

1 **High diversity of *Escherichia coli* causing invasive disease in neonates in Malawi** 2 **poses challenges for O-antigen based vaccine approach**

3 Oliver Pearse^{1,2}, Allan Zuza², Edith Tewesa³, Patricia Siyabuh³, Alice J Fraser⁴, Jennifer
4 Cornick^{2,5}, Kondwani Kawaza^{2,6}, Patrick Musicha^{2,4,7}, Nicholas R Thomson^{7,8}, Nicholas A
5 Feasey^{1,2,9*}, Eva Heinz^{1,4,10*}

6 *Equally contributing

7 **Affiliations:**

8 1. Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, UK

9 2. Malawi-Liverpool-Wellcome Programme, Kamuzu University of Health Sciences, Blantyre,
10 Malawi

11 3. Queen Elizabeth Central Hospital, Blantyre, Malawi

12 4. Department of Vector Biology, Liverpool School of Tropical Medicine, Liverpool, UK

13 5. University of Liverpool, Institute of Infection, Veterinary and Ecological Sciences, Liverpool,
14 UK

15 6. Kamuzu University of Health Sciences, Malawi

16 7. Wellcome Sanger Institute, Parasites and Microbes Program, Hinxton, UK

17 8. London School of Tropical Medicine and Hygiene, Department of Pathogen Molecular
18 Biology, London, UK

19 9. The School of Medicine, University of St. Andrews, St. Andrews, UK

20 10. University of Strathclyde, Strathclyde Institute for Pharmacy and Biomedical Sciences,
21 Glasgow, UK

22 **Corresponding author email: oliver.pearse@lstmed.ac.uk; eva.heinz@strath.ac.uk**

23 **Keywords**

24 Neonatal infection, neonatal sepsis, neonatal meningitis, vaccines, sero-epidemiology, O:H-
25 type, H-antigen, antimicrobial resistance, sub-Saharan Africa

26 **Data summary**

27 All sequencing data is freely available under the sequencing project IDs ERP120687 (short read
28 data; accessions in Supplementary Table 1) and PRJNA1121524 (long-read data; accessions in

29 Supplementary Table 2), detailed per-isolate information is provided in Supplementary Table 3.
30 Blood culture and CSF data used to show the trends and numbers of *E. coli* cases per year is
31 available in Supplementary Table 4.

32 **Abstract**

33 *Escherichia coli* is an important cause of neonatal sepsis and the third most prevalent cause of
34 neonatal infection in sub-Saharan Africa, often with negative outcomes. Development of
35 maternally administered vaccines is under consideration, but to provide adequate protection, an
36 understanding of serotypes causing invasive disease in this population is essential. We describe
37 the genomic characteristics of a collection of neonatal *E. coli* isolates from a tertiary hospital in
38 Blantyre, Malawi, with specific reference to potential protection by vaccines under development.
39 Neonatal blood or cerebrospinal fluid cultures from 2012-2021 identified 205 *E. coli* isolates, and
40 170 could be recovered for sequencing. There was very high diversity in sequence types, LPS
41 O-antigen-type and fimbrial H-type, which all showed temporal fluctuations and previously
42 undescribed diversity, including ten putative novel O-types. Vaccines in clinical trials target the
43 O-antigen but would only protect against one third (33.7%) of neonatal sepsis cases in this
44 population (EXPEC9V, in clinical trials). An O-antigen based vaccine would require 30 different
45 O-types to protect against 80% of infections. Vaccines against neonatal sepsis in Africa are of
46 considerable potential value, but their development requires larger studies to establish the
47 diversity and stability over time of relevant O-types for this population.

48 **Introduction**

49 Neonatal infection is the third largest cause of neonatal death worldwide, currently at 18 per
50 1,000 live births globally. Sustainable Development Goal 3.2¹ aims to reduce this to 12 per
51 1,000 live births by 2030. *Escherichia coli* is the third most prevalent cause of neonatal infection
52 in sub-Saharan Africa² and is a particularly important cause of early onset sepsis³ (EoS; sepsis
53 before 72 hours of life).

54 Antimicrobial resistant (AMR) *E. coli* infection makes this problem worse and is an increasing
55 problem in countries in sub-Saharan Africa⁴, such as Malawi. Limited access to antibiotics in the
56 countries with the highest burden of infection makes these infections even more lethal⁵. In
57 Malawi for example there are few antimicrobial choices for neonates beyond the first-line
58 (benzylpenicillin and gentamicin) and second-line (ceftriaxone) therapies, for which there is
59 already resistance amongst *E. coli* isolates⁶. In this context, innovative approaches to
60 preventing infection with *E. coli* are required. One proposed approach is maternal administration

61 of vaccines to give neonates passive immunity as is already used for *Bordetella pertussis*⁷ and
62 being developed for Group B *Streptococcus*⁸. This would reduce the deaths and prolonged
63 hospital stays of neonates linked to *E. coli* sepsis and meningitis, reduce the use of
64 antimicrobials and thus the population-level spread of AMR. However, the surface exposed
65 structures that could be targeted by vaccination have high levels of diversity. There are currently
66 186 known O-polysaccharide antigens (O-types) for *E. coli*, 67 known capsular antigens (K-
67 types) and 53 known flagellar antigens (H-types)⁹. O-types and H-types are most important as
68 they are expressed by all *E. coli* and recognized by the immune system of the host. Crucially,
69 only a small proportion of these serogroups are thought to cause the majority of invasive
70 disease¹⁰ but this can vary depending on location and patient characteristics. A vaccine would
71 therefore need to target the correct surface exposed structures for the patient population in
72 question.

73 *E. coli* vaccines targeting the O-antigen are in development, including the ExPEC4V¹¹ and
74 ExPEC9V^{12, 13} in clinical trials. These 4-valent and 9-valent vaccines have been produced to
75 target the most prevalent serogroups causing invasive *E. coli* infection in older adults in high-
76 income countries. Whether these candidates would reduce neonatal sepsis in sub-Saharan
77 Africa is unknown, and addition of other types may be challenging; one O-antigen already had
78 to be removed from the initially 10-valent ExPEC9V as the functional antibody assay for the O8
79 *E. coli* strain was not able to distinguish an immunological response to vaccination^{13, 14, 15}.

80 To our knowledge there are no studies specifically examining the antigenic diversity in *E. coli*
81 causing invasive infection in neonates in sub-Saharan Africa, a population for which an
82 understanding of this diversity is crucial if vaccination is to be feasible and to achieve equity in
83 coverage across different regions. This study presents a genomic description of an unbiased
84 collection of neonatal *E. coli* isolates, collected from 2012 to 2021 from neonates in a Malawian
85 hospital, with a focus on O- and H-type diversity.

86 **Methods**

87 **Setting**

88 Queen Elizabeth Central Hospital (QECH) is a government run tertiary referral hospital for the
89 Southern Region of Malawi; care is free at the point of delivery. It directly serves urban Blantyre
90 (population ~800,000 as of 2018 census data). Chatinkha nursery receives approximately 5,000
91 admissions a year, with between 30 and 90 neonates on the ward at any one time. It admits
92 neonates that have not gone home; either those that were born in QECH or those referred from

93 another hospital. Paediatric nursery receives neonates that have been admitted from home via
94 paediatric A & E. In both of these wards there is 24-hour nursing care, and daily medical ward
95 rounds. There is access to diagnostic blood and cerebrospinal fluid (CSF) culture, continuous
96 positive airway pressure, oxygen, IV fluids, blood transfusion, radiology, and blood testing.
97 There is also access to an on-site HDU. The PICU/HDU and paediatric surgical ward are
98 located on the Mercy James Hospital, which opened in 2017. This is a surgical hospital on the
99 QECH site but it is in a separate building, managed separately and philanthropically funded. At
100 the surgical hospital there is additionally access to intensive care facilities, which allows for the
101 use of vasopressors and intubation.

102 **Microbiological sampling and processing**

103 Routine, quality assured diagnostic blood culture services have been provided to the medical
104 and paediatric wards by the Malawi-Liverpool-Wellcome Programme (MLW) since 1998. Briefly,
105 1-2mL of blood was taken from neonates (up to 28 days old) with risk factors for sepsis
106 (i.e. maternal fever during labour, prolonged rupture of membranes, tachypnoea), or clinical
107 suspicion of sepsis (fever $>38^{\circ}\text{C}$, tachypnoea, tachycardia, reduced activity, seizures). For
108 some clinical records, information on age was only described in months of age and individuals
109 whose age was entered as being '1 month' were also considered as neonates for the purposes
110 of our study. For neonates with clinical suspicion of sepsis or other clinical suspicion of
111 meningitis (raised fontanelle, abnormal neurology), a lumbar puncture was also performed.

112 Blood was collected using aseptic methods and inoculated into a single aerobic bottle
113 (BacT/Alert, bioMérieux, Marcy-L'Etoile, France), then incubated using the automated
114 BacT/Alert system. Samples that flagged positive were Gram stained and Gram-negative bacilli
115 were identified by Analytical Profile Index (bioMérieux). Antimicrobial susceptibility testing was
116 determined by the disc diffusion method (Oxoid, United Kingdom). All *E. coli* isolates were
117 tested for their susceptibility to ampicillin, cefpodoxime (as an ESBL screen), chloramphenicol,
118 ciprofloxacin, co-trimoxazole and gentamicin, and those that were resistant to cefpodoxime also
119 had their sensitivity tested against amikacin, co-amoxiclav, ceftazidime, meropenem, pefloxacin
120 and piperacillin-tazobactam. BSAC breakpoints were used until 2018, at which time EUCAST
121 breakpoints were introduced. Details for the identification of other organisms are described
122 elsewhere⁴. *E. coli* isolates were stored at -80°C on microbank beads.

123 Between 1998-2010, diagnostic results were entered into ledgers that were later digitized, and
124 from 2010, PreLink, a Laboratory Information Management system was used and information

125 stored on an SQL database. The MLW database was screened from 2000 to 2021 to identify all
126 cases of *E. coli* infection in the hospital, and all *E. coli* isolates from neonates (recorded as less
127 than 29 days old or as 1 month old on ledgers) in the period from September 2012 to March
128 2021 were selected for whole genome sequencing. This time period was chosen as this was the
129 time period for which we had consistent metadata at the time of whole genome sequencing.

130 **Whole genome sequencing**

131 A single microbank bead was removed from all selected *E. coli* isolates, which was thawed,
132 streaked on MacConkey's media and this media incubated for 18-24 hours at 37°C. Plates with
133 growth of a single colony type then had a single colony pick taken and inoculated into 15ml of
134 buffered peptone water for 18-24 hours at 37°C. These samples were then centrifuged and the
135 supernatant was discarded. The pellet was then resuspended in buffer. For the short-read
136 sequencing the DNA was extracted using the QIAasympphony machine and QIAasympphony DSP
137 kit with onboard lysis, according to the manufacturer's instructions. Quality control was done
138 using Qubit and samples with a DNA volume of less than 200ng were repeated. Samples that
139 passed QC underwent Whole Genome Sequencing (WGS) at the Wellcome Sanger Institute at
140 364 plex on the Novaseq SP generating 150bp paired-end reads.

141 For long-read sequencing of selected isolates, DNA was extracted using the MasterPure
142 Complete DNA and RNA isolation kit following the manufacturer's instructions for the purification
143 of DNA from cell samples. DNA was then quality controlled using the Qubit dsDNA Broad
144 Range assay and the TapeStation (4150) system, using the Genomic DNA Screen Tape Kit.

145 Long-read sequencing was performed on a MinION MK1B sequencing device (ONT, U.K.).
146 Library preparation was carried out according to the manufacturers protocol, using the ligation
147 sequencing kit (SQK-LSK109) and Native Barcoding Expansion Kits (EXP-NBD104; all ONT).
148 Sequencing was carried out using a FLOW-MIN106 R9.4.1 flow cell (ONT). Two samples did
149 not produce sufficient data and were re-sequenced using the Native Barcoding Kit 24 V14
150 (SQK-NBD114.24, ONT), following the manufacturer's instructions. Sequencing was then
151 carried out using a R10.4.1 Flongle flow cell (ONT).

152 **Genomic sequence analyses**

153 Species confirmation was performed using Kraken v1.1.1¹⁶, and any sample with greater than
154 5% read content other than *E. coli* or Unclassified was excluded. Annotated assemblies for the
155 short read data were produced using the pipeline described previously¹⁷. De novo assembly of

156 genome sequences was performed using SPAdes v3.14.0¹⁸, trialing different kmer lengths
157 between 41 and 127 to find the optimal kmer length. An assembly improvement step was
158 applied to the assembly with the best N50 and contigs scaffolded using SSPACE v2.0¹⁹ and
159 sequence gaps filled using GapFiller v1.11²⁰. Assembly statistics were generated using the
160 Sanger Pathogens pipeline as available on github²¹. Samples with <20 or >200 contigs and a
161 genome size of <4.4MB or >5.6MB were excluded. Isolates with greater than 5% heterozygous
162 SNPs of the total genome were also excluded due to potential within-species contamination.
163 Automated annotation was performed using PROKKA v1.5²² and genus specific databases from
164 RefSeq²³. The improved assembly step uses software developed by the Pathogen Informatics
165 team at the WSI which is freely available for download from GitHub²⁴ under an open-source
166 license, GNU GPL 3. The improvement step of the pipeline is also available as a standalone
167 Perl module from CPAN²⁵.

168 Sequence type (ST) was determined using mlst v2.23^{26,27}. AMRFinderPlus v3.10.40 was used
169 to identify AMR genes²⁸. SRST2 v0.2.0²⁹ with the EcOH database³⁰ was used to determine the
170 O- and H-types for the bacterial isolates. Where an isolate had more than one predicted O- or
171 H-type, both sets were counted. We further aimed to confirm the subtypes (O1A, O6A, O18A,
172 O25B) of relevance for the vaccine, which are not distinguished from their related (but
173 immunologically non-identical) subtypes when using automated prediction via SRST2. For the
174 distinction of O1A and O1B, sequences were compared to the specific primers and probe used
175 previously³¹. For O18, there are four distinct subtypes described: O18A, O18A1, O18B and
176 O18B1. We distinguished O18ac (~O18A/A1) from O18ab (~O18B/B1) using the
177 presence/absence of an IS element inserted at a location immediately upstream of the *wzz*
178 gene³². To distinguish O25A and O25B, we assessed the operon structure as described
179 previously to distinguish these types³³. We were not able to identify any description on the
180 genetic differences of O6A and other O6 sub-types (referred to hereafter as O6?).

181 To further investigate their antigenic structure we performed long read sequencing using the
182 Oxford Nanopore platform for 14 selected isolates. Basecalling and demultiplexing on raw long-
183 reads was performed with guppy v 2.6.1³⁴ using the super-accurate model for basecalling,
184 adapters removed with porechop v 0.2.4³⁵, and low-quality reads were removed with filtlong
185 v0.2.2³⁶ before assembly. Long-read-first hybrid assemblies from isolates sequenced on the
186 Oxford Nanopore platform were produced using Flye v 2.9.3³⁷, then visualised with Bandage
187 v0.8.1³⁸. Long-read polishing was performed using Medaka v1.8.0;

188 <https://github.com/nanoporetech/medaka>³⁹, then short-read polished with Polypolish v0.6.0⁴⁰
189 and Pypolca v0.3.0⁴¹.

190 Assembled genomes were annotated using prokka as described above, and the O:H loci were
191 investigated initially using the ECTyper⁴² which uses assemblies as input, which yielded
192 comparable results to the srst2 EcOH search (see Supplementary Table 4); we then assessed
193 these isolates further manually to determine the exact structure of these untyped, potentially
194 novel O-Ag types.

195 **Statistical analysis**

196 Statistical analysis was done in the R statistical programming language⁴³, using Rstudio⁴⁴ and
197 the packages here⁴⁵, tidyverse⁴⁶ and lubridate⁴⁷. Graphical data representation was performed
198 using the packages ggplot2⁴⁸, ghibli⁴⁹, RColorBrewer⁵⁰, pals⁵¹, MetBrewer⁵², gggenes⁵³ and
199 ggpubr⁵⁴.

200 *E. coli* sample positivity was calculated as the number of samples which were positive for *E. coli*
201 per 1,000 samples taken (blood culture or CSF) according to the equation below.

$$P = n \div N \times 1,000$$

202 Where P is the sample positivity rate, n is the number of *E. coli* infections and N is the total
203 number of blood culture or CSF samples taken.

204 The change in AMR over time was estimated by regressing resistance pattern of isolates
205 against year of occurrence using a generalized linear model in R (glm function) with the
206 binomial family and a logit link statement. Isolates were given a binary categorization of 1 if they
207 were resistant or had intermediate resistance to an antibiotic and 0 if they were sensitive. Plots
208 of AMR trends over time had lines of best fit that utilized a linear model.

209 For the rarefaction curves in the main text, lines showing hypothetical coverage for vaccines
210 based on the n most frequent O-types or H-types were compared to lines showing the
211 hypothetical coverage based on the EXPEC4V and EXPEC9V vaccine O-types; the
212 supplementary data furthermore shows our analysis for the original ExPEC10V composition. For
213 isolates with more than one allele for H-type, both were included in the rarefaction curve. For H-
214 types the rarefaction curve goes above 1 as there were multiple isolates with more than one H-
215 type. For O-types there was only one O-type per isolate (there was one O-type with both O8
216 and O160 sections, but it is unclear whether this expresses both sugar molecules or is a hybrid

217 O-antigen) so the rarefaction curve only goes to 1. The analysis was repeated for the
218 supplementary materials but only counting each isolate with multiple H-types once (the H-type
219 that had the highest population frequency was chosen). The other allele was ignored, as
220 theoretical protection from the vaccine was assumed by the H-type or O-type that was most
221 prevalent.

222 **Ethics statement**

223 This study was ethically approved by the Kamuzu University of Health Sciences College of
224 Medicine Research Ethics Committee (COMREC P.06.20.3071). The ID numbers used in this
225 manuscript are specimen IDs generated by the MLW diagnostic laboratory and not patient IDs;
226 only the research team and members of the clinical staff in hospital that have access to the
227 password protected laboratory information management system at MLW would be able to make
228 the link to an individual.

229 **Results**

230 **Clinical characteristics**

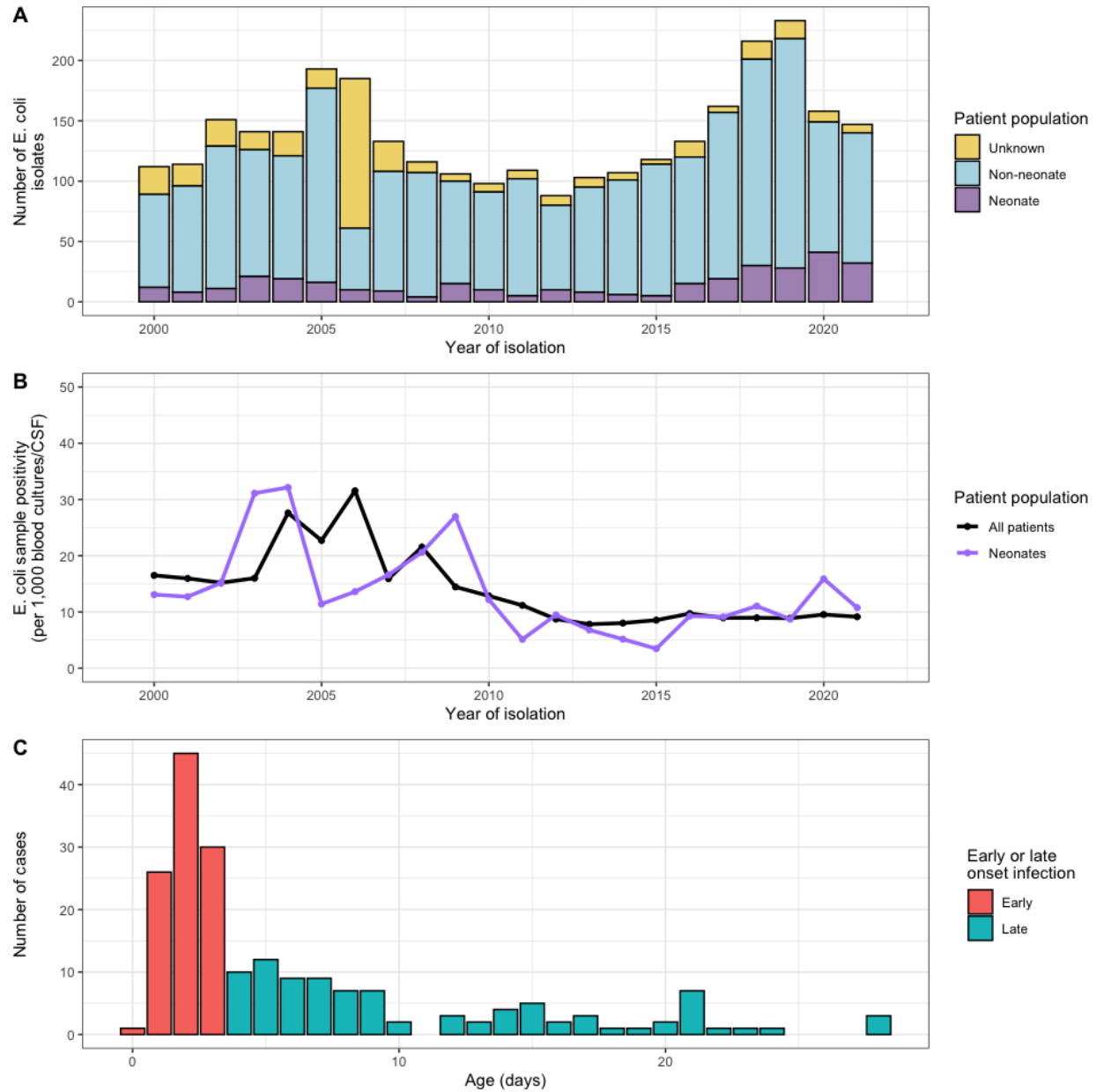
231 There were 3394 *E. coli* isolated from 264692 blood culture and CSF tests over the period from
232 2000 - 2021. The number of cases of *E. coli* per year for all ages ranged from 88 to 233, with an
233 average of 139 cases per year. The number of cases of *E. coli* per year for neonates ranged
234 from 4 to 41, with an average of 15 cases per year. (Figure 1A). The number of blood culture or
235 CSF samples taken per year for all ages ranged from 2796 to 26230 with an average of 11028.8
236 in a year. The number of blood culture or CSF samples taken per year for neonates ranged from
237 77 to 3211 with an average of 1227.7 per year. The positivity rate per 1,000 blood culture or
238 CSF samples for all age groups was highest in 2006 and lowest in 2013, with an average
239 positivity rate of 14.1/1000 samples/year. For neonates it was highest in 2004 and lowest in
240 2015, with a similar average positivity rate of 13.7/1,000 samples (Figure 1B).

241 We identified 201 *E. coli* isolated from neonates in the period from September 2012 to March
242 2021 (Figure 2C); 95/201 (47.3%) were female, with a median age of 3 [IQR 2 - 8] days (Figure
243 1D). Early onset sepsis accounted for 109/201 (54.2%) of cases, with late onset sepsis
244 accounting for the rest (Figure 1C). Of these isolates 163/201 (81.1%) were cultured from blood
245 and 38/201 (18.9%) from CSF.

246 There were 110/201 (54.7%) cases from the Chatinkha nursery, 62/201 (30.8%) from paediatric
247 nursery, 12/201 (6%) were from the paediatric A&E, 12/201 (6%) were from the PICU/HDU and
248 110/201 (54.7%) were from the paediatric surgical ward (Supplementary Figure 1).

249 Of these 201 isolates, 192 were recovered for WGS and 170 passed QC (22 failed; Figure 2C).
250 17 isolates were excluded due to contamination, and 17 were excluded due to failure of
251 assembly or poor assembly metrics (twelve isolates failed both, and five of each just failed due
252 to one reason). There were no isolates that were excluded due to potential within-species
253 contamination. There was one duplicate isolate (two sequencing runs of the same isolate from
254 the same sample), of which one isolate was removed from analysis, so the total number of
255 isolates analysed was 169.

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).



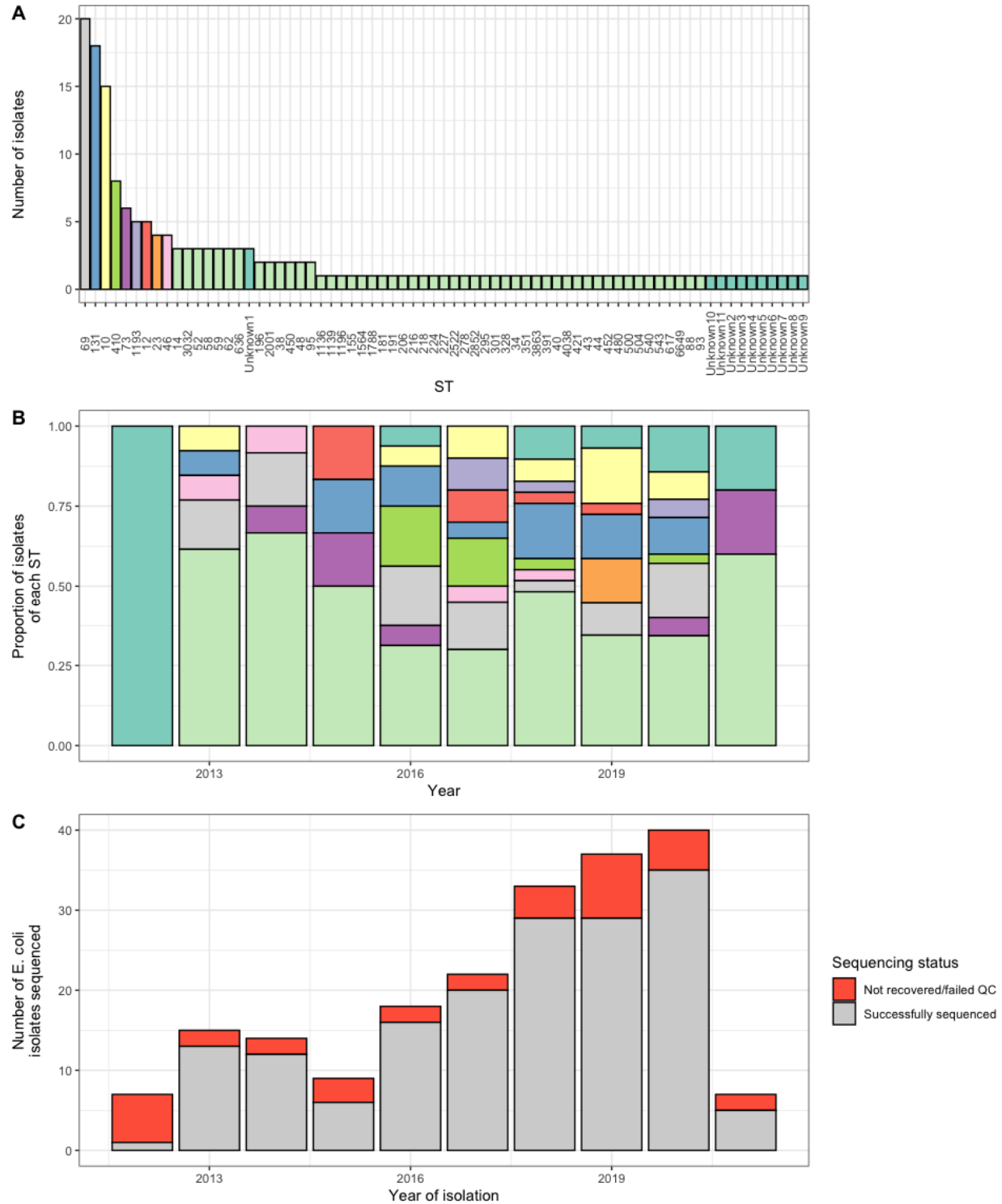
256

257 *Figure 1A) Numbers of E. coli cases per year at QECH. Bars represent the crude frequency of*
258 *E. coli infection for each year from 2000 - 2021, with the different colours representing the*
259 *different age groups of the patients. B) Blood culture and CSF positivity rate (per 1,000 blood*
260 *culture or CSF samples) of E. coli in neonates and the entire patient population (including*
261 *neonates). C) Age range of neonates in the current study, colours highlighting early (< 72 hours*
262 *of life) or late (>72 hours of life) onset infection.*

263 **Population structure**

264 There were 71 different STs represented in the collection (Figure 2A & B; 60 typed and 11
265 untyped). The most frequently isolated STs were ST69 with 20/169 (11.8%) isolates, ST131
266 with 18/169 (10.7%) isolates, ST10 with 15/169 (8.9%) isolates and ST410 with 8/169 (4.7%)
267 isolates. There were 13/169 (7.7%) isolates with untyped STs using MLST which represented
268 eleven different MLST allele patterns (ten singles and three of the same allele pattern). Over
269 half of the observed STs were only represented by a single isolate (38/71, 53.5%) showing that
270 neonates tested here were exposed to and infected by a highly diverse pool of *E. coli* that span
271 the species phylogeny. ST410 was disproportionately found in the CSF rather than blood culture
272 samples (6/8 [75%]), compared to ST69 (2/20 [10%]), ST131 (2/18 [11%]) and ST10 (2/14
273 [14%]) which were found primarily in blood culture samples.

274 Importantly, the ST diversity was also highly variable over time. 50/71 (70.4%) STs were only
275 found in a single year, and 7/71 (9.9%) STs were found in only two years, with each year
276 showing a similar pattern of high diversity during the entire study period. Frequently occurring
277 STs were also prominent in different years (e.g. ST410 in 2016 and 2017), only ST69 was
278 consistently isolated and was the most frequently isolated or joint most frequently isolated ST in
279 5 out of 7 of the years with more than ten isolates. Even the most prevalent STs like ST69 or
280 ST131 however fluctuated over time, with numbers between 1/29 (3%) and 6/35 (17%) for
281 ST69, and 1/20 (5%) and 5/29 (17%) for ST131, respectively, with no clear trend over time
282 observable for any of the main STs. We also note that untyped isolates are derived from a
283 range of years, including the most recent data. This indicates that these are not representing
284 older lineages that might not be covered well in databases consisting mainly of recent samples,
285 but indicating a high undescribed diversity circulating at present time. Two samples (one blood
286 culture and one CSF) were polymicrobial. They both contained two different colony
287 morphologies. The CSF sample contained one isolate of ST14 and one of ST69 and the
288 bloodstream sample contained one isolate of ST1136 and one of ST10.



289

290 *Figure 2A) Frequency of different STs in the collection. B) Frequency of the different STs by*
 291 *year. C) Number of E. coli isolates selected for WGS for this current study. 2012 and 2021 were*
 292 *years for which isolates were only selected for part of the year.*

293 **O-antigen and H-antigen diversity**

294 There were 63 O-types found in the collection, none of which were identified in more than 10%
295 of the isolates. The most frequently isolated were O15 with 15/169 (8.9%) isolates, O25B with
296 15/169 (8.9%) isolates and O8 with 13/169 (7.7%) isolates (Figure 3A). These same O-types
297 (O15, O25B and O8) were also the only ones found in greater than 75% of the years (6 out of 7
298 or more) that had more than 10 isolates (Figure 3B). Similar to ST types there was no sign of
299 the population becoming increasingly dominated by any types over time, the composition
300 between years differed strongly (Figure 2B). There were no years in which any O-type
301 represented greater than 20% of the isolates, the largest proportion of isolates belonging to a
302 single O-type per year were O11 and O8 which were both associated with 3/16 (18.8%) of all
303 cases in 2016 (Figure 3B).

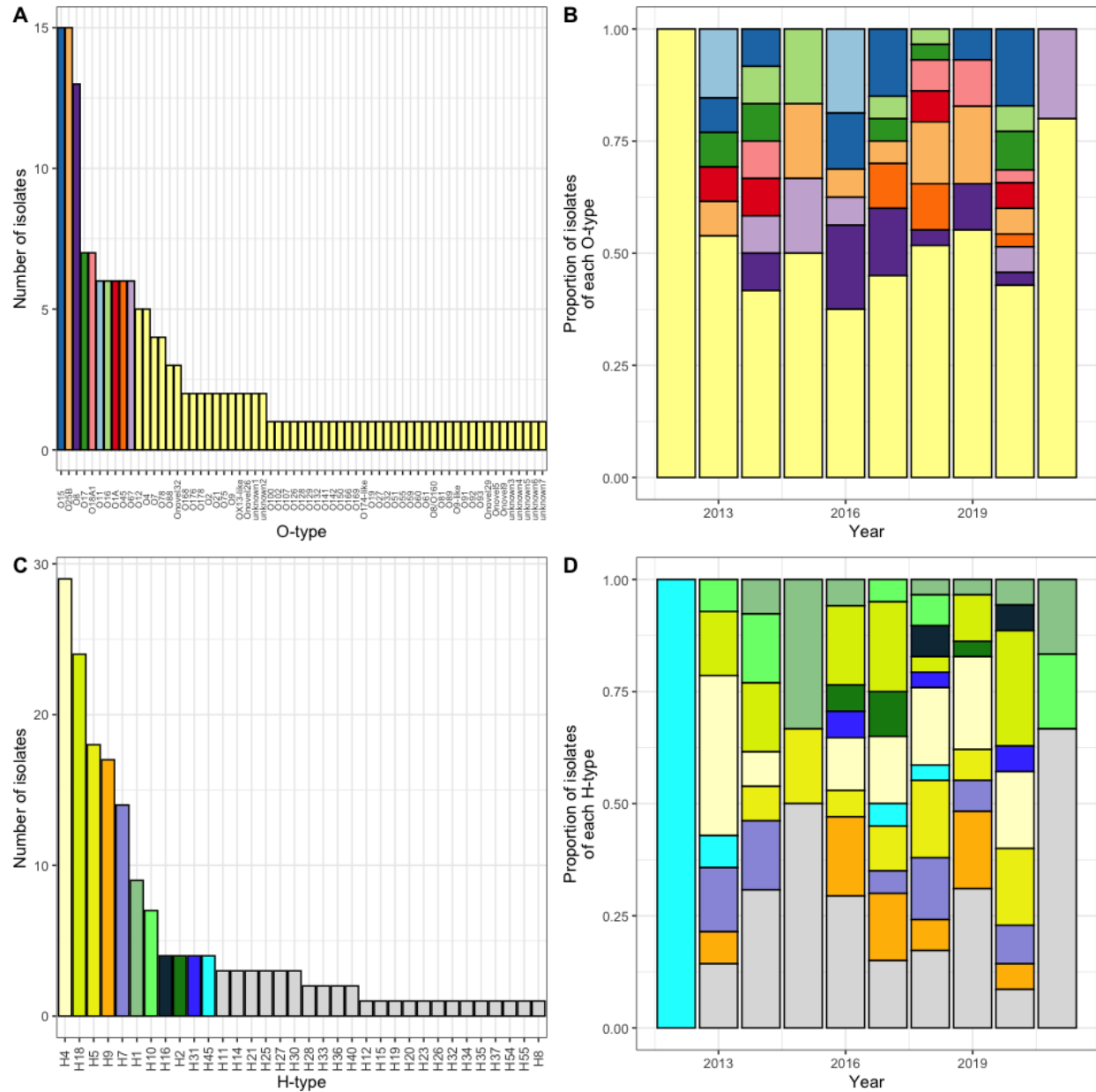
304 We performed long-read sequencing for 14 isolates which either had no O-type call or had
305 multiple O-type calls, to determine the genomic region between *galF* and *gnd* where the O-
306 antigen type locus is usually found in *E. coli* (Figure 4). Having no O-antigen is highly unusual
307 and leads to increased susceptibility to antimicrobial stress, and thus seems unlikely to be
308 present in clinical isolates. Ten different O-antigen loci were revealed in these 14 isolates, with
309 four of them represented by two isolates each. Two isolates were confirmed as O178 and one
310 appeared to be a hybrid of O8 and O160, whilst seven of these ten O-types were so far
311 undescribed (Figure 4) and one showed similarity to the OX-13 gene in *Salmonella* (BKRHXR).
312 Three of the isolates (CAAH3Y, CNS75S and BKQ37E) have loci heavily disrupted by insertion
313 elements and seem to lack the components for an export machinery (*wzm/wzt* or *wzx/wzy*). It
314 remains to be investigated whether they acquired an entirely unrelated O-antigen locus from a
315 different organism that integrated into a different part of the genome and encodes for export
316 machineries sufficiently different to not even get recognized by read-based searches, or
317 whether these isolates indeed do not encode for a classical O-antigen. One of the isolates
318 (CAAXI3) has a similarly disrupted *galF/gnd* site but encodes for a potentially functional O-
319 antigen locus on a plasmid (CAAXI3_2), raising interesting questions regarding the expression
320 of this O-antigen locus and whether this will remain plasmid-located or eventually become
321 integrated into the disrupted chromosomal location.

322 There were 34 H-types found in the collection, of which 4 H-types were each identified in more
323 than 10% of isolates. These were H4 with 29/173 (16.8%), H18 with 24/173 (13.9%), H5 with
324 18/173 (10.4%) isolates and H9 with 17/173 (9.8%) isolates (Figure 3C). Five H-types were
325 found in at least 75% of the years (6 out of 7 or more) that had more than 10 isolates (H4, H18,

326 H5, H9 and H7). H4 and H18 were the only H-types responsible for greater than 20% of the
327 isolates in any year with more than 10 isolates (in two years each) but no H-type was identified
328 in over 50% of isolates in a single year. The largest proportion of isolates belonging to a single
329 H-type in a single year was H4 with 5/14 (35.7%) of cases in 2013 (Figure 3D). There were
330 4/169 (2.4%) isolates which had more than one H-type and may be able to undergo phase
331 variation for immune escape, hence the denominator of 173 above.

332 Considering the O and H types in light of body site of isolation (e.g. bloodstream or CSF), 20
333 different O-types were encoded on isolates derived from CSF. O8 was found in 8/33 (24.2%)
334 isolates from CSF, compared to 4/133 (3.0%) of bloodstream isolates ($\chi^2 p = 0.0001$). This was
335 partly due to ST410 (six isolates) which all encoded for O8, however there were also two other
336 O8 isolates (one ST58 and one ST155) that occurred in CSF. There were six other O-types that
337 occurred in two CSF samples (O78, OX13-like, O25B, O1A, O18A1 and O15) with the rest
338 occurring in only one CSF sample. Of these, only OX13-like was found in a significantly different
339 percent of CSF samples compared to bloodstream isolates (2/33 [6%] vs 0/133 [0%], $\chi^2 p =$
340 0.049). There were 17 H-types isolated from CSF. H9 was found in 8/33 (24.2%) isolates from
341 CSF, compared to 8/133 (6.0%) of bloodstream isolates ($\chi^2 p = 0.004$). This was again partly
342 due to the six ST410 isolates, however there were also two ST23 isolates that expressed H9
343 found in CSF. H5, H4, H7, H18 and H1 all occurred in more than one CSF isolate, with the other
344 H-types occurring in just one CSF isolate.

345 Excluding STs for which there was just one isolate, we examined whether multiple O-types or H-
346 types were found in isolates of a single ST. The median number of O-types per ST for the STs
347 that met this criterion was 2, with 10/23 (43%) encoding for just a single O-type and 13/23 (57%)
348 encoding for multiple O-types. The median number of H-types per ST for the STs that met our
349 criteria was 1 with 14/26 (54%) encoding for just a single H-type and 12/26 (46%) encoding for
350 multiple H-types. Three of the most frequently occurring STs encoded for multiple O-types and
351 H-types. ST10 showed the highest diversity of O-types and H-types, with 10 different O-types
352 and 7 different H-types. ST131 covered 3 different O-types (O25B, O11 and O16) and 2
353 different H-types (H4 and H5), which occurred from 2013 to 2020 and often with multiple O-
354 types or H-types in the same year. ST69 isolates included 4 different O-types and 2 different H-
355 types. ST410 on the other hand occurred frequently from 2016 to 2020 but all isolates encoded
356 for only a single O-type (O8) and a single H-type (H9).

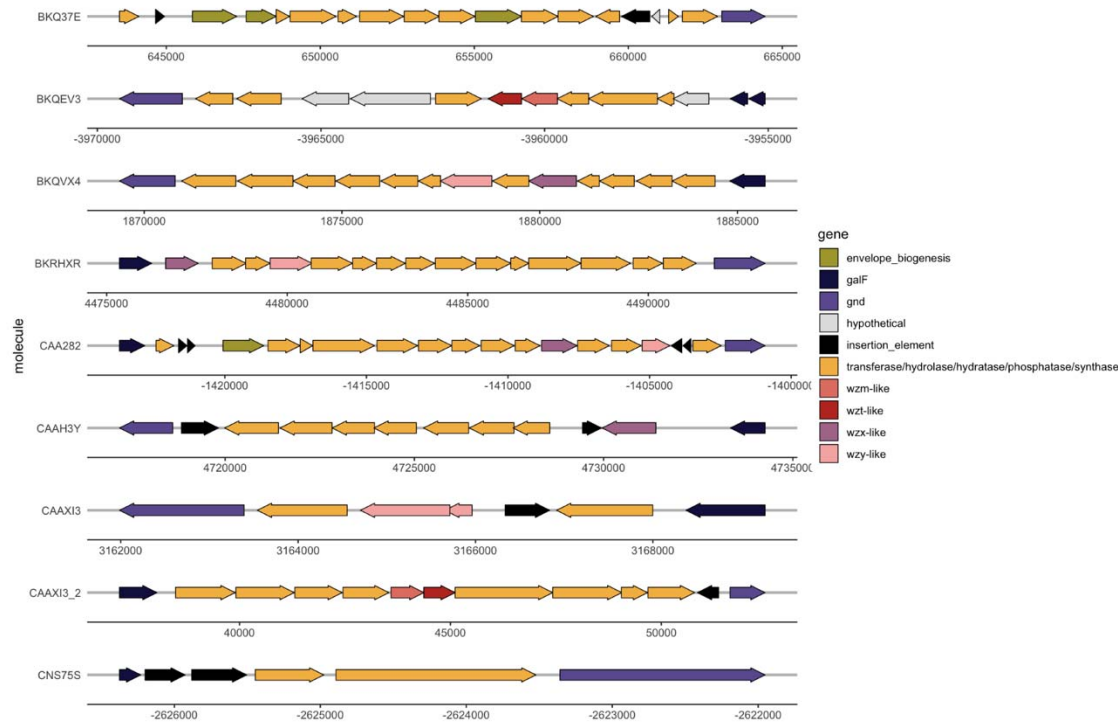


357

358 *Figure 3A) A bar chart showing the frequency of the different O-types. Where an isolate had*
359 *more than one O-type gene, this was counted twice. B) A bar chart showing the proportion of*
360 *isolates per year that had different O-types. The colours are the same as those represented in*
361 *Figure 3A.C) A bar chart showing the frequency of the different H-types. Where an isolate had*
362 *more than one H-type gene, this was counted twice. B) A bar chart showing the proportion of*
363 *isolates per year that had different H-types. The colours are the same as those represented in*
364 *Figure 3C.*

365

366



367

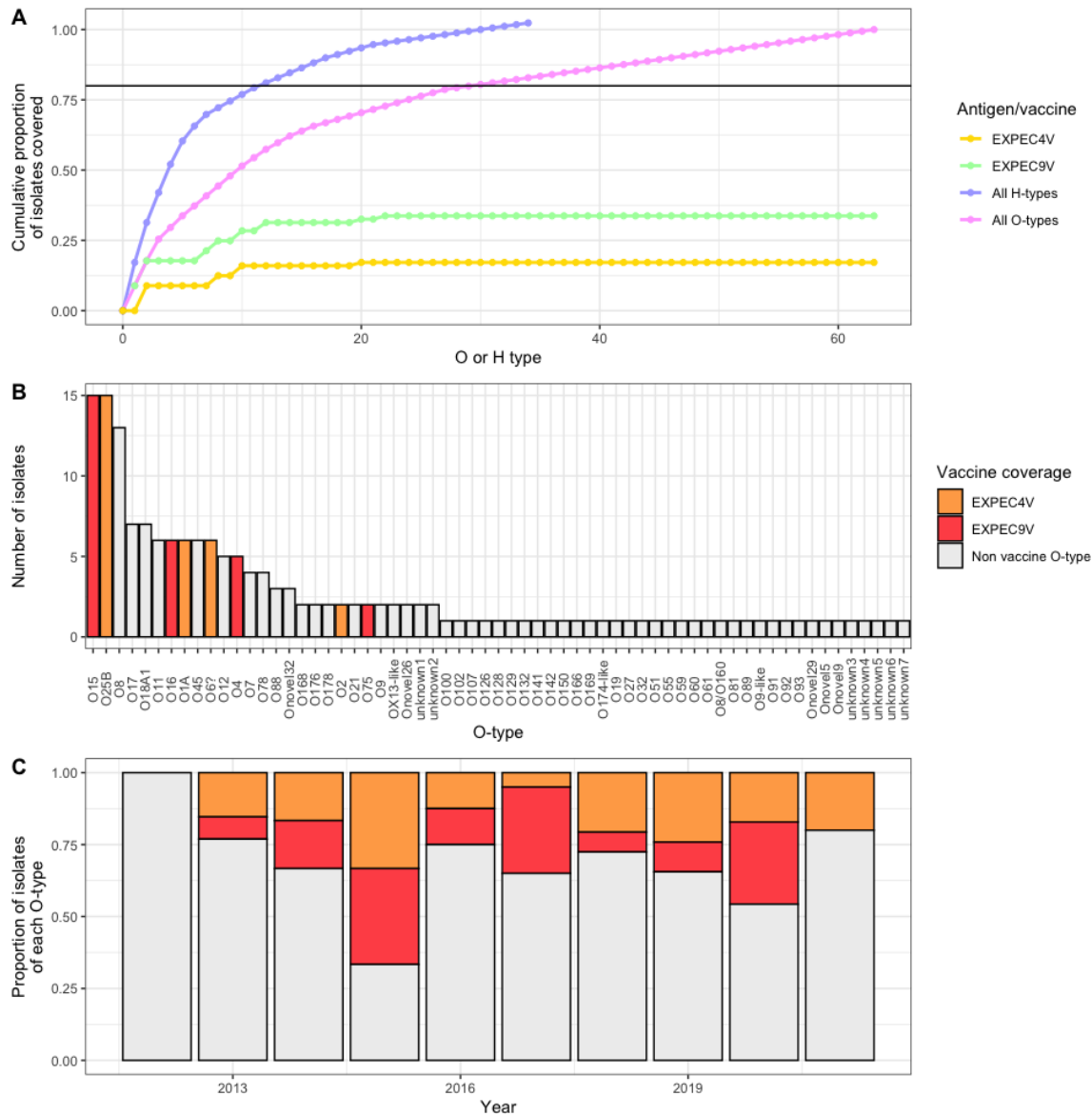
368 *Figure 4. A schematic showing the operon structure of the novel O-antigen genes. Each row*
369 *represents a different novel O-type from a single isolate. For isolates with > 0.9 sequence*
370 *homology for the O-antigen gene, only one isolate was selected for representation.*

371 The EXPEC9V conjugate vaccine (which covers the O1A, O2, O4, O6A, O15, O16, O18A,
372 O25B and O75) might be expected to confer immunity to up to 57/169 (33.7%) of these cases,
373 the original 10V composition (including O8) would have covered up to 77/169 (45.6%),
374 demonstrating a loss of 11.9% by the removal of just one O-antigen of high prevalence in our
375 setting (Figure 5A and Supplementary Figure 2). The EXPEC4V vaccination (which covers O1A,
376 O2, O6A, and O25B) would cover at maximum 29/169 (17.2%) of cases (assuming the O6
377 isolates are of the O6A subtype, something that we have not been able to confirm; Figure 5A;
378 5B; methods). Analyzing the data by year (including only years with greater than 10 isolates) the
379 EXPEC9V vaccine covered fewer than 50% of the isolates' O-types in every year, and 3 out of 7
380 years covered less than 30% of the vaccine O-types, with the lowest coverage in 2013 where
381 only 23% of the isolates were from vaccine O-types, and importantly we observe no major
382 change in coverage over time (Figure 5C). The EXPEC4V vaccine covered less than 30% of the
383 isolates' O-types in every year, and 5 out of 7 years covered less than 20% of the vaccine O-
384 types, with the lowest coverage in 2017 where only 5% of the isolates were from vaccine O-

385 types, with fluctuations over time that do not indicate any improvement in coverage in future
386 (Figure 5C). Regarding the isolates which were resistant to first- and second line antimicrobial
387 therapy (benzylpenicillin, gentamicin and ceftriaxone), the EXPEC9V vaccine would be
388 expected to confer immunity to 13/34 (38.2%) of these cases and the EXPEC4V vaccine would
389 be expected to confer immunity to 9/34 (26.5%) of these isolates.

390 To cover 80% of cases, an O-antigen based vaccine would need to offer protection against the
391 top 30 O-types. In contrast, to cover the top 80% of these cases, an H-type vaccine would only
392 need to cover the top 12 H-types (Figure 5A). If the four most frequently occurring O-types in
393 our setting were selected from our cohort for a vaccine (O15, O25B, O8 and O17) this vaccine
394 would cover 50/169 (29.6%) of cases, ranging from 21% to 40% per year, and the nine most
395 frequently occurring O-types (O15, O25B, O8, O17, O18A1, O11, O16, O1A and O45) would
396 represent just under half of the isolates 81/169 (47.9%), ranging from 33% to 67% per year, with
397 numbers fluctuating over our study period showing no indication that there would be an increase
398 in coverage in future.

399



400

401 *Figure 5A) Rarefaction curve showing the theoretical protection given against vaccines covering*
402 *the most frequently isolated H-types and O-types, as well as the potential protection given by*
403 *the EXPEC9V and EXPEC4V. The horizontal line shows the point at which 80% of isolates*
404 *would be covered. For isolates with more than one H-type both were counted, there were*
405 *multiple isolates with more than one H-type so the line for H-type goes above 1. Supplementary*
406 *Figure 3 shows the same graph but where isolates had more than one H-type called they were*
407 *only counted once. B) A bar chart showing the frequency of the different O-types. C) A bar chart*
408 *showing the proportion of isolates per year that had different O-types. The colours are the same*
409 *as those represented in Figure 5B.*

410 **Antimicrobial resistance and plasmid replicons**

411 At the time of the study the first line treatment for neonatal sepsis and meningitis in QECH was
412 benzylpenicillin and gentamicin, with second line treatment ceftriaxone. *E. coli* is intrinsically
413 resistant to benzylpenicillin and isolates with acquired resistance to gentamicin and ceftriaxone
414 were therefore difficult to treat (42/194 [21.6%]). There was occasional but limited use of
415 amikacin or meropenem for neonates with proven or high suspicion of ceftriaxone resistance or
416 who were very unwell. The use of meropenem and amikacin increased over the study period.

417 We identified AMR genes against all major classes of antibiotics and several efflux pump
418 systems, in line with the phenotypic resistances detected. The number of AMR genes varied by
419 ST, with ST410 (mean 24.0, SD 1.9) and ST131 (mean 20.6, SD 4.8) having the greatest
420 number of average AMR genes per isolate. ST10 had a lower number of AMR genes per isolate
421 (mean 8.9, SD 1.6), whilst ST69 was intermediate (mean 12.9, SD 1.5). ST410 was present
422 only from 2016 onwards which may partly explain the higher number of resistance genes, whilst
423 the other STs, including ST131 were present throughout the study period.

424 The number of *E. coli* isolates that were resistant to ampicillin was 143/190 (75.3%) and was
425 stable over the period (-0.3% change per year; $p = 0.96$; Figure 6A). *bla*_{EC} genes, which are
426 chromosomally encoded in *E. coli*, were found in all 169 isolates (*bla*_{EC-15}, *bla*_{EC-5}, *bla*_{EC-18} and
427 *bla*_{EC-8}). The most frequently occurring plasmid-encoded beta-lactamase penicillinase genes
428 included *bla*_{TEM-1} found in 121/169 (71.6%) isolates, and *bla*_{OXA-1} found in 21/169 (12.4%) of
429 isolates (Supplementary Figure 3). Whilst only 49/197 (24.9%) were resistant to gentamicin, this
430 however showed a temporal trend, increasing from 3/15 (20%) in 2013 to 16/39 (41%) in 2020
431 (22.8% change per year; $p = 0.0035$; Figure 6A). Gentamicin resistance mechanisms were
432 mainly variants of the gene *aac*(3), *aac*(3)-IId found in 23/169 (13.6%) isolates and the *aac*(3)-
433 IId gene found in 12/169 (7.1%) isolates (Supplementary Figure 3).

434 Ceftriaxone was the second-line treatment at the time of study, and 55/198 (27.8%) were
435 resistant to ceftriaxone and increased over the period from 0/7 (0%) in 2012 to 18/39 (46.2%) in
436 2020 (31.8% change per year; $p = 0.00014$; Figure 6B); and 42/55 (76.4%) of the isolates
437 resistant to ceftriaxone were also resistant to ampicillin and gentamicin, hampering the
438 effectiveness of all first- and second-line treatments. The increase in ceftriaxone resistance was
439 due to the widespread ESBL gene *bla*_{CTX-M-15}, which was detected in 39/169 (23.0%) isolates
440 (Supplementary Figure 4), whilst other alleles, *bla*_{CTX-M-14} and *bla*_{CTX-M-27}, could be identified in
441 only 1/169 (0.6%) isolate each, both from 2018 (Supplementary Figure 2). ESBL genes were

442 frequent in ST410 (8/8 [100%]) and ST131 (9/18 [50%]) as observed in other studies^{55, 56}, and
443 infrequent in ST69 (3/20 [15%]) and ST10 isolates (1/15 [7%]). The proportion of isolates from
444 late onset infection resistant to ceftriaxone was 16.9% (95% CI 3.5 - 30.2%) higher than from
445 early onset cases (37.2% vs 20.4%, $p = 0.012$); the same pattern was observed for gentamicin
446 with 15.8% (95% CI 2.8 - 28.7%) more resistant isolates in late onset infections compared to
447 early onset cases (33.6% vs 17.9%, $p = 0.016$), whereas similar proportions were observed for
448 ampicillin with 78.5% vs 73.2% in late and early, respectively ($p = 0.50$).

449 Alternatives for isolates resistant against all of the above antimicrobials are amikacin or
450 carbapenems, and so far only a small proportion, 7/67 (10.4%), were resistant to amikacin. This
451 increased from 1/6 (17%) in 2016 (amikacin was not routinely tested until 2016) to 7/21 (33%) in
452 2019 (52.7% change per year; $p = 0.042$; Figure 6C). The main gene identified was the
453 amikacin resistance gene *aac(6')-Ib-cr5* in 21/170 (12.4%) isolates (Supplementary Figure 2).
454 Of the 42 isolates resistant to all first and second-line agents 11/42 (26.2%) were causing
455 meningitis and a further 5/42 (11.9%) were resistant to amikacin, leaving only meropenem as
456 effective treatment for these (amikacin does not reliably penetrate the blood-brain barrier). In
457 line with the low levels of carbapenem resistance identified in other studies from this setting, no
458 isolates showed phenotypic meropenem resistance.

459 Other antimicrobials are not regularly used on the neonatal unit but are tested for routinely for *E.*
460 *coli*. Fluoroquinolones are still widely used in other wards, and 45/199 (22.6%) were resistant to
461 ciprofloxacin which increased over the period from 0/7 (0%) in 2012 to 12/39 (30.8%) in 2020
462 (20% change per year; $p = 0.013$; Figure 6D). There were several isolates with fluoroquinolone
463 resistance mutations, the most frequent were the *gyrA* mutations *gyrA_S38L* found in 38/169
464 (22.5%) isolates and *gyrA_D87N* found in 24/169 (14.2%) isolates, the *parC* mutation
465 *parC_S80I* found in 26/169 (15.4%) (Supplementary Figure 2). We further note 37/169 (21.9%)
466 isolates with *parE* mutations which are not described as sufficient to provide resistance for *E.*
467 *coli*, but could lead to reduced susceptibility or higher resistance levels if an additional mutation
468 is present. All ST410 isolates encoded four different, acquired fluoroquinolone resistance genes
469 (two *gyrA* mutations, one *parC* and one *parE* each). Likewise, we identified at least one
470 acquired fluoroquinolone resistance gene in all ST131 isolates *gyrA*, *parC* and *parE*).
471 Fluoroquinolone resistance genes (*gyrA*, *parC* and *qnrS1*) were found in 8/20 (40%) of ST69
472 isolates and 3/15 (20%) of ST10 isolates.

473 Chloramphenicol resistance has been observed in other isolates in this setting to decrease, and
474 in line with this we observed 41/195 (21%) isolates resistant to chloramphenicol with a

475 decreasing trend, from 2/7 (28.6%) in 2012 to 3/39 (7.7%) in 2020 (-30.5% change per year; $p <$
476 0.0001; Figure 6D). The most frequently occurring chloramphenicol genes were *catA1* found in
477 22/170 (12.9%) isolates and *catB3*, found in 20/170 (11.8%) isolates (Supplementary Figure 2).

478 Overall 54/67 (80.6%) were resistant to co-amoxiclav (not used on the neonatal unit and
479 infrequently used in other hospital wards) and the proportion decreased slightly over the period
480 (-20.7% change per year; $p = 0.33$; Figure 6D). Resistance to co-trimoxazole (used as
481 prophylaxis against *Pneumocystis pneumonia* in HIV patients) was high over the period at
482 183/200 (91.5%) and increased slightly (14.9% change per year; $p = 0.13$; Supplementary
483 Figure 2). Colistin resistance was not tested in our cohort, but genes conferring colistin
484 resistance were found relatively frequently, with *pmrB_E123D* found in 48/169 (28.4%) isolates
485 and *pmrB_Y358N* found in 36/169 (21.3%) isolates (Supplementary Figure 2).

486

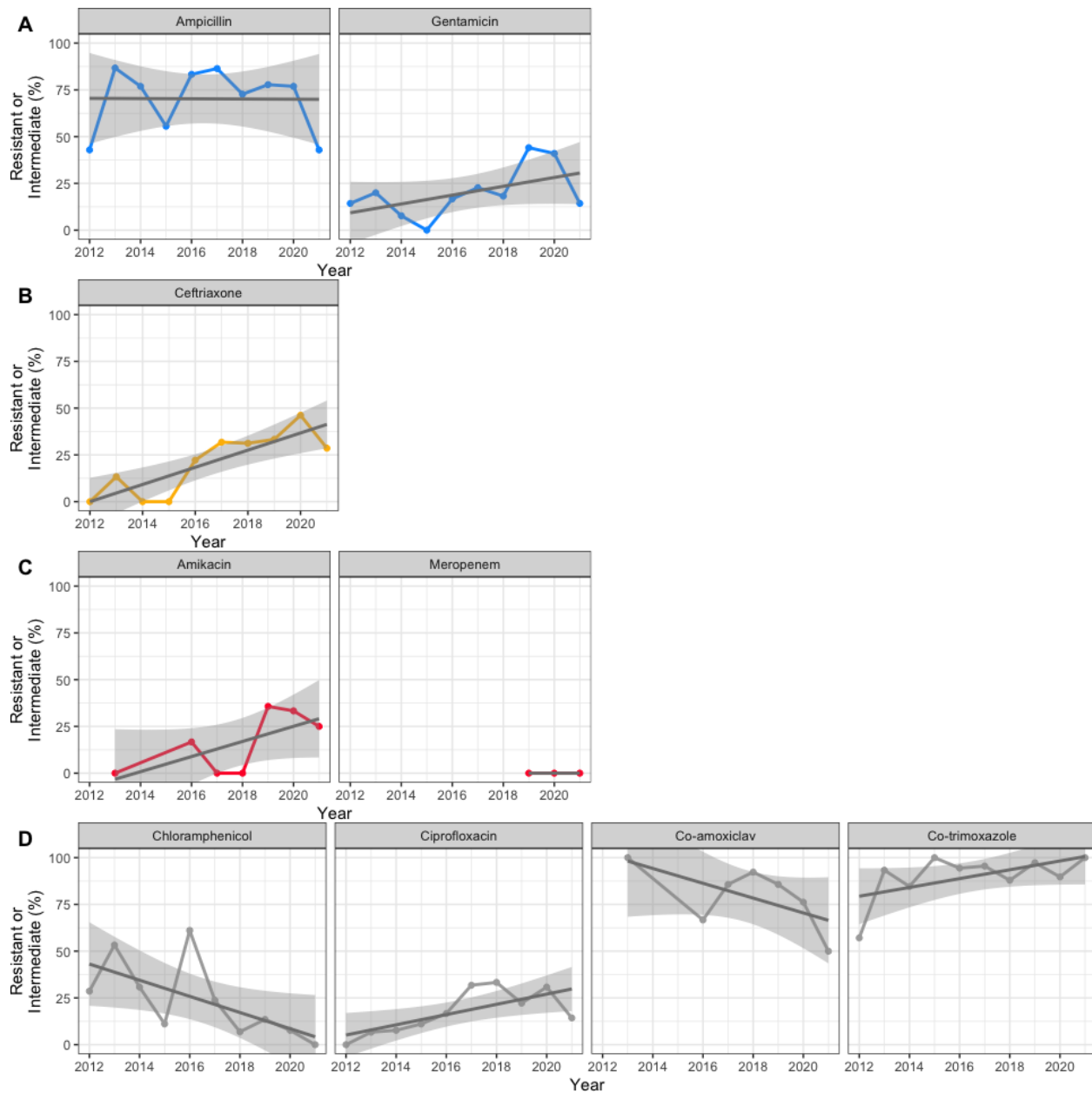
487

488

489

490

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).



491

492 *Figure 6. The proportion of E. coli isolates that were phenotypically resistant to different*
493 *antibiotics by year. A) First-line antibiotics for neonatal infection B) Second-line antibiotics for*
494 *neonatal infection C) Occasionally used antibiotics for neonatal infection D) Antibiotics not used*
495 *in neonates, but used elsewhere in the hospital or in the community.*

496 There were 34 different plasmid replicons found in the dataset. The most frequently identified
497 were the commonly detected IncF plasmid replicons that frequently carry resistance cassettes,
498 with IncFIB_AP001918 found in 107/169 (63.3%) of isolates, IncFI found in 91/169 (53.8%) of
499 isolates and IncFII_p found in 46/169 (27.2%) of isolates (Supplementary Table 1). Multiple

500 different plasmid replicons were also found of the Col, IncH and IncX types, with other types
501 found less frequently.

502 **Discussion**

503 This study highlights the challenges of controlling neonatal sepsis with vaccines in a low income
504 setting when there is a paucity of data from this region describing the nature and diversity of
505 isolates causing disease. We present the trends in predicted serogroup epidemiology of a
506 collection of *E. coli* isolated from neonates in a single, large teaching hospital in Malawi from
507 2012 to 2021. Our study reveals that O-antigen vaccines would need a high valency (30 O-
508 types) to achieve protection against greater than 80% of isolates, and vaccines in current
509 development for use in elderly populations in high-income countries would offer protection
510 against only one third of the *E. coli* isolated in this study.

511 Our study, consistent with similar studies from high income countries^{57, 58}, found approximately
512 50% of *E. coli* cases to be from early-onset sepsis (EoS) and 50% from late-onset sepsis (LoS).
513 We found higher rates of ceftriaxone and gentamicin resistance in the LoS cases compared to
514 the EoS cases (though this finding is not typical of other studies^{59, 60}), which might imply that in
515 our setting these two groups have different epidemiology (e.g. EoS cases being maternally
516 transmitted and LoS cases deriving from the hospital environment or from infection in the
517 community). Almost a fifth of our isolates were from neonatal meningitis cases, and *E. coli* is an
518 important cause of meningitis in neonates, including in low-income countries⁶¹. Neonatal
519 meningitis is particularly concerning as it is associated with greater morbidity and mortality,
520 requires longer treatment (minimum 21 days of antimicrobial therapy for *E. coli* meningitis
521 compared to 7 days for bloodstream infection) and certain drugs such as amikacin cannot be
522 used for meningitis due to probable poor blood-brain barrier penetration, leaving very few
523 options for isolates resistant against third-generation cephalosporins.

524 The numbers of *E. coli* cases per 1,000 blood culture or CSF tests in all age groups decreased
525 from the period 2000 to 2012 but was relatively stable from 2012 onwards (which is the time
526 period for which we had genomic data). There were peaks in the absolute numbers of cases in
527 2005 and 2019, although the peak in 2019 appears to be primarily related to greater patient
528 numbers as there was no increase in the number of cases per 1,000 blood culture or CSF tests
529 done. This contrasts with *K. pneumoniae* at the same site over the same period⁶² which showed
530 a large peak in numbers in 2019.

531 We saw very high ST diversity, with many STs occurring only once and eleven STs that were
532 previously unknown. This high ST diversity which was consistent over the whole study period
533 may indicate that neonates are exposed to diverse sources of *E. coli*. It also illustrates the need
534 for more studies of *E. coli* diversity from sub-Saharan Africa. Although the STs with highest
535 numbers in our study are part of globally prevalent high-risk clones⁶³ (ST131, ST10, ST69 and
536 ST410), no ST was persistently present as a major lineage in our study. Our findings are
537 comparable to previous findings from the same site in Malawi over a longer period⁶⁴ and
538 another study from Lilongwe⁶⁵ which also found these STs to be common. ST410 has been
539 associated with increased mortality in the Malawian context as one study found that all 8/8
540 (100%) of patients in the study that had bloodstream infection with ST410 died⁵⁵ vs 135/326
541 (41%) of the overall cohort. In our study ST410 was associated with meningitis, the ESBL genes
542 *bla*_{CTX-M-15} and fluoroquinolone resistance genes, leaving only meropenem as a treatment option.

543 The vaccines against *E. coli* which are in clinical trial stages target O-antigens. Whilst being
544 developed to protect against UTIs and urosepsis in high-income settings, these or similar O-
545 antigen based vaccines could feasibly be administered to mothers to prevent neonatal sepsis.
546 The choice of O-antigen glyco-conjugate vaccines is based on the knowledge that O-antigens
547 are the major cell surface component of *E. coli*⁶⁶ and appear to be essential for *E. coli* survival in
548 human serum⁶⁷. There are also multiple other glyco-conjugate vaccines have been successful
549 (including *Haemophilus influenzae* type b⁶⁸, *Streptococcus pneumoniae*⁶⁹ and *Neisseria*
550 *meningitidis*⁷⁰, though these are all based on bacterial capsule rather than O-antigen). Our study
551 indicated high O-type and H-type diversity, with large flux of both, and no O-type or H-type
552 representing the majority of cases in any year. As is the case in other collections³¹ there was
553 higher diversity of O-type than H-type, meaning vaccine approaches targeting O-type require a
554 much higher valency than those targeting H-types to protect against a similar proportion of
555 isolates (30 O-types vs 12 H-types to protect against 80% of isolates). There were also 10
556 previously undescribed O-types and no undescribed H-types in our cohort. This higher diversity
557 (including some that is previously undescribed) may call into question the practicality of using an
558 O-type based vaccine approach in our patient population and may direct efforts towards
559 exploring vaccines based on other antigens.

560 The EXPEC9V vaccine was designed for an elderly population in high-income countries, and for
561 this role, it is likely to be effective. It was initially designed as a 10-valent vaccine; however, the
562 serotype O8 was removed after the functional antibody assay did not work^{13, 14, 15}. In our cohort,
563 this change would have had a significant effect on the utility of such a vaccine (dropping the

564 proportion of isolates protected against from 45.6% to 33.7%). O8 was the third most frequently
565 occurring O-type in our cohort, it was enriched amongst our meningitis cases and was present
566 in all our ST410 isolates (which were highly AMR) and thus seems to be particularly common in
567 high consequence infections. Its removal would therefore drastically reduce the likely impact of
568 this vaccine in our setting. Other studies in the target population for this vaccine (elderly adults
569 in high-income countries) have shown good protection against invasive cases (64.7%⁷¹ -
570 67.5%⁷²). The O-types in this vaccine appear wholly appropriate for this patient population.
571 Interestingly, one of these studies⁷¹, though it was conducted across three continents and seven
572 countries reported lower O-type diversity than in our collection (49 O-types; 47 identified O-
573 types as well as two unknown O-types compared to the 63 distinct O-types found in our
574 collection). This might reflect differences in the target setting as well as the patient population.
575 Most appropriate to compare to our study is a study on a cohort of paediatric (most cases were
576 from neonates) *E. coli* meningitis cases from France, where O1, O18, O45 and O7 were the
577 most common O-types⁷³. Amongst these, O18 was the only O-type found frequently in our study
578 whilst other O-types such as O17, O12 and O11 were frequently found in our study but not in
579 that one.

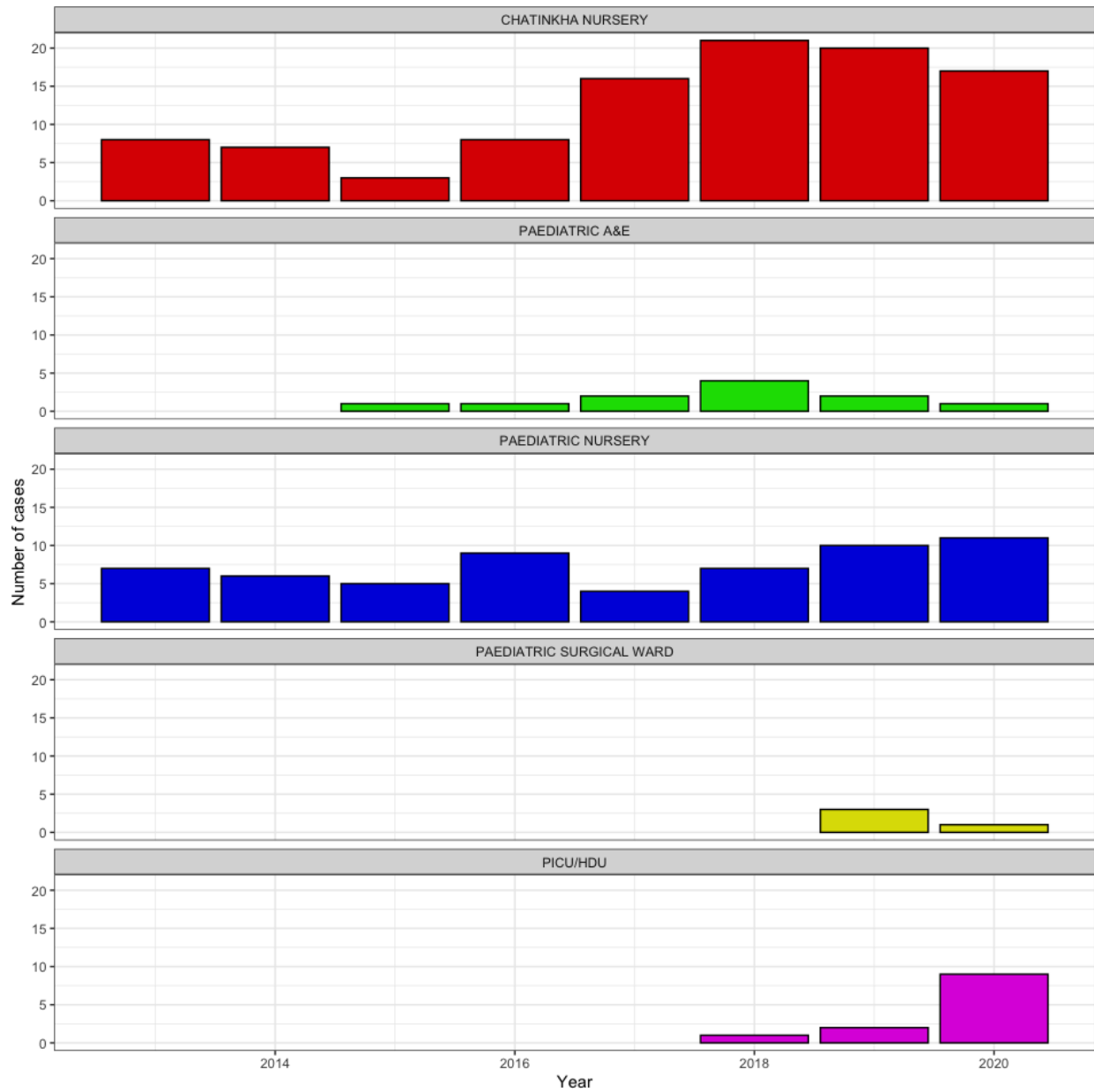
580 We also identified nine potentially novel O-types and one combined O-type that were not
581 identified by the EcOH database or ECTyper, highlighting our incomplete overview of O-antigen
582 diversity, and that we lack knowledge of how frequently new types emerge. These putative
583 novel O-types furthermore emphasize the need to perform sero-typing and WGS on more
584 isolates from sub-Saharan Africa as there might be substantial undescribed diversity. One of the
585 unknown O-types was similar to the OX-13 antigen on *Salmonella enterica*. There are some O-
586 types that are known to be shared by both *S. enterica* and *E. coli* (*E. coli* O-types O55, O111
587 and O157^{74, 75}). This is likely another O-type that is shared by both bacterial species. One O-
588 type appeared to have sections from both the O8 and O160 O-type sugar molecules, it is not
589 clear whether this is a hybrid of the two sugar molecules or whether this organism would be able
590 to express both molecules.

591 Phenotypic AMR increased for several different antibiotics over the study period (ceftriaxone,
592 co-trimoxazole, gentamicin, ciprofloxacin and amikacin, though co-trimoxazole did not have a *p*
593 value < 0.05) and was explained by a number of different genomic mechanisms. Ceftriaxone
594 had been introduced in QECH as standard therapy for many infections in 2004, so this
595 introduction alone cannot be responsible for the increase in resistance seen here. This is a
596 global trend and whilst rates of AMR for *E. coli* are lower than for *Klebsiella pneumoniae* at the

597 same site^{4,62}, they are of significant concern. We identified no isolates with carbapenem
598 resistance, which is not surprising as these genes are not widespread in Malawi, however a
599 previous study has identified a single carbapenemase carrying *E. coli*⁷⁴. Chloramphenicol and
600 co-amoxiclav resistance decreased over the study period (though only chloramphenicol had a p
601 < 0.05), this agent is however contraindicated in neonates and thus does not provide a
602 treatment alternative. The increase in AMR, particularly in a setting where watch and reserve
603 antibiotics are often unavailable due to cost, highlights the importance of prevention of neonatal
604 infection in the first place with strategies such as vaccines or investment in improvements in
605 infection, prevention and control (IPC).

606 In conclusion, the ongoing burden of neonatal sepsis combined with worsening AMR in *E. coli*
607 motivates the development of *E. coli* vaccines for this population. However, the prevalent O-
608 types in this collection from sub-Saharan Africa are highly diverse, partially novel, and different
609 to those currently covered by vaccines in clinical trials. If maternally-administered vaccines are
610 to be developed, they need to be based on robust genomic surveillance of prevalent antigens
611 and temporal trends in this population, and far more data from sub-Saharan Africa is required to
612 ensure equity of coverage compared to vaccines developed for HICs. Development of a suitable
613 vaccine will be a lengthy process, and until a successful product is available other methods of
614 preventing mortality from neonatal sepsis such as IPC strategy and early recognition of neonatal
615 infection should be urgently supported.

616 **Supplementary materials**



617

618 *Supplementary Figure 1. E. coli cases per year by ward.*

619

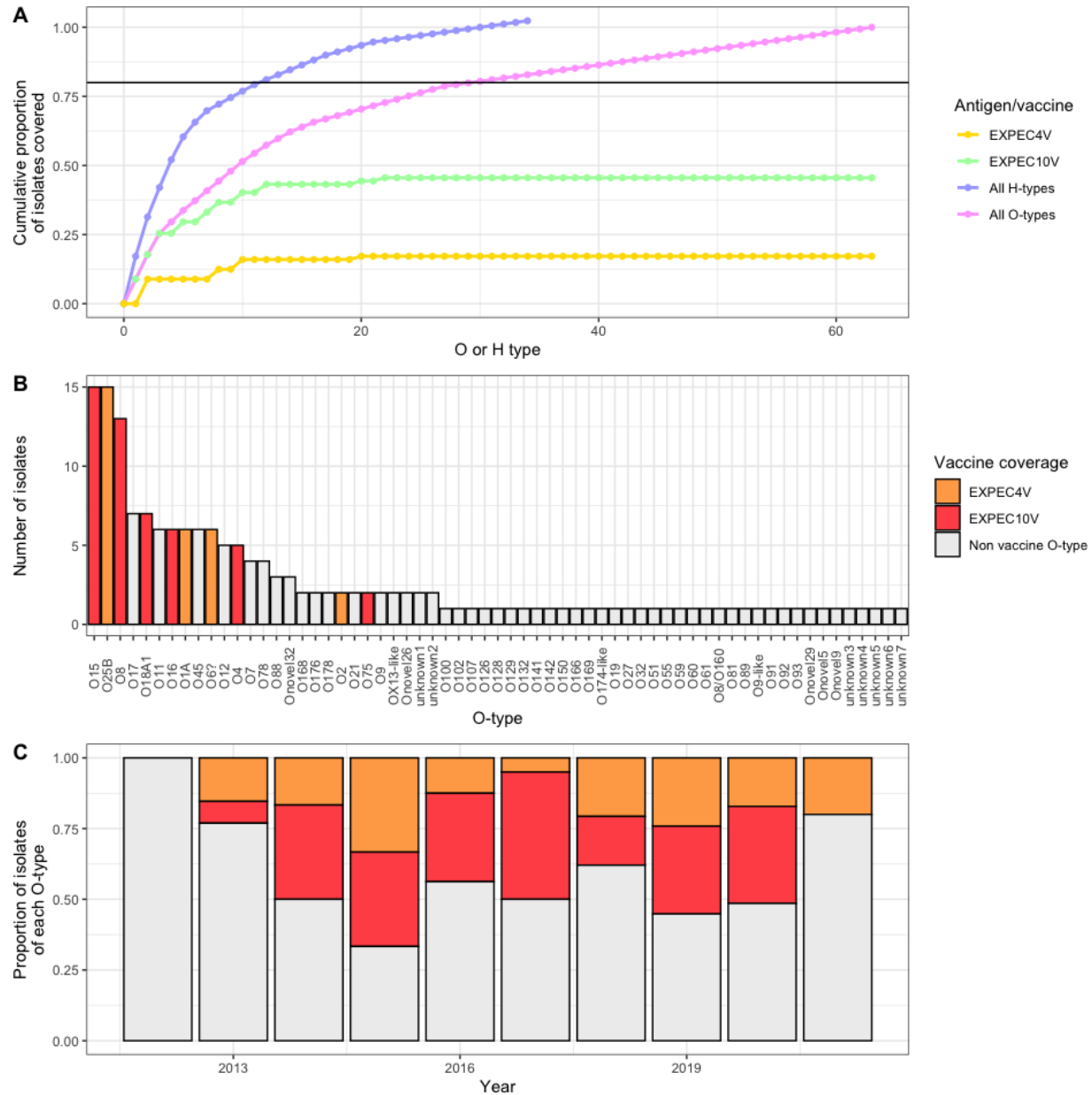
620

621

622

623

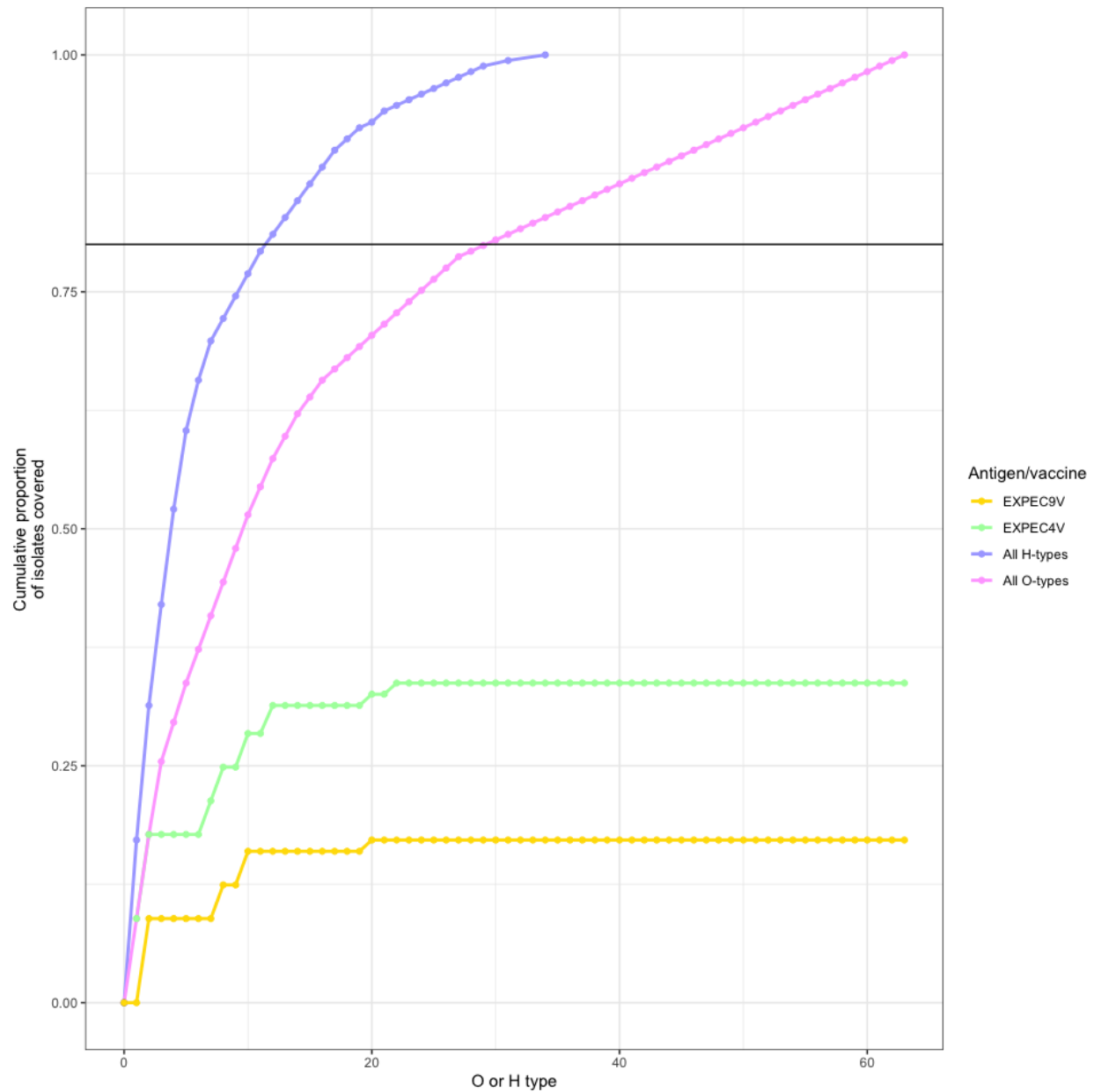
It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).



624

625 *Supplementary Figure 2) Rarefaction curve showing the theoretical protection given against*
 626 *vaccines covering the most frequently isolated H-types and O-types, as well as the potential*
 627 *protection given by the EXPEC10V and EXPEC4V. The horizontal line shows the point at which*
 628 *80% of isolates would be covered. For isolates with more than one H-type both were counted,*
 629 *there were multiple isolates with more than one H-type so the line for H-type goes above 1. B) A*
 630 *bar chart showing the frequency of the different O-types. C) A bar chart showing the proportion*
 631 *of isolates per year that had different O-types. The colours are the same as those represented*
 632 *in Figure 5B.*

633

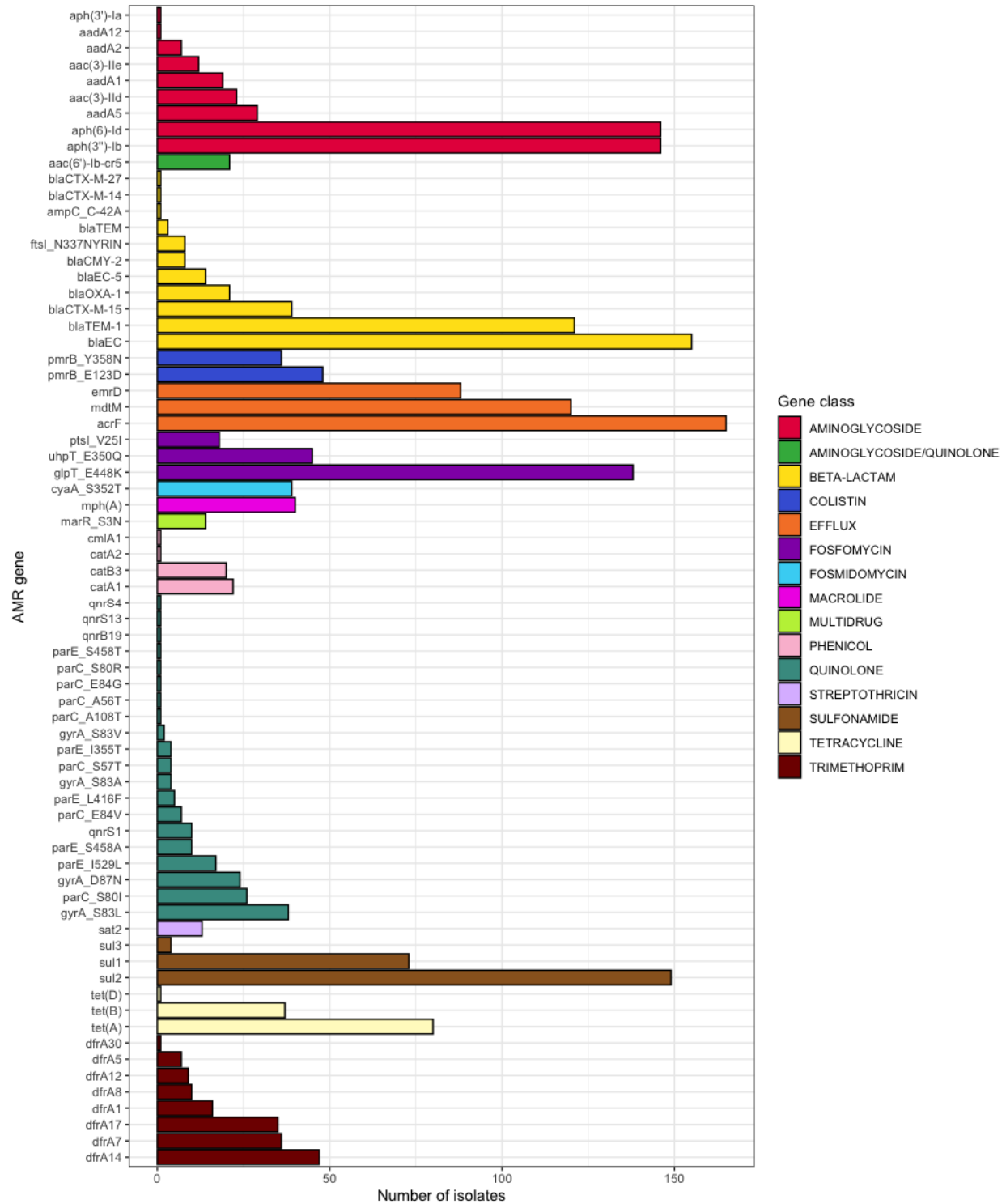


634

635 *Supplementary Figure 3. Rarefaction curve showing the theoretical protection given against*
636 *vaccines covering the most frequently isolated H-types and O-types, as well as the potential*
637 *protection given by the EXPEC9V and EXPEC4V. The horizontal line shows the point at which*
638 *80% of isolates would be covered. For isolates with more than one H-type only the most*
639 *frequently occurring O-type or H-type was counted.*

640

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).



641

642 *Supplementary Figure 3. Frequency of antimicrobial resistance (AMR) genes in the collection*
 643 *separated by class.*

644

645 **Conflicts of interest**

646 The authors declare that there are no conflicts of interest.

647 **Funding information**

648 This study was conducted with funding from the Bill & Melinda Gates Foundation (project grant
649 number INV-005692 to NAF) and Wellcome core institutional grants for MLW (206545/Z/17/Z)
650 and the Wellcome Sanger Institute (220540/Z/20/A). EH acknowledges funding from the
651 BBSRC (BB/V011278/1, BB/V011278/2). For the purpose of Open Access, the author has
652 applied a CC BY public copyright license to any Author Accepted Manuscript version arising
653 from this submission.

654 **Author contributions**

655 The study was conceived by NAF and OP. NAF, NRT and EH were responsible for funding
656 acquisition. Investigation and methodology development was carried out by AZ, ET, PS, AJF,
657 JC, PM and EH. Data curation and project administration was carried out by OP and EH.
658 Resources were managed by AZ, KK and OP. Formal analysis, validation and visualization
659 were carried out by OP, AJF and EH. Supervision was carried out by NAF and EH. The writing
660 of the original draft of the manuscript was by OP, NAF and EH. It was edited and revised by all
661 authors. All authors read and agreed on the final manuscript.

662 **Acknowledgements**

663 We would like to acknowledge the clinical team at QECH and the Mercy James hospital for
664 caring for the babies who were affected by *E. coli* in this study. We would also like to
665 acknowledge the MLW microbiology laboratory team, who isolated and identified the *E. coli*. We
666 would also like to acknowledge the Pathogen Informatics teams at the Wellcome Sanger
667 Institute for their expert support.

668 **References**

669 1. Goal 3 Department of Economic and Social Affairs. (n.d.).

670 https://sdgs.un.org/goals/goal3#targets_and_indicators.

671 2. Okomo, U., Akpalu, E. N. K., Le Doare, K., Roca, A., Cousens, S., Jarde, A., Sharland, M.,
672 Kampmann, B., & Lawn, J. E. (2019). Aetiology of invasive bacterial infection and antimicrobial
673 resistance in neonates in sub-Saharan Africa: A systematic review and meta-analysis in line

- 674 with the STROBE-NI reporting guidelines. *The Lancet. Infectious Diseases*, 19(11), 1219–1234.
675 [https://doi.org/10.1016/S1473-3099\(19\)30414-1](https://doi.org/10.1016/S1473-3099(19)30414-1)
- 676 3. Stoll, B. J., Puopolo, K. M., Hansen, N. I., Sánchez, P. J., Bell, E. F., Carlo, W. A., Cotten, C.
677 M., D’Angio, C. T., Kazzi, S. N. J., Poindexter, B. B., Van Meurs, K. P., Hale, E. C., Collins, M.
678 V., Das, A., Baker, C. J., Wyckoff, M. H., Yoder, B. A., Watterberg, K. L., Walsh, M. C., ...
679 Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal
680 Research Network. (2020). Early-Onset Neonatal Sepsis 2015 to 2017, the Rise of *Escherichia*
681 *coli*, and the Need for Novel Prevention Strategies. *JAMA Pediatrics*, 174(7), e200593.
682 <https://doi.org/10.1001/jamapediatrics.2020.0593>
- 683 4. Musicha, P., Cornick, J. E., Bar-Zeev, N., French, N., Masesa, C., Denis, B., Kennedy, N.,
684 Mallewa, J., Gordon, M. A., Msefula, C. L., Heyderman, R. S., Everett, D. B., & Feasey, N. A.
685 (2017). Trends in antimicrobial resistance in bloodstream infection isolates at a large urban
686 hospital in Malawi (1998-2016): A surveillance study. *The Lancet. Infectious Diseases*, 17(10),
687 1042–1052. [https://doi.org/10.1016/S1473-3099\(17\)30394-8](https://doi.org/10.1016/S1473-3099(17)30394-8)
- 688 5. Murray, C. J. L., Ikuta, K. S., Sharara, F., Swetschinski, L., Aguilar, G. R., Gray, A., Han, C.,
689 Bisignano, C., Rao, P., Wool, E., Johnson, S. C., Browne, A. J., Chipeta, M. G., Fell, F.,
690 Hackett, S., Haines-Woodhouse, G., Hamadani, B. H. K., Kumaran, E. A. P., McManigal, B., ...
691 Naghavi, M. (2022). Global burden of bacterial antimicrobial resistance in 2019: A systematic
692 analysis. *The Lancet*, 399(10325), 629–655. [https://doi.org/10.1016/S0140-6736\(21\)02724-0](https://doi.org/10.1016/S0140-6736(21)02724-0)
- 693 6. Iroh Tam, P.-Y., Musicha, P., Kawaza, K., Cornick, J., Denis, B., Freyne, B., Everett, D.,
694 Dube, Q., French, N., Feasey, N., & Heyderman, R. (2019). Emerging Resistance to Empiric
695 Antimicrobial Regimens for Pediatric Bloodstream Infections in Malawi (1998-2017). *Clinical*
696 *Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 69(1),
697 61–68. <https://doi.org/10.1093/cid/ciy834>
- 698 7. Gkentzi, D., Katsakiori, P., Marangos, M., Hsia, Y., Amirthalingam, G., Heath, P. T., &
699 Ladhani, S. (2017). Maternal vaccination against pertussis: A systematic review of the recent
700 literature. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, 102(5), F456–F463.
701 <https://doi.org/10.1136/archdischild-2016-312341>
- 702 8. Madhi, S. A., Anderson, A. S., Absalon, J., Radley, D., Simon, R., Jongihlati, B., Strehlau, R.,
703 Niekerk, A. M. van, Izu, A., Naidoo, N., Kwatra, G., Ramsamy, Y., Said, M., Jones, S., Jose, L.,
704 Fairlie, L., Barnabas, S. L., Newton, R., Munson, S., ... Jansen, K. U. (2023). Potential for

- 705 Maternally Administered Vaccine for Infant Group B Streptococcus. *New England Journal of*
706 *Medicine*, 389(3), 215–227. <https://doi.org/10.1056/NEJMoa2116045>
- 707 9. Denamur, E., Clermont, O., Bonacorsi, S., & Gordon, D. (2021). The population genetics of
708 pathogenic *Escherichia coli*. *Nature Reviews Microbiology*, 19(1), 37–54.
709 <https://doi.org/10.1038/s41579-020-0416-x>
- 710 10. Gransden, W. R., Eykyn, S. J., Phillips, I., & Rowe, B. (1990). Bacteremia due to
711 *Escherichia coli*: A study of 861 episodes. *Reviews of Infectious Diseases*, 12(6), 1008–1018.
712 <https://doi.org/10.1093/clinids/12.6.1008>
- 713 11. Frenck, R. W., Ervin, J., Chu, L., Abbanat, D., Spiessens, B., Go, O., Haazen, W., van den
714 Dobbelsteen, G., Poolman, J., Thoelen, S., & Ibarra de Palacios, P. (2019). Safety and
715 immunogenicity of a vaccine for extra-intestinal pathogenic *Escherichia coli* (ESTELLA): A
716 phase 2 randomised controlled trial. *The Lancet Infectious Diseases*, 19(6), 631–640.
717 [https://doi.org/10.1016/S1473-3099\(18\)30803-X](https://doi.org/10.1016/S1473-3099(18)30803-X)
- 718 12. Fierro, C. A., Sarnecki, M., Spiessens, B., Go, O., Day, T. A., Davies, T. A., van den
719 Dobbelsteen, G., Poolman, J., Abbanat, D., & Haazen, W. (2024). A randomized phase 1/2a
720 trial of ExPEC10V vaccine in adults with a history of UTI. *Npj Vaccines*, 9(1), 1–9.
721 <https://doi.org/10.1038/s41541-024-00885-1>
- 722 13. Fierro, C. A., Sarnecki, M., Doua, J., Spiessens, B., Go, O., Davies, T. A., van den
723 Dobbelsteen, G., Poolman, J., Abbanat, D., & Haazen, W. (2023). Safety, Reactogenicity,
724 Immunogenicity, and Dose Selection of 10-Valent Extraintestinal Pathogenic *Escherichia coli*
725 Bioconjugate Vaccine (VAC52416) in Adults Aged 60–85 Years in a Randomized, Multicenter,
726 Interventional, First-in-Human, Phase 1/2a Study. *Open Forum Infectious Diseases*, 10(8),
727 ofad417. <https://doi.org/10.1093/ofid/ofad417>
- 728 14. Fierro, C., Sarnecki, M., Doua, J., Spiessens, B., Go, O., Davies, T., Dobbelsteen, G. van
729 den, Poolman, J., Abbanat, D., & Haazen, W. (2023). IMMUNOGENICITY OUTCOMES OF A
730 RANDOMIZED, MULTICENTER, INTERVENTIONAL, FIRST-IN-HUMAN, PHASE 1/2A STUDY
731 OF VAC52416 (EXPEC10V), A VACCINE CANDIDATE FOR THE PREVENTION OF
732 INVASIVE *ESCHERICHIA COLI* DISEASE. *International Journal of Infectious Diseases*, 130,
733 S114. <https://doi.org/10.1016/j.ijid.2023.04.282>

- 734 15. Fierro, C. A., Sarnecki, M., Spiessens, B., Go, O., Day, T. A., Davies, T. A., van den
735 Dobbelsteen, G., Poolman, J., Abbanat, D., & Haazen, W. (2024). A randomized phase 1/2a
736 trial of ExPEC10V vaccine in adults with a history of UTI. *Npj Vaccines*, 9(1), 1–9.
737 <https://doi.org/10.1038/s41541-024-00885-1>
- 738 16. Wood, D. E., & Salzberg, S. L. (2014). Kraken: Ultrafast metagenomic sequence
739 classification using exact alignments. *Genome Biology*, 15(3), R46. [https://doi.org/10.1186/gb-](https://doi.org/10.1186/gb-2014-15-3-r46)
740 [2014-15-3-r46](https://doi.org/10.1186/gb-2014-15-3-r46)
- 741 17. Page, A. J., De Silva, N., Hunt, M., Quail, M. A., Parkhill, J., Harris, S. R., Otto, T. D., &
742 Keane, J. A. (2016). Robust high-throughput prokaryote de novo assembly and improvement
743 pipeline for Illumina data. *Microbial Genomics*, 2(8), e000083.
744 <https://doi.org/10.1099/mgen.0.000083>
- 745 18. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V.
746 M., Nikolenko, S. I., Pham, S., Pribelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N.,
747 Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly
748 algorithm and its applications to single-cell sequencing. *Journal of Computational Biology: A*
749 *Journal of Computational Molecular Cell Biology*, 19(5), 455–477.
750 <https://doi.org/10.1089/cmb.2012.0021>
- 751 19. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., & Pirovano, W. (2011). Scaffolding pre-
752 assembled contigs using SSPACE. *Bioinformatics (Oxford, England)*, 27(4), 578–579.
753 <https://doi.org/10.1093/bioinformatics/btq683>
- 754 20. Boetzer, M., & Pirovano, W. (2012). Toward almost closed genomes with GapFiller.
755 *Genome Biology*, 13(6), R56. <https://doi.org/10.1186/gb-2012-13-6-r56>
- 756 21. Pathogen Informatics, Wellcome Sanger Institute. (n.d.). In GitHub.
757 <https://github.com/sanger-pathogens/assembly-stats>
- 758 22. Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics (Oxford,*
759 *England)*, 30(14), 2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
- 760 23. Pruitt, K. D., Tatusova, T., Brown, G. R., & Maglott, D. R. (2012). NCBI Reference
761 Sequences (RefSeq): Current status, new features and genome annotation policy. *Nucleic*
762 *Acids Research*, 40(Database issue), D130–135. <https://doi.org/10.1093/nar/gkr1079>

- 763 24. Pathogen Informatics, Wellcome Sanger Institute. (n.d.). In GitHub.
764 <https://github.com/sanger-pathogens>.
- 765 25. Andrew Page (AJPAGE) - metacpan.org. (n.d.). <https://metacpan.org/author/AJPAGE>.
- 766 26. Jolley, K. A., & Maiden, M. C. (2010). BIGSdb: Scalable analysis of bacterial genome
767 variation at the population level. *BMC Bioinformatics*, 11(1), 595. <https://doi.org/10.1186/1471->
768 2105-11-595
- 769 27. Seemann, T. (2024). Tseemann/mlst.
- 770 28. Feldgarden, M., Brover, V., Haft, D. H., Prasad, A. B., Slotta, D. J., Tolstoy, I., Tyson, G. H.,
771 Zhao, S., Hsu, C.-H., McDermott, P. F., Tadesse, D. A., Morales, C., Simmons, M., Tillman, G.,
772 Wasilenko, J., Folster, J. P., & Klimke, W. (2019). Validating the AMRFinder Tool and
773 Resistance Gene Database by Using Antimicrobial Resistance Genotype-Phenotype
774 Correlations in a Collection of Isolates. *Antimicrobial Agents and Chemotherapy*, 63(11),
775 e00483–19. <https://doi.org/10.1128/AAC.00483-19>
- 776 29. Feldgarden, M., Brover, V., Gonzalez-Escalona, N., Frye, J. G., Haendiges, J., Haft, D. H.,
777 Hoffmann, M., Pettengill, J. B., Prasad, A. B., Tillman, G. E., Tyson, G. H., & Klimke, W. (2021).
778 AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links
779 among antimicrobial resistance, stress response, and virulence. *Scientific Reports*, 11, 12728.
780 <https://doi.org/10.1038/s41598-021-91456-0>
- 781 30. Inouye, M., Dashnow, H., Raven, L.-A., Schultz, M. B., Pope, B. J., Tomita, T., Zobel, J., &
782 Holt, K. E. (2014). SRST2: Rapid genomic surveillance for public health and hospital
783 microbiology labs. *Genome Medicine*, 6(11), 90. <https://doi.org/10.1186/s13073-014-0090-6>
- 784 31. Ingle, D. J., Valcanis, M., Kuzevski, A., Tauschek, M., Inouye, M., Stinear, T., Levine, M. M.,
785 Robins-Browne, R. M., & Holt, K. E. (2016). In silico serotyping of E. Coli from short read data
786 identifies limited novel O-loci but extensive diversity of O:H serotype combinations within and
787 between pathogenic lineages. *Microbial Genomics*, 2(7), e000064.
788 <https://doi.org/10.1099/mgen.0.000064>
- 789 32. Delannoy, S., Beutin, L., Mariani-Kurkdjian, P., Fleiss, A., Bonacorsi, S., & Fach, P. (2017).
790 The Escherichia coli Serogroup O1 and O2 Lipopolysaccharides Are Encoded by Multiple O-
791 antigen Gene Clusters. *Frontiers in Cellular and Infection Microbiology*, 7, 30.
792 <https://doi.org/10.3389/fcimb.2017.00030>

- 793 33. Iguchi, A., Iyoda, S., Kikuchi, T., Ogura, Y., Katsura, K., Ohnishi, M., Hayashi, T., &
794 Thomson, N. R. (2015). A complete view of the genetic diversity of the Escherichia coli O-
795 antigen biosynthesis gene cluster. *DNA Research: An International Journal for Rapid*
796 *Publication of Reports on Genes and Genomes*, 22(1), 101–107.
797 <https://doi.org/10.1093/dnares/dsu043>
- 798 34. Szijártó, V., Lukaszewicz, J., Gozdiewicz, T. K., Magyarics, Z., Nagy, E., & Nagy, G. (2014).
799 Diagnostic potential of monoclonal antibodies specific to the unique O-antigen of multidrug-
800 resistant epidemic Escherichia coli clone ST131-O25b:H4. *Clinical and Vaccine Immunology*:
801 *CVI*, 21(7), 930–939. <https://doi.org/10.1128/CVI.00685-13>
- 802 35. Basecalling using Guppy. (n.d.). In Long-Read, long reach Bioinformatics Tutorials.
803 https://timkahlke.github.io/LongRead_tutorials/BS_G.html.
- 804 36. Wick, R. R., Judd, L. M., Gorrie, C. L., & Holt, K. E. (2017). Completing bacterial genome
805 assemblies with multiplex MinION sequencing. *Microbial Genomics*, 3(10), e000132.
806 <https://doi.org/10.1099/mgen.0.000132>
- 807 37. Wick, R. (2024). Rrwick/Filtlong.
- 808 38. Ryan R. Wick, Mark B. Schultz, Justin Zobel, Kathryn E. Holt, Bandage: interactive
809 visualization of de novo genome assemblies, *Bioinformatics*, Volume 31, Issue 20, October
810 2015, Pages 3350–3352, <https://doi.org/10.1093/bioinformatics/btv383>
- 811 39. Nanoporetech/medaka. (2024). Oxford Nanopore Technologies.
- 812 40. Wick, R. R., & Holt, K. E. (2022). Polypolish: Short-read polishing of long-read bacterial
813 genome assemblies. *PLOS Computational Biology*, 18(1), e1009802.
814 <https://doi.org/10.1371/journal.pcbi.1009802>
- 815 41. Zimin AV, Salzberg SL (2020) The genome polishing tool POLCA makes fast and accurate
816 corrections in genome assemblies. *PLoS Comput Biol* 16(6): e1007981.
817 <https://doi.org/10.1371/journal.pcbi.1007981>
- 818 42. Bessonov, K., Laing, C., Robertson, J., Yong, I., Ziebell, K., Gannon, V. P. J., Nichani, A.,
819 Arya, G., Nash, J. H. E., & Christianson, S. (2021). ECTyper: In silico Escherichia coli serotype
820 and species prediction from raw and assembled whole-genome sequence data. *Microbial*
821 *Genomics*, 7(12), 000728. <https://doi.org/10.1099/mgen.0.000728>

- 822 43. R: The R Project for Statistical Computing. (n.d.). <https://www.r-project.org/>.
- 823 44. RStudio Team. (2020). RStudio: Integrated development environment for R [Manual].
824 RStudio, PBC.
- 825 45. Müller, K. (2024). Here: A simpler way to find your files [Manual].
- 826 46. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Golemund,
827 G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller,
828 K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the
829 tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- 830 47. Golemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal*
831 *of Statistical Software*, 40(3), 1–25.
- 832 48. Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New
833 York.
- 834 49. Henderson, E. (2024). Ghibli: Studio ghibli colour palettes [Manual].
- 835 50. Neuwirth, E. (2022). RColorBrewer: ColorBrewer Palettes.
- 836 51. Wright, K. (2023). Pals: Color Palettes, Colormaps, and Tools to Evaluate Them.
- 837 52. Mills, B. R. (2024). BlakeRMills/MetBrewer.
- 838 53. Wilkins, D. (2023). Gggenes: Draw gene arrow maps in 'Ggplot2' [Manual].
- 839 54. Kassambara, A. (2023). Ggpubr: 'ggplot2' based publication ready plots [Manual].
- 840 55. Lester R, Musicha P, Kawaza K, Langton J, Mango J, Mangochi H, Bakali W, Pearse O,
841 Mallewa J, Denis B, Bilima S, Gordon SB, Lalloo DG, Jewell CP, Feasey NA. Effect of
842 resistance to third-generation cephalosporins on morbidity and mortality from bloodstream
843 infections in Blantyre, Malawi: a prospective cohort study. (2022) *Lancet Microbe*.Dec;3(12):
844 e922-e930. doi: 10.1016/S2666-5247(22)00282-8.
- 845 56. Pitout, J. D. D., & DeVinney, R. (2017). *Escherichia coli* ST131: A multidrug-resistant clone
846 primed for global domination. *F1000Research*, 6, F1000 Faculty Rev–195.
847 <https://doi.org/10.12688/f1000research.10609.1>

- 848 57. Stoll, B. J., Hansen, N. I., Sánchez, P. J., Faix, R. G., Poindexter, B. B., Van Meurs, K. P.,
849 Bizzarro, M. J., Goldberg, R. N., Frantz, I. D., III, Hale, E. C., Shankaran, S., Kennedy, K.,
850 Carlo, W. A., Watterberg, K. L., Bell, E. F., Walsh, M. C., Schibler, K., Lptook, A. R., Shane, A.
851 L., ... for the Eunice Kennedy Shriver National Institute of Child Health and Human
852 Development Neonatal Research Network. (2011). Early Onset Neonatal Sepsis: The Burden of
853 Group B Streptococcal and E. Coli Disease Continues. *Pediatrics*, 127(5), 817–826.
854 <https://doi.org/10.1542/peds.2010-2217>
- 855 58. Bergin, S. P., Thaden, J., Ericson, J. E., Cross, H., Messina, J., Clark, R. H., Fowler, V. G.,
856 Benjamin, D. K., Hornik, C. P., & Smith, P. B. (2015). Neonatal *Escherichia coli* Bloodstream
857 Infections: Clinical Outcomes and Impact of Initial Antibiotic Therapy. *The Pediatric Infectious*
858 *Disease Journal*, 34(9), 933–936. <https://doi.org/10.1097/INF.0000000000000769>
- 859 59. Guiral, E., Bosch, J., Vila, J., & Soto, S. M. (2012). Antimicrobial resistance of *Escherichia*
860 *coli* strains causing neonatal sepsis between 1998 and 2008. *Chemotherapy*, 58(2), 123–128.
861 <https://doi.org/10.1159/000337062>
- 862 60. Lai, J., Zhu, Y., Tang, L., & Lin, X. (2021). Epidemiology and antimicrobial susceptibility of
863 invasive *Escherichia coli* infection in neonates from 2012 to 2019 in Xiamen, China. *BMC*
864 *Infectious Diseases*, 21(1), 295. <https://doi.org/10.1186/s12879-021-05981-4>
- 865 61. Furyk, J. S., Swann, O., & Molyneux, E. (2011). Systematic review: Neonatal meningitis in
866 the developing world. *Tropical Medicine & International Health*, 16(6), 672–679.
867 <https://doi.org/10.1111/j.1365-3156.2011.02750.x>
- 868 62. Heinz, E., Pearse, O., Zuza, A., Bilima, S., Msefula, C., Musicha, P., Siyabu, P., Tewesa, E.,
869 Graf, F. E., Lester, R., Lissauer, S., Cornick, J., Lewis, J. M., Kawaza, K., Thomson, N. R., &
870 Feasey, N. A. (2024). Longitudinal analysis within one hospital in sub-Saharan Africa over 20
871 years reveals repeated replacements of dominant clones of *Klebsiella pneumoniae* and stresses
872 the importance to include temporal patterns for vaccine design considerations. *Genome*
873 *Medicine*, 16(1), 67. <https://doi.org/10.1186/s13073-024-01342-3>
- 874 63. Sands, K., Carvalho, M. J., Portal, E., Thomson, K., Dyer, C., Akpulu, C., Andrews, R.,
875 Ferreira, A., Gillespie, D., Hender, T., Hood, K., Mathias, J., Milton, R., Nieto, M., Taiyari, K.,
876 Chan, G. J., Bekele, D., Solomon, S., Basu, S., ... Walsh, T. R. (2021). Characterization of
877 antimicrobial-resistant Gram-negative bacteria that cause neonatal sepsis in seven low- and

- 878 middle-income countries. *Nature Microbiology*, 6(4), 512–523. [https://doi.org/10.1038/s41564-](https://doi.org/10.1038/s41564-021-00870-7)
879 021-00870-7
- 880 64. Musicha, P., Feasey, N. A., Cain, A. K., Kallonen, T., Chaguza, C., Peno, C., Khonga, M.,
881 Thompson, S., Gray, K. J., Mather, A. E., Heyderman, R. S., Everett, D. B., Thomson, N. R., &
882 Msefula, C. L. (2017). Genomic landscape of extended-spectrum β -lactamase resistance in
883 *Escherichia coli* from an urban African setting. *The Journal of Antimicrobial Chemotherapy*,
884 72(6), 1602–1609. <https://doi.org/10.1093/jac/dkx058>
- 885 65. Tegha, G., Ciccone, E. J., Krysiak, R., Kaphatika, J., Chikaonda, T., Ndhlovu, I., van Duin,
886 D., Hoffman, I., Juliano, J. J., & Wang, J. (2021). Genomic epidemiology of *Escherichia coli*
887 isolates from a tertiary referral center in Lilongwe, Malawi. *Microbial Genomics*, 7(1),
888 mgen000490. <https://doi.org/10.1099/mgen.0.000490>
- 889 66. Huttner, A., & Gambillara, V. (2018). The development and early clinical testing of the
890 ExPEC4V conjugate vaccine against uropathogenic *Escherichia coli*. *Clinical Microbiology and*
891 *Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious*
892 *Diseases*, 24(10), 1046–1050. <https://doi.org/10.1016/j.cmi.2018.05.009>
- 893 67. Sarkar, S., Ulett, G. C., Totsika, M., Phan, M.-D., & Schembri, M. A. (2014). Role of Capsule
894 and O Antigen in the Virulence of Uropathogenic *Escherichia coli*. *PLOS ONE*, 9(4), e94786.
895 <https://doi.org/10.1371/journal.pone.0094786>
- 896 68. Agrawal, A., & Murphy, T. F. (2011). Haemophilus influenzae Infections in the H. Influenzae
897 Type b Conjugate Vaccine Era. *Journal of Clinical Microbiology*, 49(11), 3728–3732.
898 <https://doi.org/10.1128/JCM.05476-11>
- 899 69. Prymula, R., Peeters, P., Chrobok, V., Kriz, P., Novakova, E., Kaliskova, E., Kohl, I.,
900 Lommel, P., Poolman, J., Prieels, J.-P., & Schuerman, L. (2006). Pneumococcal capsular
901 polysaccharides conjugated to protein D for prevention of acute otitis media caused by both
902 *Streptococcus pneumoniae* and non-typable *Haemophilus influenzae*: A randomised double-
903 blind efficacy study. *The Lancet*, 367(9512), 740–748. [https://doi.org/10.1016/S0140-](https://doi.org/10.1016/S0140-6736(06)68304-9)
904 6736(06)68304-9
- 905 70. Pichichero, M. E. (2005). Meningococcal Conjugate Vaccine in Adolescents and Children.
906 *Clinical Pediatrics*, 44(6), 479–489. <https://doi.org/10.1177/000992280504400603>

- 907 71. Arconada Nuin, E., Vilken, T., Xavier, B. B., Doua, J., Morrow, B., Geurtsen, J., Go, O.,
908 Spiessens, B., Sarnecki, M., Poolman, J., Bonten, M., Ekkelenkamp, M., Lammens, C.,
909 Goossens, H., Glupczynski, Y., Van Puyvelde, S., & the COMBACTE-NET
910 Consortium/EXPECT Study Group. (2024). A microbiological and genomic perspective of
911 globally collected *Escherichia coli* from adults hospitalized with invasive *E. Coli* disease. *Journal*
912 *of Antimicrobial Chemotherapy*, dkae182. <https://doi.org/10.1093/jac/dkae182>
- 913 72. Weerdenburg, E., Davies, T., Morrow, B., Zomer, A. L., Hermans, P., Go, O., Spiessens, B.,
914 van den Hoven, T., van Geet, G., Aitabi, M., DebRoy, C., Dudley, E. G., Bonten, M., Poolman,
915 J., & Geurtsen, J. (2023). Global Distribution of O Serotypes and Antibiotic Resistance in
916 Extraintestinal Pathogenic *Escherichia coli* Collected From the Blood of Patients With
917 Bacteremia Across Multiple Surveillance Studies. *Clinical Infectious Diseases*, 76(3), e1236–
918 e1243. <https://doi.org/10.1093/cid/ciac421>
- 919 73. Basmaci, R., Bonacorsi, S., Bidet, P., Biran, V., Aujard, Y., Bingen, E., Béchet, S., Cohen,
920 R., & Levy, C. (2015). *Escherichia Coli* Meningitis Features in 325 Children From 2001 to 2013
921 in France. *Clinical Infectious Diseases*, 61(5), 779–786. <https://doi.org/10.1093/cid/civ367>
- 922 74. Wang, L., & Reeves, P. R. (2000). The *Escherichia coli* O111 and *Salmonella enterica* O35
923 Gene Clusters: Gene Clusters Encoding the Same Colitose-Containing O Antigen Are Highly
924 Conserved. *Journal of Bacteriology*, 182(18), 5256–5261.
- 925 75. Wang, L., Huskic, S., Cisterne, A., Rothmund, D., & Reeves, P. R. (2002). The O-Antigen
926 Gene Cluster of *Escherichia coli* O55:H7 and Identification of a New UDP-GlcNAc C4
927 Epimerase Gene. *Journal of Bacteriology*, 184(10), 2620–2625.
928 <https://doi.org/10.1128/JB.184.10.2620-2625.2002>
- 929 76. Lewis, J. M., Mphasa, M., Banda, R., Beale, M. A., Mallewa, J., Anscome, C., Zuza, A.,
930 Roberts, A. P., Heinz, E., Thomson, N. R., & Feasey, N. A. (2023). Genomic analysis of
931 extended-spectrum beta-lactamase (ESBL) producing *Escherichia coli* colonising adults in
932 Blantyre, Malawi reveals previously undescribed diversity. *Microbial Genomics*, 9(6), 001035.
933 <https://doi.org/10.1099/mgen.0.001035>