrug resistant.

58

59 **INTRODUCTION**

60 *Escherichia coli* is ubiquitously present in the gut microbiota of humans and other vertebrates,
61 and also represents the primary cause of bloodstream infections (BSIs) and urinary tract
62 infections (UTIs) (1, 2). The increasing and progressive annual rate of BSIs episodes at a global
63 level made the WHO to consider sepsis as a Global Health Threat in 2017 (https://www.global-
64 sepsis-alliance.org/news/2017/5/26/wha-adopts-resolution-on-sepsis). The problem affects
65 more than 30 million of people and increases 9-13% in Western countries where constitutes the
66 third cause of mortality (3–6). The current knowledge suggests differences between countries
67 but such information is highly fractionated in multicentric studies(3), mostly focused on antibiotic
68 resistant BSI isolates (7), and performed at variable periods of time using different sampling
69 criteria (3–8). Only studies from the UK, where surveillance of BSI is compulsory since 2011,
70 provide a long-term comprehensive analysis of BSI and the population structure and the
71 dynamics of *E. coli* causing BSI (6, 9).

72 The gut microbiota is often the origin of extraintestinal infections caused by *E. coli* (10) and vary
73 between humans of different age and lifestyle (10, 11). Among the 7 major phylogenomic
74 groups of the species, the B2 phylogenetic group is currently predominantly detected in clinical
75 and fecal isolates from adults and children of Western areas (10–12). Moreover, the seven *E.*
76 *coli* phylogenetic groups (A, B1, B2, C, D, E, F) and also the 10 B2-*E. coli* subgroups, are
77 consistently overrepresented by certain sequence type complexes (STcs) [e.g.B2-I (STc131),
78 B2-II (STc73), B2-IX (STc95), D (STc69) or F (STc648)], in isolates of different origins although
79 their distribution vary between human populations (13–16).

80 Despite the apparent persistent structure of *E. coli* in the gut of Western individuals, clonal
81 expansions of emerging STcs periodically occurred (15, 17). The STc131 represents the most
82 emblematic example of a global emerging clone nowadays (17). A multicentric study in the UK
83 describes how an emerging and pandemic clone STc131 contributes but only transiently alter
84 the persistent population structure of *E. coli* (9). Nonetheless, the factors behind such *E. coli*

85 clonal expansions are not well understood and different evolutionary pathways including the
86 acquisition of antibiotic resistance, exist for different STcs (18).

87 The number of works related to the *E. coli* population structure contrasts with the scarcity of
88 longitudinal studies necessary for the understanding of long-term population dynamics of the
89 lineages, and the local evolutionarily stable strategies for each lineage that led to local lineage
90 conservation and coexistence (19). Local settings are important sources of information because
91 can reflect the dimensions of human populations structure (age, sex, interconectedness), and
92 offer stability in terms of the intensity of selection (e.g. common policies to control antimicrobial
93 resistance such as antibiotic use, infection control strategies), models of healthcare delivery,
94 and measuring approaches (diagnostic tools) (3, 8).

95 We retrospectively studied a randomized sample of 649 isolates drawn from a collection of 7165
96 *E. coli* isolates, which represented all BSI episodes registered at our institution between 1996
97 and 2016. This period coincided with the global emergence and amplification of various bla_{ESBLs}
98 genes and B2 *E. coli* clones and with the overall increase in the frequency of *E. coli* BSIs. The
99 aim of this study was to infer the local diversity and dynamics of *E. coli* causing BSIs, focusing
100 on the B2 phylogroup and the most abundant contemporary B2 subtype, the pandemic B2-I
101 STc131.

102 **METHODS**

103 **Study design.** Ramón y Cajal University Hospital is a tertiary-level public health center with
104 1155 beds that provides attention to 600,000 habitants in the Northern area of Madrid (Spain),
105 which reflects a pyramid-age of "declining type", has full accessibility to primary attention care
106 and have a predominant medium-high socioeconomic level. Of the total 21,695 positive blood
107 cultures detected between January 1996 and December 2016, we identified 7165 *E. coli* that
108 represented 1 isolate per patient and per BSI episode. A sample of nearly 10% of this *E. coli*
109 collection, stratified by sex, age, and antimicrobial resistance pattern was sorted by statistical
110 randomization and was further analyzed (649 *E. coli* isolates from 339 females and 310 males;
111 < 1-98 years of age). The study was approved by the Ethics Committee of our hospital.

112 BSIs are classified as hospital acquired (HA), community acquired (CA), and community-onset
113 healthcare-associated (HCA), according to the date of the sample collection after patient
114 admission and the patient exposure to hospitals before the BSI episode (8). Due to the
115 inaccessibility to all the medical records of patients enrolled in this study, and their advanced
116 age, we classified the episodes into HA (if the blood culture was obtained at the ICU, surgical,
117 or medical areas after 48 h of admission) and CA+HCA (if the blood culture was obtained at the
118 hospital emergency wards or at the day care centers) categories. UTIs were considered as the
119 origin of BSI if *E. coli* was recovered from both the urine and blood samples, with a difference of
120 ±24 h.

121 **Characterization of the bacterial isolates.** Blood culture isolates of *E. coli* are routinely frozen
122 and stocked in skimmed milk at −70°C, and were subcultured onto brain-heart infusion agar
123 prior to analysis. Bacterial susceptibility against 13 antibiotics (ampicillin, amoxicillin-clavulanic
124 acid, cefotaxime, ceftazidime, meropenem, nalidixic acid, ciprofloxacin, streptomycin,
125 kanamycin, gentamicin, tetracycline, chloramphenicol, and trimethoprim/sulfamethoxazole) was
126 performed by the disk-diffusion method (20).

127 Multiplex PCR assays allowed classifying the *E. coli* isolates into major phylogenetic groups A,
128 B1, B2, C, D, E, and F (21); B2 subgroups (B2 I-X) (22); and B2-I-STc131 *E. coli* serotypes

129 (O16/O25b), clades (H41, H22, H30), and the H30 subclades C0 [H30, fluoroquinolone
130 susceptible (FQ^S)], C1 [H30-R, fluoroquinolone resistant (FQ^R)], C2 (H30Rx, FQ^R+$bla_{CTX-M-15}$),
131 and C1-M27 (H30-Rx, FQ^R+ $bla_{CTX-M-27}$) (22, 23). STc131 isolates were further analyzed by
132 pulsed field gel electrophoresis (PFGE). The presence of 29 virulence-associated genes
133 (VAGs) was determined for all B2 isolates by PCR (2, 24).

134 **Statistical analysis.** To calculate statistical significance, the chi-squared test, a 2-sample t-test
135 for normally distributed variables, and Kendall's correlation were used, considering p-values
136 <0.05 to be statistically significant.

137

138 **RESULTS**

139 **Epidemiology of the 7165 *E. coli* isolates causing BSIs at Hospital Ramón y Cajal**

140 The incidence of BSIs caused by *E. coli* doubled from 1996 to 2016 in our institution (5.5 to 10.8
141 BSI episodes/1000 hospitalizations). Although the overall number of BSI episodes in the
142 hospital and community settings was similar at the beginning of the study in the mid 1990s, we
143 observed a steady increase in both HA-BSIs and CA+HCA-BSIs from 1995 to 2002 (CA/HA
144 ratio >1–2) followed by waves of alternative predominance of either CA-BSIs or HA-BSIs. The
145 increases of BSI episodes in the community predated those in the hospital setting and would
146 explain the wave dynamics between the hospital and the community-based populations and the
147 overall shift in the ratio of BSI acquired in the community and the hospital (Figure 1). The
148 analysis of antimicrobial resistance records in our department for the blood *E. coli* isolates
149 revealed a coincidental increase in the rise of BSI infections and the increasing trends of *E. coli*
150 resistant to fluoroquinolones (FQ^R) from 1994, and resistant to third-generation cephalosporins
151 (3GC) from 2003. Rates of resistance to other antibiotics remained stable during the period of
152 study (data not shown).

153 UTIs were identified as the origin of the BSI in one-third of the cases. This finding was more
154 frequent in women than in men (42.33% vs. 26.77%, $p \leq .005$) but not significantly different
155 between patients of different age groups ($p > .05$). The proportion of polymicrobial BSIs was 9%
156 (n = 57/649), and was similar for men and women (11.8% vs. 7.3% of the BSI cases,
157 respectively).

158 **Clonal diversity of *E. coli* causing BSI**

159 Figure 2 shows the occurrence of various *E. coli* populations, namely phylogenetic groups, B2
160 subgroups, and B2-I (STc131) subclades. The predominant phylogenetic group was B2
161 (348/649; 53.06%), followed at a much lower frequency by D (11,4%), B1 (7,24%), A (6.3%), C
162 (5.,1%), F (6.8%), E (0.9%) and Clades I and II (0.9%). Half of the B2 isolates corresponded to
163 subgroups B2-I (25.6%) and B2-II (25.1%), which were followed by B2-IX (14.2%), B2-VI (9.5%)
164 and others. Almost 10% of the isolates were catalogued as The STc131 isolates, clearly

165 predominant within the B2-I subgroup, represents 12.3% of the total BSI *E. coli* isolates and
166 21.8% of the B2 phylogroup. The STc131 isolates (serogroups O25b and O16 representing
167 92% and 8%, respectively) split in clade A (7/82, 9%), clade B (16/82, 20%), and clade C (59/82,
168 71%). Clade C comprised isolates of subclades C1 (29/82, 35.4%), C2 (22/82, 26.8%), C0 (6/82,
169 7.3%), and C1-M27 (2/82, 2.4%).

170 *Distribution by Age*. All phylogenetic groups were present in patients of all age groups, although
171 differences were observed for subgroups B2-I (STc131), B2-II, B2-VI, and B2-IX. This
172 distribution implies the clear predominance of B2-I in the elderly (>80 years) and B2-II in
173 individuals younger than 45 years. Others showed a bimodal distribution as B2-IX and B2-VI
174 (Figure S1).

175 *Acquisition of the BSIs*. Most B2 isolates were community based although intertwined waves of
176 CA+HCA-BSI and HA-BSI episodes of strains of both B2 and non-B2 phylogroups, which
177 occurred synchronously until 2008, when B2 clearly overtook non-B2 BSI episodes and STc131
178 become transiently predominant (Figure S2). Almost half of patients with STc131 (46%) were
179 admitted to the medical emergency ward with an established BSI, suggesting the frequent
180 presence of this clone in the community. As an example, we detected among these patients
181 several coincidences of strains with the same PFGE pattern corresponding to clades H22
182 (2000), C1 (2004, 2005), and C2 (2008-2011) (data not shown).

183 *Temporal variation*. Except for phylogroups B2-I (ST131) and phylogroup D, the trends of the
184 phylogroups did not significantly change during the period analyzed. However, the occurrence
185 of major subgroups (B2-II, B2-IX, B2-VI, B-IV, B2-VII) greatly varied through the years
186 suggesting episodes of transmission with transient amplifications as occurred for B2-I (STc131)
187 (Figure 3).

188 The STc131 *E. coli* was the only group increasingly recovered coinciding with its global
189 amplification (17). The strains of ST131 clade B were initially detected in 1996, and remained
190 steadily identified since then. In the current study, the ST131 of clade A was first detected in
191 2004, but we retrospectively identified STc131 clade A in clinical isolates of TEM-4 and TEM-24

192 producers from 1991 and 2000, respectively (25, 26) For the predominant ST131 clade C, the
193 subclade C0 (H30, FQ^S) emerged in 2000, followed by C1 (H30-R, FQ^R) in 2004, C2 (H30-Rx,
194 FQ^R $bla_{CTX-M-15}$) in 2006, and C1-M27 (H30-Rx, FQ^R $bla_{CTX-M-27}$) in 2016 (Figure 4). The number
195 of *E.coli* strains of phylogroup D decreased from the beginning of the study, but it varied
196 significantly through the years.

197 **Antimicrobial susceptibility**.

198 Figure 5 shows the antimicrobial susceptibility patterns in the sample of 649 *E. coli* isolates,
199 revealing that FQ^R appears in between 35% and 50% of the isolates of each phylogroup, with
200 B2 being less prevalent (17%), although it is of note that most FQ^R B2 isolates correspond to
201 STc131 (70% vs. 3.5% in non-ST131 B2). Only 6.1% of the total number of B2 isolates showed
202 a $3GC^R$ phenotype compatible with the production of an extended-spectrum beta-lactamase
203 (ESBL), further identified as CTX-M-15, CTX-M-14, and CTX-M-27, and was only detected
204 among STc131 isolates. Resistance to ampicillin (70.4%), streptomycin (39.6%), nalidixic acid
205 (32.8%), and cotrimoxazole (45%–50%) was frequent among the isolates of the various
206 phylogroups. Remarkable differences were observed for amoxicillin-clavulanic acid, kanamycin,
207 gentamicin, tetracycline, and chloramphenicol, mostly due to the phylogroup C isolates.
208 Susceptibility to the 13 antimicrobials tested was observed in 17% of the total number of
209 isolates tested, with the frequency of susceptible isolates higher among those of phylogroups A,
210 B1, B2, and F (15% each) than those of phylogroups D (10%), E (1%), and C (0%). Within B2, a
211 pan-susceptibility pattern was more frequent for non-STc131 than for STc131 isolates (26.8%
212 vs. 3.7%). The B2-STc131 strains showed a multiresistant profile (resistance against ≥4
213 antimicrobial agents) more frequently than other B2 members (56.6% vs. 11.6%; $p < .001$).

214

215 **Virulence-associated gene profiling in the B2 phylogroup**

216 According to the content of the VAG genes, the B2 *E. coli* strains were clustered into 2 large
217 groups and 10 subgroups, each comprising numerous gene combinations and showing variable
218 redundancies (Figure 6). Despite some VAG variability within each B2 subgroup, the results

219 suggest a conserved genetic structure related to virulence and colonization in the B2 strains
220 analyzed in this study. More than half (56%) of the B2-I (STc131) strains clustered in VAG 1.5.
221 Most (54%–75%) of the *E. coli* phylogenetic subgroup isolates B2-II, B2-III, B2-V, B2-VII, B2-IX,
222 and B2-X clustered in the VAG 2.6, 2.2, 2.9, 2.6, 2.7, and 2.4 groups, respectively (ordered by
223 frequency). A detailed analysis of virulence content is provided in Supplementary Text and
224 Figures S3, S4, S5, S6.

225

226 **DISCUSSION**

227 This work, one of the very few studies -and the first in Spain- to analyse the dynamics of *E. coli*
228 strains for a long period of time at a single centre, reveals intertwined waves of CA+HCA-BSI
229 and HA-BSI episodes of strains of both B2 and non-B2 phylogroups, which occurred
230 synchronously until B2 overtook non-B2 BSI episodes and the B2-I-STc131 became transiently
231 predominant and predate new waves of other major B2 subgroups. This dynamics is similar to
232 that observed in the UK (18) but further suggest that the burden of CA-BSI and HA-BSI
233 infections are closely related. Besides the risk of cross-infections/colonizations in the hospital
234 that led to local increases in BSIs in particular periods, species or clones which are part of the
235 microbiota of healthy individuals in the community might occasionally have an epidemic
236 structure, in our case providing a source of *E. coli* strains influencing the incidence of hospital
237 BSIs.

238 *E. coli* bacteremia, which was not considered common at the beginning of the 20th century, has
239 steadily increased from the past to the present, according to long surveys performed at Boston
240 City Hospital between 1935 and 1972 (27), at St. Thomas hospital in London between 1969 and
241 1986 (28), and more recently, at our and other hospitals in Western countries (most cross-
242 sectional or longitudinal multicentric studies) (3, 6, 8, 9, 29). The local survey performed during
243 the 1980s (28) reflected, for the very first time, a relatively constant proportion of *E. coli* among
244 all cases of BSI with where the incidence peak was coincidental with a community-based clonal
245 outbreak which led to the introduction of an *E. coli* O15:H12 clone in the hospital setting and a
246 subsequent increase in nosocomial BSI cases (28). Shortly afterwards, community-based and
247 hospital clonal outbreaks by *E. coli* of phylogroup D were extensively reported during the 1990s,
248 namely ST69 in the US and ST393 O15:H12 in the UK and Spain (15). During the 2000s, the
249 increasing of CA outbreaks of B2 *E. coli* lineages as STc73, STc95 and more recently, STc131,
250 became common (9, 14). Massot et al demonstrated that the shift in the population structure of
251 commensal *E. coli* from A and D groups in 1970s to B2 and F in 2000s could be influenced by
252 human interventions as antibiotic use (12). However, clonal expansions at local level have been

253 poorly analyzed and always under the perspective of antimicrobial resistance. Significant
254 changes in healthcare delivery and in the human population structure (age, interconnectedness,
255 diet), have been recently suggested to explain both the transmission (amplification) of
256 microorganisms and the acquisition of resistance even in the absence of personal history of
257 antibiotic consumption and other known risk factors for antimicrobial resistance (30–32).

258 The most comprehensive recent analysis, using European Antimicrobial Resistance
259 Surveillance System (EARSS) data from 2002 to 2008, highlighted a remarkable average
260 annual increase of 29.9% in the number of reported bacteremias caused by *E. coli* isolates
261 resistant to 3GCs. However, such increases were less noticeable in countries from Southern
262 Europe (3), probably because of the overrepresentation of major *E. coli* in hospitals predates
263 the acquisition of genes coding for resistance to first-line beta-lactam antibiotics (9, 13). What
264 we should be aware of the steady increase in the occurrence of FQ^R in all non-B2 groups since
265 the early 1990s that would reflect the selection, amplification and stabilization of certain *E. coli*
266 lineages as some of the phylogroups D, and more recently, of STc131 of phylogroup B2. These
267 lineages would have also been circulating in the community for decades before the emergence
268 of STc131 clade C, a FQ^R lineage that, in contrast with others, is predominant in individuals
269 >80y. FQ^R has recently been considered an advantage for the fitness of high-risk clones of *E.*
270 *coli* and other species (33, 34). Some FQ^R clones are predominant in the elderly, with a
271 sustained pattern in hospital (or day-care-center) admissions and personal clinical history that
272 facilitates the acquisition of other antibiotic resistances (35). An increasing cumulative "elderly
273 microbiota reservoir" of antimicrobial-resistant subpopulations of various bacterial species would
274 be congruent with the overrepresentation of community-based individuals with a BSI by STc131
275 at admission, with the age of the patients in our sample, and with the identification of identical
276 PFGE types in groups of unrelated patients. However, this does not fully explain the
277 overrepresentation of certain B2 lineages which are amplified time to time as it reflects the
278 waves of different B2 subgroups in Figure 3.

279 B2-I (STc131) was the only lineage that significantly increased during the period analyzed, and
280 this transient increase was due to the expansion of both C1 and C2 subclades, and probably, to
281 CTX-M-27 in the last years not included in the study according to the emergence of this
282 subclade in the community (36). Although the epidemiology of ST131 has extensively been
283 documented (17), this work and that by Kallonen (9) are the only ones that reveals the evolution
284 and coexistence of all STc131 subclades which, after a transient amplification, are incorporated
285 into the *E. coli* population structure (17, 36). The other predominant B2 subgroups varied
286 greatly through the period studied, suggesting some peaks of prevalent B2-II, B2-IX, B2-VI
287 clones and also others as B2-VII and B2-X and non-I/II which are poorly documented in the
288 literature. The correlation between VAG profiles and B2 subgroups reflects an apparent
289 structure of B2 populations in agreement with the niche specialization hypothesis (10, 37–39).
290 Although the concordance of specific VAGs with isolates of B2-II and B2-IX subgroups is
291 congruent with their categorization as uropathogenic *E. coli*, such a way explaining their
292 bimodal distribution in different age groups, most B2-I strains (STc131 clade C) exhibit a low
293 number of classical EC-VAGs while retaining some traits that would facilitate colonization and
294 persistence in the gut (40, 41). To explain the expansion of the STc131 and other emerging
295 lineages, such as STc648 (phylogroup F), studies that have applied mathematical models to a
296 high number of genomes suggest that negative frequency-dependent selection of previously
297 rare populations might have favored their increase (16, 41), and will foster the expansion of
298 some others in the future. However, these studies are fully focused on the microorganisms and
299 do not allow us to associate the observed changes with the type of hosts located either in
300 hospital or community compartments.

301 We are aware that the size of the sample analyzed (despite of the high number of isolates
302 analyzed it represents the 10% of the total number of BSI in our institution), and the lack of
303 whole genome comparative analysis might limit our conclusions about the clonal dynamics.
304 However, the long period analyzed at a single institution helps identifying relevant
305 epidemiological conditions that explain the expansion or particular *E. coli* populations observed

306 in our and other areas during recent years. Currently, the B2 *E. coli* strains are relevant units of
307 BSI pathogenicity, which should correlate with their success in particular microecological
308 landscapes, in part determined by interventions exerted on particular human populations
309 (particularly elderly patients). This view is in agreement with the concept of incorporating
310 ecological features in the identification of the fundamental units of bacterial diversity (38) and
311 pathogenicity. In fact, particular clones and lineages are differentially represented in a "human
312 microbiota reservoir" (elderly being a reservoir of high-risk B2 lineages) flowing from the
313 community to the hospital and vice versa, where they can either be selected or coexist as
314 predicted by an evolutionarily stable strategy (38). The evolutionary trajectories of recently
315 amplified major lineages indicate the relevance of undetected selective events, which could be
316 further amplified by the acquisition of antimicrobial resistance in settings under (or not) antibiotic
317 pressure (31, 32, 42). The early detection of the abundance and diversity of B2 subgroups of
318 clinical significance for BSI in metagenomic samples and the analysis of common causes that
319 enhance their selection, either in the hospital, nursing homes or the community, are priority
320 research challenges that warrant attention in an era that favor the pandemics of microorganisms
321 and antimicrobial resistance (43).
322

323 **FUNDING**

324 This work was supported by the European Commission (grants ST131TS AC00043/17 and
325 EVOTAR-FP7-HEALTH-282004) and the Instituto de Salud Carlos III of Spain, co-financed by
326 the European Development Regional Fund (A Way to Achieve Europe program; Spanish
327 Network for Research in Infectious Diseases grant RD12/0015/0004 and RD16/0016/0011, an
328 CIBER CB06/02/0053). During the performance of this study, I.R. was a recipient of a
329 postdoctoral contract "Sara Borrell" (Ref. CD12/00492) from the Instituto de Salud Carlos III
330 (Spain), MS and ASF were recipients of an ErasmusPlus fellowship, and PM was funded by the
331 Programa Operativo de Empleo Juvenil (PEJ-2017-AI/BMD-7200).
332

333 **CONFLICTS OF INTEREST**

334 None.

335

336 **AUTHORS' CONTRIBUTIONS**

337 I.R. performed the wetlab work of the isolates collected from 1996–2012, assisted by C.R., and
338 participated in the study design, in the data analysis, and the writing of the initial draft of the
339 manuscript. A.S.F and M.S., assisted by PM, performed the wetlab work of the isolates
340 collected from 2012–2016, the categorization of all ST131s included in the study, the data
341 analysis, and the revision of the manuscript. V.F.L performed the bioinformatic and statistical
342 analysis of the results, most of the figures, and the revision of the manuscript. J.Z, E.L., and R.C.
343 participated in the study design and the revision of the manuscript. C.J.B. provided expertise
344 and participated in the revision of the manuscript. F.B. and T.M.C. participated in the study
345 design, data analysis, and wrote the manuscript. All the authors have read and approved the
346 final version of this manuscript.

347

348 **FIGURE CAPTIONS**

349 **Figure 1**. Incidence and Origin of BSI episodes caused by *Escherichia coli a*t the HRyC (1994-
350 2016). (A) Incidence of E. coli BSI (episodes/1000 hospitalizations); (B) Occurrence of
351 HCA+CA and HA episodes of BSI; (C) Ratio of HCA+CA/HA BSI episodes.

352 Abbreviations: HA = hospital acquired; CA= community acquired; HCA = community-onset
353 healthcare-associated.

354 **Figure 2**. Distribution of *E. coli* phylogroups and B2 subgroups. (A) Occurrence of the *E. coli*
355 phylogroups; *(*B) Occurrence of *E. coli* B2 subgroups A, B, C; (C) Occurrence of B2-I (STc131)
356 subclades C0/H30 (FQ^S), C1/H30-R (FQ^R), C2/H30Rx (FQ^R+$bla_{CTX\text{-}M\text{-}15}$), and C1-M27/H30-R
357 (FQ^R+ $bla_{CTX\text{-}M\text{-}27}$).

358 **Figure 3**. Temporal Distribution of *E. coli* populations at HRyC (1996-2016). (A and B)
359 Temporal distribution of *E. coli* phylogroups; (C and D) Temporal distribution of *E. coli* B2

360 subgroups; (C) Distribution of STc131 clades; (D) Timeline of the STc131 subclades C0, C1.

361 **Figure 4**. Temporal Distribution of *E. coli* populations at HRyC (1996-2016). (A) Distribution of
362 STc131 clades; (B) Timeline of the STc131 subclades C0, C1.

363 **Figure 5**. Antibiotic susceptibility of B2-*Escherichia coli* causing BSI. Abbreviations: AMC,
364 amoxicillin-clavulanic acid; AMP, ampicillin; CAZ, ceftazidime; CHL, chloramphenicol; CIP,
365 ciprofloxacin; CTX, cefotaxime; GEN, gentamicin; KAN, kanamycin; MEM, meropenem; NAL,
366 nalidixic acid; STR, streptomycin; TET, tetracycline; TMP/SXT, trimethoprim-sulfamethoxazole;
367 Susceptible, susceptible to all 13 antibiotics analyzed.

368 **Figure 6**. Heatmap of 29 virulence factors (presence/absence) for the 348 B2 phylogenetic
369 group strains causing BSI at Ramón y Cajal University Hospital (1996–2016). VAGs: Virulence
370 Associated groups. VAGs were determined by hierarchical clustering using the Ward method
371 and Jaccard similarity distance.

372 Abbreviations of the VAGs tested. *Toxins*: *hlyA* (α-Hemolysin), *sat* (secreted autotransporter
373 toxin, a serine protease), *cnf1* (cytotoxic necrotizing factor 1), *cdtB* (cytolethal distending
374 toxin), *tsh* (temperature-sensitive hemagglutinin); *siderophores*: *iroN* (salmochelin receptor),
375 *iutA* (aerobactin synthesis, receptor), *ireA* (iron-regulated element, catecholate siderophore),
376 *fyuA* (yersiniabactin receptor); *adhesins***:** fimbriae type P (*pap*GI, *pap*GII, *pap*GIII, *pap*A, *pap*C,
377 *pap*EF), *sfa/focDE* (type S fimbriae, sfa/foc S and F1C fimbriae); *afa/draBC,* Adhesinsd afa/dra
378 Dr antigen-binding adhesins (AFA I-III, Dr, F1845; *bmaE* (blood group M-specific adhesin; *iha*
379 (iron-regulated-gene-homologue adhesin); *protectins*: kpsMT II (group II capsule synthesis,
380 e.g., K1, K5, K12). kpsMT III Group III capsule synthesis (e.g., K3, K10, K54), *traT* (surface
381 exclusion; serum resistance-associated), *invasins*: ibeA-C (invasion of brain endothelium IbeA);
382 *miscellanea*: *cvaC* (microcin/colicin V; on plasmids with *traT, iss, iuc/iut*), *ompT* (outer
383 membrane protein T), *usp* (uropathogenic-specific protein, bacteriocin), PAI (malX, a PAI
384 CFT073 marker), *iss* (increased serum survival, outer membrane protein).

385 **Supplementary Figure S1**. Diversity of *E. coli* per age group. (A) Per phylogroup; (B) per B2

386 subgroup. Half of the B2 isolates corresponded to B2-II and B2-IX (33% and 21%) for patients
387 between 15 and 45 years, B2-I and B2-II (23% and 25%–28%) for patients between 45 and 80
388 years, and B2-I and B2-IX (36% and 20%) for patients >80 y

389 **Supplementary Figure S2.** Evolution in time of B2 and non-B2 lineages in hospital acquired
390 (HA) and community acquired (CA) BSI E. coli isolates.

391 The first increase in BSI (1994/1996–1999) occurred in the hospital and the community. While
392 the CA-BSI cases involve predominantly B2 but also non-B2 strains, the HA-BSI increases at
393 the expenses of episodes due to non-B2 strains. This was followed by an increase of CA-BSIs
394 (1999–2002) that involved predominantly non-B2 strains and an increase of HA-BSIs (2003–
395 2006) caused by non-B2 and B2 strains. Between 2006 and 2008, the ocurrence of CA-BSIs
396 reached that of HA-BSIs. The increase in B2 strains from the community predates a new
397 increase of nosocomial BSI mostly caused by B2 strains in the hospital (2009–2011) and
398 afterward, another convergence in frequency of HA and CA-BSI involved an increase in B2
399 strains in the community. Finally, the predominance of BSI in the community in 2016 involved
400 the increase of episodes caused by both B2 and non-B2 strains.

401 **Supplementary Figure S3.** Heatmap displaying the relative abundance of B2 phylogenetic
402 groups and VAG profiles. Variables showing goodness-of-fit are shown in red, whereas dark
403 blue regions indicate poor frequencies. Clustering of the B2 subgroups and VAGs and the
404 assembly statistics were performed using a hierarchical clustering method in R (hclust).
405 Abbreviations of the VAGs tested. *Toxins*: *hlyA* (α-Hemolysin), *sat* (secreted autotransporter
406 toxin, a serine protease), *cnf1* (cytotoxic necrotizing factor 1), *cdtB* (cytolethal distending
407 toxin), *tsh* (temperature-sensitive hemagglutinin); *siderophores*: *iroN* (salmochelin receptor),
408 *iutA* (aerobactin synthesis, receptor), *ireA* (iron-regulated element, catecholate siderophore),
409 *fyuA* (yersiniabactin receptor); *adhesins*: fimbriae type P (*pap*GI, *pap*GII, *pap*GIII, *pap*A, *pap*C,
410 *pap*EF), *sfa/focDE* (type S fimbriae, sfa/foc S and F1C fimbriae); *afa/draBC,* Adhesinsd afa/dra
411 Dr antigen-binding adhesins (AFA I-III, Dr, F1845; *bmaE* (blood group M-specific adhesin; *iha*
412 (iron-regulated-gene-homologue adhesin); *protectins*: kpsMT II (group II capsule synthesis,

413 e.g., K1, K5, K12). kpsMT III Group III capsule synthesis (e.g., K3, K10, K54), *traT* (surface
414 exclusion; serum resistance-associated), *invasins*: ibeA-C (invasion of brain endothelium IbeA);
415 *miscellanea*: *cvaC* (microcin/colicin V; on plasmids with *traT, iss, iuc/iut*), *ompT* (outer
416 membrane protein T), *usp* (uropathogenic-specific protein, bacteriocin), PAI (malX, a PAI
417 CFT073 marker), *iss* (increased serum survival, outer membrane protein).

418 **Supplementary Figure S4.** Heatmap displaying the relative abundance of B2 phylogenetic
419 groups and VAGs. Variables showing goodness-of-fit are shown in red, whereas dark blue
420 regions indicate poor performance. Clustering of the B2 subgroups and VAGs and the
421 assembly statistics were performed using a hierarchical clustering method in R (hclust).

422 Abbreviations of the VAGs tested. *Toxins*: *hlyA* (α-Hemolysin), *sat* (secreted autotransporter
423 toxin, a serine protease), *cnf1* (cytotoxic necrotizing factor 1), *cdtB* (cytolethal distending
424 toxin), *tsh* (temperature-sensitive hemagglutinin); *siderophores*: *iroN* (salmochelin receptor),
425 *iutA* (aerobactin synthesis, receptor), *ireA* (iron-regulated element, catecholate siderophore),
426 *fyuA* (yersiniabactin receptor); *adhesins***:** fimbriae type P (*pap*GI, *pap*GII, *pap*GIII, *pap*A, *pap*C,
427 *pap*EF), *sfa/focDE* (type S fimbriae, sfa/foc S and F1C fimbriae); *afa/draBC,* Adhesinsd afa/dra
428 Dr antigen-binding adhesins (AFA I-III, Dr, F1845; *bmaE* (blood group M-specific adhesin; *iha*
429 (iron-regulated-gene-homologue adhesin); *protectins*: kpsMT II (group II capsule synthesis,
430 e.g., K1, K5, K12). kpsMT III Group III capsule synthesis (e.g., K3, K10, K54), *traT* (surface
431 exclusion; serum resistance-associated), *invasins*: ibeA-C (invasion of brain endothelium IbeA);
432 *miscellanea*: *cvaC* (microcin/colicin V; on plasmids with *traT, iss, iuc/iut*), *ompT* (outer
433 membrane protein T), *usp* (uropathogenic-specific protein, bacteriocin), PAI (malX, a PAI
434 CFT073 marker), *iss* (increased serum survival, outer membrane protein).

435 **Supplementary Figure S5.** Heatmap the relative abundance of each VAG in each VAG group.
436 Variables showing goodness-of-fit. Abbreviations of the VAGs tested. *Toxins*: *hlyA* (α-
437 Hemolysin), *sat* (secreted autotransporter toxin, a serine protease), *cnf1* (cytotoxic necrotizing
438 factor 1), *cdtB* (cytolethal distending toxin), *tsh* (temperature-sensitive hemagglutinin);
439 *siderophores*: *iroN* (salmochelin receptor), *iutA* (aerobactin synthesis, receptor), *ireA* (iron-

440 regulated element, catecholate siderophore)*, fyuA* (yersiniabactin receptor); *adhesins***:** fimbriae
441 type P (*pap*GI, *pap*GII, *pap*GIII, *pap*A, *pap*C, *pap*EF), *sfa/focDE* (type S fimbriae, sfa/foc S and
442 F1C fimbriae); *afa/draBC,* Adhesinsd afa/dra Dr antigen-binding adhesins (AFA I-III, Dr, F1845;
443 *bmaE* (blood group M-specific adhesin*; iha* (iron-regulated-gene-homologue adhesin);
444 *protectins*: kpsMT II (group II capsule synthesis, e.g., K1, K5, K12). kpsMT III Group III capsule
445 synthesis (e.g., K3, K10, K54), *traT* (surface exclusion; serum resistance-associated), *invasins*:
446 ibeA-C (invasion of brain endothelium IbeA); *miscellanea*: *cvaC* (microcin/colicin V; on
447 plasmids with *traT, iss, iuc/iut*), *ompT* (outer membrane protein T), *usp* (uropathogenic-specific
448 protein, bacteriocin), PAI (malX, a PAI CFT073 marker), *iss* (increased serum survival, outer
449 membrane protein).

450 **Supplementary Figure S6.** Heatmap of 29 VAGs (presence/absence) for 82 B2 ST131
451 causing BSIs at Ramón y Cajal University Hospital (1996–2016). The heatmap was annotated
452 with the FimH allele, ST131 serotype, patient age, and sex. Abbreviations of the VAGs tested.
453 *Toxins*: *hlyA* (α-Hemolysin), *sat* (secreted autotransporter toxin, a serine protease), *cnf1*
454 (cytotoxic necrotizing factor 1), *cdtB* (cytolethal distending toxin), *tsh* (temperature-sensitive
455 hemagglutinin); *siderophores*: *iroN* (salmochelin receptor)*, iutA* (aerobactin synthesis,
456 receptor)*, ireA* (iron-regulated element, catecholate siderophore)*, fyuA* (yersiniabactin
457 receptor); *adhesins***:** fimbriae type P (*pap*GI, *pap*GII, *pap*GIII, *pap*A, *pap*C, *pap*EF), *sfa/focDE*
458 (type S fimbriae, sfa/foc S and F1C fimbriae); *afa/draBC,* Adhesinsd afa/dra Dr antigen-binding
459 adhesins (AFA I-III, Dr, F1845; *bmaE* (blood group M-specific adhesin*; iha* (iron-regulated-
460 gene-homologue adhesin); *protectins*: kpsMT II (group II capsule synthesis, e.g., K1, K5, K12).
461 kpsMT III Group III capsule synthesis (e.g., K3, K10, K54), *traT* (surface exclusion; serum
462 resistance-associated), *invasins*: ibeA-C (invasion of brain endothelium IbeA); *miscellanea*:
463 *cvaC* (microcin/colicin V; on plasmids with *traT, iss, iuc/iut*), *ompT* (outer membrane protein T),
464 *usp* (uropathogenic-specific protein, bacteriocin), PAI (malX, a PAI CFT073 marker), *iss*
465 (increased serum survival, outer membrane protein).

466

467 **REFERENCES**

468 1. **Tenaillon O**, **Skurnik D**, **Picard B**, **Denamur E**. 2010. The population genetics of
469 commensal Escherichia coli. Nat. Rev. Microbiol.

470 2. **Johnson JR**, **Russo TA**. 2018. Molecular Epidemiology of Extraintestinal Pathogenic
471 Escherichia coli. EcoSal Plus **8**.

472 3. **de Kraker MEA**, **Jarlier V**, **Monen JCM**, **Heuer OE**, **van de Sande N**, **Grundmann H**.
473 2013. The changing epidemiology of bacteraemias in Europe: trends from the European
474 Antimicrobial Resistance Surveillance System. Clin. Microbiol. Infect. **19**:860–8.

475 4. **Jarlier V**, **Diaz Högberg L**, **Heuer OE**, **Campos J**, **Eckmanns T**, **Giske CG**,
476 **Grundmann H**, **Johnson AP**, **Kahlmeter G**, **Monen J**, **Pantosti A**, **Rossolini GM**, **van**
477 **de Sande-Bruinsma N**, **Vatopoulos A**, **Żabicka D**, **Žemličková H**, **Monnet DL**,
478 **Simonsen GS**, **Ears-Net Participants**. 2019. Strong correlation between the rates of
479 intrinsically antibiotic-resistant species and the rates of acquired resistance in Gram-
480 negative species causing bacteraemia, EU/EEA, 2016. Euro Surveill. **24**.

481 5. **Laupland KB**, **Gregson DB**, **Church DL**, **Ross T**, **Pitout JDD**. 2008. Incidence, risk
482 factors and outcomes of Escherichia coli bloodstream infections in a large Canadian
483 region. Clin. Microbiol. Infect. **14**:1041–1047.

484 6. **Bou-Antoun S**, **Davies J**, **Guy R**, **Johnson AP**, **Sheridan EA**, **Hope RJ**. 2016.
485 Descriptive epidemiology of Escherichia coli bacteraemia in England, April 2012 to March
486 2014. Euro Surveill. **21**.

487 7. **Rodríguez-Baño J**, **Picón E**, **Gijón P**, **Hernández JR**, **Ruíz M**, **Peña C**, **Almela M**,
488 **Almirante B**, **Grill F**, **Colomina J**, **Giménez M**, **Oliver A**, **Horcajada JP**, **Navarro G**,
489 **Coloma A**, **Pascual A**. 2010. Community-Onset Bacteremia Due to Extended-Spectrum
490 β-Lactamase–Producing Escherichia coli: Risk Factors and Prognosis . Clin. Infect. Dis.
491 **50**:40–48.

492 8. **Laupland KB**, **Church DL**. 2014. Population-based epidemiology and microbiology of

493 community-onset bloodstream infections. Clin. Microbiol. Rev. **27**:647–664.

494 9. **Kallonen T**, **Brodrick HJ**, **Harris SR**, **Corander J**, **Brown NM**, **Martin V**, **Peacock SJ**,
495 **Parkhill J**. 2017. Systematic longitudinal survey of invasive Escherichia coli in England
496 demonstrates a stable population structure only transiently disturbed by the emergence
497 of ST131. Genome Res. **27**:1437–1449.

498 10. **Le Gall T**, **Clermont O**, **Gouriou S**, **Picard B**, **Nassif X**, **Denamur E**, **Tenaillon O**. 2007.
499 Extraintestinal virulence is a coincidental by-product of commensalism in B2 phylogenetic
500 group Escherichia coli strains. Mol. Biol. Evol. **24**:2373–84.

501 11. **Nowrouzian FL**, **Clermont O**, **Edin M**, **Östblom A**, **Denamur E**, **Wold AE**, **Adlerberth I**.
502 2019. Escherichia coli phylogenetic B2 subgroups in the infant gut microbiota:
503 predominance of uropathogenic lineages in Swedish infants and enteropathogenic
504 lineages in Pakistani infants . Appl. Environ. Microbiol.

505 12. **Massot M**, **Daubié A-S**, **Clermont O**, **Jauréguy F**, **Couffignal C**, **Dahbi G**, **Mora A**,
506 **Blanco J**, **Branger C**, **Mentré F**, **Eddi A**, **Picard B**, **Denamur E**, **The Coliville Group**.
507 2016. Phylogenetic, virulence and antibiotic resistance characteristics of commensal
508 strain populations of Escherichia coli from community subjects in the Paris area in 2010
509 and evolution over 30 years. Microbiology **162**:642–650.

510 13. **Manges AR**, **Geum HM**, **Guo A**, **Edens TJ**, **Fibke CD**, **Pitout JDD**. 2019. Global
511 Extraintestinal Pathogenic Escherichia coli (ExPEC) Lineages. Clin. Microbiol. Rev. **32**.

512 14. **Alhashash F**, **Weston V**, **Diggle M**, **McNally A**. 2013. Multidrug-resistant Escherichia
513 coli bacteremia. Emerg. Infect. Dis.

514 15. **Olesen B**, **Scheutz F**, **Menard M**, **Skov MN**, **Kolmos HJ**, **Kuskowski MA**, **Johnson JR**.
515 2009. Three-decade epidemiological analysis of Escherichia coli O15:K52:H1. J. Clin.
516 Microbiol. **47**:1857–1862.

517 16. **Schaufler K**, **Semmler T**, **Wieler LH**, **Trott DJ**, **Pitout J**, **Peirano G**, **Bonnedahl J**,
518 **Dolejska M**, **Literak I**, **Fuchs S**, **Ahmed N**, **Grobbel M**, **Torres C**, **McNally A**, **Pickard**
519 **D**, **Ewers C**, **Croucher NJ**, **Corander J**, **Guenther S**. 2019. Genomic and Functional
520 Analysis of Emerging Virulent and Multidrug-Resistant Escherichia coli Lineage

521 Sequence Type 648. Antimicrob. Agents Chemother. **63**.

522 17. **Nicolas-Chanoine MH**, **Bertrand X**, **Madec JY**. 2014. Escherichia coli st131, an
523 intriguing clonal group. Clin. Microbiol. Rev. **27**:543–574.

524 18. **Kallonen T**, **Brodrick HJ**, **Harris SR**, **Corander J**, **Brown NM**, **Martin V**, **Peacock SJ**,
525 **Parkhill J**. 2017. Systematic longitudinal survey of invasive Escherichia coli in England
526 demonstrates a stable population structure only transiently disturbed by the emergence
527 of ST131.

528 19. **Swenson SA**. 2015. “From Man to Bacteria”: W.D. Hamilton, the theory of inclusive
529 fitness, and the post-war social order. Stud. Hist. Philos. Sci. Part C Stud. Hist. Philos.
530 Biol. Biomed. Sci. **49**:45–54.

531 20. **National Committee for Clinical Laboratory Standards (CLSI)**. 2013. Performance
532 Standards for Antimicrobial susceptibility Testing; Twenty-Third Informational Supplement.
533 NCCLS, Wayne, PA, USA.

534 21. **Clermont O**, **Christenson JK**, **Denamur E**, **Gordon DM**. 2013. The Clermont
535 Escherichia coli phylo-typing method revisited: Improvement of specificity and detection
536 of new phylo-groups. Environ. Microbiol. Rep. **5**:58–65.

537 22. **Clermont O**, **Christenson JK**, **Daubié AS**, **Gordon DM**, **Denamur E**. 2014.
538 Development of an allele-specific PCR for Escherichia coli B2 sub-typing, a rapid and
539 easy to perform substitute of multilocus sequence typing. J. Microbiol. Methods **101**:24–
540 27.

541 23. **Johnson JR**, **Clermont O**, **Johnston B**, **Clabots C**, **Tchesnokova V**, **Sokurenko E**,
542 **Junka AF**, **Maczynska B**, **Denamur E**. 2014. Rapid and specific detection, molecular
543 epidemiology, and experimental virulence of the O16 subgroup within Escherichia coli
544 sequence type 131. J. Clin. Microbiol. **52**:1358–1365.

545 24. **Russo TA**, **Johnson JR**. 2000. Proposal for a New Inclusive Designation for
546 Extraintestinal Pathogenic Isolates of Escherichia coli: ExPEC . J. Infect. Dis. **181**:1753–
547 1754.

548 25. **Rodríguez I**, **Novais Â**, **Lira F**, **Valverde A**, **Curião T**, **Martínez JL**, **Baquero F**, **Cantón**

549 **R**, **Coque TM**. 2015. Antibiotic-resistant Klebsiella pneumoniae and Escherichia coli
550 high-risk clones and an IncFII(k) mosaic plasmid hosting Tn1 (blaTEM-4) in isolates from
551 1990 to 2004. Antimicrob. Agents Chemother. **59**:2904–8.
552 26. **Novais A**, **Baquero F**, **Machado E**, **Cantón R**, **Peixe L**, **Coque TM**. 2010. International
553 spread and persistence of TEM-24 is caused by the confluence of highly penetrating
554 enterobacteriaceae clones and an IncA/C2 plasmid containing Tn1696::Tn1 and IS5075-
555 Tn21. Antimicrob. Agents Chemother. **54**:825–34.
556 27. **McGowan JE**, **Barnes MW**, **Finland M**. 1975. Bacteremia at Boston City Hospital:
557 occurrence and mortality during 12 selected years (1935-1972), with special reference to
558 hospital acquired cases. J. Infect. Dis. **132**:316–335.
559 28. **WR G**, **SJ E**, **I P**, **B R**. 1990. Bacteremia Due to Escherichia Coli: A Study of 861
560 Episodes. Rev. Infect. Dis. **12**.
561 29. **Laupland KB**. 2013. Incidence of bloodstream infection: a review of population-based
562 studies. Clin. Microbiol. Infect. **19**:492–500.
563 30. **Olesen SW**, **Lipsitch M**, **Grad YH**. 2019. The potential for “spillover” in outpatient
564 antibiotic stewardship interventions among US states. bioRxiv 536714.
565 31. **Low M**, **Neuberger A**, **Hooton TM**, **Green MS**, **Raz R**, **Balicer RD**, **Almog R**. 2019.
566 Association between urinary community-acquired fluoroquinolone-resistant Escherichia
567 coli and neighbourhood antibiotic consumption: a population-based case-control study.
568 Lancet Infect. Dis. **19**:419–428.
569 32. **Gottesman B-S**, **Low M**, **Almog R**, **Chowers M**. 2019. Quinolone Consumption by
570 Mothers Increases Their Children’s Risk of Acquiring Quinolone-Resistant Bacteriuria.
571 Clin. Infect. Dis.
572 33. **Fuzi M**, **Szabo D**, **Csercsik R**. 2017. Double-serine fluoroquinolone resistance mutations
573 advance major international clones and lineages of various multi-drug resistant bacteria.
574 Front. Microbiol. Frontiers Media S.A.
575 34. **Redgrave LS**, **Sutton SB**, **Webber MA**, **Piddock LJV**. 2014. Fluoroquinolone
576 resistance: Mechanisms, impact on bacteria, and role in evolutionary success. Trends

577 Microbiol. Elsevier Ltd.

578 35. **Yelin I**, **Snitser O**, **Novich G**, **Katz R**, **Tal O**, **Parizade M**, **Chodick G**, **Koren G**, **Shalev**
579 **V**, **Kishony R**. 2019. Personal clinical history predicts antibiotic resistance of urinary tract
580 infections. Nat. Med. **25**:1143–1152.

581 36. **Merino I**, **Hernández-García M**, **Turrientes M-C**, **Pérez-Viso B**, **Ló Pez-Fresne~ Na N**,
582 **Diaz-Agero C**, **Maechler F**, **Fankhauser-Rodriguez C**, **Kola A**, **Schrenzel J**, **Harbarth**
583 **S**, **Bonten M**, **Gastmeier P**, **Canton R**, **Ruiz-Garbajosa P**. Emergence of ESBL-
584 producing Escherichia coli ST131-C1-M27 clade colonizing patients in Europe.

585 37. **Selander RK**, **Levin BR**. 1980. Genetic diversity and structure in Escherichia coli
586 populations. Science (80-.). **210**:545–547.

587 38. **Holt RD**. 2009. Bringing the Hutchinsonian niche into the 21st century: Ecological and
588 evolutionary perspectives. Proc. Natl. Acad. Sci. U. S. A. **106**:19659–19665.

589 39. **Thänert R**, **Reske KA**, **Hink T**, **Wallace MA**, **Wang B**, **Schwartz DJ**, **Seiler S**, **Cass C**,
590 **Burnham CAD**, **Dubberke ER**, **Kwon JH**, **Dantas G**. 2019. Comparative genomics of
591 antibiotic-resistant uropathogens implicates three routes for recurrence of urinary tract
592 infections. MBio **10**.

593 40. **Sarkar S**, **Hutton ML**, **Vagenas D**, **Ruter R**, **Schüller S**, **Lyras D**, **Schembri MA**,
594 **Totsika M**. 2018. Intestinal Colonization Traits of Pandemic Multidrug-Resistant
595 Escherichia coli ST131. J. Infect. Dis. **218**:979–990.

596 41. **McNally A**, **Kallonen T**, **Connor C**, **Abudahab K**, **Aanensen DM**, **Horner C**, **Peacock**
597 **SJ**, **Parkhill J**, **Croucher NJ**, **Corander J**. 2019. Diversification of colonization factors in
598 a multidrug-resistant escherichia coli lineage evolving under negative frequency-
599 dependent selection. MBio **10**.

600 42. **Amarsy R**, **Guéret D**, **Benmansour H**, **Flicoteaux R**, **Berçot B**, **Meunier F**, **Mougari F**,
601 **Jacquier H**, **Pean de Ponfilly G**, **Clermont O**, **Denamur E**, **Teixeira A**, **Cambau E**.
602 2019. Determination of Escherichia coli phylogroups in elderly patients with urinary tract
603 infection or asymptomatic bacteriuria. Clin. Microbiol. Infect. **25**:839–844.

604 43. **Anderson RM**. 1999. The pandemic of antibiotic resistance. Nat. Med. **5:**147–149**.**
605

606 ## Supplementary Text

607 **Virulence-associated gene profiling in the B2 phylogroup of *Escherichia coli***

608 The presence of 29 virulence-associated genes (VAGs) was determined for all B2 isolates by
609 PCR [1,2]. The B2 *E. coli* strains were clustered into 2 large groups and 10 subgroups on the
610 basis of the content of the VAG genes. Each group comprises numerous gene combinations
611 and showed variable redundancies. Most *E. coli* isolates analyzed here harbored *fyuA,* an outer
612 membrane protein acting as a receptor for ferric-yersiniabactin (*fyu*) and a marker of PAIJ86IV
613 (a "high pathogenicity island") [3,4]; *malX,* a phosphotransferase system enzyme and marker of
614 PAICFT073 [5]; *ompT* (an extracellular protease active against host cationic peptides located on
615 *E. coli* plasmids or chromosomes [6,7]; and *traT*, only located on F plasmids, which act as a
616 surface exclusion protein, conferring serum resistance. [8]. However, the isolates differ in the
617 presence of type P fimbriae, all of which contain various combinations of siderophores, toxins,
618 and types S and I fimbriae, which are usually located on emblematic *E. coli* pathogenicity
619 islands (PAIs). Despite some VAG variability within each B2 subgroup, more than half (56%) of
620 the B2-I (STc131) strains clustered in VAG 1.5. Most (54%–75%) of the *E. coli* phylogenetic
621 subgroup isolates B2-II, B2-III, B2-V, B2-VII, B2-IX, and B2-X clustered in the VAG 2.6, 2.2, 2.9,
622 2.6, 2.7, and 2.4 groups, respectively (ordered by frequency) (Figures 6, S3, S4, S5, S6).

623 The VAG 1.5 group only comprised isolates of the B2-I (STc131) and B2-VII subgroups
624 (representing 54% and 33% of the isolates of these B2 subgroups, respectively). They
625 contained the genes *iha, sat, iutA, kpsMT,* and sporadically *hlyA*, all frequently located on
626 PAICFT073-pheV*,* and to a lesser extent, *afa/draBC* (25%) or *ibeA*. Some B2-I (STc131)
627 isolates were also associated with VAG 1.8 (lacking traits of PAICFT073-pheV but containing
628 *ibeA* and often *iroN*), VAG 2.4 (containing *iroN*, *ibeA,* and *papABCED*), and VAG2.9
629 (*papABCED)*. The analysis of STc131 isolates in Figure S6 reflects the diversity of VAGs within
630 the ST131 groups H30 (or subclade C), H22 (or subclade B), and H41 (or subclade A); many of
631 the isolates exhibiting VAGs could not be associated with any of the virotypes proposed by
632 Dahbi et al. [9]

633 The VAG 2.7 was predominant in isolates of the B2-IX group (63%, p <.005) and was only
634 associated with this B2 subgroup. This VAG profile was characterized by *papACEF* genes, *ireA*
635 (encoding a protein involved in iron acquisition, adhesion to epithelial cells, resistance to stress,
636 and a high pH, a trait frequently associated with isolates causing UTIs and BSIs), and a variable
637 presence of genes in PAIU189 (*iss*, *cvaC*, and K1-capsule).

638 The VAG 2.6 profile comprised more than half of B2-II and B2-VII isolates and was linked to the
639 presence of type P fimbriae and with various traits associated with PAIJ96 (*hlyA, cnf1, sfa,*
640 *ΔpapGIII*), PAICFT073-serX *(sfa/focDE, iroN*), and PAICFT073-pheV (*iutA, iha, iroN, sat,*
641 *kpsMTK5*).

642 The VAG 2.9 group was mostly associated with B2-V (50% of the isolates), and differs from
643 VAG 2.6 in the absence of *iroN* and *K5*. VAG 2.2 comprises most B2-III (73 of the isolates) and
644 VAG 2.4, most B2-X isolates. With the exception of B2-I and B2-IV, genes of most strains of B2-
645 II, B2-IX, and B2-IV have traditionally been associated with UTIs or BSIs.

646

647
648 **References (Supplementary Text)**

649 [1] Johnson JR, Russo TA. Molecular Epidemiology of Extraintestinal Pathogenic
650 Escherichia coli. EcoSal Plus 2018;8. https://doi.org/10.1128/ecosalplus.esp-0004-2017.

651 [2] Russo TA, Johnson JR. Proposal for a New Inclusive Designation for Extraintestinal
652 Pathogenic Isolates of Escherichia coli: ExPEC. J Infect Dis 2000;181:1753–4.
653 https://doi.org/10.1086/315418.

654 [3] Clermont O, Bonacorsi S, Bingen E. The Yersinia high-pathogenicity island is highly
655 predominant in virulence-associated phylogenetic groups of Escherichia coli . FEMS
656 Microbiol Lett 2001;196:153–7. https://doi.org/10.1111/j.1574-6968.2001.tb10557.x.

657 [4] Hancock V, Ferrières L, Klemm P. The ferric yersiniabactin uptake receptor FyuA is
658 required for efficient biofilm formation by urinary tract infectious Escherichia coli in human
659 urine. Microbiology 2008;154:167–75. https://doi.org/10.1099/mic.0.2007/011981-0.

660 [5] Östblom A, Adlerberth I, Wold AE, Nowrouzian FL. Pathogenicity island markers,
661 virulence determinants malX and usp, and the capacity of Escherichia coli to persist in
662 infants' commensal microbiotas. Appl Environ Microbiol 2011;77:2303–8.
663 https://doi.org/10.1128/AEM.02405-10.

664 [6] Stumpe S, Schmid R, Stephens DL, Georgiou G, Bakker EP. Identification of OmpT as
665 the protease that hydrolyzes the antimicrobial peptide protamine before it enters growing
666 cells of Escherichia coli. J Bacteriol 1998;180:4002–6.

667 [7] Dan M, Yair Y, Samosav A, Gottesman T, Yossepowitch O, Harari-Schwartz O, et al.
668 Escherichia coli isolates from patients with bacteremic urinary tract infection are
669 genetically distinct from those derived from sepsis following prostate transrectal biopsy.
670 Int J Med Microbiol 2015;305:464–8. https://doi.org/10.1016/j.ijmm.2015.04.003.

671 [8] Harrison JL, Taylor IM, Platt K, O'Connor CD. Surface exclusion specificity of the TraT
672 lipoprotein is determined by single alterations in a five-amino-acid region of the protein.

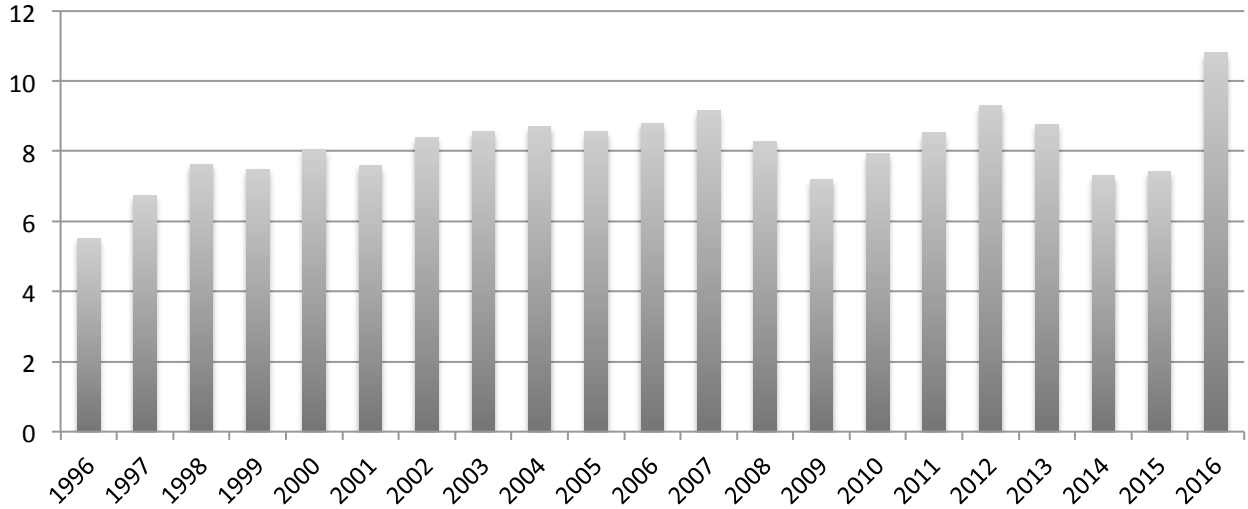
673 Mol Microbiol 1992;6:2825–32. https://doi.org/10.1111/j.1365-2958.1992.tb01462.x.

674 [9] Dahbi G, Mora A, Mamani R, López C, Alonso MP, Marzoa J, et al. Molecular
675 epidemiology and virulence of Escherichia coli O16: H5-ST131: Comparison with H30
676 and H30-Rx subclones of O25b: H4-ST131. Int J Med Microbiol 2014;304:1247–57.
677 https://doi.org/10.1016/j.ijmm.2014.10.002.

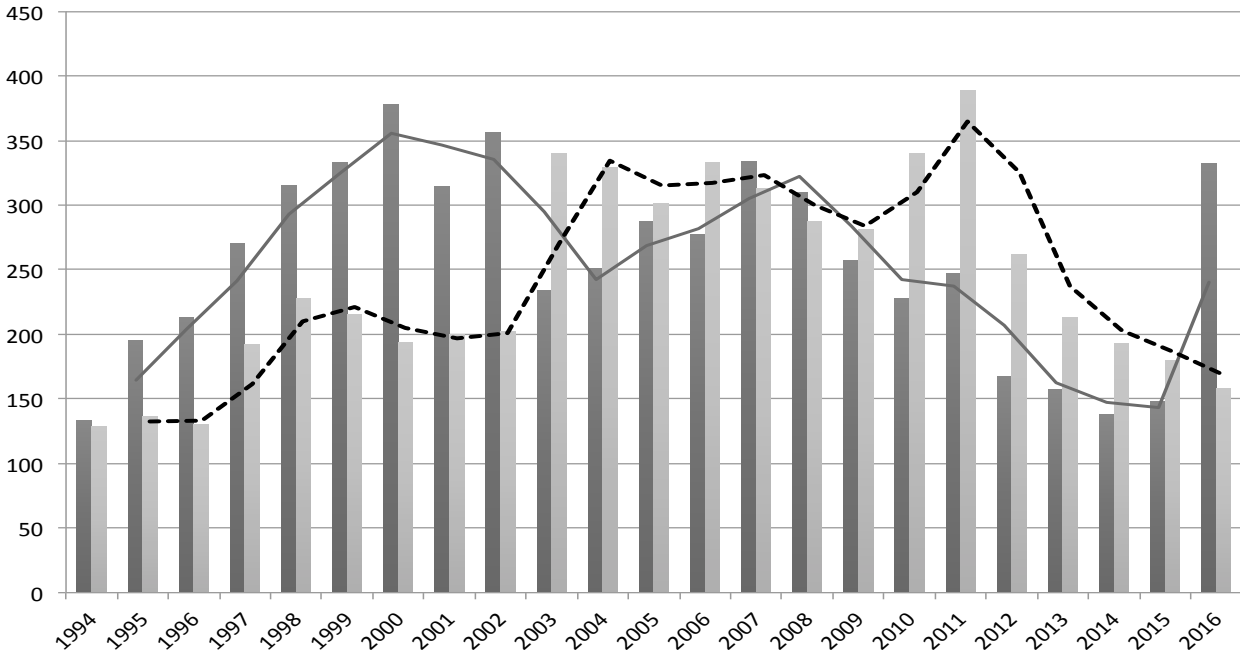
678

679

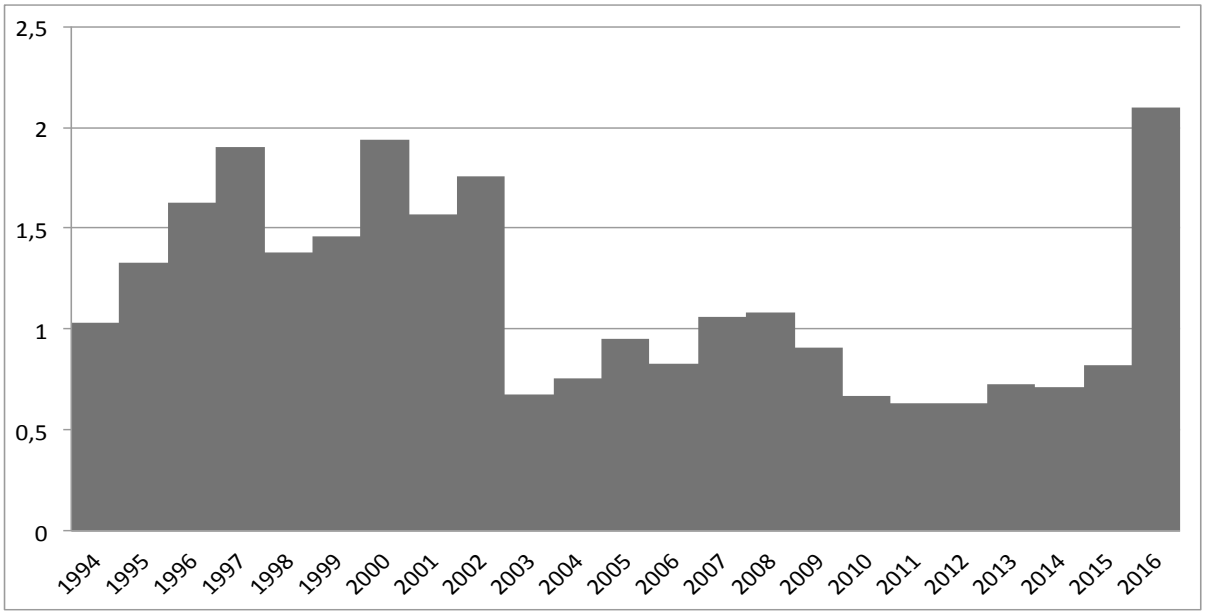
Figure 1



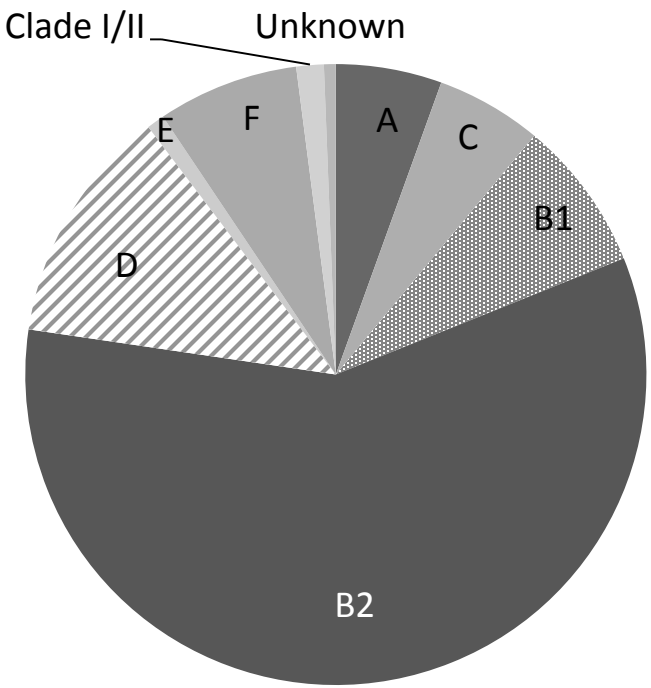
1A)



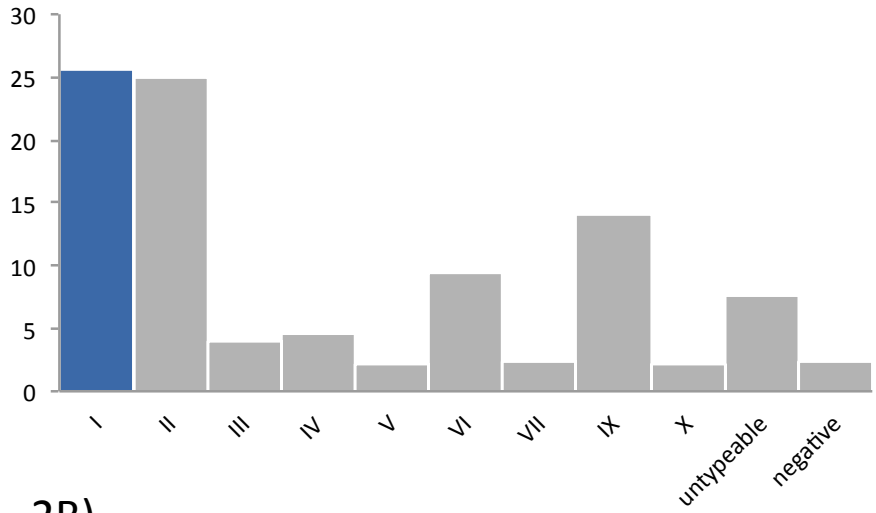
1B)



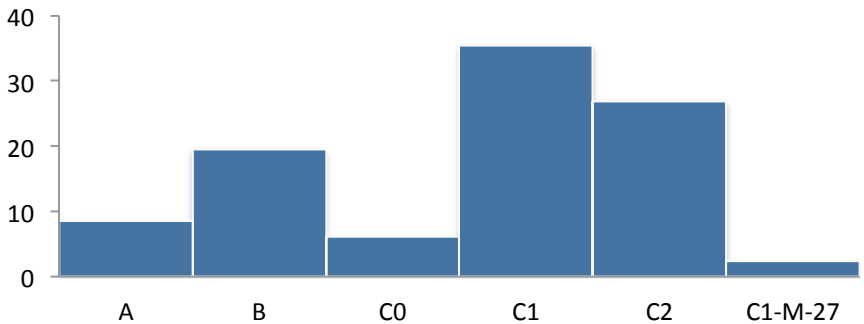
1C)



2A)

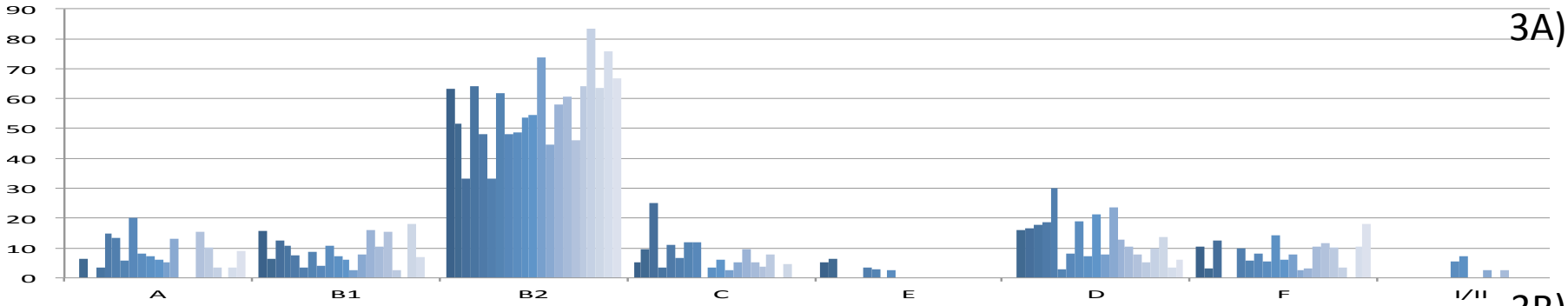


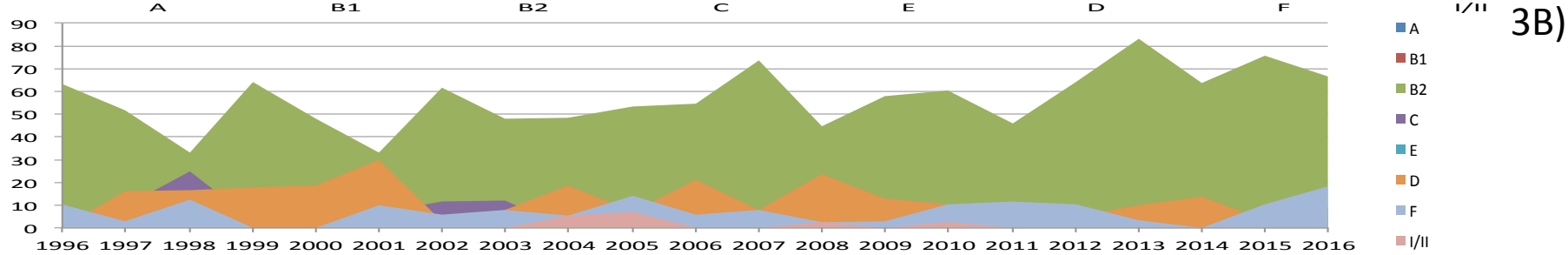
2B)

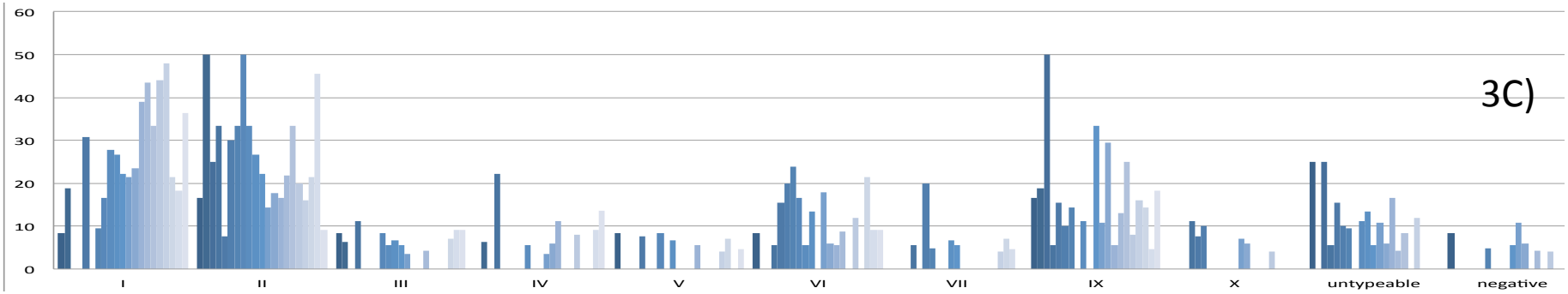


2C)

Figure 2


3A)
90
80
70
60
50
40
30
20
10
0
A
B1
B2
C
E
D
F
I/II


3B)
90
80
70
60
50
40
30
20
10
0
1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016
A
B1
B2
C
E
D
F
I/II


3C)
60
50
40
30
20
10
0
I
II
III
IV
V
VI
VII
IX
X
untypeable
negative

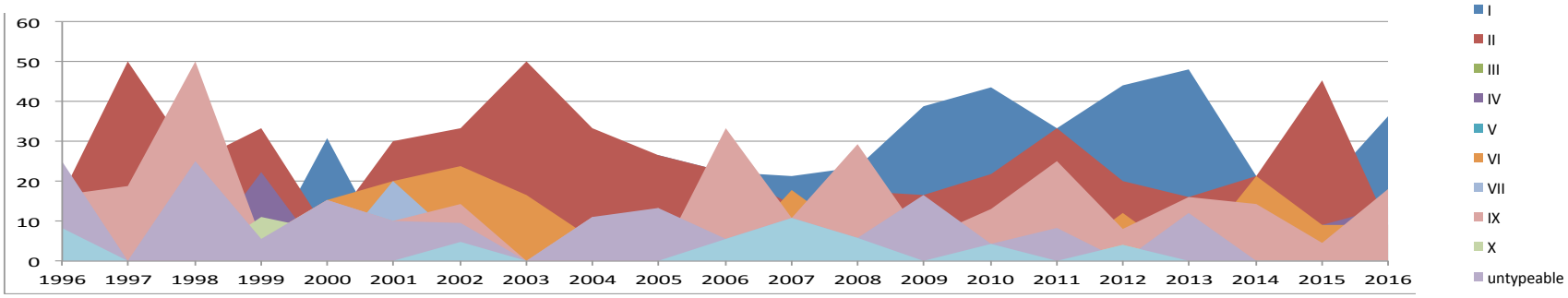
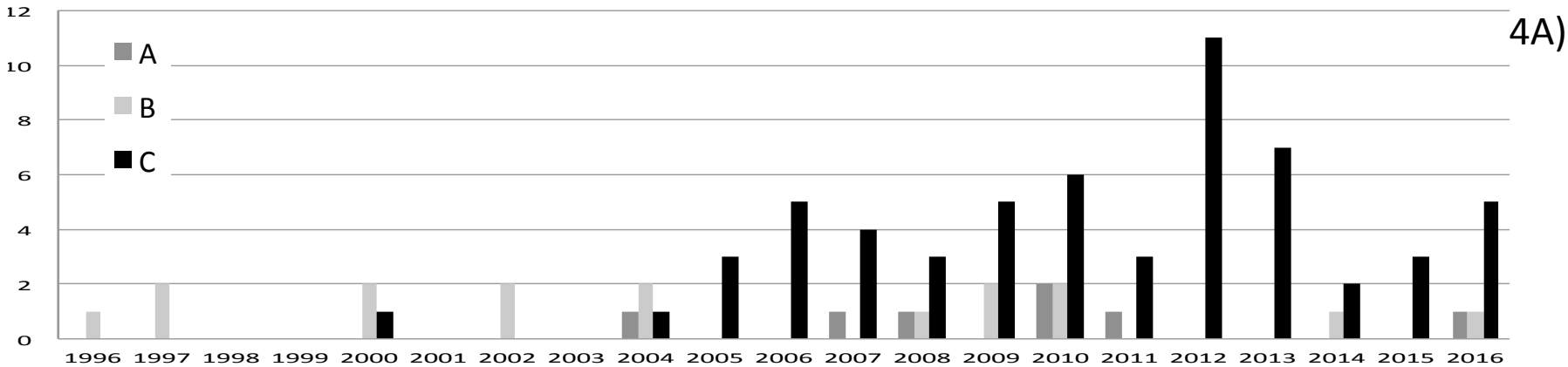
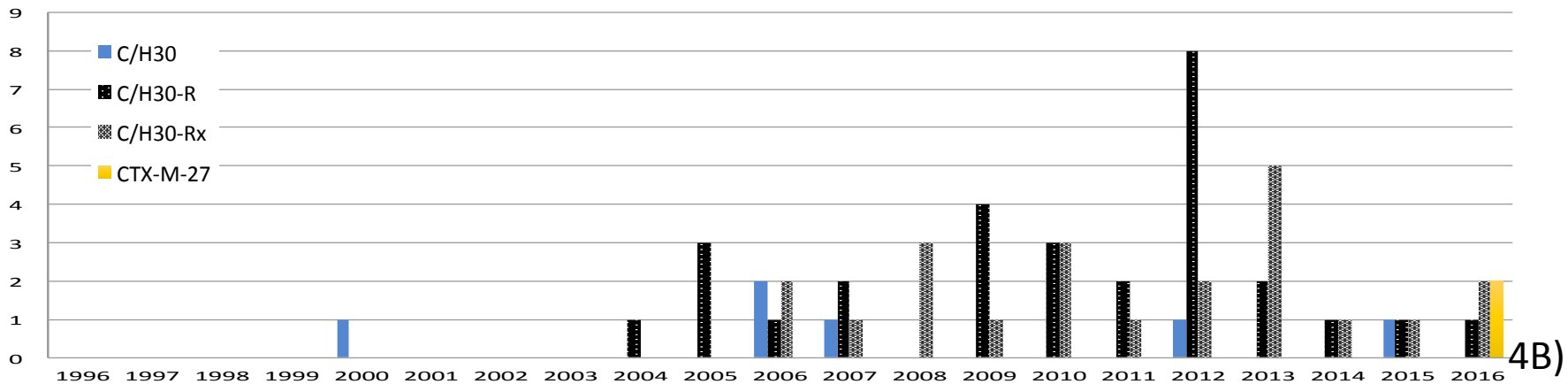

3D)
60
50
40
30
20
10
0
1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016
I
II
III
IV
V
VI
VII
IX
X
untypeable

Figure 4


4A)
A
B
C
0
2
4
6
8
10
12
1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016


C/H30
C/H30-R
C/H30-Rx
CTX-M-27
0
1
2
3
4
5
6
7
8
9
1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016
4B)

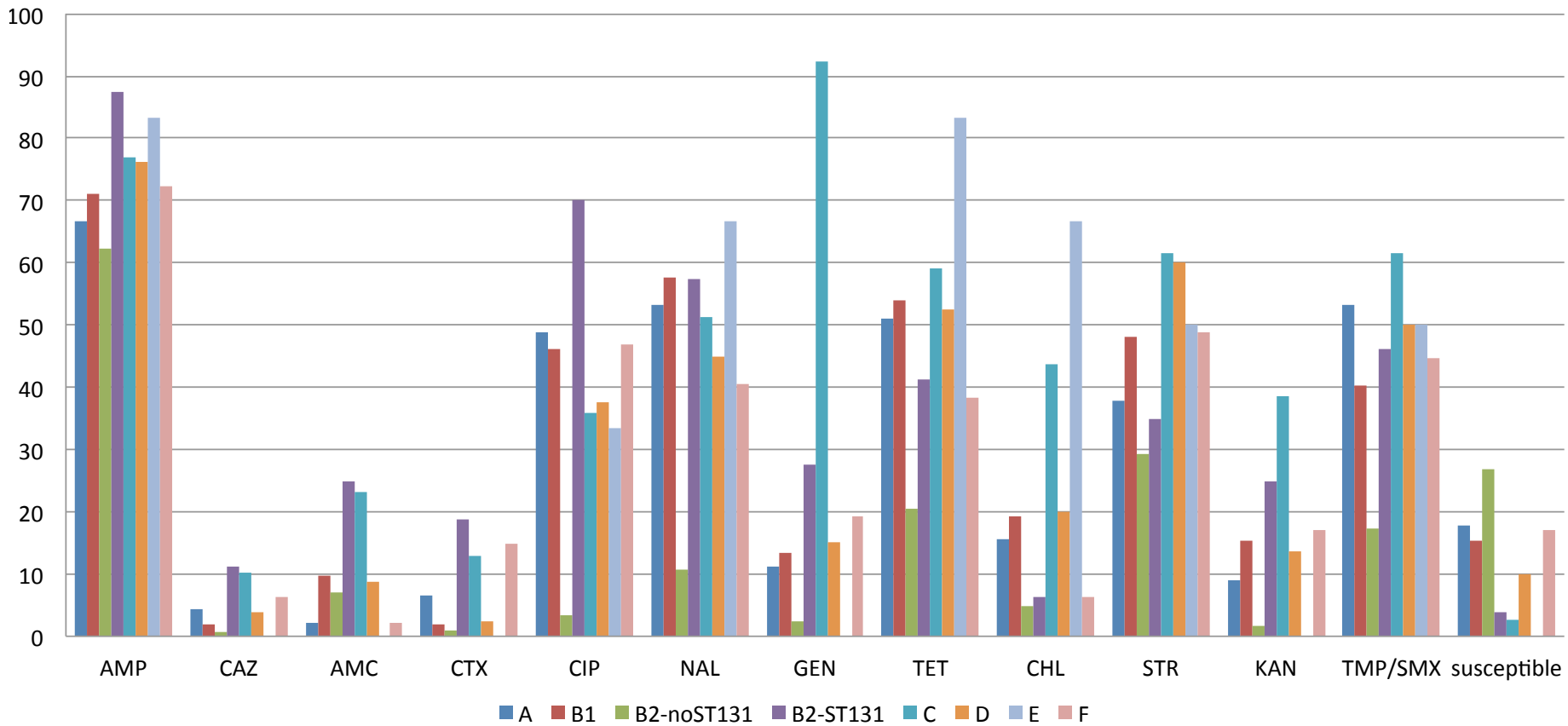

100
90
80
70
60
50
40
30
20
10
0
AMP
CAZ
AMC
CTX
CIP
NAL
GEN
TET
CHL
STR
KAN
TMP/SMX
susceptible
A
B1
B2-noST131
B2-ST131
C
D
E
F

Figure 5

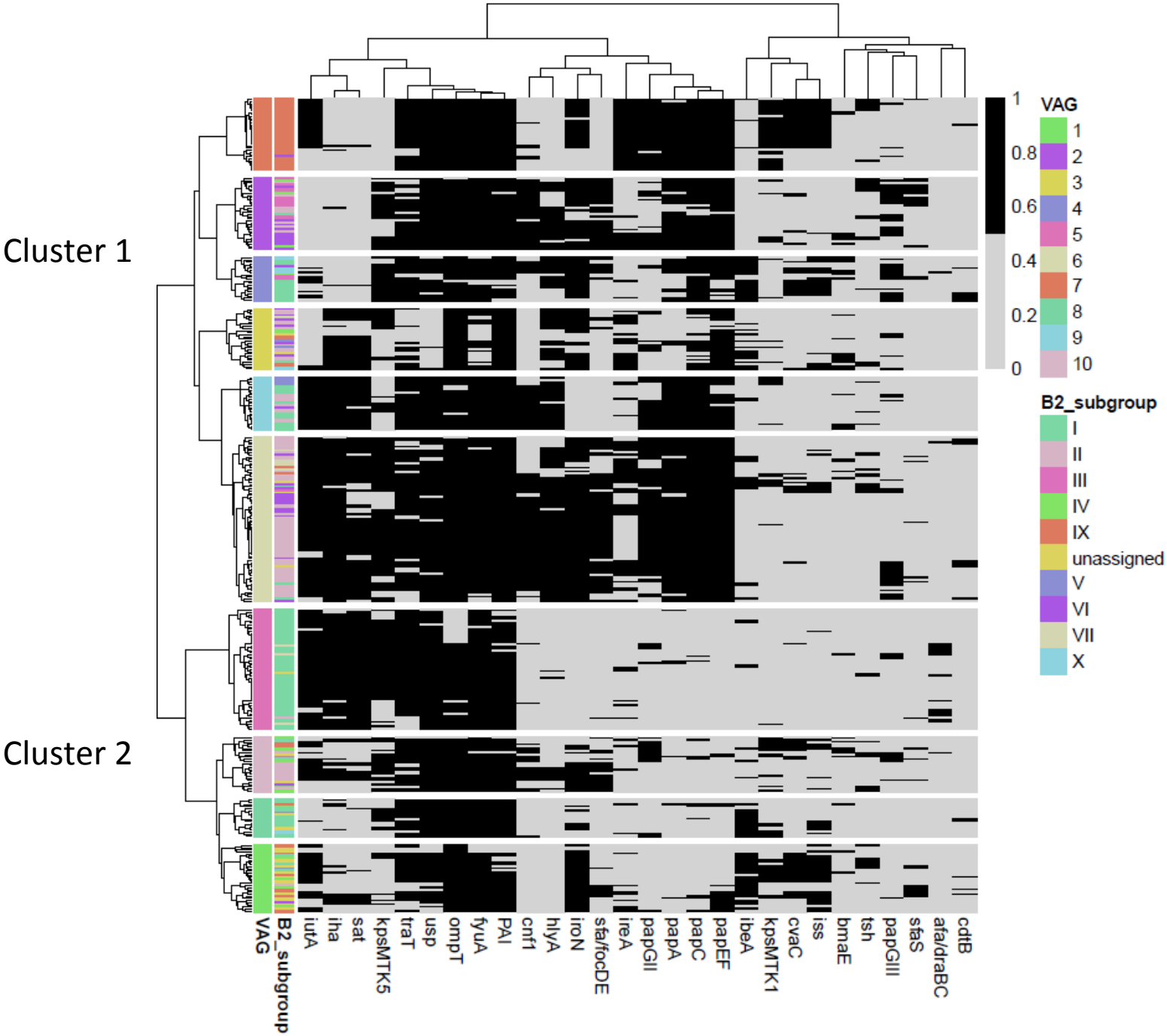
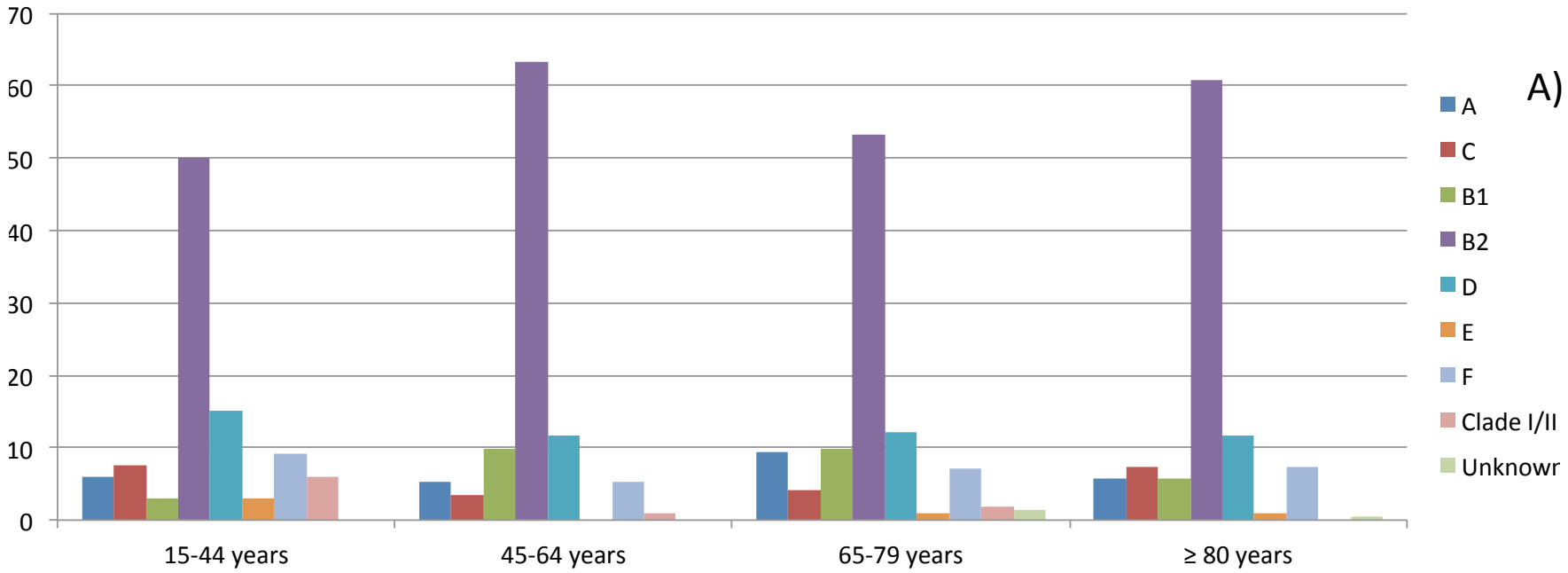

Cluster 1
Cluster 2
VAG
B2_subgroup
iutA
iha
sat
kpsMTK5
traT
usp
ompT
fyuA
PAI
cnf1
hlyA
iroN
sfa/focDE
ireA
papGII
papA
papC
papEF
ibeA
kpsMTK1
cvaC
iss
bmaE
tsh
papGIII
sfaS
afa/draBC
cdtB
1
0.8
0.6
0.4
0.2
0
VAG
1
2
3
4
5
6
7
8
9
10
B2_subgroup
I
II
III
IV
IX
unassigned
V
VI
VII
X

Figure 6


A)
70
60
50
40
30
20
10
0
15-44 years
45-64 years
65-79 years
≥ 80 years
A
C
B1
B2
D
E
F
Clade I/II
Unknowr

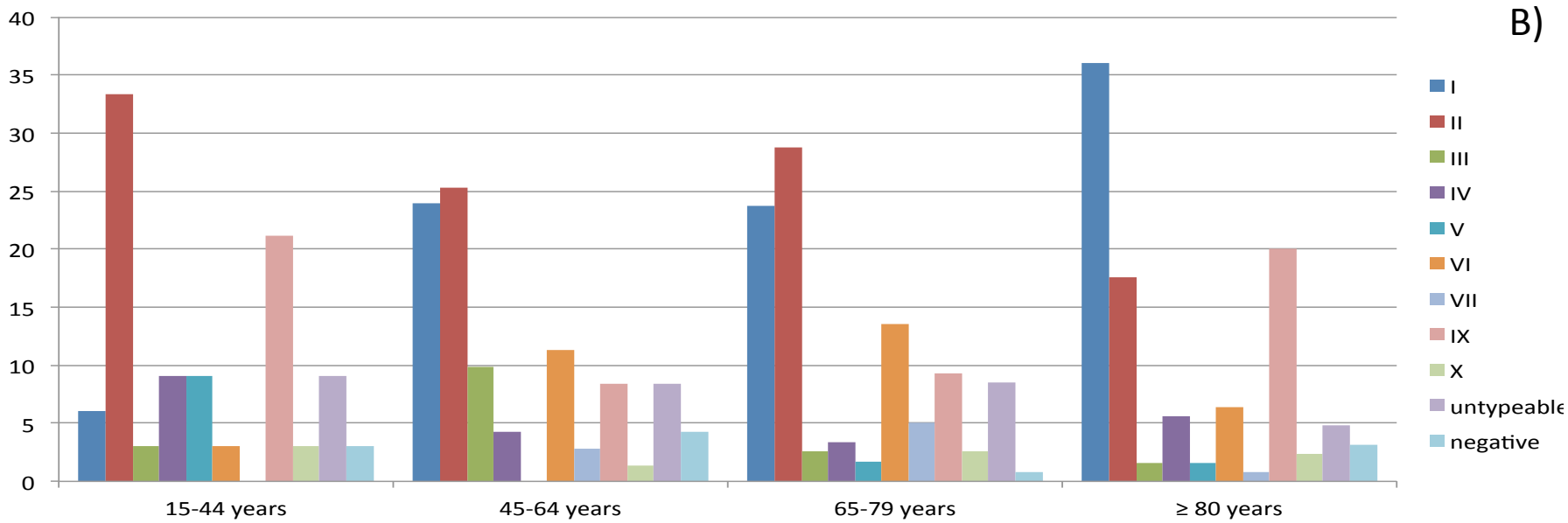

B)
40
35
30
25
20
15
10
5
0
15-44 years
45-64 years
65-79 years
≥ 80 years
I
II
III
IV
V
VI
VII
IX
X
untypeable
negative

Figure S1

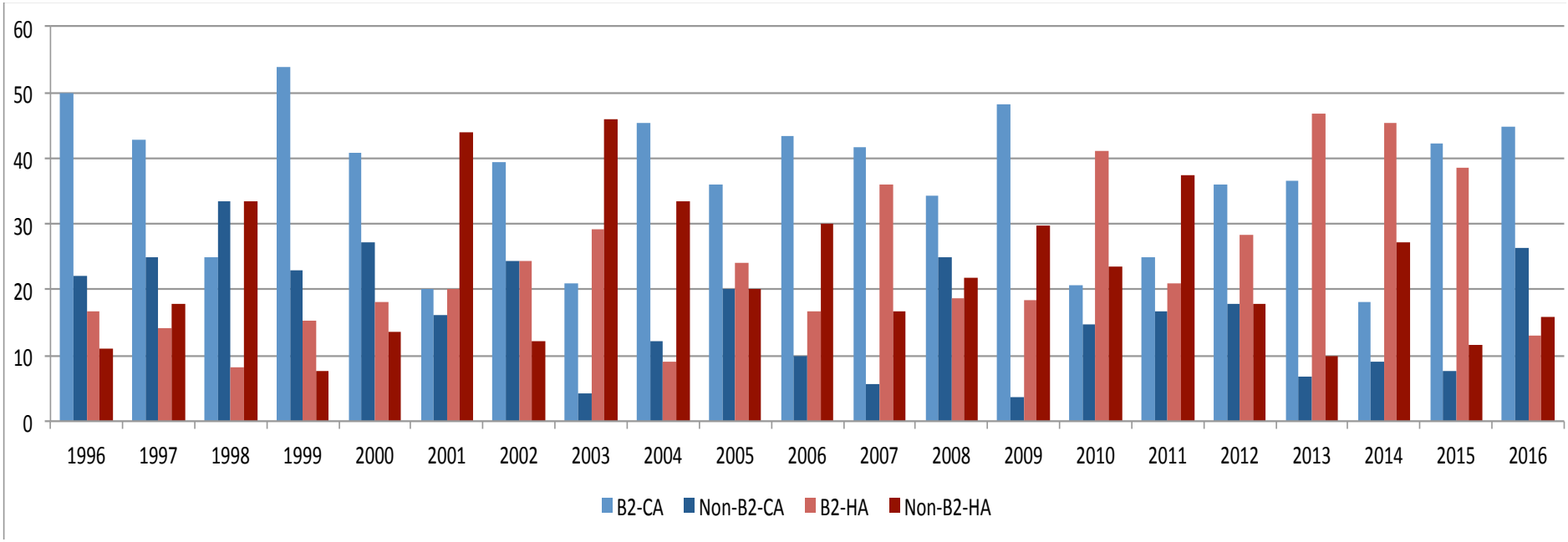

60
50
40
30
20
10
0
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
B2-CA
Non-B2-CA
B2-HA
Non-B2-HA

Figure S2

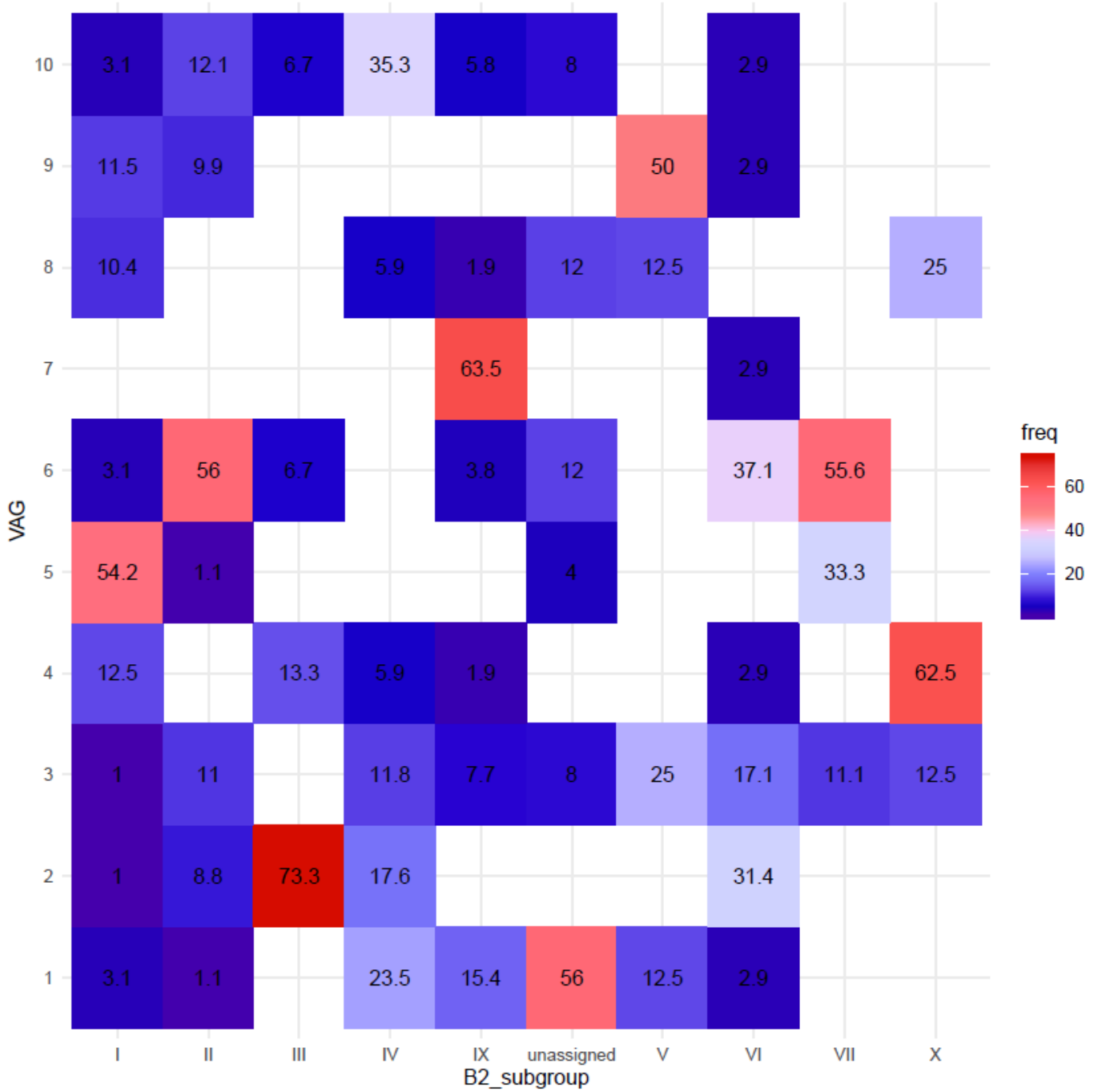

VAG
10
9
8
7
6
5
4
3
2
1
3.1 12.1 6.7 35.3 5.8 8 2.9
11.5 9.9 50 2.9
10.4 5.9 1.9 12 12.5 25
63.5 2.9
3.1 56 6.7 3.8 12 37.1 55.6
54.2 1.1 4 33.3
12.5 13.3 5.9 1.9 2.9 62.5
1 11 11.8 7.7 8 25 17.1 11.1 12.5
1 8.8 73.3 17.6 31.4
3.1 1.1 23.5 15.4 56 12.5 2.9
I
II
III
IV
IX
unassigned
V
VI
VII
X
B2_subgroup
freq
60
40
20

Figure S3

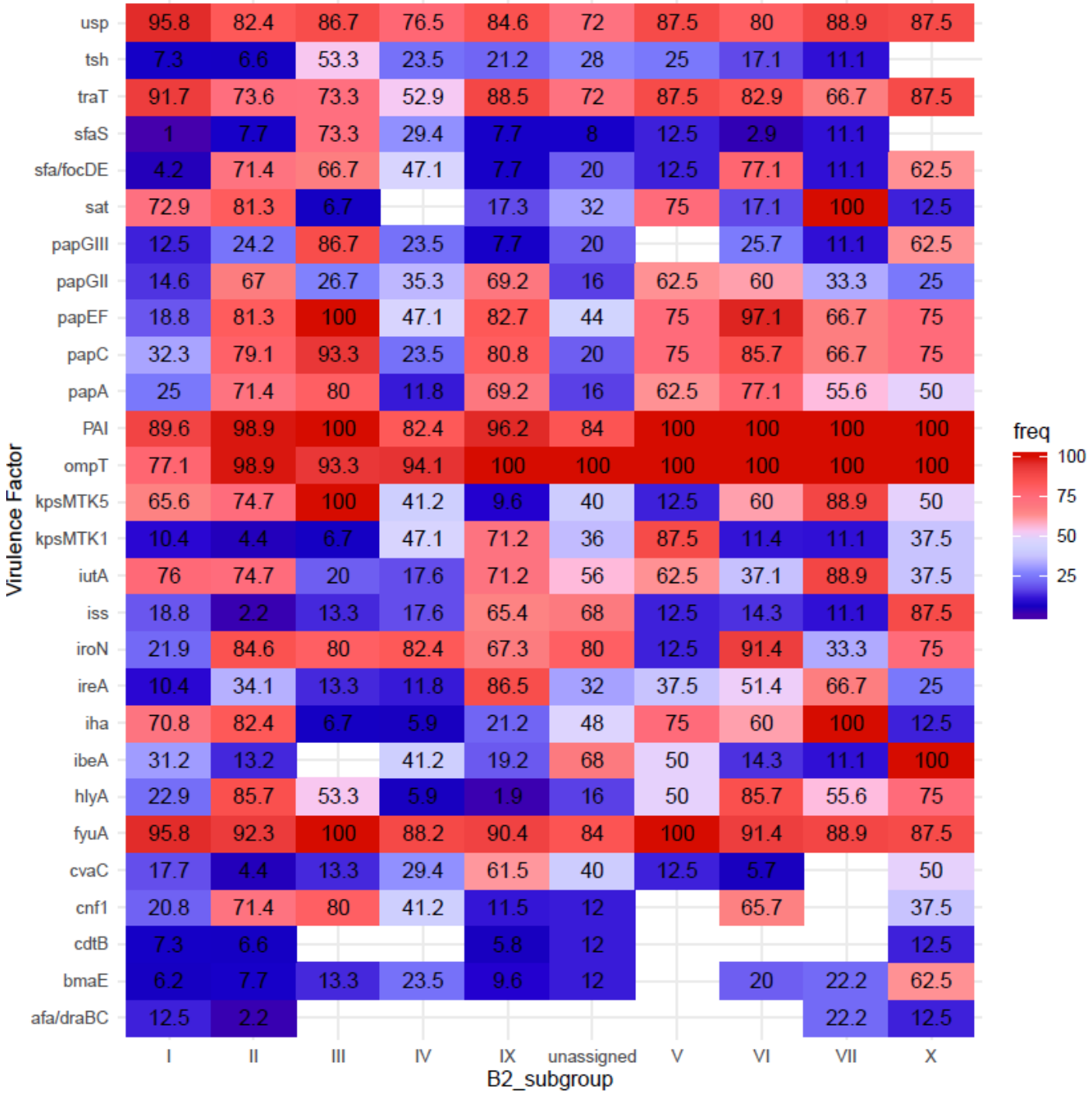

Virulence Factor
usp 95.8 82.4 86.7 76.5 84.6 72 87.5 80 88.9 87.5
tsh 7.3 6.6 53.3 23.5 21.2 28 25 17.1 11.1
traT 91.7 73.6 73.3 52.9 88.5 72 87.5 82.9 66.7 87.5
sfaS 1 7.7 73.3 29.4 7.7 8 12.5 2.9 11.1
sfa/focDE 4.2 71.4 66.7 47.1 7.7 20 12.5 77.1 11.1 62.5
sat 72.9 81.3 6.7 17.3 32 75 17.1 100 12.5
papGIII 12.5 24.2 86.7 23.5 7.7 20 25.7 11.1 62.5
papGII 14.6 67 26.7 35.3 69.2 16 62.5 60 33.3 25
papEF 18.8 81.3 100 47.1 82.7 44 75 97.1 66.7 75
papC 32.3 79.1 93.3 23.5 80.8 20 75 85.7 66.7 75
papA 25 71.4 80 11.8 69.2 16 62.5 77.1 55.6 50
PAI 89.6 98.9 100 82.4 96.2 84 100 100 100 100
ompT 77.1 98.9 93.3 94.1 100 100 100 100 100 100
kpsMTK5 65.6 74.7 100 41.2 9.6 40 12.5 60 88.9 50
kpsMTK1 10.4 4.4 6.7 47.1 71.2 36 87.5 11.4 11.1 37.5
iutA 76 74.7 20 17.6 71.2 56 62.5 37.1 88.9 37.5
iss 18.8 2.2 13.3 17.6 65.4 68 12.5 14.3 11.1 87.5
iroN 21.9 84.6 80 82.4 67.3 80 12.5 91.4 33.3 75
ireA 10.4 34.1 13.3 11.8 86.5 32 37.5 51.4 66.7 25
iha 70.8 82.4 6.7 5.9 21.2 48 75 60 100 12.5
ibeA 31.2 13.2 41.2 19.2 68 50 14.3 11.1 100
hlyA 22.9 85.7 53.3 5.9 1.9 16 50 85.7 55.6 75
fyuA 95.8 92.3 100 88.2 90.4 84 100 91.4 88.9 87.5
cvaC 17.7 4.4 13.3 29.4 61.5 40 12.5 5.7 50
cnf1 20.8 71.4 80 41.2 11.5 12 65.7 37.5
cdtB 7.3 6.6 5.8 12 12.5
bmaE 6.2 7.7 13.3 23.5 9.6 12 20 22.2 62.5
afa/draBC 12.5 2.2 22.2 12.5
I II III IV IX unassigned V VI VII X
B2_subgroup
freq
100
75
50
25

Figure S4

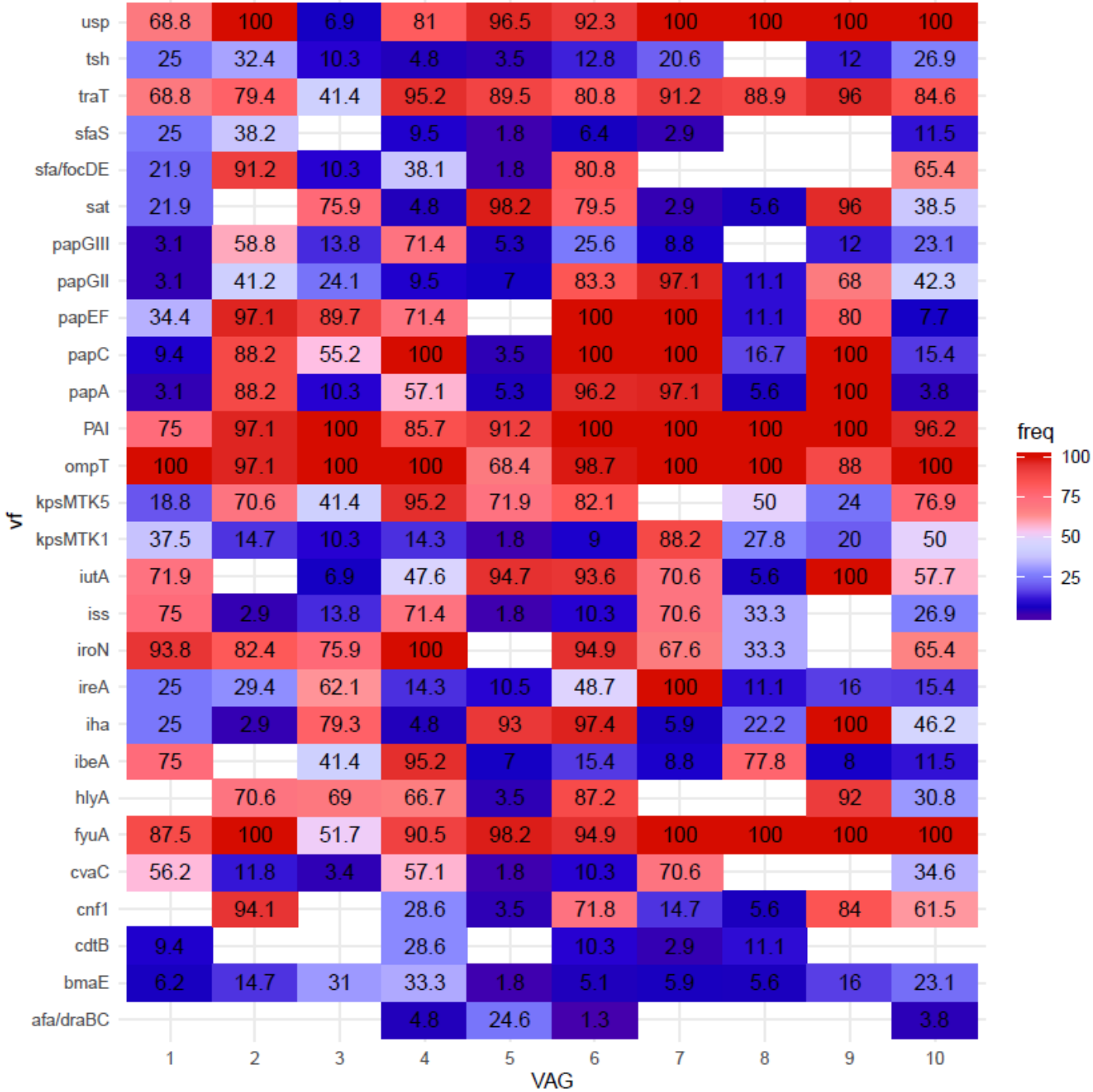

usp 68.8 100 6.9 81 96.5 92.3 100 100 100 100
tsh 25 32.4 10.3 4.8 3.5 12.8 20.6 12 26.9
traT 68.8 79.4 41.4 95.2 89.5 80.8 91.2 88.9 96 84.6
sfaS 25 38.2 9.5 1.8 6.4 2.9 11.5
sfa/focDE 21.9 91.2 10.3 38.1 1.8 80.8 65.4
sat 21.9 75.9 4.8 98.2 79.5 2.9 5.6 96 38.5
papGIII 3.1 58.8 13.8 71.4 5.3 25.6 8.8 12 23.1
papGII 3.1 41.2 24.1 9.5 7 83.3 97.1 11.1 68 42.3
papEF 34.4 97.1 89.7 71.4 100 100 11.1 80 7.7
papC 9.4 88.2 55.2 100 3.5 100 100 16.7 100 15.4
papA 3.1 88.2 10.3 57.1 5.3 96.2 97.1 5.6 100 3.8
PAI 75 97.1 100 85.7 91.2 100 100 100 100 96.2
ompT 100 97.1 100 100 68.4 98.7 100 100 88 100
kpsMTK5 18.8 70.6 41.4 95.2 71.9 82.1 50 24 76.9
kpsMTK1 37.5 14.7 10.3 14.3 1.8 9 88.2 27.8 20 50
iutA 71.9 6.9 47.6 94.7 93.6 70.6 5.6 100 57.7
iss 75 2.9 13.8 71.4 1.8 10.3 70.6 33.3 26.9
iroN 93.8 82.4 75.9 100 94.9 67.6 33.3 65.4
ireA 25 29.4 62.1 14.3 10.5 48.7 100 11.1 16 15.4
iha 25 2.9 79.3 4.8 93 97.4 5.9 22.2 100 46.2
ibeA 75 41.4 95.2 7 15.4 8.8 77.8 8 11.5
hlyA 70.6 69 66.7 3.5 87.2 92 30.8
fyuA 87.5 100 51.7 90.5 98.2 94.9 100 100 100 100
cvaC 56.2 11.8 3.4 57.1 1.8 10.3 70.6 34.6
cnf1 94.1 28.6 3.5 71.8 14.7 5.6 84 61.5
cdtB 9.4 28.6 10.3 2.9 11.1
bmaE 6.2 14.7 31 33.3 1.8 5.1 5.9 5.6 16 23.1
afa/draBC 4.8 24.6 1.3 3.8
vf
1 2 3 4 5 6 7 8 9 10
VAG
freq
100
75
50
25

Figure S5

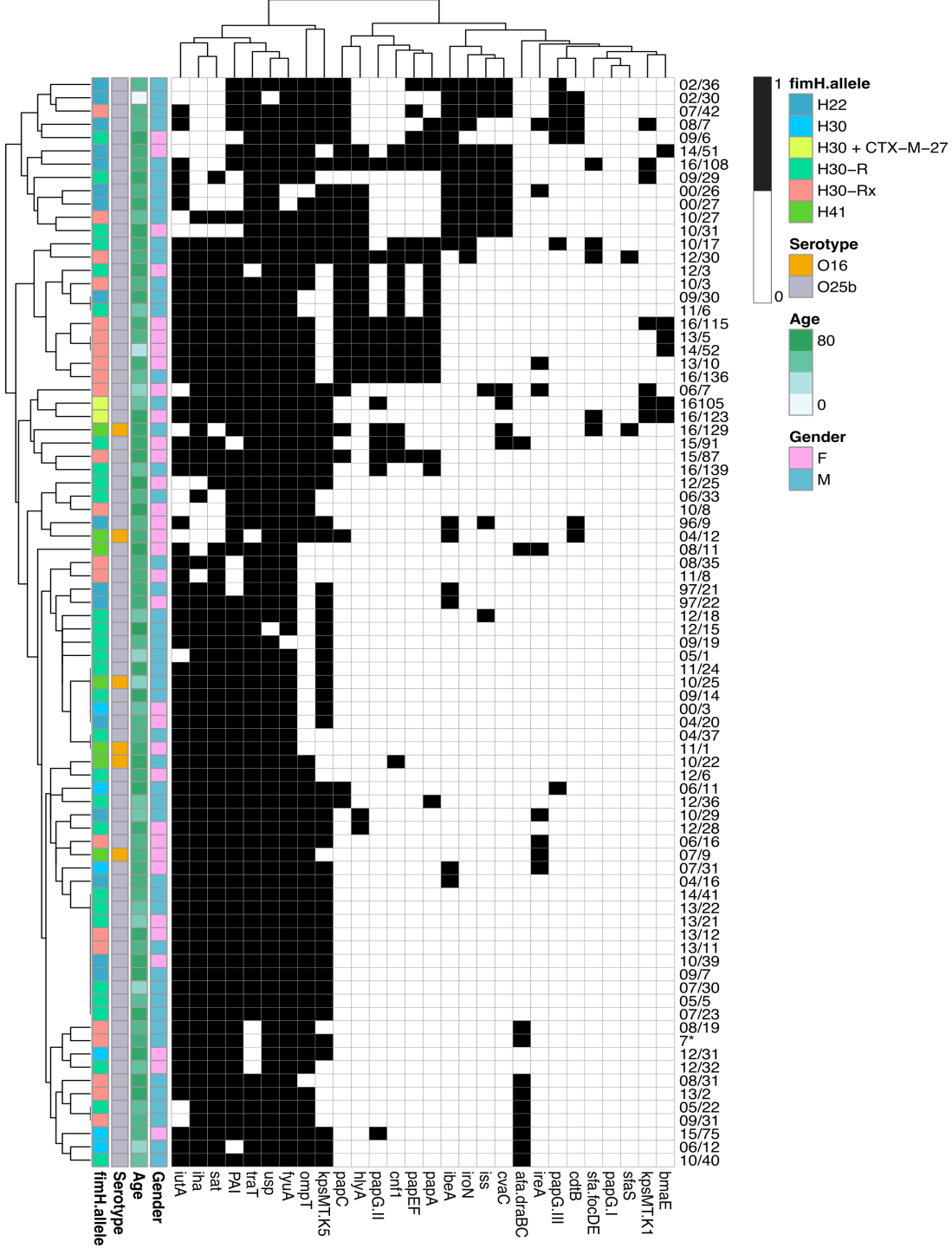

fimH.allele
Serotype
Age
Gender
iutA
iha
sat
PAI
traT
usp
fyuA
ompT
kpsMT.K5
papC
hlyA
papG.II
cnf1
papEF
papA
ibeA
iroN
iss
cvaC
afa.draBC
ireA
papG.III
cdtB
sfa.focDE
papG.I
sfaS
kpsMT.K1
bmaE
1
0
fimH.allele
H22
H30
H30 + CTX−M−27
H30−R
H30−Rx
H41
Serotype
O16
O25b
Age
80
0
Gender
F
M

Figure S6