

1 **Bacterial strain sharing between humans, animals, and the environment among urban households**

2 Daehyun D. Kim¹, Jenna M. Swarthout², Colin J. Worby³, Benard Chieng⁴, John Mboya¹, Ashlee M. Earl³, Sammy
3 M. Njenga⁴, Amy J. Pickering^{1,5,6,*}

4

5 ¹Department of Civil and Environmental Engineering, University of California, Berkeley, CA, USA

6 ²Department of Civil and Environmental Engineering, Tufts University, Medford, MA, USA

7 ³Infectious Disease & Microbiome Program, Broad Institute, Cambridge, MA, USA

8 ⁴Kenya Medical Research Institute, Nairobi, Kenya

9 ⁵Chan Zuckerberg Biohub – San Francisco

10 ⁶Blum Center for Developing Economies, University of California, Berkeley, Berkeley, CA 94720

11

12 *Address correspondence to Amy J. Pickering, pickering@berkeley.edu

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29 **Abstract**

30 Identifying bacterial transmission pathways is crucial to inform strategies aimed at curbing the spread of pathogenic
31 and antibiotic-resistant bacteria, especially in rapidly urbanizing low- and middle-income countries. In this study, we
32 assessed bacterial strain-sharing and dissemination of antibiotic resistance across humans, domesticated poultry,
33 canines, household soil, and drinking water in urban informal settlements in Nairobi, Kenya. We collected 321 samples
34 from 50 households and performed Pooling Isolated Colonies-seq (PIC-seq) by sequencing pools of up to five
35 *Escherichia coli* colonies per sample to capture strain diversity, strain-sharing patterns, and overlap of antibiotic-
36 resistant genes (ARGs). Bacterial strains isolated from the household environment carried clinically relevant ARGs,
37 reinforcing the role of the environment in antibiotic resistance dissemination. Strain-sharing rates and resistome
38 similarities across sample types were strongly correlated within households, suggesting clonal spread of bacteria is a
39 main driver of dissemination of ARGs in the domestic urban environment. Within households, *E. coli* strain-sharing
40 was rare between humans and animals but more frequent between humans and drinking water. *E. coli* contamination
41 in stored drinking water was also associated with higher strain-sharing between humans in the same household. Our
42 study demonstrates that contaminated drinking water facilitates human to human strain sharing and water treatment
43 can disrupt transmission.

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58 **Introduction**

59 Bacterial enteric infections cause 9.2% of annual deaths worldwide in children under five.^{1,2} Bacterial infections can
60 trigger acute gastrointestinal illness, while asymptomatic infections are associated with child growth faltering.^{3,4} The
61 incidence and disease burden of bacterial infections are highest in rapidly urbanizing low- and middle-income
62 countries (LMICs),⁵ where over 50% of the urban population lives in informal urban settings with limited access to
63 safe drinking water, sanitation, and appropriate hygiene (WASH) infrastructure.⁶ Furthermore, in these informal urban
64 settings, families often engage in small-scale animal husbandry within confined spaces, increasing the risk of
65 zoonosis.^{7,8} The high prevalence of infectious diseases in urban informal settlements, coupled with a lack of regulated
66 antibiotic use, promotes both the emergence of new antibiotic resistance gene (ARGs) variants through the
67 proliferation of antibiotic resistant bacteria and the spread of antibiotic resistance between bacterial strains via
68 horizontal gene transfer.^{6,9} A recent meta-analysis of fecal metagenomes from 26 countries¹⁰ identified a higher
69 abundance of ARGs in urban areas in Africa and South-East Asia compared to rural areas.

70

71 Enteric pathogens are transmitted to humans by direct contact with infected hosts or via exposure to fecally-
72 contaminated environments.¹¹ Exposure to animals has recently garnered attention as a risk factor for enteric infections
73 in LMICs with reports of strain-sharing between humans and domestic or wild animals that are in close proximity.^{12,13}
74 The previous studies have mainly focused on examining bacterial strain sharing between humans and animals,^{14,15}
75 without investigating the environmental exposure pathways through which transmission occurs.¹⁶ However, humans
76 are regularly exposed to environmental microbiota through consuming drinking water, hand-to-mouth contact with
77 soil, and inhaling soil dust.^{7,14} Understanding transmission pathways across humans, animals, and the environment,
78 along with their spatial range, is important to determine how bacterial transmission could be interrupted and the scale
79 at which interventions need to be implemented.¹⁷

80

81 The epidemiological study of bacterial transmission has traditionally been conducted by sequencing whole genomes
82 of individual isolates selectively cultured from each host.¹⁹ Sequencing individual isolates has the advantage that it
83 can facilitate the comparison of complete genomes at high-resolution, using metrics such as core genome multilocus
84 sequence typing (cgMLST)²⁰ or core genome single nucleotide polymorphisms (cgSNP).²¹ However, due to the labor

85 and cost burden of culturing and sequencing isolates, studies often include only one isolate per sample, potentially
86 missing a diverse population.^{7,12} Sequencing of DNA extracted directly - metagenomics - can describe the diversity
87 of complex microbial communities without a culture step. Several software tools have been recently developed for
88 tracking strains within metagenomes even when they are present at low relative abundance within a complex
89 community.^{22,23} However, reconstructing accurate sequences of ARGs and placing them within their genomic context
90 is often challenging due to low sequencing coverage and high heterogeneity, which can confound assembly.^{24,25} In
91 addition, ARGs are often carried on plasmids, which makes linking these ARGs to their host organisms challenging.²⁶
92 Plate-sweep metagenomics is a recent intermediate approach between isolated colonies and metagenomics that
93 includes a culture step to select for target viable organisms of interest from a sample followed by sequencing the total
94 biomass from the plate. While the overall microbial complexity of a sample is reduced, sequencing effort can still be
95 wasted as most selective media is not species specific.^{18,27}

96
97 In this study, we aimed to elucidate bacterial strain-sharing and ARG dissemination between humans, domesticated
98 animals, and the household environment. Given its ubiquity and role as a common human commensal as well as
99 clinically important and often antibiotic resistant pathogen, we selected *Escherichia coli* as the target organism.^{28,29} In
100 order to capture strain diversity within each sample and achieve high depth of sequencing coverage, we pooled up to
101 5 presumptive *E. coli* isolates cultured from the same sample prior to sequencing. Herein, we call our approach “PIC-
102 seq” (Pooling Isolated Colonies-seq). Leveraging matched sample sets of human stool, poultry cloacal swabs, canine
103 feces, stored drinking water, and soil from 50 households in low-income informal settlements in Nairobi, Kenya, we
104 used PIC-seq to determine strain sharing patterns. In addition, we reconstructed the resistomes of pooled *E. coli*
105 isolates at the sample level to explore the role of strain dissemination *versus* horizontal gene transfer in the spread of
106 ARGs.

107

108 **Results**

109 **Characteristics of enrolled households**

110 Poultry-owning households (n=50) were enrolled in the study across two subcounties (Dagoretti South: n=25 and
111 Kibera: n=25) in Nairobi, Kenya. In Kibera, all households collected drinking water from chlorinated piped water, as
112 confirmed by chlorine residual testing at the tap during the study. In contrast, water sources in Dagoretti South varied,

113 with households collecting water from boreholes (44%; 11/25 households), piped water (44%; 11/25 households), and
114 water tanks (12%; 3/25 households), none of which underwent chlorination (Supplementary Table 1). Additional
115 household-level water treatment of the collected water was slightly more common in Kibera (24%; 6/25 households)
116 than in Dagoretti South (12%; 3/25 households).

117
118 Households from Dagoretti South and Kibera had similar wealth indices, as determined by household assets including
119 availability of electricity, TV, mobile phone, and a stove (Supplementary Table 2). Urban informal settlements in
120 Kenya are often organized into compounds consisting of households sharing a common courtyard, which may include
121 a shared latrine. Households in Dagoretti South were more likely to have latrines inside their compound (76%; 19/25
122 households) compared to Kibera (20%; 5/25 households), where the majority relied on public facilities located >5
123 meters away outside of the household compound (64%; 16/24 households) (Supplementary Table 3). The purpose for
124 poultry ownership across both subcounties was primarily nutrient provision via meat (Dagoretti South: 88%, 22/25
125 households; Kibera: 92%, 23/25 households) and eggs (Dagoretti South: 88%, 22/25 households; Kibera: 68%, 17/25
126 households), or for income generation (Dagoretti South: 44%, 11/25 households; Kibera: 80%, 20/25 households)
127 (Supplementary Table 4). Over half (56%; 28/50 households) of respondents reported allowing poultry and canines
128 into the household (Supplementary Table 4). Study staff observed less animal feces near household soil sampling
129 areas in Dagoretti South (24%; 6/25 households) than in Kibera (48%; 12/25 households).

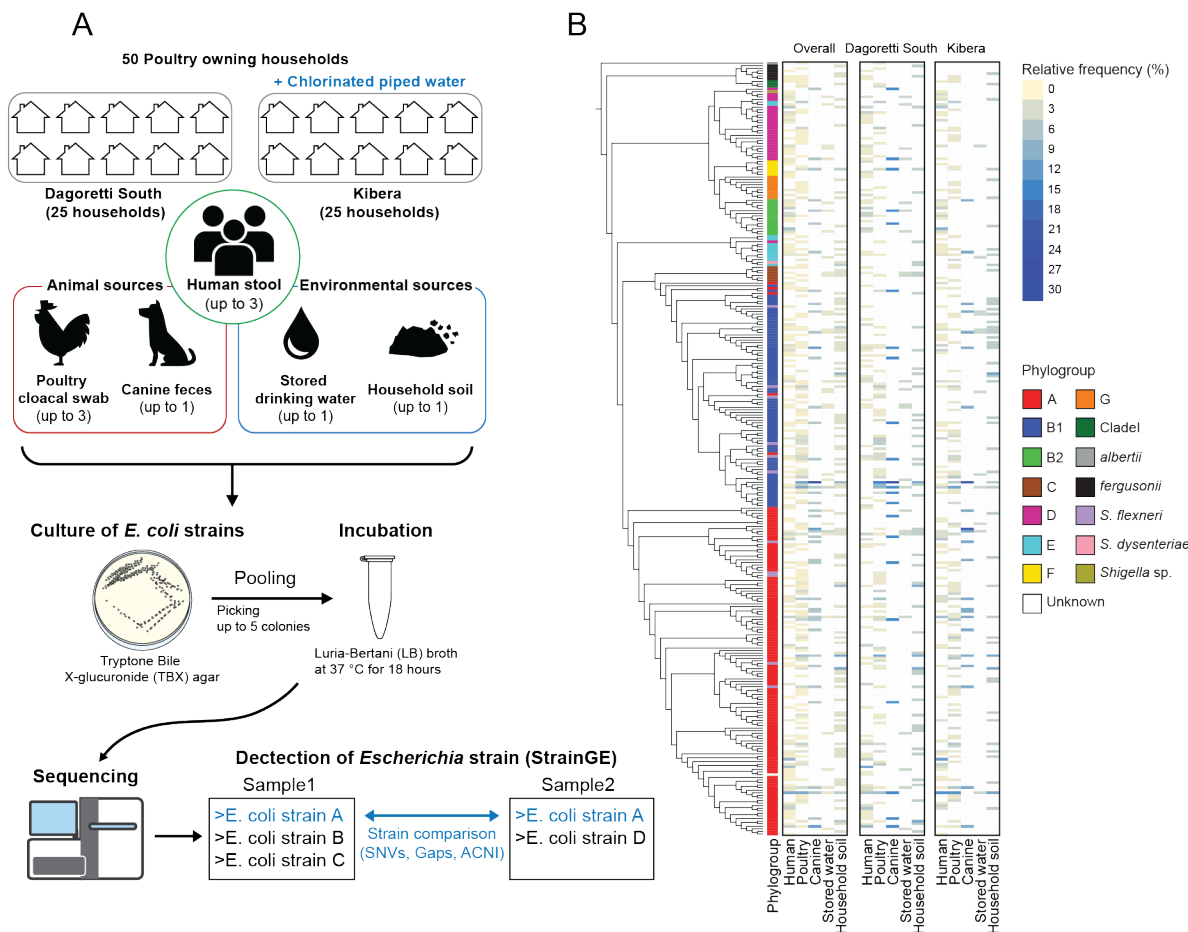
130 131 **Strain identification of colonies by sample type**

132 A total of 321 samples (132 human stool, 111 poultry cloacal swabs, 17 canine feces, 22 stored drinking water, and
133 39 household soil) were successfully collected at the 50 enrolled households, and from each sample, up to 5 isolates
134 were pooled together for PIC-seq. A total of 1,516 colonies were included. Each sample pool was sequenced and
135 analyzed with StrainGE²² to estimate the number of strains present in each sample, and their corresponding
136 representative genomes, using a database of 2,496 *Escherichia* and *Shigella* reference genomes, clustered at 99% k-
137 mer similarity (Methods). We identified 800 total strains that corresponded to 299 unique representative reference
138 strains (Fig. 1A). The identified reference strains encompassed 279 *Escherichia coli*, 1 *Escherichia albertii*, 6
139 *Escherichia fergusonii*, and 13 strains within the genus *Shigella* (Fig. 1B). Out of the 13 identified *Shigella* strains, 11
140 were assigned to *Shigella flexneri*, and the remaining two were assigned to *Shigella dysenteriae* and *Shigella* sp..

141
142 The number of identified strains in each sample was compared to assess sample strain diversity across sample types.
143 Stored drinking water had significantly lower counts of strains compared to other sample types, and household soil
144 had significantly higher strain counts than human stool (Supplementary Fig. 1A); no other significant differences were
145 observed. We then investigated how distinctively strains are distributed within sample types, based on their
146 representative reference strains. To account for different sample sizes of each sample type, we subsampled an equal
147 number ($n = 10$) of samples with 1,000 iterations and compared the number of unique reference strains (Supplementary
148 Fig. 1B). Of all sample types, household soil samples exhibited the highest number of unique reference strains while
149 stored drinking water samples exhibited the lowest. The number of unique reference strains from poultry cloacal swabs
150 and canine feces was greater than from human stool samples. In human stool, the number of unique reference strains
151 increased with age (Supplementary Fig. 1B).

152
153 We compared the phylogroup compositions of the representative reference *E. coli* strains identified across sample
154 types. Phylogroup A was the most abundant with 31.4% to 46.4% of all identified strains across all sample types
155 (Supplementary Fig. 2). The next most common phylogroup was B1, although its frequency among strains identified
156 from humans was noticeably lower (17.2%) than in other sample types (30.4% - 37.1%). Phylogroups B2 and D were
157 found to be enriched in humans (phylogroup B2: 12.3%; phylogroup D: 11.7%), stored drinking water (phylogroup
158 B2: 8.6%; phylogroup D: 11.4%), and B2 in canine samples (6.3%), compared to other sample types (phylogroup B2:
159 0.3% - 0.9%; phylogroup D: 3.4% - 4.8%). Additionally, *S. flexneri* was more frequently found among the strains
160 from poultry (7.5%) and canine (4.2%) samples, compared to other sample types, which ranged from 0% to 2.6%.

161

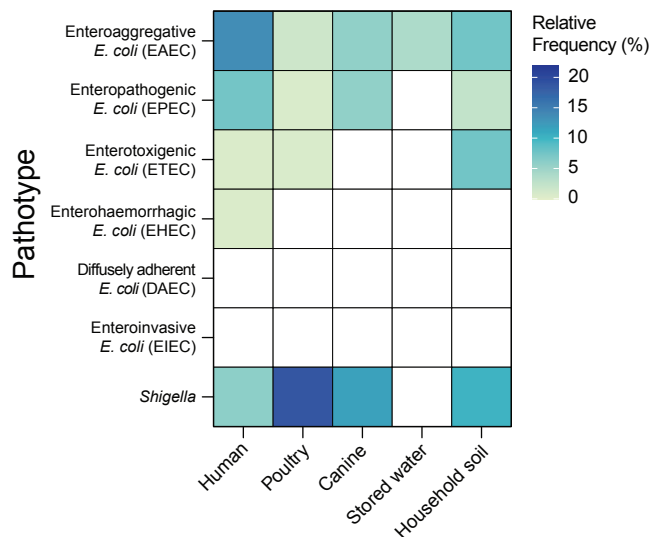


162
163 **Fig. 1** (A) Schematic of sample collection and pooling *E. coli* isolates cultured for sequencing. (B) Cladogram of all
164 isolates built using maximum likelihood based on the alignment of core genes from genomes of reference strains with
165 1000 bootstraps. The heatmap shows relative frequency of reference strain detection in each sample type, either from
166 overall households or households within each subcounty.

167
168 **Prevalence of *E. coli* pathotypes and *Shigella***

169 To examine the prevalence of *E. coli* pathotypes across sample types, we investigated the presence of known virulence
170 genes on the assembled contigs for each sample (Methods). Humans carried the most diverse pathotypes of *E. coli*,
171 with Enteroaggregative *E. coli* (EAEC) (13.6% of human samples) and Enteropathogenic *E. coli* (EPEC) (7.6%) most
172 frequently detected (Fig. 2). EAEC and EPEC were also frequently detected in canine feces (EAEC: 5.9% and EPEC:
173 5.9% of canine samples). In poultry cloacal swabs, EAEC (1.8% of poultry samples), EPEC (0.9%), and
174 Enterotoxigenic *E. coli* (ETEC) (0.9%) were detected. In stored drinking water, EAEC (4.1% of drinking water
175 samples) was the only pathotype detected while household soil had more diverse pathotypes present including EAEC

176 (7.7% of soil samples), ETEC (7.7%), and EPEC (2.6%). Diffusely adherent *E. coli* (DAEC) and Enteroinvasive *E.*
177 *coli* (EIEC) were not detected in any of the samples. We also determined the prevalence of *Shigella* across samples,
178 which was highest in poultry cloacal swabs (19.0% of poultry samples), followed by canine feces (11.8% of canine
179 samples), household soil (10.3% of soil samples), and human stool (6.1% of human samples) (Fig. 2). No *Shigella*
180 strains were detected in stored drinking water.
181

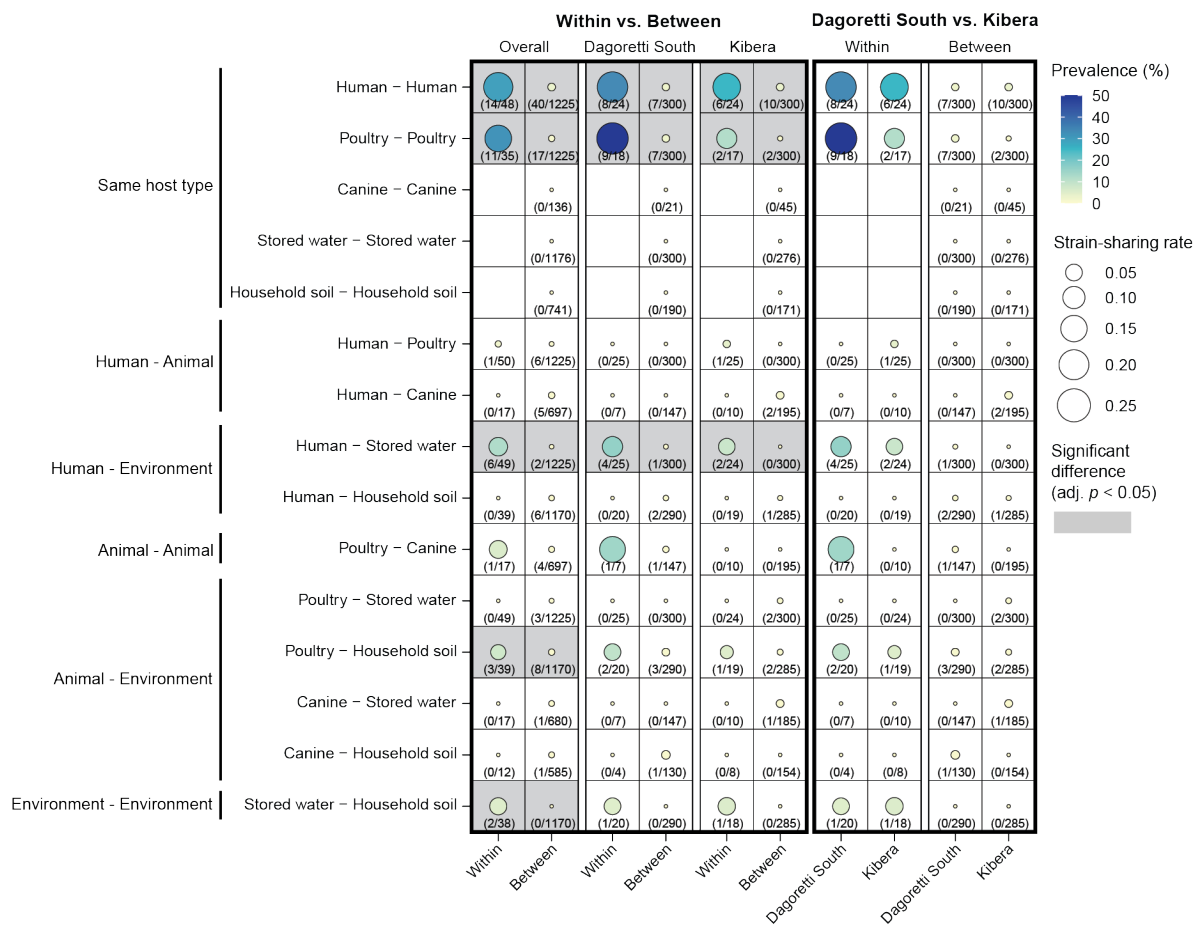


182
183 **Fig. 2 Prevalence of *E. coli* pathotypes and *Shigella* strains across sample types.** Detection of *E. coli* pathotypes
184 based on specific sets of virulence genes associated with each pathotype. The presence of *Shigella* was identified by
185 StrainGE.

186 187 **Strain-sharing patterns**

188 We examined strain-sharing patterns across humans, poultry, canine, soil, and stored drinking water *within* and
189 *between* households. Sample pairs with strains with over 99.95% average callable nucleotide identity (ACNI) by
190 StrainGE were considered to share strains based on previous benchmarking,²² and the average strain-sharing rate (the
191 average strain-sharing events/total possible pairs in each household or household pair) *within* or *between* households
192 were compared across different sample type pairs. *Within* households, strain-sharing was most frequent between the
193 same host type (human-human rate: 0.19 cases per sample pair; poultry-poultry rate: 0.15 per sample pair) with rates
194 significantly higher than the rates *between* households (human-human rate: 0.0051, $p < 0.001$; poultry-poultry rate:

195 0.0023, $p < 0.001$) (Fig. 3). Direct human-animal strain-sharing was nearly absent; a strain shared between human and
 196 poultry was detected in just one household, with the rates observed *between* households not significantly different
 197 from *within* households. Strain-sharing with environmental sources (human-stored water rate: 0.065; poultry-soil rate:
 198 0.038) was frequently observed *within* households and at significantly higher rates than *between* households (human-
 199 stored water rate: 0.00030, $p < 0.001$; poultry-soil rate: 0.0024, $p = 0.0048$). Additionally, the strain-sharing rate
 200 between stored drinking water and soil *within* households (rate: 0.053) was significantly higher than *between*
 201 households, where no strain-sharing was detected ($p < 0.001$). Unlike *within* households, where strain-sharing was
 202 most common between human and stored drinking water, significantly higher than that between human and animal (p
 203 = 0.027), human-stored drinking water strain-sharing was one of the lowest *between* households (rate: 0.00030) (Fig.
 204 3 and Supplementary table 5).
 205



206
 207 **Fig. 3 Strain-sharing patterns within and between households.** The size of each circle corresponds to the strain-
 208 sharing rate (average strain-sharing events/total possible pairs in each household or household pair). The color gradient

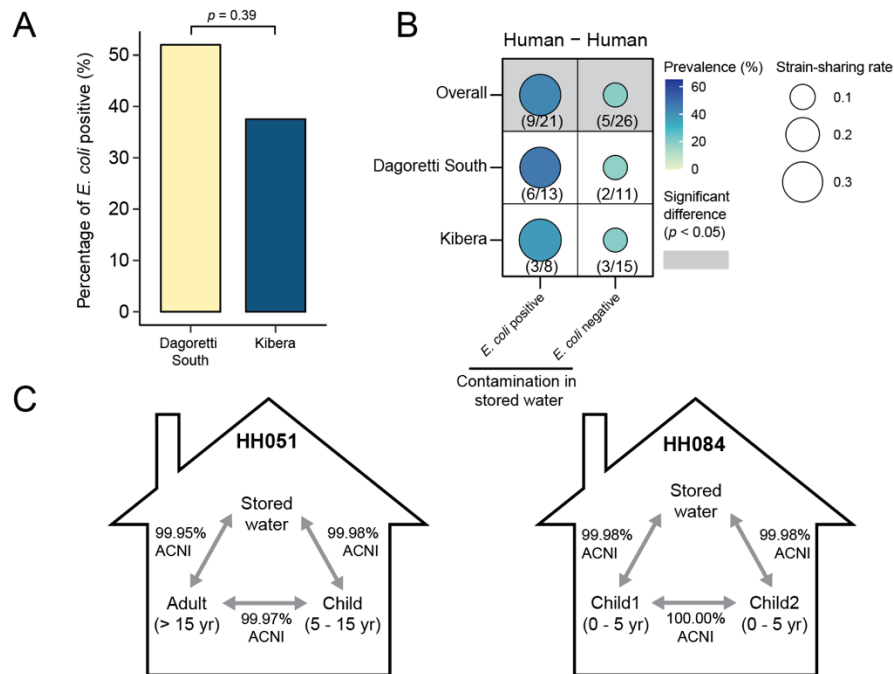
209 of each circle represents the prevalence of strain-sharing, calculated as the percentage of households or pairs of
210 households where strain-sharing was detected relative to the total number of households or household pairs, with
211 actual numbers provided as texts below each circle. Permutation tests with 1,000 bootstraps were used for testing for
212 significant differences between mean strain-sharing rates, and all resulting p -values were adjusted using Benjamini-
213 Hochberg correction.

214

215 **Contaminated drinking water facilitates human-human strain sharing**

216 In Kibera, households had access to piped chlorinated drinking water. We observed poorer water quality in Dagoretti
217 South; 52.0% of households (13 out of 25) in Dagoretti South had *E. coli* contamination in stored drinking water while
218 37.5% of households had *E. coli*-contaminated water in Kibera (9 out of 24) (Fig. 4A). We hypothesized that
219 chlorinated water could reduce human-drinking water strain-sharing *within* households. Indeed, we observed less
220 human-drinking water strain-sharing in Kibera (rate: 0.049) than Dagoretti South (rate: 0.080) although the difference
221 was not statistically significant (Fig. 3). Given that humans *within* households share the same drinking water source,
222 contaminated stored drinking water may play a role in strain-sharing between humans *within* households. Indeed,
223 human-human strain-sharing was observed in 42.9% (9/21) of the households with *E. coli*-contaminated stored
224 drinking water, compared to just 19.2% (5/26) of the households without *E. coli* contaminated water (Fig. 4B). The
225 human-human strain sharing (rate: 0.32) *within* households with *E. coli*-contaminated stored water was significantly
226 higher than households with clean water (rate: 0.090, $p = 0.026$) (Fig. 4B). No other strain-sharing sample type pairings
227 showed significant differences when households were stratified by stored water contamination (data not shown).
228 Further, we observed identical strains being shared across more than one human and stored drinking water in two
229 households (Household ID: 051 and 084), lending evidence for strain-sharing between humans through the
230 consumption of contaminated stored drinking water (Fig. 4C).

231



232

233 **Fig. 4 Strain sharing within households stratified by presence of *E. coli* contamination in stored drinking water.**

234 (A) Proportion of households with *E. coli* contaminated stored water in Dagoretti South and Kibera. (B) Comparison

235 of human-human strain-sharing within households with and without *E. coli* contamination in stored drinking water.

236 (C) Observation of strain-sharing patterns in two households with the same strain shared across humans and stored

237 drinking water within households.

238

239 PIC-seq outperforms metagenomics

240 We additionally evaluated the performance of PIC-seq in investigating *E. coli* strain-sharing patterns by comparing

241 the results with metagenome sequence data generated without the culture enrichment step. As part of another study,

242 total DNA was extracted from a subset of matched human stool samples in this study (110 samples) and sequenced.

243 StrainGE detected a slightly lower number of strains across the total metagenome samples (230 strains) vs. the

244 corresponding PIC-seq samples (248 strains). Further, while the PIC-Seq approach detected strain-sharing in 50

245 sample pairs, strain-sharing was detected in only 7 metagenomic sample pairs (5 pairs overlapping between the two

246 approaches), possibly due to insufficient coverage depth for detecting genomic substitutes in *E. coli* between samples

247 (Supplementary Fig. 3).

248

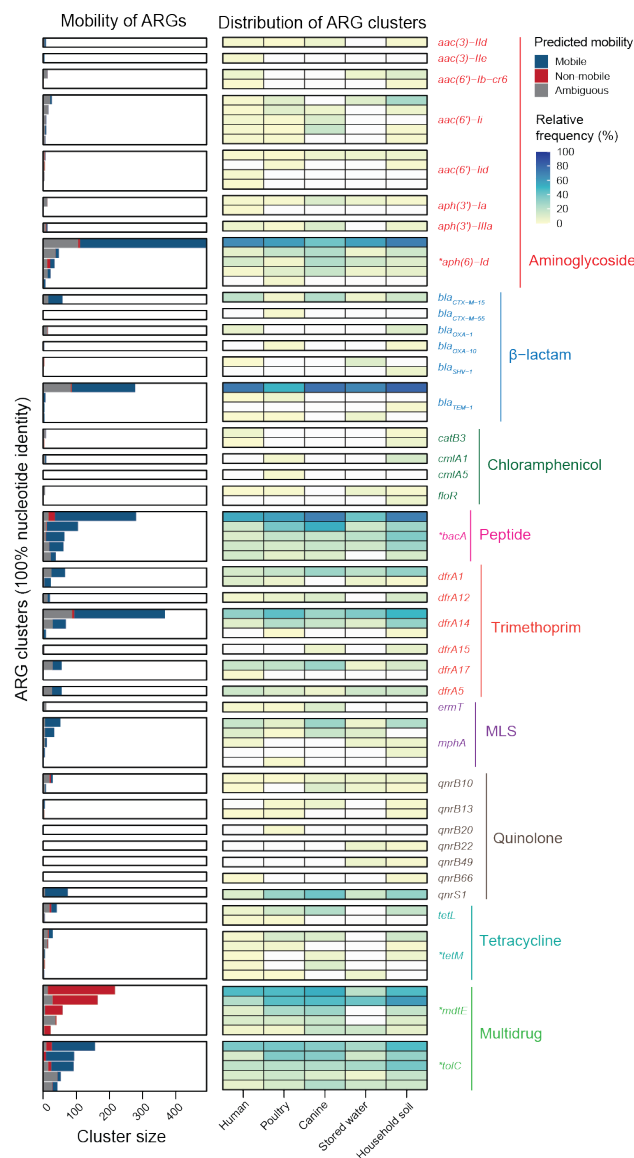
249 **Distribution of clinically relevant ARGs across humans, animals, and the environment**

250 To investigate ARG transmission, we examined the distribution and diversity of ARG variants based on the assembled
251 contigs for each sample type. To assess if long-read sequencing provided higher resolution insights into ARG
252 transmission, we benchmarked using simulated pooled *E. coli* datasets on contigs assembled from Oxford Nanopore
253 Technology long reads versus Illumina short reads (Supplementary methods). While contigs assembled from long
254 read data consistently generated longer contigs than those from short reads across all simulated datasets - in terms of
255 the proportion of the recovered genome fraction, the length of the largest alignment, and N50 - they exhibited the
256 lowest accuracy in their sequence identity to the original genomes (Supplementary Fig. 4). Hybrid assembly utilizing
257 both short and long reads marginally improved contig length but resulted in lower accuracy compared to short-read
258 assemblies. Thus, we opted to analyze ARGs using only contigs assembled from short reads. Additionally, we
259 benchmarked our assembly approach, which uses reference-based binning before *de novo* assembly to reconstruct
260 resistomes, against sole *de novo* assembly (Supplementary Methods). Our approach significantly outperformed sole
261 *de novo* assembly, reconstructing 81.6% to 89.2% of the original resistome compared to just 15.1% to 27.8%
262 (Supplementary Fig. 5).

263
264 We grouped all identified ARGs on the contigs into clusters based on their nucleotide sequences at 100% identity and
265 compared the number of unique ARG clusters observed in each sample type. Similar to strain diversity patterns,
266 household soil had the highest number of unique ARG clusters (median: 266) and stored drinking water the lowest
267 (median: 152) (Supplementary Fig. 6). No significant difference was observed between humans (median: 200.5) and
268 animals (poultry median: 231; canine median: 238).

269
270 We investigated the diversity and distribution of high-risk ARGs, classified as Rank I ARGs by Zhang *et al.*³⁰, across
271 our sample types. Rank I ARGs refer to gene families that have been reported to be significantly enriched in human-
272 associated environments, associated with gene mobility, and are known to be present in ESKAPE pathogens
273 (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas*
274 *aeruginosa*, and *Enterobacter* sp.). Humans had significantly higher numbers of Rank I ARG clusters conferring
275 resistance to the drug classes of beta-lactam and macrolide-lincosamide-streptogramin (MLS) than poultry, while
276 Rank I ARG clusters of trimethoprim and quinolone were found to be higher in poultry (Supplementary Fig. 7). We

277 assessed ARG mobility in our samples based on adjacency to mobile genetic elements (MGEs); if ARGs were
278 associated with contigs under 5000 bp, they were determined to have ambiguous mobility (Methods). For ARGs where
279 mobility could be determined, an average of 64.2% of ARGs within Rank I ARG clusters were predicted to be mobile,
280 except for *mdtE*, where none of ARGs within the clusters were predicted to be mobile (Fig. 5 and Supplementary
281 Table 6). Prevalence of ARG clusters across humans, animals, and the environment was similar, with several ARG
282 clusters prevalent in all sample types. *aph(6)-Id*, which confers resistance to the antibiotic class of aminoglycoside,
283 had the largest cluster size and was the most frequently present in all groups, including in 41.2% of canine samples.
284 Among β -lactamase genes, a *bla*_{TEM-1} cluster was the most prevalent in all groups, with detection in 51.4% of poultry
285 samples. Another β -lactamase gene cluster found in all sample types was *bla*_{CTX-M-15}, which confers resistance to third
286 generation cephalosporins (Fig. 5). Its prevalence was comparable in humans (20.5%), canines (23.5%), and
287 household soil (15.8%); it had lower prevalence in poultry (4.5%) and stored water (5.3%). We identified several other
288 ARG clusters in all sample types; genes conferring resistance to trimethoprim (*dhfrA*), MLS (*mphA*), and quinolone
289 (*qnrS*) were prominent. We also observed widespread presence across all samples of the bacitracin resistance gene
290 *bacA*, a peptide antibiotic, and the *tolC* gene, associated with multidrug resistance as part of the efflux pump system.
291



292
 293 **Fig. 5 Distribution and mobility of clinically relevant ARG clusters at high-risk (Rank I) identified across**
 294 **humans, animals, and the environment.** The asterisks on ARG annotations indicate that only the top five most
 295 abundant ARG clusters are depicted in the figure.

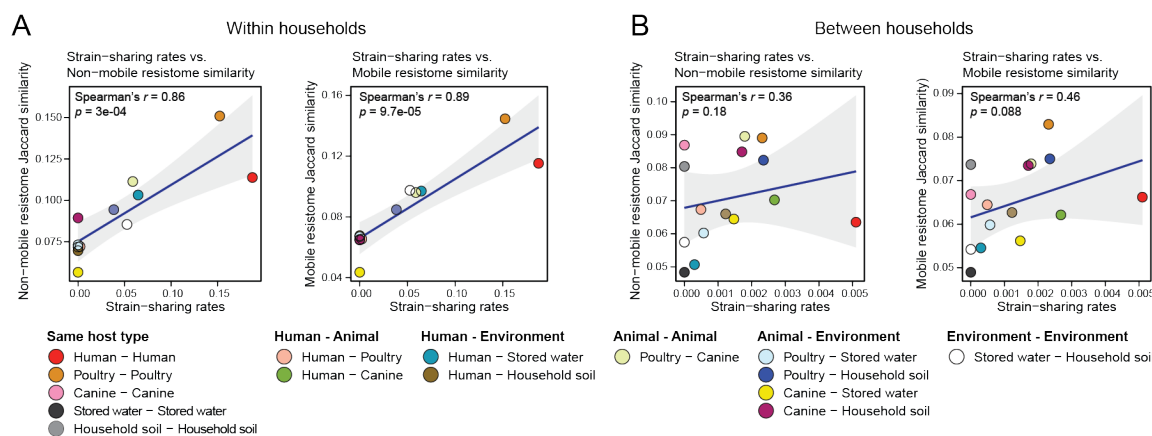
296
 297 **Strain-sharing disseminates ARGs**

298 To investigate whether spatial proximity is associated with similarity in ARG profiles, sample resistomes were
 299 compared *within* versus *between* households. Resistome similarity closely mirrored strain-sharing rates, with
 300 significantly higher resistome similarity *within* households compared to *between* households for human-human,
 301 poultry-poultry, human-water, and water-soil (Supplementary Fig. 8). In addition, poultry-canine pairs shared

302 significantly higher resistome similarity *within* households than *between* household when stratified by predicted
 303 mobility (Supplementary Fig. 8).

304
 305 To further investigate the influence of strain-sharing on resistome composition, we performed a correlation analysis
 306 between mean strain-sharing rates and mean resistome similarities between sample types (Fig. 6). We hypothesized
 307 that if ARGs are not specific to certain strains due to frequent horizontal gene transfer and are already prevalent among
 308 samples, a weak correlation between strain-sharing rates and resistome similarities would be expected. Both non-
 309 mobile and mobile resistome similarities were strongly correlated with strain-sharing rates *within* households (non-
 310 mobile vs. rate: Spearman's $r = 0.86$; $p = 0.0003$ and mobile vs. rate: Spearman's $r = 0.89$; $p = 0.000097$), suggesting
 311 frequent strain-sharing *within* households was a strong factor for sharing similar resistomes (Fig. 6). However, no
 312 significant correlation was observed *between* households, presumably due to the insufficient number of strain-sharing
 313 events to affect mean resistome similarities. A bootstrapped analysis revealed different correlation coefficient
 314 distributions *within* versus *between* households (Supplementary Fig. 9).

315



316
 317 **Fig. 6 Spearman's correlation analyses of non-mobile and mobile resistome similarity patterns with strain-**
 318 **sharing patterns.** Left and right panels are the analyses (A) *within* and (B) *between* households.

319

320 Discussion

321 Here, we observed high rates of *E. coli* strain sharing between humans and stored drinking water *within* households
 322 compared to *between* households, suggesting the importance of drinking water as a transmission route of bacterial
 323 strains to humans.^{31,32} *E. coli* contaminated drinking water was associated with higher human-human strain-sharing

324 rates in households and we captured instances of the same strain shared between humans and water in the same
325 household. Frequent strain exchange between drinking water and humans could explain the similarities in phylogroup
326 profiles of strains between humans and stored drinking water, including the comparable proportion of phylogroups
327 B2 and D.³³ Our results are consistent with bacterial strains ingested through water consumption readily colonizing
328 the human gastrointestinal tract.^{34,35} We observed lower levels of human-water strain-sharing rates and stored drinking
329 water *E. coli* contamination in Kibera compared to Dagoretti South (although not statistically significant), suggesting
330 access to chlorinated water in Kibera disrupted strain-sharing via drinking water.³⁶ Water quality data collected by the
331 parent study documented significantly lower levels of stored drinking water *E. coli* contamination in Kibera (22.6%)³⁷
332 compared to Dagoretti South (68.2%) with a larger set of households (120 households) (Fisher's exact test; p value <
333 0.001).³⁷ A previous randomized controlled trial of community-wide passive chlorination of drinking water in urban
334 Bangladesh found a 23% reduction in 7-day child diarrhea prevalence, suggesting the effectiveness of chlorinated
335 water in disrupting bacterial transmission.³⁶ Considering the complex bacterial transmission routes in densely
336 populated urban informal settings,⁸ our findings highlight chlorinated drinking water as a key strategy to disrupt
337 household strain-sharing.

338
339 It has been hypothesized that domestic animal exposure increases the risk of zoonosis.^{31,38} While we rarely observed
340 direct zoonotic sharing events in our study population, we documented that poultry-soil strain sharing was more
341 frequent *within* households compared to *between* households. In our urban study context, rare poultry-human strain
342 sharing within households could be explained by infrequent exchange of strains between humans and soil (Fig 3).
343 Previous research in rural and/or urban regions of Kenya,³⁹ Bangladesh,^{31,40} and India⁴¹ has indicated that exposure to
344 domesticated animals increases environmental fecal bacteria contamination, evidenced by the higher prevalence of
345 animal versus human fecal markers in soil and drinking water. Given that environment-associated strain-sharing rates
346 with humans or animals were significantly higher *within* households versus *between* households, our findings highlight
347 the critical role of the environment in bacterial transmission. A few human-animal strain-sharing events were observed
348 *between* households at a rate not significantly different from that *within* households, consistent with a previous study
349 in Kenya, which found that sharing between humans and livestock was not confined to the same households.¹³ Taken
350 together with previous evidence, our results suggest that human-animal strain-sharing can occur at the community
351 scale.

352

353 Our results are consistent with clonal spread of bacteria being the main driver shaping resistomes of pooled *E. coli*
354 strains. ARG variants associated with MGEs have the potential for horizontal gene transfer, which would result in
355 overlapping ARGs between non-identical or phylogenetically distant strains. In our study, the significant correlation
356 of mobile (and non-mobile) ARG similarity with strain-sharing rates between samples *within* households, indicates
357 that close proximity strain-sharing had a greater influence on resistome similarities than ARG exchange via horizontal
358 gene transfer. This observation aligns with a previous study in Kenya,⁴³ where the multilocus sequence type of *E. coli*
359 isolates from humans, domestic animals, and wild animals best explained the variance of the mobile ARG assemblages
360 among them. Within households, human-water pairs showed higher resistome similarity (due to clonal transmission)
361 than human-animal pairs, underscoring the role of frequent bacterial strain-sharing in ARG spread. Between
362 households, where strain-sharing events were rare, no correlations were observed between resistome similarities and
363 strain-sharing rates across sample types.

364

365 Environmental sample resistomes possessed many high-risk ARGs.³⁰ Rank I ARG variants within environmental
366 samples had similar or even higher diversity than human and animal hosts across antibiotics classes, and our results
367 suggest new ARG variants can be transferred to humans and animals through drinking water and consumption of soil
368 particles. ARGs like *bla*_{CTX-M-15}, conferring resistance to third-generation cephalosporins, were detected consistently
369 across all sample types. Use of antibiotics could then exert selection pressure in the gut enriching these ARGs.
370 Quinolones and trimethoprim have been reported to be widely used by poultry farmers in Nairobi, Kenya, and this
371 would explain the higher number of Rank I ARG clusters of these antibiotics observed in poultry compared to
372 humans.^{44,45} Beta-lactam antibiotics have also been reported for use in animals in Nairobi;⁴⁴ however, the observed
373 higher diversity in humans could be attributed to frequent human consumption of these antibiotics.⁴⁶ Prevalence of
374 plasmid-mediated ARGs like *bla*_{TEM-1}⁴⁷ and *aph(6)-Ia*⁴⁸ across all sample types corroborated that certain ARGs and
375 MGEs would be more prone to dissemination.⁴⁹ Our profiling of ARG mobility was not always possible due to
376 insufficient contiguity of partial genomic segments assembled from short reads. Oxford Nanopore sequencing
377 technology, which generates longer sequences has the potential to identify the genomic context of ARGs.⁵⁰ In our
378 benchmarking, however, the contigs assembled with short reads in this study outperformed those with long reads in
379 their sequence accuracy (Supplementary Fig. 4). Hybrid assembly, leveraging both short and long reads, has been

380 suggested as an alternative approach with better assembly quality,⁵¹ however, the improvement in sequence contiguity
381 and accuracy for hybrid assembly was marginal in our benchmarking.

382
383 Our results demonstrate the value of PIC-seq for examining strain sharing and the prevalence of ARGs. PIC-seq
384 demonstrated higher performance in capturing more strain-sharing events (50 sharing events) compared to those
385 identified through traditional metagenomic sequencing (only 7 sharing events) in a subset of matching stool
386 metagenomes sequenced for another study (Supplementary Fig. 3). Further, our integration of additional *de novo*
387 assembly with reference-based binning assembly enabled the recovery of ARGs located in genomic regions of pooled
388 organisms that differed from reference genomes, outperforming sole *de novo* assembly (Supplementary Fig. 10).
389 ARGs recovered from unbinned reads were mostly associated with MGEs, such as plasmids and genomic islands,
390 which can circulate among strains via horizontal gene transfer.⁵² While these ARGs could not be tied to a specific
391 strain, we were able to better capture the resistome of pooled isolates from each sample. Compared to sequencing of
392 single isolates or sequencing the entire microbiome for strain tracking, our approach offers a distinct advantage in that
393 it allows for simultaneous capture of strain identity and sensitive detection of ARGs with genomic context. PIC-seq
394 is a scalable method for examining strain-sharing patterns that could be replicated for other bacterial species.

395
396 **Methods**

397 **Study sites and sample collection**

398 To investigate *E. coli* strain-sharing between humans, domesticated animals, and the environment, 50 poultry-owning
399 households with at least one child under the age of 5 years were enrolled in two subcounties, Dagoretti South (n = 25)
400 and Kibera (n = 25), in Nairobi, Kenya, during June – August 2019. One household was enrolled from each compound
401 that owned poultry. Written informed consent was obtained from each adult participant prior to stool sampling and
402 child assent and parental written consent were obtained for each child participant. A household survey was performed
403 along with the sample collection to collect demographic information and household characteristics including water,
404 sanitation, and hygiene access and practices.

405
406 Each household was visited twice for the sample collection. During the first visit, we collected up to 3 poultry cloacal
407 swabs (depending on the number of poultry present), one canine fecal sample, one household soil, and one stored

408 drinking water sample from each household. A trained veterinary student administered the poultry cloacal swabs and
409 placed them in storage tubes filled with Cary-Blair transport medium, kept in a cooler for transport to the lab. To
410 collect canine feces, the top layer of fresh feces from the center of the pile was transferred into a 50 mL centrifuge
411 tube with a sterile plastic scoop. For household soil collection, a 30 cm x 50 cm square area closest to the entrance of
412 the household was marked, and the top layer of the entire surface of the marked area was scraped with a sterile
413 disposable plastic scoop. Approximately 150 g of soil was collected and transferred into a 500 mL sterile Whirlpak
414 bag. The stored water within each household was collected by withdrawing 350 mL of the water from the bottom of
415 the storage container with a sterile serological pipette. The water was then transferred into a 500 mL sterile Whirlpak
416 bag containing 10 mg of sodium thiosulfate. All samples were immediately placed in a cooler filled with ice and
417 transported to the Kenya Medical Research Institute (KEMRI).

418
419 Approximately 1 week from the initial visit, households were revisited to collect human stool from one household
420 member in the following three age groups: child aged 0 - 4 years, child aged 5 - 14 years, and adult aged 15 years or
421 older. A stool collection kit was provided during the first visit, which included a 50 mL plastic pot with a sterile scoop
422 for each member with instructions on how to collect the sample. Caretakers were instructed to collect feces on
423 aluminum foil, then, using sterile gloves and scoops, to transfer the feces to the plastic pot for collection. The primary
424 caretaker of each household was informed by mobile phone one day prior to the revisit to collect stool from the
425 previous night or the morning of the revisit day. If stool samples were not available on the visit day, households were
426 revisited up to three times to collect the remaining stool samples. After pick-up, the received stool samples were placed
427 in a cooler filled with ice and transported to KEMRI.

428
429 The study received ethical approval from the Kenya Medical Research Institute (KEMRI) Scientific and Ethics Review
430 Unit (12/3823) and the Tufts Health Sciences Institutional Review Board (13205). Additionally, a research permit was
431 granted by the Kenyan National Commission for Science, Technology, and Innovation.

432

433 **Pooling Isolated Colonies for sequencing (PIC-seq)**

434 Here we describe our method of pooling presumptive isolated *E. coli* colonies for sequencing (PIC-seq). *E. coli*
435 colonies were first cultured from the collected samples within 6 hours of sample collection. A sterile swab was inserted

436 into 1 g of each collected human stool and canine feces sample, and the swab was streaked on Tryptone Bile X-
437 glucuronide (TBX) agar plates. Poultry cloacal swabs were directly streaked on the agar plates. Household soil and
438 stored water samples were membrane filtered and cultured on TBX. Collected soil was first screened through 2 mm-
439 pore mesh sieves to remove rocks and leaves, then 5 g were homogenized with 50 mL of distilled water. After allowing
440 soil particles to settle for 1 minute, 100 μ L of supernatant was diluted with 10 mL of distilled water and vacuumed
441 through a 0.45- μ m membrane filter, and plated. 100 mL of stored water was directly filtered, then plated on TBX. All
442 plates were incubated overnight (18 – 24 hours) at 44 °C.

443
444 Post incubation, up to 5 presumptive *E. coli* colonies (blue-green color) were selected from each agar plate for pooling.
445 The selected colonies from each plate were transferred into a single 2 mL tube containing 1 mL of Luria-Bertani (LB)
446 medium. The pooled *E. coli* colony cultures were incubated overnight (18 hours) at 37 °C with shaking at 200 rpm.
447 After incubation, 100 μ L of the cultures were mixed with 1 mL of 30 % glycerol-70% Tryptone Soy Broth (TSB)
448 solution and stored at – 80 °C. The samples were then shipped on dry ice to Tuft University (Medford, MA).

449
450 Pooled isolates were thawed on ice and inoculated into 5 mL of fresh TSB medium with a sterile inoculating loop.
451 The cultures were incubated overnight at 37 °C with 200 rpm shaking, then 1.5 mL of the culture was pelleted by
452 centrifugation at 5,000 x g for 10 min. The DNA was then extracted from pellets using DNeasy PowerSoil Pro Kit
453 (Qiagen, Hilden, Germany) following the protocol provided by the manufacturer. The concentration and purity of the
454 extracted DNA were measured by a Qubit 4.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA), a sequencing
455 library was prepared using Nextera XT kit (Illumina, San Diego, CA), and shotgun sequencing was performed on
456 Illumina NovaSeq 6000 (Illumina) with target sequencing depth of 1.5 Gb at the Broad Institute (Cambridge, MA).

457

458 **Identification of *E. coli* strains and calculation of household-wise strain-sharing rates**

459 The quality of raw sequence reads was examined using FastQC,⁵³ and low-quality reads were filtered out using
460 Trimmomatic v0.39⁵⁴ with parameters set to LEADING:3, TRAILING:3, SLIDINGWINDOW:4:15, and
461 MINLEN:70. The subsequent reads were used for the identification and comparison of *Escherichia* strains using the
462 software StrainGE v1.3.3, the software specifically designed for accurate calling and comparison of strains in the
463 complex metagenome data based on k-mer comparison, bypassing assembly process.²² The database for StrainGE was

464 constructed using the complete genome sequences (2,496 genomes) of the genus of *Escherichia* or *Shigella*
465 downloaded from NCBI RefSeq database on August 20, 2022, by using the built-in database construction pipeline
466 ('ncbi-genome-download', 'kmerize', 'kmersim', 'cluster', and 'createdb'). The reference genomes in the database
467 were clustered if they shared 99% of their k-mer content with a Jaccard similarity over 0.90, and to minimize
468 redundancy, only the representative genome sequences of each cluster were retained, resulting in a final set of 885
469 genomes.

470

471 After strain identification, we utilized StrainGR, the second module of StrainGE, to call sequence variants among
472 samples.²² Sequence reads were aligned to the genomes of the identified strains in each sample using bwa mem v0.7.17
473 with default parameters.⁵⁵ For sample pairs with the same reference strain, the average callable nucleotide identity
474 (ACNI) and gap similarity were calculated based on the callable regions, using 'call' and 'compare' command in
475 StrainGR.²² The threshold for defining the same strain was set to > 99.95% ACNI.

476

477 The strain-sharing rates between sample types were calculated, using the results from StrainGR. Within each
478 household, we normalized the number of sample pairs sharing the same strain by the total number of combinations
479 between the certain sample types. We then computed the average rates based on the rates from all households for
480 comparison. For strain-sharing rates between different households, we normalized the number of sample pairs with
481 the same strain by the total number of possible sample pairs, between all combinations of different households. We
482 then calculated the average rates from these inter-household rates for comparisons between certain types of samples.
483 Stored water samples without *E. coli* contamination were included in calculating water-related total number of possible
484 pairs.

485

486 **Phylogeny and phylogroup of detected *E. coli* strains**

487 To infer the phylogenetic relationships among the strains identified from the samples, we utilized the reference genome
488 sequences included in the StrainGE database. Prokka v1.14.5 was used to annotate genes, with the genus parameter
489 set to *Escherichia*.⁵⁶ The core genes that exist in 99% of all reference genomes were identified using Roary v3.13.0.⁵⁷
490 The concatenated core gene alignment was generated using MAFFT, and SNP-sites v2.5.1⁵⁸ was used to extract the
491 alignment regions with single nucleotide polymorphisms (SNPs) to reduce the size of the alignment. The reduced

492 alignment was uploaded to CIPRES Science Gateway v3.3⁵⁹ (www.phylo.org), and a maximum likelihood tree was
493 constructed using RAxML v8.2.12⁶⁰ with the nucleotide substitution model set to GTRGAMMA with 500 bootstraps.
494 *E. coli* phylogroup was predicted using ClermonTyping v20.03⁶¹ with default parameters.

495

496 **Annotations of antibiotic resistance genes and mobile genetic elements**

497 The integrated assembly pipeline, which combines reference-based binning and *de novo* assembly, was used to
498 generate contigs for ARGs and mobile genetic elements (MGEs) annotation (Supplementary Fig. 11). Plasmid
499 sequences were assembled from sequencing reads using metaplasmidSPAdes v3.15.4⁶² with default parameters.
500 Subsequently, the original reads were aligned and binned to the set of the assembled plasmid sequences and the
501 reference genomes of strains identified through StrainGST, using mSWEEP v.2.0.0²⁷ and mGEMs v.1.3.0¹⁸ with
502 default parameters. For each reference strain, the reads binned to their genome sequences were individually assembled
503 *de novo* using metaSPAdes v3.15.4⁶³ with default parameters. The sequence reads that did not align to any sequences
504 were separately assembled *de novo* using metaSPAdes. The gene-coding sequences in the assembled contigs were
505 predicted using Prodigal v2.6.3⁶⁴ with default parameters. The translated amino acid sequences of the predicted gene-
506 coding sequences were annotated using DIAMOND blastp v2.0.14.152⁶⁵ against the Comprehensive Antibiotic
507 Resistance Database (CARD) v3.2.4⁶⁶ with the parameters set to ‘--id 90’, ‘--min-orf 25’, and ‘--query-cover 80’. The
508 annotations with the highest bitscore for each gene-coding sequence were retained for the downstream analysis.

509

510 To predict the mobility of ARGs, we examined the presence of any MGEs in the flanking regions of the annotated
511 ARGs. We extracted up to 5000 base pairs from both upstream and downstream of ARGs on the contigs using a
512 custom bash script. The mobileOG-pl pipeline was then used to annotate MGEs with mobileOG-db release beatrix-
513 1.6 as the reference database,⁶⁷ which is a manually curated database that includes sequences from eight different
514 databases (ICEBerg, ACLAME, GutPhage Database, Prokaryotic viral orthologous groups, COMPASS, NCBI
515 Plasmid RefSeq, immedb, and ISfinder), along with homologs of the manually curated sequences. ARGs with
516 identified MGEs in their flanking region were presumed to be mobile, while those without MGEs were categorized as
517 non-mobile. In instances where the contig length was insufficient (< 5000 base pairs) to detect MGEs in the flanking
518 region, the mobility of these ARGs was considered ambiguous.

519

520 **Identification of pathogenic *Escherichia* and *Shigella* strains**

521 The identification of pathogenic *E. coli* strains was inferred based on the presence of pathotype-associated virulence
522 genes detected in the assembled contigs for each sample. These virulence genes were annotated from the contigs using
523 ABRicate, in combination with the Virulence Factor Database (VFDB) (download in May, 2023).⁶⁸ The presence of
524 different *E. coli* pathotypes was based on the detection of the following gene sets: *agg*, *aat*, and *aai* for
525 Enteroaggregative *E. coli* (EAEC); *eae* and *bfp* for Enteropathogenic *E. coli* (EPEC); *elt* and *est* for Enterotoxigenic
526 *E. coli* (ETEC); *stx* for Enterohaemorrhagic *E. coli* (EHEC); *afa* and *drg* for Diffusely adherent *E. coli* (DAEC); and
527 *ipa* for Enteroinvasive *E. coli* (EIEC).^{69–71} Presence of *Shigella* was additionally investigated based on StrainGST
528 results.

529

530 **Statistical analysis**

531 All statistical analyses in this study were carried out using R version 4.2.1.⁷² The differences in mean strain-sharing
532 rates were compared using a nonparametric permutation test, with 1000 iterations of resampling. Resistome similarity
533 was estimated by Jaccard similarity of ARG cluster compositions, defined as the number of unique ARG clusters
534 shared divided by the total number of unique ARG clusters from both samples. The pairwise Mann-Whitney U test
535 was used for comparing resistome similarities between sample types, using the 'wilcox.test' command from the 'stats'
536 package in R.⁷² All *p*-values were adjusted accordingly using Benjamini-Hochberg correction. Finally, the correlation
537 analyses among strain-sharing and resistome similarity patterns were performed using Spearman's rank correlation
538 method, as implemented in the 'cor.test' function from the 'stats' package in R.⁷²

539

540 **Data availability**

541 All raw sequence data were deposited in the NCBI database under BioProject accession number PRJNA1126668.

542

543 **Acknowledgments**

544 This work was financially supported by the Bill and Melinda Gates Foundation (Grant number OPP1200651); the
545 Tufts Clinical and Translational Science Institute (CTSI) Pilot Studies Program; a Tufts Center for Integrated
546 Management of Antimicrobial Resistance (CIMAR) Pilot Award; and with federal funds from the National Institute
547 of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under

548 grant number U19AI110818 to the Broad Institute. We thank Marieke H. Rosenbaum, Christine Imali, Anne Okello,
549 Fredrick Okuku Owuor, Eric Kipchirchir Tanui, Glory Wairimu, Henry Emonje Mukabwa, Frank Odhiambo, Erica
550 R. Fuhrmeister, Maya Nadimpalli, Tim Julian, Gretchen Walch, Deeksha Bathini, Gracie Hornsby, Jeremy Lowe, and
551 the many KEMRI interns for their research assistance.

552

553 **References**

- 554 1. Bryce, J., Boschi-Pinto, C., Shibuya, K. & Black, R. E. WHO estimates of the causes of death in children. *Lancet*
555 365, 1147–1152 (2005).
- 556 2. Liu, L. et al. Global, regional, and national causes of child mortality in 2000–13, with projections to inform post-
557 2015 priorities: an updated systematic analysis. *Lancet* 385, 430–440 (2015).
- 558 3. Robertson, R. C., Manges, A. R., Finlay, B. B. & Prendergast, A. J. The Human Microbiome and Child Growth –
559 First 1000 Days and Beyond. *Trends Microbiol.* 27, 131–147 (2019).
- 560 4. Kotloff, K. L. et al. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries
561 (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet* 382, 209–222 (2013).
- 562 5. Ikimiukor, O. O., Odih, E. E., Donado-Godoy, P. & Okeke, I. N. A bottom-up view of antimicrobial resistance
563 transmission in developing countries. *Nat. Microbiol.* 7, 757–765 (2022).
- 564 6. Nadimpalli, M. L. et al. Urban informal settlements as hotspots of antimicrobial resistance and the need to curb
565 environmental transmission. *Nat. Microbiol.* 5, 787–795 (2020).
- 566 7. Swarthout, J. M., Chan, E. M., Garcia, D., Nadimpalli, M. L. & Pickering, A. J. Human colonization with antibiotic-
567 resistant bacteria from nonoccupational exposure to domesticated animals in low-and middle-income countries: a
568 critical review. *Environ. Sci. Technol.* 56, 14875–14890 (2022).
- 569 8. Penakalapati, G. et al. Exposure to animal feces and human health: a systematic review and proposed research
570 priorities. *Environ. Sci. Technol.* 51, 11537–11552 (2017).
- 571 9. Nadimpalli, M. L. et al. Drinking water chlorination has minor effects on the intestinal flora and resistomes of
572 Bangladeshi children. *Nat. Microbiol.* 7, 620–629 (2022).
- 573 10. Fuhrmeister, E. R. et al. Evaluating the relationship between community water and sanitation access and the global
574 burden of antibiotic resistance: an ecological study. *Lancet Microbe* 4, e591–e600 (2023).

- 575 11. Pehrsson, E. C. et al. Interconnected microbiomes and resistomes in low-income human habitats. *Nature* 533, 212–
576 216 (2016).
- 577 12. Muloj, D. M. et al. Population genomics of *Escherichia coli* in livestock-keeping households across a rapidly
578 developing urban landscape. *Nat. Microbiol.* 7, 581–589 (2022).
- 579 13. Hassell, J. M., Begon, M., Ward, M. J. & Fèvre, E. M. Urbanization and disease emergence: dynamics at the
580 wildlife–livestock–human interface. *Trends Ecol. Evol.* 32, 55–67 (2017).
- 581 14. Banerjee, S. & Van Der Heijden, M. G. A. Soil microbiomes and one health. *Nat. Rev. Microbiol.* 21, 6–20 (2023).
- 582 15. Essack, S. Y. Environment: the neglected component of the One Health triad. *Lancet Planet. Health* 2, e238–e239
583 (2018).
- 584 16. Huijbers, P. M. et al. Role of the environment in the transmission of antimicrobial resistance to humans: a review.
585 *Environ. Sci. Technol.* 49, 11993–12004 (2015).
- 586 17. Rousham, E. K., Unicomb, L. & Islam, M. A. Human, animal and environmental contributors to antibiotic
587 resistance in low-resource settings: integrating behavioural, epidemiological and One Health approaches. *Proc. R.*
588 *Soc. B* 285, 20180332 (2018).
- 589 18. Mäklin, T. et al. Bacterial genomic epidemiology with mixed samples. *Microb. Genom.* 7, (2021).
- 590 19. Gardy, J. L. & Loman, N. J. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat.*
591 *Rev. Genet.* 19, 9–20 (2018).
- 592 20. Schürch, A., Arredondo-Alonso, S., Willems, R. & Goering, R. Whole genome sequencing options for bacterial
593 strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene–based
594 approaches. *Clin. Microbiol. Infect.* 24, 350–354 (2018).
- 595 21. Treangen, T. J., Ondov, B. D., Koren, S. & Phillippy, A. M. The Harvest suite for rapid core-genome alignment
596 and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* 15, 1–15 (2014).
- 597 22. van Dijk, L. R. et al. StrainGE: a toolkit to track and characterize low-abundance strains in complex microbial
598 communities. *Genome Biol.* 23, 74 (2022).
- 599 23. Olm, M. R. et al. inStrain profiles population microdiversity from metagenomic data and sensitively detects shared
600 microbial strains. *Nat. Biotechnol.* 39, 727–736 (2021).
- 601 24. Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J. & Segata, N. Shotgun metagenomics, from sampling to
602 analysis. *Nat. Biotechnol.* 35, 833–844 (2017).

- 603 25. Kaminski, J. et al. High-specificity targeted functional profiling in microbial communities with ShortBRED. *PLoS*
604 *Comput. Biol.* 11, e1004557 (2015).
- 605 26. Munk, P. et al. Genomic analysis of sewage from 101 countries reveals global landscape of antimicrobial resistance.
606 *Nat. Commun.* 13, 7251 (2022).
- 607 27. Mäklin, T. et al. High-resolution sweep metagenomics using fast probabilistic inference. *Wellcome Open Res.* 5,
608 (2020).
- 609 28. Van Elsas, J. D., Semenov, A. V., Costa, R. & Trevors, J. T. Survival of *Escherichia coli* in the environment:
610 fundamental and public health aspects. *ISME J.* 5, 173–183 (2011).
- 611 29. Sarowska, J. et al. Virulence factors, prevalence and potential transmission of extraintestinal pathogenic
612 *Escherichia coli* isolated from different sources: recent reports. *Gut Pathog.* 11, 1–16 (2019).
- 613 30. Zhang, A.-N. et al. An omics-based framework for assessing the health risk of antimicrobial resistance genes. *Nat.*
614 *Commun.* 12, 4765 (2021).
- 615 31. Ercumen, A. et al. Animal feces contribute to domestic fecal contamination: evidence from *E. coli* measured in
616 water, hands, food, flies, and soil in Bangladesh. *Environ. Sci. Technol.* 51, 8725–8734 (2017).
- 617 32. Schmidt, W.-P. & Cairncross, S. Household water treatment in poor populations: is there enough evidence for
618 scaling up now? *Environ. Sci. Technol.* 43, 986–992 (2009).
- 619 33. Daly, S. W. & Harris, A. R. Modeling exposure to fecal contamination in drinking water due to multiple water
620 source use. *Environ. Sci. Technol.* 56, 3419–3429 (2022).
- 621 34. Hernando-Amado, S., Coque, T. M., Baquero, F. & Martínez, J. L. Defining and combating antibiotic resistance
622 from One Health and Global Health perspectives. *Nat. Microbiol.* 4, 1432–1442 (2019).
- 623 35. Walsh, T. R., Weeks, J., Livermore, D. M. & Toleman, M. A. Dissemination of NDM-1 positive bacteria in the
624 New Delhi environment and its implications for human health: an environmental point prevalence study. *Lancet*
625 *Infect. Dis.* 11, 355–362 (2011).
- 626 36. Pickering, A. J. et al. Effect of in-line drinking water chlorination at the point of collection on child diarrhoea in
627 urban Bangladesh: a double-blind, cluster-randomised controlled trial. *Lancet Glob. Health* 7, e1247–e1256 (2019).
- 628 37. Chan, E. M. et al. Poultry Ownership in Urban Kenya is Associated with Increased Fecal Contamination in
629 Household Soil. *medRxiv* 2023.09.16.23295449 (2023) doi:10.1101/2023.09.16.23295449.

- 630 38. Zambrano, L. D., Levy, K., Menezes, N. P. & Freeman, M. C. Human diarrhea infections associated with domestic
631 animal husbandry: a systematic review and meta-analysis. *Trans. R. Soc. Trop. Med. Hyg.* 108, 313–325 (2014).
- 632 39. Mosites, E. et al. Microbiome sharing between children, livestock and household surfaces in western Kenya. *PLoS*
633 *One* 12, e0171017 (2017).
- 634 40. Swarthout, J. M. et al. Addressing Fecal Contamination in Rural Kenyan Households: The Roles of Environmental
635 Interventions and Animal Ownership. *Environ. Sci. Technol.* (2024).
- 636 41. Harris, A. R. et al. Ruminants contribute fecal contamination to the urban household environment in Dhaka,
637 Bangladesh. *Environ. Sci. Technol.* 50, 4642–4649 (2016).
- 638 42. Schriewer, A. et al. Human and animal fecal contamination of community water sources, stored drinking water
639 and hands in rural India measured with validated microbial source tracking assays. *Am. J. Trop. Med. Hyg.* 93, 509
640 (2015).
- 641 43. Hassell, J. M. et al. Epidemiological connectivity between humans and animals across an urban landscape. *Proc.*
642 *Natl. Acad. Sci. U.S.A.* 120, e2218860120 (2023).
- 643 44. Odundo, F., Ngigi, A. & Magu, M. Sulfonamides and β -lactam antibiotic residues and human health risk
644 assessment in commercial chicken meat sold in Nairobi City, Kenya. *Heliyon* 9, (2023).
- 645 45. Mathenge, S. G., Wanjau, R. N. & Kenji, G. M. Trimethoprim and Sulfamethoxazole Residues in Untreated
646 Wastewater used for Irrigation in Peri-urban Farms in Nairobi County, Kenya. *Nat. Environ. Pollut. Technol.* 16,
647 989–994 (2017).
- 648 46. Omulo, S. et al. Evidence of superficial knowledge regarding antibiotics and their use: Results of two cross-
649 sectional surveys in an urban informal settlement in Kenya. *PLoS One* 12, e0185827 (2017).
- 650 47. Brown, N. G., Shanker, S., Prasad, B. V. & Palzkill, T. Structural and biochemical evidence that a TEM-1 β -
651 lactamase N170G active site mutant acts via substrate-assisted catalysis. *J. Biol. Chem.* 284, 33703–33712 (2009).
- 652 48. Khezri, A., Avershina, E. & Ahmad, R. Plasmid identification and plasmid-mediated antimicrobial gene detection
653 in Norwegian isolates. *Microorganisms* 9, 52 (2020).
- 654 49. Ellabaan, M. M., Munck, C., Porse, A., Imamovic, L. & Sommer, M. O. Forecasting the dissemination of antibiotic
655 resistance genes across bacterial genomes. *Nat. Commun.* 12, 2435 (2021).
- 656 50. Peter, S. et al. Tracking of antibiotic resistance transfer and rapid plasmid evolution in a hospital setting by
657 Nanopore sequencing. *mSphere* 5, e00525-20 (2020).

- 658 51. Brown, C. L. et al. Critical evaluation of short, long, and hybrid assembly for contextual analysis of antibiotic
659 resistance genes in complex environmental metagenomes. *Sci. Rep.* 11, 3753 (2021).
- 660 52. Redondo-Salvo, S. et al. Pathways for horizontal gene transfer in bacteria revealed by a global map of their
661 plasmids. *Nat. Commun.* 11, 3602 (2020).
- 662 53. Andrews, S. FastQC: a quality control tool for high throughput sequence data. (2010).
- 663 54. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data.
664 *Bioinformatics* 30, 2114–2120 (2014).
- 665 55. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint
666 arXiv:1303.3997 (2013).
- 667 56. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069 (2014).
- 668 57. Page, A. J. et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693 (2015).
- 669 58. Page, A. J. et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb. Genom.* 2,
670 (2016).
- 671 59. Miller, M. A., Pfeiffer, W. & Schwartz, T. Creating the CIPRES Science Gateway for inference of large
672 phylogenetic trees. in 1–8 (IEEE, 2010).
- 673 60. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies.
674 *Bioinformatics* 30, 1312–1313 (2014).
- 675 61. Beghain, J., Bridier-Nahmias, A., Le Nagard, H., Denamur, E. & Clermont, O. ClermonTyping: an easy-to-use
676 and accurate in silico method for *Escherichia* genus strain phylotyping. *Microb. Genom.* 4, (2018).
- 677 62. Antipov, D., Raiko, M., Lapidus, A. & Pevzner, P. A. Plasmid detection and assembly in genomic and
678 metagenomic data sets. *Genome Res.* 29, 961–968 (2019).
- 679 63. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler.
680 *Genome Res.* 27, 824–834 (2017).
- 681 64. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC*
682 *Bioinformatics* 11, 1–11 (2010).
- 683 65. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat.*
684 *Methods* 18, 366–368 (2021).

- 685 66. Alcock, B. P. et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance
686 database. *Nucleic Acids Res.* 48, D517–D525 (2020).
- 687 67. Brown, C. L. et al. mobileOG-db: a manually curated database of protein families mediating the life cycle of
688 bacterial mobile genetic elements. *Appl. Environ. Microbiol.* 88, e00991-22 (2022).
- 689 68. Liu, B., Zheng, D., Zhou, S., Chen, L. & Yang, J. VFDB 2022: a general classification scheme for bacterial
690 virulence factors. *Nucleic Acids Res.* 50, D912–D917 (2022).
- 691 69. Pakbin, B., Brück, W. M. & Rossen, J. W. Virulence factors of enteric pathogenic *Escherichia coli*: A review. *Int.*
692 *J. Mol. Sci.* 22, 9922 (2021).
- 693 70. Boisen, N. et al. Redefining enteroaggregative *Escherichia coli* (EAEC): Genomic characterization of
694 epidemiological EAEC strains. *PLoS Negl. Trop. Dis.* 14, e0008613 (2020).
- 695 71. Denamur, E., Clermont, O., Bonacorsi, S. & Gordon, D. The population genetics of pathogenic *Escherichia coli*.
696 *Nat. Rev. Microbiol.* 19, 37–54 (2021).
- 697 72. R Core Team, R. R: A language and environment for statistical computing. (2013).
- 698
- 699