Plasmid overlap and evolution between *Enterobacterales* isolates from bloodstream infections and non-human compartments

William Matlock*[,1], Samuel Lipworth*[,1,2], Kevin K. Chau[1], Manal Abu Oun[3], Leanne Barker[1], James Kavanagh[1], Monique Andersson[2], Sarah Oakley[2], Marcus Morgan[2], Derrick W. Crook[1,2,4,5], Daniel S. Read[6], Muna Anjum[3], Liam P. Shaw[#,7]/Nicole Stoesser[#,1,2,4,5], REHAB Consortium

[*] Contributed equally as first authors

[#] Contributed equally as senior authors

[1] Nuffield Department of Medicine, University of Oxford, Oxford, UK

[2] Oxford University Hospitals NHS Foundation Trust, Oxford, UK

[3] Animal and Plant Health Agency, Addlestone, UK

[4] NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at University of Oxford in partnership with Public Health England, Oxford, UK

[5] NIHR Biomedical Research Centre, Oxford, UK

[6] Centre for Ecology and Hydrology, Wallingford, UK

[7] Department of Zoology, University of Oxford, Oxford, UK

Correspondence: William Matlock (william.matlock@ndm.ox.ac.uk) and Nicole Stoesser (nicole.stoesser@ndm.ox.ac.uk)

Keywords: plasmid, *Enterobacterales*, antimicrobial resistance (AMR), bloodstream infections (BSI), One Health, genomic epidemiology, *Escherichia coli*

Abstract

Horizontal gene transfer in *Enterobacterales* allows mobile genetic elements to move between strains, species, and genera. In particular, the movement of plasmids is a known route for rapid dissemination of antimicrobial resistance (AMR) genes. However, it is difficult to establish to what extent plasmids are shared between *Enterobacterales* causing human infections and those from non-human sources, such as livestock or the environment. While some previous studies have found only limited evidence for genetic overlap, these have often been limited in size, restricted to drug-resistant isolates, and have used fragmented genome assemblies. Here, we report a collection of geographically and temporally restricted isolates from human bloodstream infections (BSI), environmental soils, livestock (cattle, pigs, poultry, sheep), wastewater (influent, effluent), and rivers. Isolates were all collected between 2008-2018 from sampling sites <60km apart. The combined dataset contains 1,458 complete *Enterobacterales* genomes, including 3,697 circularised plasmids of which one third were unclassifiable. We find eight groups of near-identical plasmids seen in both human BSI and non-human isolates, of which two are conjugative F-type plasmids carrying AMR genes. We cluster plasmids based on alignment-free distances and find that 73/247 (30%) plasmid clusters contain plasmids from both human BSI and non-human isolates. Pangenome-style analyses of the 69 most prevalent clusters (1,832/3,697 plasmids) reveals sets of shared core genes alongside accessory gene repertoires. Core-gene phylogenies suggest an intertwined ecology where well-conserved putative plasmid 'backbones' carry diverse accessory functions, potentially linked to niche adaptation. Furthermore, we show that closely related human and non-human plasmids are frequently found across distantly related bacterial hosts. Our findings underline the importance of diverse sampling in 'One Health' approaches for AMR management.

50

51 **Introduction**

52 *Enterobacterales* are found both in human compartments (e.g., hospital patients[1–3], healthy
53 individuals[4–6], and wastewater[7,8]) and non-human compartments (e.g., livestock[9–11],
54 waterways[12,13], and soil[9,14]). In recent decades, widespread carriage of antimicrobial
55 resistance (AMR) genes has complicated the treatment of both human and livestock
56 *Enterobacterales* infections[15,16]. The dissemination of AMR genes between *Enterobacterales*
57 occurs in a 'Russian-doll'-style hierarchy of nested, mobilisable genetic structures[17]: genes
58 can not only move between bacterial hosts on mobilisable or conjugative plasmids[18,19] but
59 can also be transferred within and between plasmids and chromosomes by smaller mobile
60 genetic elements (MGEs)[20] such as insertion sequences[19,21,22]. Although this gene movement
61 takes place within plasmids, many plasmids have been shown to have a persistent structure
62 which encodes replication and transfer machinery[23,24] which acts as a stable 'backbone' for
63 diverse accessory gene repertoires, allowing plasmids to be grouped into related taxonomic
64 units[25]. AMR gene dissemination therefore has complex dynamics, similar to other
65 mobilisable genes encoding adaptive functions such as virulence traits[26,27] and heavy metal
66 resistance[28].

67

68 Many plasmids can transfer between species and are seen across different niches[29] but the
69 extent to which they are shared between human and non-human compartments remains
70 poorly understood. While previous studies have investigated this topic, they have often been
71 both limited in size given the genetic diversity in these compartments[30,31], and restricted to
72 single species[32] or phenotypes (e.g., drug-resistant isolates[33,34]). Further, fragmented genome
73 assemblies in many cases make recovering complete plasmids, and other MGEs,
74 impossible[35,36]. The conclusion of several studies has been that transmission between human

75 and non-human compartments (i.e., 'One Health'-associated dissemination) has minimal
76 relevance to AMR, but this may be driven by sampling strategies. On the other hand, there
77 are known instances where clinically relevant AMR genes are believed to have emerged from
78 non-human compartments and subsequently spread within human compartments, stressing
79 the role of inter-compartmental transfer in AMR gene dissemination (e.g., *mcr-1*[21,37], $bla_{\text{CTX-M}}$
80 alleles[38], and $bla_{\text{NDM-1}}$[39]).
81
82 To explore *Enterobacterales* plasmid diversity and sharing across compartments in a
83 geographically and temporally restricted context, we studied hybrid assemblies (see Methods)
84 of large *Enterobacterales* isolate collections from (i) human bloodstream infections (BSI;
85 Oxfordshire, UK, 2008-2018), and (ii) livestock (cattle, pigs, poultry, sheep), and waterways
86 (influent, effluent, and rivers) surrounding wastewater treatment works (WwTWs;
87 Oxfordshire, 2017).
88
89 **Results**
90 **A large sample of completely reconstructed human and non-human *Enterobacterales***
91 **isolate genomes, including plasmids**
92 We combined genomes from *Enterobacterales* isolates sampled from across Oxfordshire, UK
93 (Fig. 1a). All sampling sites were both geographically (Fig. 1b) and temporally restricted
94 (Fig. 1c): all sites <60km apart and sampled between 2008 and 2018. We included *n*=738
95 sequentially collected human BSI *Enterobacterales* genomes (nosocomial, community, and
96 other healthcare associated onset), including *n*=1,880 associated plasmids (as previously
97 described in Lipworth *et al.*, 2021[40]; see Methods). We also included *n*=558 *Enterobacterales*
98 genomes, including *n*=1,515 plasmids, sampled from livestock faeces (cattle, pig, sheep),
99 surrounding environmental soils, and wastewater influent and effluent surrounding WwTWs

100 (as previously described in Shaw *et al.*, 2021[22]; see Methods). In this study, we also
101 assembled an additional *n*=162 *Enterobacterales* genomes, including *n*=302 plasmids, from
102 poultry, as well as soils surrounding all livestock farms (see Methods). This resulted in a total
103 of *n*=1,458 chromosomes and *n*=3,697 plasmids (Table 1), with median 2 plasmids carried
104 per isolate (IQR=1-4, range=0-16). All genome assemblies included were 'complete': that is,
105 the chromosome and all plasmids were circularised. Isolates mostly represented the genera
106 *Enterobacterales* (see Methods for details of taxonomic assignment): *n*=125 *Citrobacter*
107 (*n*=122 *C. freundii*, *n*=2 *C. sedlakii*, *n*=1 *C. tructae*), *n*=1 *Cronobacter* (*n*=1 *Cronobacter*
108 sp.), *n*=63 *Enterobacter* (*n*=1 *E. cancerogenus*, *n*=62 *E. cloacae*), *n*=1,044 *Escherichia*
109 (*n*=1,044 *E. coli*), *n*=211 *Klebsiella* (*n*=4 *K. aerogenes*, *n*=1 *K. huaxiensis, n*=44 *K. oxytoca*,
110 *n*=163 *K. pneumoniae*), *n*=1 *Lelliottia* (*n*=1 *L. nimipressuralis*), *n*=6 *Proteus* (*n*=4 *P.*
111 *mirabilis*, *n*=1 *P. vulgaris*), and *n*=5 *Salmonella* (*n*=5 *S. enterica*). From the BSI isolates, we
112 also included *n*=2 *Aeromonas* spp., a non-*Enterobacterales* genus from the wider
113 *Gammaproteobacteria* class.
114

Table 1. Isolate compartment breakdown.			
Compartment	**Sample type(s)**	**No. isolates**	**No. plasmids**
BSI	Community, nosocomial, other healthcare associated infections	738	1,880
Environment	Soil surrounding livestock farms	128	190
Livestock	Cattle faeces	133	215
	Sheep faeces	113	286
	Pig faeces	104	352
	Poultry faeces	34	112
WwTW	Influent/upstream waterways	113	373
	Effluent/downstream waterways	95	289
Total		**1,458**	**3,697**

115

Table 2. Isolate genus breakdown.						
Compartment	**Isolate genus**					**Total**
	Citrobacter	***Enterobacter***	***Escherichia***	***Klebsiella***	**Other**	
BSI	6	11	547	161	13	**738**
Environment	28	10	85	5	0	**128**
Livestock	26	0	348	9	1	**384**
WwTW	65	42	64	37	0	**208**
Total	**125**	**63**	**1,044**	**212**	**14**	**1,458**

116

117 We found a significant association between sampling compartment and isolate genus
118 (Fisher's test, *p*-value<5e-4; see Table 2). *Klebsiella* isolates were disproportionately derived
119 from BSI samples: 76% (161/212) of all *Klebsiella* sequenced were from BSIs, versus 51%
120 (738/1,458) of all isolates. *Citrobacter* and *Enterobacter* were disproportionately derived
121 from WwTW samples: 52% (65/125) and 67% (42/63) of sequenced *Citrobacter* and
122 *Enterobacter* were from WwTWs samples versus 14% (208/1,458) of all isolates. We
123 produced chromosomal Mash trees (see Methods) for the two most common isolate species in
124 the dataset, *E. coli* (72% [1,044/1,458]; see Fig. S1) and *K. pneumoniae* (11% [163/1,458];
125 see Fig. S2). Both trees demonstrated intermixing of human and non-human isolates within
126 clades.
127
128 **One third of plasmids in our dataset are unclassified using conventional methods,**
129 **highlighting the extent of plasmid diversity**
130 We contextualised our plasmids within known plasmid diversity using COPLA to determine
131 each plasmid's 'plasmid taxonomic unit' (PTU; see Methods), which is designed to be
132 equivalent to a 'species' concept for plasmids[41]. Briefly, COPLA classifies query plasmids
133 based on average nucleotide identity (ANI) against a non-redundant reference plasmid
134 database where most plasmids have been assigned to a reference PTU[40]. Within our sample,
135 64% (2,369/3,697) plasmids were assigned a PTU and 4% (135/3,697) a putative PTU (i.e.,
136 the query plasmid was clustered with 3 unclassified reference plasmids). This is consistent
137 with a previous COPLA analysis of 1,000 *Enterobacterales* plasmids which found that 63%
138 were classified into a PTU [41]. The remaining 32% (1,193/3,697) of plasmids were
139 unclassified (i.e., connected set with less than 4 plasmids) highlighting the previously
140 unsampled plasmid diversity within our dataset. In total, we found *n*=67 known PTUs,
141 containing a median 9 plasmids (IQR=4-30, range=1-556), where the largest assigned PTU

142 (556/2,504) was PTU-F_E, corresponding to F-type *Escherichia* plasmids[42,43]. The proportion
143 of unclassified plasmids was higher in environmental/livestock samples (33%; 385/1,155)
144 versus BSI samples (26%; 485/1,880), emphasising the underrepresentation of non-human
145 plasmids in reference plasmid databases.
146
147 **Near-identical plasmid sharing observed between human and non-human**
148 ***Enterobacterales***
149 We identified instances where a near-identical plasmid was seen in multiple isolates by
150 grouping plasmids based on a low Mash[44] distance and highly similar lengths (d<0.0001 and
151 longer plasmid ≤1% longer than shorter plasmid; see Methods). For whole genomes, Mash
152 distance has a strong positive correlation with ANI [45]. We found n=225 such near-identical
153 groups of at least 2 members, covering 19% (712/3,697) plasmids. By bootstrapping b=1,000
154 accumulation curves (ACs) for near-identical plasmids groups and singletons as a function of
155 the number of isolates, we found a highly 'open' accumulation (Heap's parameter γ=0.97,
156 Fig. S3) suggesting further isolate sampling would detect more unique plasmids
157 approximately linearly. Restricting to the BSI or environmental/livestock isolates alone, we
158 found similar curves for both compartments (Heap's parameters: BSI γ=0.98,
159 environmental/livestock γ=0.94), suggesting they had similar levels of plasmid diversity.
160
161 Near-identical groups containing two plasmids were the most common, representing 71%
162 (159/225) of groups (group size IQR=2-3, range=2-32). Within near-identical groups,
163 plasmid length differed by a median 0% (IQR=0-0%, range=0-0.89%), consistent with our
164 approach incorporating plasmid length as part of clustering. Near-identical groups
165 represented multiple bacterial host STs (25% [56/225]), species (4% [9/225]), and genera
166 (4% [9/225]), consistent with plasmids capable of inter-lineage, -species, and -genus transfer.

167 Further, 15% (34/225) of near-identical groups contained plasmids found in multiple
168 sampling compartments, indicating inter-compartmental transfer (i.e., 'cross-compartmental
169 groups'; Fig. 2a). Within cross-compartmental groups, we found *n*=4/34 contained plasmids
170 found in multiple bacterial species (Fig. 2b), and that most cross-compartmental groups were
171 made up of conjugative plasmids (*n*=16/34 conjugative, *n*=11/34 mobilisable, *n*=7/34 non-
172 mobilisable; Fig. 2c). AMR genes were carried in plasmids in *n*=13/34 cross-compartmental
173 groups (Fig. 2d), with *n*=10/13 of these groups containing plasmids with at least one beta-
174 lactamase.
175
176 As a subset of cross-compartmental groups, we specifically investigated groups that
177 supported plasmid sharing between human (BSI) and non-human compartments (all
178 environmental/livestock niches sampled in this study). Plasmids from both human BSI and
179 environmental/livestock isolates were represented in *n*=8/34 cross-compartmental groups,
180 comprising *n*=45 plasmids. On investigation of these groups, *n*=2/8 contained non-
181 mobilisable Col-type plasmids; *n*=4/8 contained mobilisable Col-type plasmids, of which
182 *n*=1/4 carried the AMR genes *aph(3")-Ib*, *aph(6)-Id*, *dfrA14*, and *sul2* (see Methods). The
183 remaining 2/8 groups contained conjugative FIB-type plasmids. One group carried the AMR
184 genes *aph(3")-Ib*, *aph(6)-Id*, bla_{TEM-1}, *dfrA5*, *sul2*, and the other group carried the MDR
185 efflux pump protein *robA*.
186
187 We attempted to estimate how the number of non-human isolates affected conclusions about
188 plasmid sharing between human BSI and non-human compartments. We bootstrapped the
189 number of near-identical BSI plasmid matches against the number of environmental/livestock
190 isolates sampled (*b*=1,000). The rarefaction curve had an estimated Heap's parameter of
191 γ=0.95 (Fig. S4), suggesting many more such cross-compartmental plasmids would be found

192 with greater sampling. Assuming we sampled only *n*=100 environmental/livestock isolates
193 (i.e., around one fifth of our actual sample size of *n*=512 isolates), we estimated a 39%
194 (392/1000) chance that we would have detected fewer than 5 matches with BSI plasmids.
195
196 Of the plasmids within the near-identical groups, 71% (502/712) were assigned a PTU or
197 putative PTU. We used homogeneity (*h*) and completeness (*c*) to evaluate the relationship
198 between the groups and non-putative PTUs (see Methods). Briefly, a grouping has *h*=1 if all
199 groups only contain a single PTU, and *c*=1 if all instances of a PTU fall within the same
200 group[46]. Groups had a perfect completeness (*c*=1.00) but lower homogeneity (*h*=0.54),
201 indicating that our groups contained plasmids from multiple PTUs, but plasmids with the
202 same PTUs always fell within the same cluster.
203
204 **Plasmid clustering reveals a diverse population structure**
205 Beyond near-identical groups, which are the strongest suggestion of recent transfer events,
206 we also wanted to examine the wider population structure of the plasmids. To do this, we
207 agnostically clustered plasmids based on their genetic content, both because our dataset
208 contained previously unsampled diversity as seen by the PTU analysis, and because
209 reference-based classifications such as MOB and replicon typing schemes are known to be
210 incongruent[47] or unreliable: 16% (602/3,697) of our plasmids had an unidentifiable replicon
211 type, which is not uncommon[42]. The evolutionary histories of plasmids can incorporate
212 multiple gain, loss, and rearrangement events in addition to mutations[48], and as such,
213 traditional measures of genetic relatedness (e.g., single nucleotide variant [SNV] thresholds)
214 used for genomic epidemiology of whole genomes are likely less appropriate here.
215

216 First, we calculated the pairwise plasmid similarities using the Jaccard index (*JI*) of their 21-
217 mer content (ranges from 0 to 1, 1 being identical; see Methods). These similarities formed
218 the edge-weights in a plasmid-plasmid network, which was subsequently thresholded to
219 sparsify the network and allow the detection of clusters. The threshold was chosen at JI=0.5,
220 meaning that edges between plasmids with JI<0.5 were deleted from the network. From this
221 threshold onwards, both the number of connected components and the number of singletons
222 steadily increased at a similar rate (Fig S5). This regime indicates an approximately stable
223 non-singleton structure from JI=0.5 onwards.
224
225 At this threshold, we detected plasmid clusters using the Louvain algorithm, which optimises
226 the modularity of the network whilst accounting for plasmid similarities (see Methods).
227 Plasmid clusters are groups of n≥3 plasmids that have high within-cluster-similarities and low
228 between-cluster-similarities. Though non-deterministic, the Louvain algorithm showed low
229 variation in cluster distribution over 50 runs, consistent with reproducible segregation of
230 plasmids in clusters (range of clusters detected: 245-247; Fig. S6). We defined n=247
231 plasmid clusters with a median of 5 plasmid members on average (IQR=3-10, range=3-123)
232 recruiting 71% (2,627/3,697) of the plasmids. The remainder plasmids were either singletons
233 (i.e., single, unconnected plasmids; 19% [718/3,697]) or doubletons (i.e., pairs of connected
234 plasmids; 10% [352/3,697]). By bootstrapping b=1,000 ACs for plasmid clusters, doubletons,
235 and singletons found against number of isolates sampled (Fig. S7; see Methods), we
236 estimated that the rarefaction curve had a Heap's parameter γ=0.75, suggesting further isolate
237 sampling would likely detect more plasmid diversity.
238
239 Of the plasmid clusters, n=69/247 plasmid clusters had at least 10 members and covered 50%
240 (1,832/3,697) of all plasmids (Fig. 3a). We used these prevalent plasmid clusters for

241 subsequent pangenome-style analyses (see Methods). We found 30% (73/247) of clusters
242 contained both BSI and environmental/livestock plasmids, representing *n*=38 unique plasmid
243 replicon haplotypes (i.e., combinations of replication proteins) of which only 24% (9/38)
244 were Col-type plasmids, which are often well-conserved and carry few genes[42,49].
245
246 Near-identical plasmid groups were also included in the wider cluster analysis, as many were
247 cross-compartmental and found across bacterial hosts (see earlier, Fig. 2). Of the *n*=194/225
248 groups which were clustered, 100% (194/194) had all members fall within the same plasmid
249 cluster, with *n*=30/247 clusters containing multiple near-identical plasmid groups. Only 6%
250 (14/247) of plasmid clusters comprised exclusively of near-identical plasmid groups,
251 suggesting that near-identical groups of plasmids often have nearby genetically related
252 plasmids.
253
254 Overall, plasmid clusters scored high completeness (*c*) but low homogeneity (*h*) with respect
255 to biological and ecological characteristics (PTU [*c*=0.99, *h*=0.66]; replicon haplotype
256 [*c*=0.92, *h*=0.69]; bacterial host ST [*c*=0.84, *h*=0.14] in Fig. 3b; predicted mobility [*c*=0.93,
257 *h*=0.20] in Fig 3c). This indicated that characteristics varied between similar plasmids, but
258 plasmids with the same features usually fell within the same cluster. This is to be anticipated
259 as replicon haplotypes and mobilities can vary within plasmid families[47], and plasmid
260 families can have diverse host ranges[29]. Examining the PTU distribution within clusters, most
261 contained at least one unclassified plasmid (51%; 127/247) or plasmid assigned a putative
262 PTU (9%; 23/247). However, many clusters exclusively contained just one known PTU
263 (42%; 105/247).
264

265 Plasmids carrying AMR genes were found in 21% (52/247) of the plasmid clusters (i.e.,
266 'AMR-carrying clusters'), representing *n*=550 plasmids (Fig. 3d). Of the AMR-carrying
267 clusters, 92% (48/52) contained at least one beta-lactamase-carrying plasmid (*n*=437
268 plasmids in total). AMR genes were present in a median proportion 67% of AMR-carrying
269 cluster members (IQR=28-100%, range=3-100%). This highlights that AMR genes are not
270 necessarily widespread on genetically similar plasmids and can be potentially acquired
271 multiple different times through the activity of smaller MGEs, such as transposons, or
272 recombination. For example, cluster 12 was a group of *n*=42 conjugative, PTU-F_E plasmids
273 found in BSI, wastewater, and waterways. Of these, 31% (13/42) carried the AMR gene
274 bla_{TEM-1}, and in a range of genetic contexts: *n*=9/13 bla_{TEM-1} genes were found within Tn*3* and
275 *n*=4/13 were carried without a transposase, of which *n*=2/4 were found with the additional
276 AMR genes *aph(6)-Id*, *aph(3'')-Ib*, and *sul2*. AMR genes were disproportionately carried by
277 F-type plasmids (61% [337/550] AMR-carrying cluster plasmids versus 34% [891/2627] of
278 the total clustered plasmids), further underlining their known role in AMR gene
279 dissemination[43,50,51]. More broadly, bla_{TEM-1} was found in 15% (38/247) of clusters over
280 *n*=292 plasmids, suggesting a diversity of plasmid backgrounds.
281
282 **An intertwined ecology of plasmids across human and non-human compartments**
283 Plasmids can change their genetic content, particularly when subject to new selective
284 pressures[52–54]. There is evidence that the structure of some plasmids can be separated into a
285 'backbone' of conserved core genes and a 'cargo' of variable accessory genes[23,24]. It is then
286 possible that cross-compartmental plasmids which have been recently transferred (within
287 years) could carry the same backbone with minimal mutational evolution but have different
288 accessory gene content. This could be driven by any of the following: the loss of accessory

289 genes in the absence of the old niche's selective pressure, the gain of new genes as adaptation
290 to the new niche, or neutral structural gain/loss as a form of genomic drift.
291
292 To investigate the genetic structure of the plasmid clusters, we conducted a pangenome-style
293 analysis (see Methods). We only considered the *n*=69/247 plasmid clusters with at least 10
294 members. For each cluster, we determined a set of core genes (i.e., genes found in ≥95% of
295 plasmids), alongside an accessory gene repertoire (i.e., genes found in <95% of plasmids).
296 Within clusters, we found a median 9 core genes (IQR=4-53, range=0-219), and a median 9
297 accessory genes (IQR=3-145, range=0-801) (Fig. 3e). Core genes comprised a median
298 proportion 42.2% of the total pangenome sizes (IQR=20.9-66.7%). The median length of
299 plasmids within a cluster was positively correlated with number of core genes (*R*=0.85,
300 *t*=13.4, *p*-value<2.2e-16) and total pangenome size (*R*=0.87, *t*=14.6, *p*-value<2.2e-16). At an
301 individual plasmid level, core genes shared by a cluster comprised a median proportion
302 62.5% of each plasmid's gene repertoire (IQR=37.4-83.3%; Fig. 3e). Further, putatively
303 conjugative plasmids carried a significantly higher proportion of accessory genes in their
304 repertoires than mobilisable or non-mobilisable plasmids (Kruskal-Wallis test [*H*(2)=193.01,
305 *p*-value<2.2e-16] followed by Dunn's test).
306
307 Using multiple sequence alignments of the core genes within each cluster, we produced
308 maximum likelihood phylogenies (see Supp. File 1; see Methods). For this step, we only
309 considered the *n*=62/69 clusters where each plasmid had at least 1 core gene. With the
310 *n*=27/62 clusters that contained both environmental/livestock and BSI plasmids, we measured
311 the phylogenetic signal for plasmid sampling compartment using Fritz and Purvis' D[55].
312 Briefly, if *D*=1, the distribution of the trait is random with respect to phylogenetic structure
313 and if *D*=0, the distribution of the trait is conserved (*D* can also fall outside this range; see

314 Methods). We considered the binary 'trait' of human or non-human compartment and
315 estimated *D* with *n*=10,000 permutations. We found 42% (11/26) clusters had *D*>0.5 (see
316 Table 3). However, only 23% (6/26) of phylogenies were significantly different (*p*-
317 value<0.05) from the conserved null model, compared to 50% (13/26) significantly different
318 from the random null model. Overall, this indicates that the evolutionary history of plasmid
319 clusters is neither strictly segregated by sampling compartment nor completely intermixed,
320 but something intermediate.

321

Table 3. Fritz and Purvis' *D* estimates for plasmid phylogenies.			
Cluster	**Estimated *D***	**Probability estimated *D* is significantly different from 1**	**Probability estimated *D* is significantly different from 0**
2	0.2407	0.0000	0.2256
3	0.6070	0.0060	0.0031
7	0.0073	0.3994	0.3220
9	0.0794	0.0000	0.3421
10	-0.9101	0.1478	0.5357
11	-0.4134	0.0000	0.7482
17	0.5072	0.0175	0.0378
23	0.0045	0.0001	0.4881
24	0.9269	0.3759	0.0074
25	-1.9546	0.3326	0.2656
27	0.4150	0.0994	0.2113
31	-0.1561	0.0035	0.5225
33	2.7016	0.7929	0.1393
34	-1.6007	0.0000	0.9043
35	1.0622	0.5421	0.0319
37	0.9959	0.3749	0.3269
38	0.5876	0.2934	0.2432
39	0.2680	0.0498	0.2827
43	-0.2927	0.0102	0.7268
44	-4.4685	0.0000	0.7459
47	0.0631	0.0488	0.4227
48	2.6723	0.7522	0.0738
52	-0.1550	0.0205	0.5342
55	1.3161	0.5377	0.2235
58	1.4901	0.8494	0.0208
61	1.1262	0.5284	0.0481
64	-3.0852	0.2255	0.7278

322 Alongside the core gene phylogenies, we generated plasmid cluster gene repertoire heatmaps
323 (example cluster 2 in Fig. 4a-b; all clusters and heatmaps in Supp. File 1). By visualising the
324 genes in a consensus synteny order (see Methods), the putative backbone within each plasmid
325 cluster is shown alongside its accessory gene and transposase repertoire. This highlights how
326 plasmids might gain and lose accessory functions within a persistent backbone. Log-
327 transformed linear regression revealed a significant relationship between Jaccard distance of
328 accessory genes presence against core gene cophenetic distance (y=0.080log(x)+0.978,
329 R^2=0.47, F(1,52988)=4.75e4, p-value< 2.2e-16; see Fig. S8).
330
331 **Plasmid evolution between human and non-human compartments is not structured by**
332 **bacterial host**
333 We wanted to examine the congruence between the evolution of plasmids within a cluster and
334 their respective bacterial hosts. Alongside vertical inheritance, conjugative and mobilisable
335 plasmids are capable of inter-host transference, crossing between bacterial lineages, species,
336 up to phyla[29]. Bacterial surveillance often tracks clonally evolving lineages[56], including
337 clinically relevant sequence types such as *E. coli* ST131[57,58] and ST10[59,60], but this might not
338 account for clinically relevant AMR genes moving on plasmids. Phylogenetic analysis can
339 determine whether plasmid evolution between human and non-human compartments is driven
340 by host clonal expansion or by other means of spread.
341
342 We chose to inspect plasmid cluster 2, since it was the largest cluster (n=100 members) that
343 contained both human and non-human plasmids. Within cluster 2, all plasmids carried at least
344 one F-type replicon and were all putatively conjugative, with 75% (75/100) assigned PTU-F_E
345 and 25% (25/100) assigned a putative PTU. Further, 48% (48/100) plasmids carried bla_{TEM-1},
346 and 51% (51/100) carried more than one AMR gene, so were of clinical significance. All host

347 chromosomes were *E. coli* except OX-BSI-481_2, which was sequenced from a *S. enterica*
348 ST 2998 isolate and omitted from the phylogenetic analysis. The n=99 E. coli isolates all fell
349 into one of six phylogroups: A (5/99), B1 (18/99), B2 (52/99), C (14/99), D (7/99), and G
350 (3/99; see Methods).
351
352 Figure 4b-c shows the plasmid core gene phylogeny (T_{plasmid}) and the *E. coli* host core gene
353 phylogeny ($T_{\text{chromosome}}$). The *E. coli* phylogeny was structured by six clades corresponding to
354 the six phylogroups (all had 100% support; see Methods). We found low congruence between
355 the plasmid core-gene phylogeny and the chromosomal core-gene phylogeny as seen in the
356 central 'tanglegram' (i.e., lines connecting pairs of plasmid and chromosome tips from the
357 same isolate). Additionally, we calculated a Robinson-Foulds distance $RF(T_{\text{plasmid}},$
358 $T_{\text{chromosome}})$=162, pointing to a high number of structural differences between the phylogenies
359 (see Methods). Colouring of the tip labels by sampling compartment demonstrated some
360 evidence of structuring by compartment (D=0.24; see earlier).
361
362 Within the plasmid phylogeny, we detected a clade of n=44 plasmids (support 100%; circled
363 in grey in Figure 4b) containing both human and non-human plasmids, which were within
364 median 4 core gene SNPs of each other (IQR=2-8, range=0-59). Placing plasmid evolution at
365 a rate of one SNP per year (see Methods) we estimated the median time of divergence within
366 the clade at 4 years prior to sampling, consistent with recent transference between human and
367 non-human compartments. This plasmid clade was mainly present in phylogroup B2 (20/44),
368 but also A (3/44), B1 (9/44), C (8/44), and D (4/44), pointing to recent plasmid movement
369 between bacterial lineages. Further, 77% (34/44) of plasmids within the clade carried bla_{TEM-1}
370 (BSI: 25/34, Environmental: 4/34, Livestock: 4/34, WwTW: 1/34), and 82% (36/44) carried
371 more than one AMR gene, pointing to potential clinical significance.

372

373 To examine the evolution of entire plasmid sequences within the clade, we represented all
374 *n*=44 plasmids as a 'pangraph'[61] (Figure 4d; see Methods). Briefly, pangraph converts input
375 sequences into a consensus graph, where each sequence is a path along a set of homologous
376 sequence alignments i.e., 'blocks', which in series form 'pancontigs'. First, we removed
377 duplicated blocks from the pangraph. Then, filtering for 'core blocks' (i.e., those found in at
378 least 95% plasmids), we found 4 pancontigs (40 blocks in total), with the longest at 98,269bp
379 (total length 125,369bp), indicating a putative plasmid backbone (Fig. 4e). Further, filtering
380 for 'accessory blocks' (i.e., those found in less than 95% plasmids), we found 18 pancontigs
381 (39 blocks in total), with median length 2,380bp (total length 63,753bp), forming the
382 accessory gene repertoire (Fig. 4f). This points to a persistent plasmid backbone structure
383 with loss/gain events at particular 'hotspots' as well as rearrangements.

384

385 **Discussion**

386 We have presented a dataset of *Enterobacterales* plasmids sampled from a geographically
387 and temporally restricted context from both human and non-human compartments. By
388 analysing this plasmid diversity, our analysis demonstrates how plasmid ecology is
389 intertwined across compartments, including instances of plasmids carrying genes of clinical
390 relevance such as AMR genes. Our results suggest the need for broader sampling frames to
391 fully understand plasmid evolution.

392

393 By segregating plasmids into near-identical groups, we detected 8 instances of plasmid
394 groups likely very recently shared between human and non-human compartments, 3 of which
395 were MDR plasmids. At first glance, this may seem small. However, accumulation curve
396 analysis suggested that increasing the size of the dataset would have led to further BSI near-

397 identical matches at an approximately linear rate. This suggests even a dataset of this size
398 captures only a small fraction of the true extent of plasmid sharing between human and non-
399 human compartments. Furthermore, plasmids observed in BSI isolates represent a restricted
400 and small proportion of human *Enterobacterales* diversity, meaning many more sharing
401 events may occur in the human gut[62,63] which was only sampled incompletely here using
402 wastewater influent as a proxy. The human colon contains around 10^{14} bacteria[64], with large
403 ranges of *Enterobacteriaceae* abundance (in one study, 10^{2}-10^{9} CFU per gram of stool[65]).
404 Detailed study of within-niche diversity suggest that these populations can be highly diverse
405 and fluctuate over short periods of time[66]. Even if the proportion of plasmids which are
406 shared appears small, it may still be epidemiologically important particularly given known
407 examples of clinically relevant AMR genes emerging from non-human compartments into
408 human compartments. Future studies need to carefully consider the limitations of sampling
409 frames in detecting any genetic overlap, given both substantial diversity and the effects of
410 niche and geography[22,31].
411
412 To capture broader evidence of ecological overlap of plasmids between compartments, we
413 grouped plasmids based on their *k*-mer content into plasmid clusters, finding that just under
414 one third (73/247) of these plasmid clusters were present in both human and non-human
415 compartments. Over one fifth (52/247) of plasmid clusters contained plasmids carrying AMR
416 genes, totalling 550 plasmids. Most (48/52) of these 'AMR-carrying clusters' contained at
417 least one beta-lactamase producing plasmid. Further, within AMR-carrying clusters, AMR
418 genes were carried by a median proportion of two thirds (67%) of member plasmids. This
419 underlines that plasmid surveillance should focus on both AMR carrying-plasmids and their
420 close genetic relatives, in part due to transposon-mediated spread of AMR genes.
421

422 Using a pangenome-style analysis on plasmid clusters, we showed that plasmids within a
423 cluster can share sets of near-identical core genes alongside diverse accessory gene
424 repertoires. Building phylogenies from alignments of core genes, we demonstrated that
425 plasmids with more distantly related core genes tended to have dissimilar accessory gene
426 content. In contrast, plasmids with more closely related core genes can have a wide range of
427 accessory gene content. This suggests how persistent 'backbone' structures might gain and
428 lose accessory functions during recent evolutionary history (on timescales of years), across
429 human and non-human compartments. This analysis should motivate the need for
430 evolutionary models for plasmids which can accommodate well-conserved backbone
431 evolution alongside accessory structural changes and gain/loss events. However, estimating
432 appropriate rates is a challenge. We assumed that plasmids with a few SNPs in their core
433 'backbones' were likely to have common ancestors on timescales of years, but little is known
434 about appropriate values for mutation rates in plasmids and even less for non-mutational
435 processes such as gene gain and loss. By examining plasmid relatedness compared to that of
436 their bacterial hosts, we also demonstrated that cross-compartmental plasmid spread is not
437 driven by clonal lineages.
438
439 Our study had several limitations. Our non-human isolates were not as temporally varied as
440 the BSI isolates, meaning we could not fully explore linked temporal evolution. Isolate-based
441 methodologies are incomplete and biased approaches to evaluating the true diversity of these
442 compartments; complete or targeted metagenomic approaches would be appropriate for future
443 studies. Further, the 'origin' location of an isolate is poorly defined for wastewater and river
444 isolates. WwTWs agglomerate multiple sources together, so they were not included when
445 evaluating the similarities between BSI and environmental/livestock plasmids. We only
446 analysed plasmids from complete genomes i.e., where the chromosome and all plasmids were

447 circularised, meaning we disregarded proportions of the BSI dataset, REHAB dataset, and
448 new poultry/soil assemblies (approximately 23%, 33%, and 34% of isolates, respectively).
449 We have focused only on plasmids as horizontally transmissible elements here; detailed study
450 of other smaller mobile genetic elements across-compartments would represent interesting
451 future work. We have also investigated *Enterobacterales*: however, plasmid sharing is also
452 known to occur between *Enterobacterales* and other bacterial families and would have been
453 missed by this study. Lastly, our isolate culture methods for REHAB may not have been as
454 sensitive for the identification of *Klebsiella* spp. as for other *Enterobacterales* such as
455 *Escherichia*, as we used chromogenic agars designed to identify multiple *Enterobacterales*
456 genera and did not use enrichment and selective culture on Simmons citrate agar with
457 inositol[67].
458
459 In conclusion, this study presents to our knowledge the first large-scale geographically and
460 temporally restricted analysis of *Enterobacterales* plasmid overlap between human and non-
461 human compartments. Although we found few instances of near-identical plasmids present
462 across these compartments, we do not believe this should be interpreted as conclusive
463 evidence that human and non-human compartments have distinct and separate plasmid
464 evolution. Considering the broader evolutionary dynamics of plasmids, the epidemic spread
465 of a fixed and unchanging plasmid – whether vertically within a clonal host or horizontally
466 between hosts – is only one mechanism by which AMR genes can spread. Indeed, by
467 clustering plasmids, we have shown instances of plasmids between compartments which are
468 highly related in their core 'backbone', suggestive of recent shared ancestry, but diverse in
469 their accessory gene content including AMR genes. This would be consistent with an
470 intertwined ecology of plasmids across human and non-human compartments, where putative
471 'backbone' plasmid structures can rapidly gain and lose accessory genes following cross-

472 compartmental spread. Without broad sampling strategies, future One Health studies may
473 remain unable to explore this aspect of plasmid evolution – an aspect which involves not
474 merely the 'numerator' of AMR-carrying plasmids but also a much larger 'denominator' of
475 related plasmid diversity.
476
477 **Methods**
478 **BSI isolates**
479 Human BSI *Enterobacterales* isolates (nosocomial, community, and other healthcare
480 associated onset) were sampled from patients presenting to *n*=4 hospitals within Oxfordshire,
481 UK, from September 2008 to December 2018 (as previously described in Lipworth *et al.*,
482 2021[40]). Although all patients were sampled in Oxfordshire, a total of *n*=505/738 patients
483 originated from Oxfordshire postcodes, *n*=133/738 from surrounding counties, and
484 *n*=100/738 had location information omitted. This study aimed to capture overall plasmid
485 diversity as well as locally important MDR outbreak plasmids; therefore, assemblies which
486 were not complete (*n*=215/953 total assembled) i.e., chromosome and all plasmids
487 circularised, were omitted.
488 **REHAB Isolates**
489 *Enterobacterales* isolates from the REHAB project sampled faeces and nearby environmental
490 soils from *n*=14 farms (*n*=5 cattle, *n*=4 pig, *n*=5 sheep), and wastewater influent and effluent
491 surrounding *n*=5 WwTWs, across 3 seasonal timepoints in 2017 (as previously described in
492 Shaw *et al.*, 2021[22]). Assemblies which were not complete (*n*=269/827 total assembled) i.e.,
493 chromosome and all plasmids circularised, were omitted.
494 **Additional poultry and environmental soil isolates**
495 An additional *n*=247 environmental soil and poultry *Enterobacterales* isolates were
496 assembled for this study, originally collected for the REHAB project[22] from within and

497 nearby *n*=5 poultry farms. DNA extraction and sequencing was performed as in Shaw *et al.*,
498 2021[22]). From these, *n*=162/247 complete assemblies were reconstructed (plasmids and
499 chromosomes). Genomes were hybrid assemblies, using Illumina 150bp paired-end short
500 reads and Oxford Nanopore/PacBio long reads, and were recovered using Unicycler[68] (v.
501 0.4.9; default hybrid assembly parameters except min_component_size 500 and --
502 min_dead_end_size 500). Briefly, Unicycler produces a short read assembly graph using
503 SPAdes[69], then bridges the gaps with miniasm[70] and Racon[71]. Finally, assemblies are
504 polished with Pilon[72] using alignments of the short reads.

505 **Taxonomic assignment**

506 Chromosome sequence types (STs) were determined with mlst[73] (v. 2.19.0), which makes use
507 of the PubMLST database[74]. For the *n*=11/1,458 chromosomes which could not be typed with
508 mlst, species were determined with the PubMLST 'species ID' web-tool[75], for which all had a
509 support=100, except for *L. nimipressuralis* which had support=83. Of these, *n*=5/11 were
510 from BSI, *n*=4/11 from livestock, and *n*=2/11 from effluent/downstream of WwTWs. Of the
511 plasmids found in these isolates, only OX-BSI-707_4, from a *Proteus mirabilis* isolate, was
512 clustered in cluster 6, and OX-WTW-105_2, from a *Klebsiella huaxiensis* isolate, was
513 clustered in cluster 115.

514 **Chromosome trees**

515 Trees for *E. coli* and *K. pnemoniae* chromosomes were produced using Mashtree[76] on
516 'accurate' mode (--mindepth 0 --numcpus 12).

517 **PTU classification**

518 Plasmids were assigned a plasmid taxonomic unit (PTU) using COPLA[41] (default parameters
519 except -t circular, -k Bacteria, -p Pseudomonadota, -c Gammaproteobacteria, and
520 -o Enterobacterales), based on the classification system developed in Redondo-Salvo *et al.*,
521 2021[29]. COPLA compares query plasmids to a database of PTU reference plasmids, assigning

522 a PTU when both (i) the ANI>0.7 along 50% of the length of the smallest plasmid in the
523 comparison, and (ii) a graph-neighbouring condition to existing PTU clusters is satisfied. The
524 COPLA reference database contains over 10,000 curated, non-redundant plasmids retrieved
525 from the 84th NCBI RefSeq database in 2017[77].

526 **Plasmid annotation**

527 All plasmids were annotated with Prokka[7880] (v. 1.14.5) with default parameters. For replicon
528 typing, Abricate[79] (v. 1.0.0) was used with the PlasmidFinder[80], ISfinder[81], and BacMet[82]
529 databases with default parameters and output filtered for 80% minimum coverage. For
530 annotating AMR genes, NCBI Antimicrobial Resistance Gene Finder (AMRFinderPlus)[83] (v.
531 3.10.18) was used with default parameters. For putative plasmid mobilities, we used MOB-
532 typer from MOB-suite[84] (v. 3.03) with default parameters, which predicts mobility based on
533 annotations of MOB-typer predicts mobility based on of annotations of relaxase (*mob*),
534 mating pair formation (MPF) complex, and *oriT* genes. Briefly, a plasmid is putatively
535 labelled conjugative if it has both relaxase and MPF, mobilisable if it has either relaxase
536 or *oriT* but no MPF, and non-mobilisable if it has no relaxase and *oriT*.

537 **Near-identical plasmid screening**

538 Groups of near-identical plasmids were detected as connected components in a plasmid-
539 plasmid network with Mash distance[44] (v. 2.3; default parameters except sketch size -s
540 1000000) weighted edges, at a threshold d<0.0001. Briefly, Mash distance estimates an
541 evolutionary distance on a reduced-length MinHash sketch of the sequences. Since Mash is a
542 probabilistic estimate of evolutionary distance, we confirmed the probability of seeing any of
543 our pairwise Mash distances in the near-identical groups by chance was 0. We required the
544 shortest plasmid to be within 1% length (bp) of the longest plasmid, to account for assembly
545 errors. Network analysis was performed using the igraph[85] library (v. 1.2.7) in R.

546 **Accumulation and rarefaction curves**

547 To generate an accumulation curve, isolates were sampled without replacement in a random
548 order. For each isolate, the new plasmid diversity was recorded. For Fig. S3, we recorded the
549 number of new near-identical plasmid groups and singletons. For Fig. S4, we recorded the
550 number of near-identical matches with BSI plasmids from only environmental/livestock
551 isolates. For Fig. S7, we recorded the number of new clusters, doubletons, and singletons. A
552 bootstrapped average of b=1,000 accumulation curves was plotted for the rarefaction curve.
553 The bootstraps were also used to estimate Heap's parameter (γ) by fitting a linear regression
554 to log-log transformed data. For $\gamma<0$, it is possible to sample the entire diversity, and for
555 $1>\gamma>0$, the diversity will increase with every additional sample[86].

556 **Plasmid similarity**

557 Plasmid Jaccard index (*JI*) was calculated using Mash[44] (v. X; default parameters except
558 sketch size -s 1000000). The Jaccard index (*JI*), given by

559
$$JI(A,B) = \frac{|A \cap B|}{|A \cap B|}$$

560 where A, B are the sets of k-mers of plasmids a, b, respectively. This measures extent of k-
561 mer sharing between plasmids, range=0-1, where 1 indicates an identical k-mer repertoire.
562 Since the sketch size was larger than the plasmid lengths (except for one plasmid in the
563 dataset, OX-ENV-67_2, which was larger than 1Mbp at 1,310,597bp and was not clustered;
564 the next smallest was OX-WTW-80_2 at 394,284bp), the calculated Jaccard indices were
565 almost always exact.

566 **Plasmid network and clustering**

567 The determination of the plasmid-plasmid network, threshold, and clusters could be achieved
568 with several alternative methodologies. Plasmid networks have previously been constructed
569 by full sequence alignments[87], annotated genes[88], and alignment-free Mash distances[43,89,90].
570 We chose to use the Jaccard index of entire plasmid *21*-mer distributions to capture coding
571 sequences, their immediate contexts[91,92], and intergenic regions[93,94], all of which have known

importance to bacterial evolution. Network thresholding to some extent depends subjectively on the dataset, with trade-offs between successfully revealing the underlying structure of plasmid relationships without excessively separating relatives. We chose a data-driven threshold as adopted by Ledda *et al.*, 2018[88] for their plasmid network, which examined the evolution of connected components within the network. This ensured the threshold was chosen where the regime of connected component evolution approximately stabilises, minimising excessive network breakup. Plasmid clusters were then detected with the Louvain algorithm which optimises the network modularity by iterative expectation-maximisation[95]. This aims to maximise the density of edges within clusters against edges between clusters. The algorithm was implemented using the python-louvain (v. 0.16) Python module. Although the algorithm is non-deterministic, multiple runs demonstrated minimal variation at our chosen network threshold. Overall, these approaches add to the growing literature describing suitable methodologies for clustering plasmids.

Cluster homogeneity and completeness

Homogeneity (h) and completeness (c) are dual conditional entropy-based measures, independent of cluster and metadata label distributions[46]. A clustering satisfies homogeneity (h =1) if all cluster members have the same metadata label-type. Consider a network with N nodes, partitioned by a set of metadata labels, $M = \{m_i | i = 1, \dots, n\}$, and a set of communities, $C = \{c_j | j = 1, \dots, m\}$. Let $A = \{a_{ij}\}$ represent the ij^{th} entry in the contingency table of partitions. Hence, a_{ij} counts the number of nodes with label m_i in community c_j. We then say

$$\mathrm{h} = \begin{cases} 1 & \text{if } H(M,C) = 0 \\ 1 - \dfrac{H(M \mid C)}{H(M)} & \text{else} \end{cases}$$

where

$$H(M\,|\,C) = -\sum_{c=1}^{|C|}\sum_{m=1}^{|M|}\frac{a_{mc}}{N}\log\frac{a_{mc}}{\sum_{c=1}^{|M|}a_{mc}}$$

and

$$H(M) = -\sum_{m=1}^{|M|}\frac{\sum_{c=1}^{|C|}a_{mc}}{n}\log\frac{\sum_{c=1}^{|C|}a_{mc}}{n}$$

are the conditional entropy of the metadata given the clusters and the entropy of the clusters, respectively $H(M|C) = 0$ when the cluster partition coincides with the metadata partition, and no new information is added. A cluster partition satisfies completeness (c =1) if all instances of a metadata label-type are assigned the same cluster. Completeness is defined dually by

$$c = \begin{cases} 1 & \text{if } H(C,M) = 0 \\ 1 - \dfrac{H(C\,|\,M)}{H(C)} & \text{else} \end{cases}$$

The measures were calculated using the clver library (v. 0.1.1) in R.

Cluster pangenome analysis

Cluster pangenomes were generated using Panaroo[96] (v. 1.2.9) with parameters default except --clean-mode sensitive, --aligner mafft, -a core, and --core_threshold 0.95. For core gene alignments, the threshold was set at minimum 95% presence amongst clustered plasmids, whereby they were aligned using MAFFT[97] (v. 7.407) with default parameters. An identical approach was taken for the host chromosome phylogeny in Figure 4.

Plasmid core gene phylogenies

Maximum likelihood core-gene phylogenies were generated using IQ-Tree[98] (v. 2.0.6) with parameters -m GTR+F+I+G4 -keep-ident -T 2 -B 1000. The substitution model used was general time reversible (GTR) using empirical base frequencies form the alignment (F), allowing for invariable sites (I) and variable rates of substitution (G4). We used n=1000 ultrafast bootstraps (B 1000; see Minh *et al.*, 2013[99]) to visually inspect larger clades for support. Briefly, 95% support approximates a 95% probability that the clade is genuine. Only

618 the n=62/69 clusters (excluding 6, 8, 26,2 9,3 2, 40, and 65) where every plasmid carried at
619 least 1 core gene were analysed. Phylogenies were primarily plotted using the R library
620 ggtree[100].

621 **Fritz and Purvis' *D***

622 Fritz and Purvis' D measures phylogenetic signal for binary traits[55]. First, we calculate the
623 character state changes required to observe our phylogeny (d_{obs}). To account for phylogeny
624 size and prevalence, d_{obs} is standardised under the two null models (i) tip labels are random
625 permuted (d_r), and (ii) tip labels are distributed under the expectation of a Brownian motion
626 model of evolution (d_b). Then, we define

627 $$D = (d_{obs} - \overline{d_b})/(\overline{d_r} - \overline{d_b}).$$

628 Hence, for $D \approx 1$, d_{obs} follows d_r more closely, and for $D \approx 0$, d_{obs} follows d_b more closely. We
629 calculated d_{obs} n=10,000 times and averaged the result, as well as calculate p-values for
630 significant deviation from d_r or d_b. D was implemented using the R library caper[101]. Fritz and
631 Purvis' D is normally used for cross-species analysis so is not benchmarked for plasmids.
632 Results for phylogenies with less than 25 tips should be viewed more conservatively due to
633 reduced statistical power in these instances.

634 **Consensus gene synteny heatmaps**

635 For each cluster, we first generated a list of every possible pair of genes in the pangenome.
636 Then for each plasmid, we counted the distance between these pairs, modulo the number of
637 genes in the plasmid. If a gene was absent in a plasmid, NA was used. We then calculated the
638 median of these values across all plasmids in the cluster. We then built a dendrogram from a
639 hierarchal clustering of the median distances. The order of the tip labels in the dendrogram
640 were then used as the 'consensus gene synteny'.

641 **Accessory gene distances**

642 Plasmid accessory gene distances were calculated using pairwise Jaccard distances on gene
643 presence-absences matrices. For plotting the cluster-wise plasmid core gene cophenetic
644 distance against accessory gene presence-absence Jaccard distance, only the n=26/62 clusters
645 with at least 50 accessory genes were plotted.

646 **Chromosome core gene phylogeny**

647 An identical approach was taken to the plasmid phylogenies. *E. coli* phylogroups were typed
648 using EzClermont[102] (v. 0.7.0) with default parameters. Robinson-Foulds distance was
649 calculated using the R library phangorn[103].

650 **Plasmid mutation rate**

651 Mutation rates per base pair in microbes typically arise from DNA replication and tend to be
652 below $m=10^{-9}$ per site per generation[104] or perhaps as low as 10^{-10} per site per
653 generation[105,106]. For a plasmid of size L, one therefore expects L x m mutations per plasmid
654 per generation. For example, if the plasmid has $L=10^5$ then in each generation 1 in 10,000
655 plasmids will gain a mutation. The generation time of *E. coli* per day in the human gut has
656 been estimated to be between 6 and 20 generations per day[107]. For large plasmids that exist at
657 a copy number of ~1, the plasmid generation time is the cell generation time. More generally,
658 for a plasmid copy number p the number of replications of the plasmid expected for a given
659 number of cell generations g will be p x g (assuming that plasmid copies are simply and
660 linearly related to the realised number of replications per cell). A crude estimate for the
661 expected mutation rate per time period for a plasmid is therefore given by L x m x p x g. For a
662 plasmid of L=100 kbp and p=1, assuming m=[0.1-1]x10^{-9} per site per generation and g=[6-
663 20]x365 per year, one would expect it to accumulate ~0.5 mutations a year (between ~0.02-
664 0.7 depending on assumptions). One obtains the same result for L=10 kbp and p=10. There is
665 a strong inverse correlation between plasmid size and copy number. This suggests that a
666 suitable upper bound for the expected number of mutations for a typical plasmid per year

667 (under neutral evolution) is of the order of magnitude of 1 SNP a year. This rough 'SNPs and
668 years' rule-of-thumb appears consistent with known empirical results. For example: 100kbp
669 I1-type *Shigella* plasmids isolated between 2007-2010 in Vietnam were separated by at most
670 2 SNPs[108]; 30kbp X4-type plasmids carrying *mcr-1* isolated between 2016-2018 in China
671 were separated by most 4 SNPs[34] (analysis not shown).

672 **Pangraph analysis**

673 We used pangraph (v. 0.5.0) to build a pangraph of the clade within plasmid cluster 2, using
674 the --circular flag and otherwise default parameters. We used pangraph export (--edge-
675 minimum length 0, default parameters) to export the graph to GFA format and then visualised
676 this using Bandage[109].

677 **Data visualisation**

678 Plots were primally produced using the R library ggplot2[110], with additional graphics in
679 BioRender[111].

680 **Data availability**

681 Study metadata is provided in Table S1. Accessions for poultry and environmental soil isolate
682 reads are given in Table S2, and assemblies will shortly be made available on NCBI.
683 Accessions for existing BSI and REHAB reads and assemblies can be found in Lipworth *et*
684 *al.*, 2021[40] and Shaw *et al.*, 2021[22], respectively.

685 **Code availability**

686 Analysis scripts can be found in the GitHub repository
687 https://github.com/wtmatlock/oxfordshire-overlap.

688 **REHAB Consortium.**

689 Manal AbuOun[2], Muna F. Anjum[2], Mark J. Bailey[3], Brett H[8], Mike J. Bowes[3], Kevin K.
690 Chau[1], Derrick W. Crook[1,6,7], Nicola de Maio[1], Nicholas Duggett[2], Daniel J. Wilson[1,9],
691 Daniel Gilson[2], H. Soon Gweon[3,4], Alasdair Hubbard[10], Sarah J. Hoosdally[1], William

692 Matlock[1], James Kavanagh[1], Hannah Jones[2], Timothy E. A. Peto[1,6,7], Daniel S. Read[3], Robert
693 Sebra[5], Liam P. Shaw[1], Anna E. Sheppard[1,6], Richard P. Smith[2], Emma Stubberfield[2], Nicole
694 Stoesser[1,6,7], Jeremy Swann[1], A. Sarah Walker[1,6,7], Neil Woodford[11]
695 [1] Nuffield Department of Medicine, University of Oxford, Oxford, UK
696 [2] Animal and Plant Health Agency, Weybridge, Addlestone, UK
697 [3] UK Centre for Ecology & Hydrology, Wallingford, UK
698 [4] University of Reading, Reading, UK
699 [5] Icahn Institute of Data Science and Genomic Technology, Mt Sinai, NY, USA
700 [6] NIHR HPRU in healthcare-associated infection and antimicrobial resistance, University of
701 Oxford, Oxford, UK
702 [7] NIHR Oxford Biomedical Research Centre, University of Oxford, Oxford, UK
703 [8]Thames Water Utilities, Clearwater Court, Vastern Road, Reading, UK
704 [9]Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive,
705 Oxford, UK
706 [10]Department of Tropical Disease Biology, Liverpool School of Tropical Medicine,
707 Liverpool, UK
708 [11]Antimicrobial Resistance and Healthcare Associated Infections (AMRHAI) Reference Unit,
709 National Infection Service, Public Health England, London, United Kingdom
710 **Competing interest declarations**. The authors declare no competing interests.
711 **Funding**
712 This work was funded by the Antimicrobial Resistance Cross-council Initiative supported by
713 the seven research councils [grant NE/N019989/1]. The UKCEH component of the REHAB
714 consortium was supported by the Natural Environment Research Council (NERC)
715 [grant NE/N019660/1]. DWC, SG, TEAP and NS are supported by the National Institute for
716 Health Research Health Protection Research Unit (NIHR HPRU) in Healthcare-Associated

717 Infections and Antimicrobial Resistance at the University of Oxford in partnership with
718 Public Health England (PHE) [grant HPRU-2012–10041 and NIHR200915]. DWC and
719 TEAP are also supported by the NIHR Oxford Biomedical Research Centre. The
720 computational aspects of this research were funded from the NIHR Oxford BRC with
721 additional support from a Wellcome Trust Core Award Grant [grant 203141/Z/16/Z]. The
722 views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the
723 Department of Health or Public Health England. WM and KKC are supported by a
724 scholarship from the Medical Research Foundation National PhD Training Programme in
725 Antimicrobial Resistance Research (MRF-145-0004-TPG-AVISO). NS is an Oxford Martin
726 Fellow and a Senior NIHR BRC Oxford Fellow. LPS is a Sir Henry Wellcome Postdoctoral
727 Fellow funded by Wellcome (grant 220422/Z/20/Z).
728

729 **Bibliography**

730 1. Linh, T. D. *et al.* Expansion of KPC-producing Enterobacterales in four large hospitals
731 in Hanoi, Vietnam. *Journal of Global Antimicrobial Resistance* **27**, (2021).
732 2. Rivera-izquierdo, M. *et al.* Oxa-48 carbapenemase-producing enterobacterales in
733 spanish hospitals: An updated comprehensive review on a rising antimicrobial
734 resistance. *Antibiotics* vol. 10 (2021).
735 3. Kraftova, L. *et al.* Evidence of an epidemic spread of KPC-producing Enterobacterales
736 in Czech hospitals. *Scientific Reports* **11**, (2021).
737 4. Onduru, O. G., Mkakosya, R. S., Rumisha, S. F. & Aboud, S. Carriage prevalence of
738 extended-spectrum β-lactamase producing enterobacterales in outpatients attending
739 community health centers in Blantyre, Malawi. *Tropical Medicine and Infectious*
740 *Disease* **6**, (2021).
741 5. Fröding, I., Valik, J. K., Bolinder, L., Nauclér, P. & Giske, C. G. Prediction of
742 bloodstream infection caused by extended-spectrum β-lactamase-producing
743 Enterobacterales in patients with suspected community-onset sepsis. *International*
744 *Journal of Antimicrobial Agents* **53**, (2019).
745 6. Weinstein, E. J. *et al.* A clinical prediction tool for extended-spectrum cephalosporin
746 resistance in community-onset enterobacterales urinary tract infection. *Open Forum*
747 *Infectious Diseases* **6**, (2019).
748 7. Asaduzzaman, M. Environmental transmission of antibiotic resistant bacteria through
749 waste water run-off in Bangladesh. *International Journal of Infectious Diseases* **101**,
750 82 (2020).
751 8. Cahill, N. *et al.* Hospital effluent: A reservoir for carbapenemase-producing
752 Enterobacterales? *Science of the Total Environment* **672**, (2019).

753 9. Subramanya, S. H. *et al.* Detection and characterization of ESBL-producing
754 Enterobacteriaceae from the gut of subsistence farmers, their livestock, and
755 the surrounding environment in rural Nepal. *Scientific Reports* **11**, (2021).
756 10. Abuoun, M. *et al.* A genomic epidemiological study shows that prevalence of
757 antimicrobial resistance in enterobacterales is associated with the livestock host, as
758 well as antimicrobial usage. *Microbial Genomics* **7**, (2021).
759 11. Fournier, C., Aires-de-Sousa, M., Nordmann, P. & Poirel, L. Occurrence of CTX-M-
760 15- and MCR-1-producing Enterobacterales in pigs in Portugal: Evidence of direct
761 links with antibiotic selective pressure. *International Journal of Antimicrobial Agents*
762 **55**, (2020).
763 12. Lazăr, V. *et al.* Antibiotic resistance profiles in cultivable microbiota isolated from
764 some romanian natural fishery lakes included in Natura 2000 network. *BMC*
765 *Veterinary Research* **17**, (2021).
766 13. Díaz-Gavidia, C. *et al.* Isolation of Ciprofloxacin and Ceftazidime-Resistant
767 Enterobacterales From Vegetables and River Water Is Strongly Associated With the
768 Season and the Sample Type. *Frontiers in Microbiology* **12**, (2021).
769 14. Koutsoumanis, K. *et al.* Role played by the environment in the emergence and spread
770 of antimicrobial resistance (AMR) through the food chain. *EFSA Journal* **19**, (2021).
771 15. Buchy, P. *et al.* Impact of vaccines on antimicrobial resistance. *International Journal*
772 *of Infectious Diseases* vol. 90 (2020).
773 16. Ruppé, E. *et al.* From genotype to antibiotic susceptibility phenotype in the order
774 Enterobacterales: a clinical perspective. *Clinical Microbiology and Infection* **26**,
775 (2020).

776 17. Sheppard, A. E. *et al.* Nested Russian doll-like genetic mobility drives rapid
777 dissemination of the carbapenem resistance gene blakpc. *Antimicrobial Agents and*
778 *Chemotherapy* **60**, (2016).
779 18. Madec, J. Y. & Haenni, M. Antimicrobial resistance plasmid reservoir in food and
780 food-producing animals. *Plasmid* vol. 99 (2018).
781 19. Che, Y. *et al.* Conjugative plasmids interact with insertion sequences to shape the
782 horizontal transfer of antimicrobial resistance genes. *Proc Natl Acad Sci U S A* **118**,
783 (2021).
784 20. Leekitcharoenphon, P. *et al.* Genomic evolution of antimicrobial resistance in
785 Escherichia coli. *Scientific Reports* **11**, (2021).
786 21. Wang, R. *et al.* The global distribution and spread of the mobilized colistin resistance
787 gene mcr-1. *Nat Commun* **9**, 1–9 (2018).
788 22. Shaw, L. P. *et al.* Niche and local geography shape the pangenome of wastewater-and
789 livestock-associated Enterobacteriaceae. *Science Advances* **7**, (2021).
790 23. Orlek, A. *et al.* Plasmid classification in an era of whole-genome sequencing:
791 Application in studies of antibiotic resistance epidemiology. *Frontiers in Microbiology*
792 vol. 8 (2017).
793 24. Norberg, P., Bergström, M., Jethava, V., Dubhashi, D. & Hermansson, M. The IncP-1
794 plasmid backbone adapts to different host bacterial species and evolves through
795 homologous recombination. *Nature Communications* **2**, (2011).
796 25. Garcillán-Barcia, M. P., Alvarado, A. & de la Cruz, F. Identification of bacterial
797 plasmids based on mobility and plasmid population biology. *FEMS Microbiology*
798 *Reviews* vol. 35 (2011).

799 26. Mairi, A. *et al.* OXA-48-producing Enterobacterales in different ecological niches in
800 Algeria: Clonal expansion, plasmid characteristics and virulence traits. *Journal of*
801 *Antimicrobial Chemotherapy* **74**, (2019).
802 27. Amaretti, A. *et al.* Antibiotic resistance, virulence factors, phenotyping, and
803 genotyping of non–escherichia coli enterobacterales from the gut microbiota of healthy
804 subjects. *International Journal of Molecular Sciences* **21**, (2020).
805 28. Cerdeira, L. *et al.* Genomic insights of Klebsiella pneumoniae isolated from a native
806 Amazonian fish reveal wide resistome against heavy metals, disinfectants, and
807 clinically relevant antibiotics. *Genomics* **112**, (2020).
808 29. Redondo-Salvo, S. *et al.* Pathways for horizontal gene transfer in bacteria revealed by
809 a global map of their plasmids. *Nature Communications* **11**, (2020).
810 30. Mounsey, O. *et al.* Limited phylogenetic overlap between fluoroquinolone-resistant
811 Escherichia coli isolated on dairy farms and those causing bacteriuria in humans living
812 in the same geographical region. *Journal of Antimicrobial Chemotherapy* **76**, (2021).
813 31. Hanage, W. P. Two health or not two health? That is the question. *MBio* **10**, e00550-19
814 (2019).
815 32. Ludden, C. *et al.* One health genomic surveillance of escherichia coli demonstrates
816 distinct lineages and mobile genetic elements in isolates from humans versus livestock.
817 *mBio* **10**, (2019).
818 33. Mughini-Gras, L. *et al.* Attributable sources of community-acquired carriage of
819 Escherichia coli containing β-lactam antibiotic resistance genes: a population-based
820 modelling study. *The Lancet Planetary Health* **3**, (2019).
821 34. Shen, C. *et al.* Dynamics of mcr-1 prevalence and mcr-1-positive Escherichia coli after
822 the cessation of colistin use as a feed additive for animals in China: a prospective

823 cross-sectional and whole genome sequencing-based molecular epidemiological study.
824 *The Lancet Microbe* **1**, e34–e43 (2020).
825 35. Hilpert, C., Bricheux, G. & Debroas, D. Reconstruction of plasmids by shotgun
826 sequencing from environmental DNA: Which bioinformatic workflow? *Briefings in*
827 *Bioinformatics* **22**, (2021).
828 36. Antipov, D., Raiko, M., Lapidus, A. & Pevzner, P. A. Plasmid detection and assembly
829 in genomic and metagenomic data sets. *Genome Research* **29**, (2019).
830 37. Liu, Y.-Y. *et al.* Emergence of plasmid-mediated colistin resistance mechanism MCR-
831 1 in animals and human beings in China: a microbiological and molecular biological
832 study. *The Lancet infectious diseases* **16**, 161–168 (2016).
833 38. Cantón, R., González-Alba, J. M. & Galán, J. C. CTX-M enzymes: Origin and
834 diffusion. *Frontiers in Microbiology* vol. 3 (2012).
835 39. Sekizuka, T. *et al.* Complete sequencing of the bla NDM-1-positive IncA/C plasmid
836 from Escherichia coli ST38 isolate suggests a possible origin from plant pathogens.
837 *PLoS ONE* **6**, (2011).
838 40. Lipworth, S. *et al.* Ten-year longitudinal molecular epidemiology study of Escherichia
839 coli and Klebsiella species bloodstream infections in Oxfordshire, UK. *Genome*
840 *Medicine* **13**, (2021).
841 41. Redondo-Salvo, S. *et al.* COPLA, a taxonomic classifier of plasmids. *BMC*
842 *Bioinformatics* **22**, (2021).
843 42. Rozwandowicz, M. *et al.* Plasmids carrying antimicrobial resistance genes in
844 Enterobacteriaceae. *Journal of Antimicrobial Chemotherapy* **73**, (2018).
845 43. Matlock, W. *et al.* Genomic network analysis of environmental and livestock F-type
846 plasmid populations. *ISME Journal* **15**, (2021).

847 44. Ondov, B. D. *et al.* Mash: Fast genome and metagenome distance estimation using
848 MinHash. *Genome Biology* **17**, (2016).
849 45. Figueras, M. J., Beaz-Hidalgo, R., Hossain, M. J. & Liles, M. R. Taxonomic affiliation
850 of new genomes should be verified using average nucleotide identity and multilocus
851 phylogenetic analysis. *Genome Announcements* **2**, (2014).
852 46. Rosenberg, A. & Hirschberg, J. V-Measure: A conditional entropy-based external
853 cluster evaluation measure. in *EMNLP-CoNLL 2007 - Proceedings of the 2007 Joint*
854 *Conference on Empirical Methods in Natural Language Processing and*
855 *Computational Natural Language Learning* (2007).
856 47. Orlek, A. *et al.* Ordering the mob: Insights into replicon and MOB typing schemes
857 from analysis of a curated dataset of publicly available plasmids. *Plasmid* **91**, (2017).
858 48. Kizny Gordon, A. *et al.* Genomic dynamics of species and mobile genetic elements in
859 a prolonged blaIMP-4-associated carbapenemase outbreak in an Australian hospital.
860 *Journal of Antimicrobial Chemotherapy* **75**, (2020).
861 49. Riley, M. A. & Gordon, D. M. A survey of Col plasmids in natural isolates of
862 Escherichia coli and an investigation into the stability of Col-plasmid lineages. *Journal*
863 *of General Microbiology* **138**, (1992).
864 50. Lyimo, B. *et al.* IncF plasmids are commonly carried by antibiotic resistant
865 Escherichia coli isolated from drinking water sources in northern Tanzania.
866 *International Journal of Microbiology* **2016**, (2016).
867 51. Irrgang, A. *et al.* CTX-M-15-producing E. coli isolates from food products in Germany
868 are mainly associated with an IncF-type plasmid and belong to two predominant clonal
869 E. coli lineages. *Frontiers in Microbiology* **8**, (2017).

870 52. Rodríguez-Beltrán, J., DelaFuente, J., León-Sampedro, R., MacLean, R. C. & San
871 Millán, Á. Beyond horizontal gene transfer: the role of plasmids in bacterial evolution.
872 *Nature Reviews Microbiology* vol. 19 (2021).
873 53. Navon-Venezia, S., Kondratyeva, K. & Carattoli, A. Klebsiella pneumoniae: A major
874 worldwide source and shuttle for antibiotic resistance. *FEMS Microbiology Reviews*
875 vol. 41 (2017).
876 54. Pesesky, M. W., Tilley, R. & Beck, D. A. C. Mosaic plasmids are abundant and
877 unevenly distributed across prokaryotic taxa. *Plasmid* **102**, (2019).
878 55. Fritz, S. A. & Purvis, A. Selectivity in mammalian extinction risk and threat types: A
879 new measure of phylogenetic signal strength in binary traits. *Conservation Biology* **24**,
880 (2010).
881 56. David, S. *et al.* Integrated chromosomal and plasmid sequence analyses reveal diverse
882 modes of carbapenemase gene spread among Klebsiella pneumoniae. *Proc Natl Acad*
883 *Sci U S A* **117**, (2020).
884 57. Pitout, J. D. D. & DeVinney, R. Escherichia coli ST131: A multidrug-resistant clone
885 primed for global domination. *F1000Research* vol. 6 (2017).
886 58. Liu, C. M. *et al.* Escherichia coli ST131-H22 as a foodborne uropathogen. *mBio* **9**,
887 (2018).
888 59. Mohsin, J. *et al.* Plasmid-Mediated Colistin Resistance Gene mcr-1 in an Escherichia
889 coli ST10 Bloodstream Isolate in the Sultanate of Oman. *Microbial Drug Resistance*
890 **24**, (2018).
891 60. García-Meniño, I. *et al.* Swine enteric colibacillosis in Spain: Pathogenic potential of
892 mcr-1 ST10 and ST131 E. Coli Isolates. *Frontiers in Microbiology* **9**, (2018).
893 61. Nicholas Noll, M. M. R. A. N. PanGraph: scalable bacterial pan-genome graph
894 construction. *bioRxiv* (2022).

895 62. Richter, T. K. S. *et al.* Responses of the Human Gut Escherichia coli Population to
896 Pathogen and Antibiotic Disturbances . *mSystems* **3**, (2018).
897 63. Forster, S. C. *et al.* A human gut bacterial genome and culture collection for improved
898 metagenomic analyses. *Nature Biotechnology* **37**, (2019).
899 64. Sender, R., Fuchs, S. & Milo, R. Revised Estimates for the Number of Human and
900 Bacteria Cells in the Body. *PLoS Biology* **14**, (2016).
901 65. Martinson, J. N. v. & Walk, S. T. Escherichia coli Residency in the Gut of Healthy
902 Human Adults . *EcoSal Plus* **9**, (2020).
903 66. Martinson, J. N. V. *et al.* Rethinking gut microbiome residency and the
904 Enterobacteriaceae in healthy human adults. *ISME Journal* **13**, (2019).
905 67. Rodrigues, C. *et al.* High Prevalence of Klebsiella pneumoniae in European Food
906 Products: a Multicentric Study Comparing Culture and Molecular Detection Methods.
907 *Microbiology Spectrum* **10**, (2022).
908 68. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: Resolving bacterial
909 genome assemblies from short and long sequencing reads. *PLoS Computational*
910 *Biology* **13**, (2017).
911 69. Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its applications
912 to single-cell sequencing. *Journal of Computational Biology* **19**, (2012).
913 70. Li, H. Minimap and miniasm: Fast mapping and de novo assembly for noisy long
914 sequences. *Bioinformatics* **32**, (2016).
915 71. Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Fast and accurate de novo genome
916 assembly from long uncorrected reads. *Genome Research* **27**, (2017).
917 72. Walker, B. J. *et al.* Pilon: An integrated tool for comprehensive microbial variant
918 detection and genome assembly improvement. *PLoS ONE* **9**, (2014).
919 73. Torsten Seeman. mlst. (2017).

920 74. Jolley, K. A. & Maiden, M. C. J. BIGSdb: Scalable analysis of bacterial genome
921 variation at the population level. *BMC Bioinformatics* **11**, (2010).
922 75. Jolley, K. A., Bray, J. E. & Maiden, M. C. J. Open-access bacterial population
923 genomics: BIGSdb software, the PubMLST.org website and their applications [version
924 1; referees: 2 approved]. *Wellcome Open Research* **3**, (2018).
925 76. Katz, L. *et al.* Mashtree: a rapid comparison of whole genome sequence files. *Journal*
926 *of Open Source Software* **4**, (2019).
927 77. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): A
928 curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic*
929 *Acids Research* **35**, (2007).
930 78. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, (2014).
931 79. Torsten Seeman. Abricate. (2015).
932 80. Carattoli, A. *et al.* In Silico detection and typing of plasmids using plasmidfinder and
933 plasmid multilocus sequence typing. *Antimicrobial Agents and Chemotherapy* **58**,
934 (2014).
935 81. Siguier, P., Perochon, J., Lestrade, L., Mahillon, J. & Chandler, M. ISfinder: the
936 reference centre for bacterial insertion sequences. *Nucleic Acids Res* **34**, (2006).
937 82. Pal, C., Bengtsson-Palme, J., Rensing, C., Kristiansson, E. & Larsson, D. G. J.
938 BacMet: Antibacterial biocide and metal resistance genes database. *Nucleic Acids*
939 *Research* **42**, (2014).
940 83. Feldgarden, M. *et al.* AMRFinderPlus and the Reference Gene Catalog facilitate
941 examination of the genomic links among antimicrobial resistance, stress response, and
942 virulence. *Scientific Reports* **11**, (2021).
943 84. Robertson, J. & Nash, J. H. E. MOB-suite: software tools for clustering, reconstruction
944 and typing of plasmids from draft assemblies. *Microb Genom* **4**, (2018).

945 85. Csardi, G. & Nepusz, T. The igraph software package for complex network research.
946 *InterJournal Complex Systems* **Complex Sy**, (2006).
947 86. Tettelin, H., Riley, D., Cattuto, C. & Medini, D. Comparative genomics: the bacterial
948 pan-genome. *Current Opinion in Microbiology* vol. 11 (2008).
949 87. Yamashita, A., Sekizuka, T. & Kuroda, M. Characterization of antimicrobial resistance
950 dissemination across plasmid communities classified by network analysis. *Pathogens*
951 **3**, (2014).
952 88. Branger, C. *et al.* Extended-spectrum β-lactamase-encoding genes are spreading on a
953 wide range of escherichia coli plasmids existing prior to the use of third-generation
954 cephalosporins. *Microbial Genomics* **4**, (2018).
955 89. Acman, M., van Dorp, L., Santini, J. M. & Balloux, F. Large-scale network analysis
956 captures biological features of bacterial plasmids. *Nature Communications* **11**, (2020).
957 90. Jesus, T. F. *et al.* Plasmid ATLAS: Plasmid visual analytics and identification in high-
958 Throughput sequencing data. *Nucleic Acids Research* **47**, (2019).
959 91. Matlock, W. *et al.* Flanker: a tool for comparative genomics of gene flanking regions.
960 *Microb Genom* **7**, (2021).
961 92. Arcilla, M. S. *et al.* Dissemination of the mcr-1 colistin resistance gene. *The Lancet*
962 *infectious diseases* **16**, 147–149 (2016).
963 93. Zhi, S. *et al.* Assessing host-specificity of Escherichia coli using a supervised learning
964 logic-regression-based analysis of single nucleotide polymorphisms in intergenic
965 regions. *Molecular Phylogenetics and Evolution* **92**, (2015).
966 94. Delihas, N. Intergenic regions of Borrelia plasmids contain phylogenetically conserved
967 RNA secondary structure motifs. *BMC Genomics* **10**, (2009).

968 95. Blondel, V. D., Guillaume, J. L., Lambiotte, R. & Lefebvre, E. Fast unfolding of
969 communities in large networks. *Journal of Statistical Mechanics: Theory and*
970 *Experiment* **2008**, (2008).
971 96. Tonkin-Hill, G. *et al.* Producing polished prokaryotic pangenomes with the Panaroo
972 pipeline. *Genome Biology* **21**, (2020).
973 97. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7:
974 Improvements in performance and usability. *Molecular Biology and Evolution* **30**,
975 (2013).
976 98. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic
977 Inference in the Genomic Era. *Molecular Biology and Evolution* **37**, (2020).
978 99. Minh, B. Q., Nguyen, M. A. T. & von Haeseler, A. Ultrafast approximation for
979 phylogenetic bootstrap. *Molecular Biology and Evolution* **30**, (2013).
980 100. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. Y. ggtree: an r package for
981 visualization and annotation of phylogenetic trees with their covariates and other
982 associated data. *Methods in Ecology and Evolution* **8**, (2017).
983 101. Orme, D. The caper package : comparative analysis of phylogenetics and evolution in
984 R. *R package version 0.5, 2* (2013).
985 102. Waters, N. R., Abram, F., Brennan, F., Holmes, A. & Pritchard, L. Easy phylotyping
986 of Escherichia coli via the EzClermont web app and command-line tool. *Access*
987 *Microbiology* **2**, (2020).
988 103. Schliep, K. P. phangorn: Phylogenetic analysis in R. *Bioinformatics* **27**, (2011).
989 104. Drake, J. W. A constant rate of spontaneous mutation in DNA-based microbes. *Proc*
990 *Natl Acad Sci U S A* **88**, (1991).

991 105. Foster, P. L., Lee, H., Popodi, E., Townes, J. P. & Tang, H. Determinants of
992 spontaneous mutation in the bacterium Escherichia coli as revealed by whole-genome
993 sequencing. *Proc Natl Acad Sci U S A* **112**, (2015).
994 106. Wielgoss, S. *et al.* Mutation rate dynamics in a bacterial population reflect tension
995 between adaptation and genetic load. *Proc Natl Acad Sci U S A* **110**, (2013).
996 107. Ghalayini, M. *et al.* Evolution of a dominant natural isolate of Escherichia coli in the
997 human gut over the course of a year suggests a neutral evolution with reduced effective
998 population size. *Applied and Environmental Microbiology* **84**, (2018).
999 108. Holt, K. E. *et al.* Tracking the establishment of local endemic populations of an
1000 emergent enteric pathogen. *Proc Natl Acad Sci U S A* **110**, (2013).
1001 109. Wick, R. R., Schultz, M. B., Zobel, J. & Holt, K. E. Bandage: Interactive visualization
1002 of de novo genome assemblies. *Bioinformatics* **31**, (2015).
1003 110. Gómez-Rubio, V. ggplot2 - Elegant Graphics for Data Analysis (2nd Edition) .
1004 *Journal of Statistical Software* **77**, (2017).
1005 111. Munday, E. BioRender. *N/a* (2021).

1006

1007

1008 **Fig. 1. A diverse sample of geographically and temporally restricted *Enterobacterales***

1009 **(a)** Number of chromosomes and plasmids by compartment, stratified by isolate genus. **(b)**

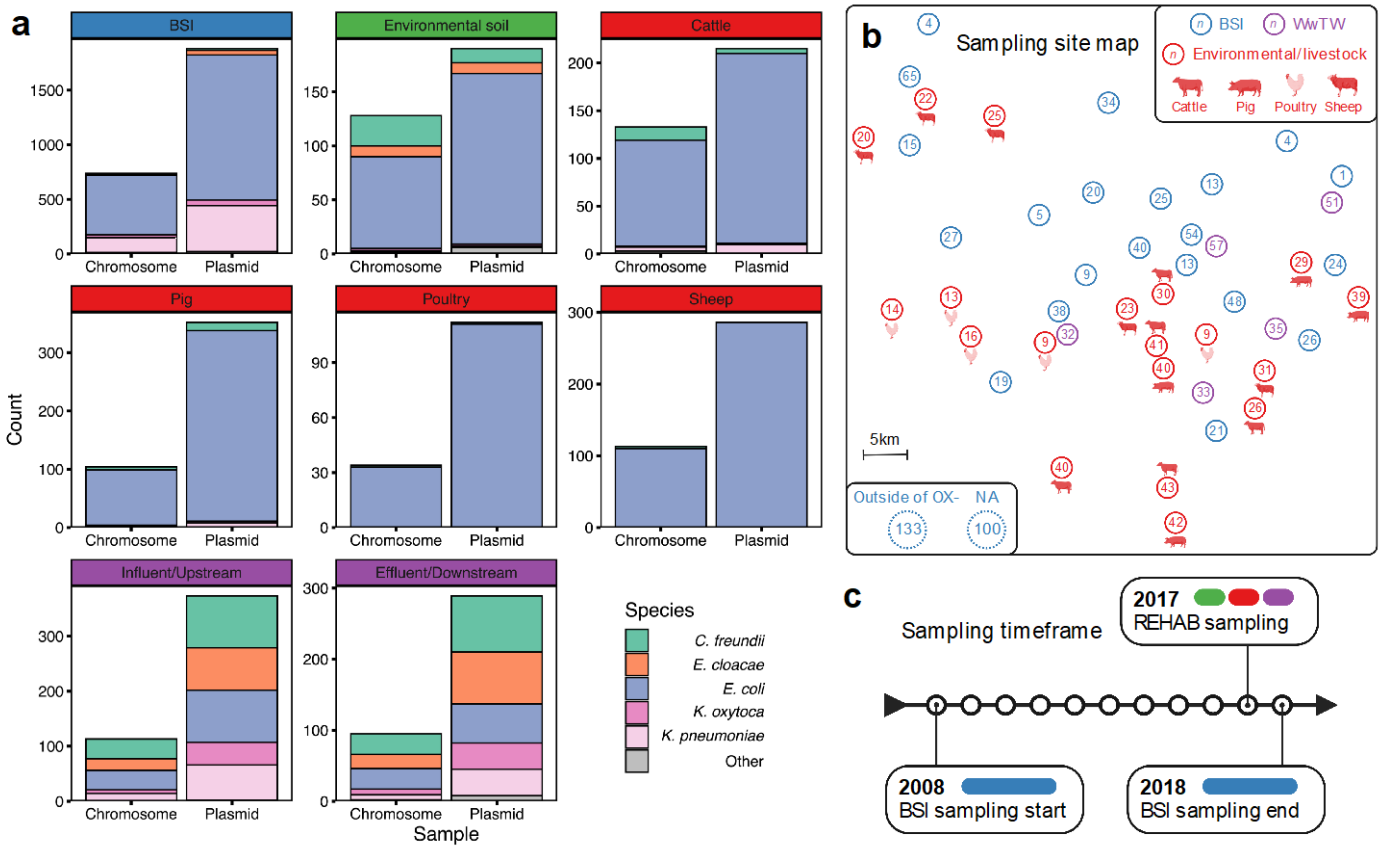
1010 Map of approximate, relative distances between sampling sites, coloured by compartment

1011 (human bloodstream infection [BSI], cattle, pig, poultry, sheep, and wastewater treatment

1012 works [WwTW]). Environmental soils were collected nearby livestock sites. Number in

1013 circles indicates how many of the *n*=1,458 isolates are from that location. **(c)** Sampling

1014 timeframe for BSI and REHAB isolates.



1015 **Fig. 2 Cross-compartmental, near-identical plasmids.**

1016 **(a)** Count of plasmids in near-identical plasmid groups, coloured by compartment (total

1017 *n*=712 plasmids). Median size of plasmid within groups increases from left to right. **(b)**

1018 Proportion of plasmid host species by group. **(c)** Predicted mobility of plasmid. **(d)** AMR

1019 gene carriage in plasmid.

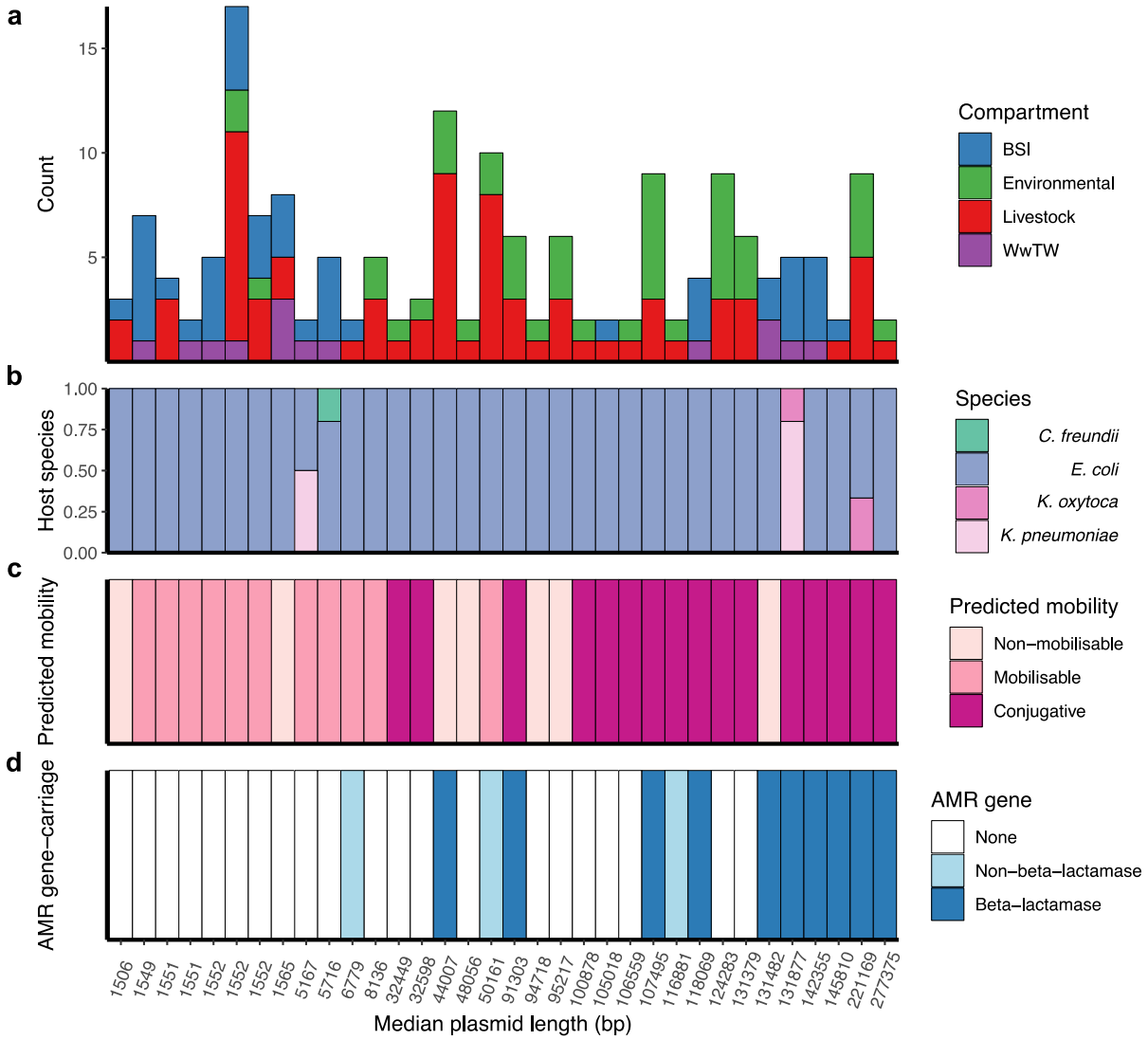


Fig. 3. Genetically similar plasmids share between compartments

(a) Size of plasmid clusters with at least 10 members, coloured by compartment. Size of clusters decreases from left to right. **(b)** Proportion of plasmid host species by cluster. **(c)** Plasmid mobility class and size: Left hand axis shows proportion of plasmids with a predicted mobility class by cluster. Right hand axis shows plasmid length boxplots by cluster. **(d)** Proportions of AMR gene carriage by cluster. **(e)** Plasmid core and accessory genomes: Left hand axis shows the count of core and accessory coding sequences (CDS) by cluster. Right hand axis shows plasmid core gene proportion (i.e., plasmid core CDS/total plasmid CDS) boxplots by cluster.

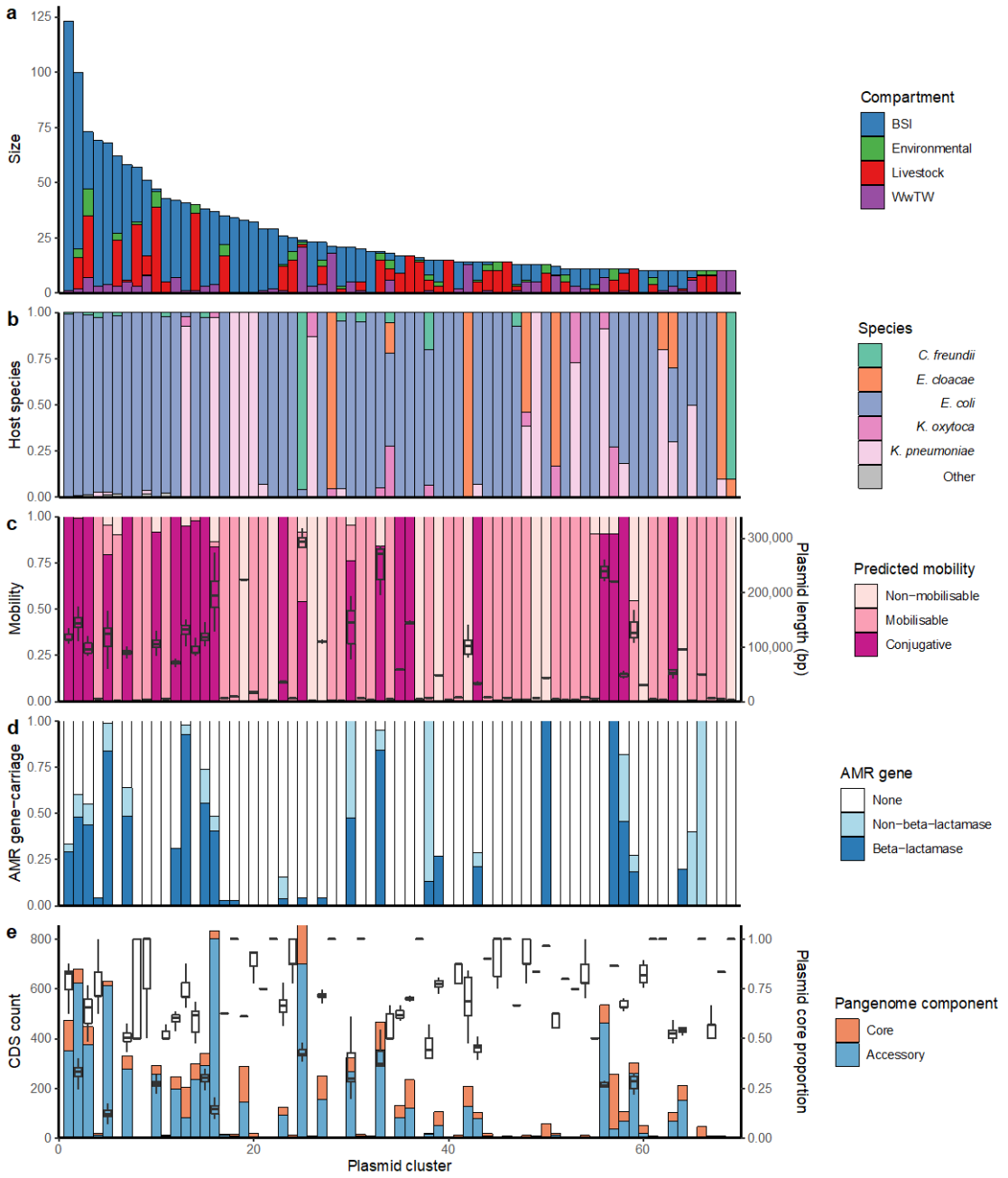
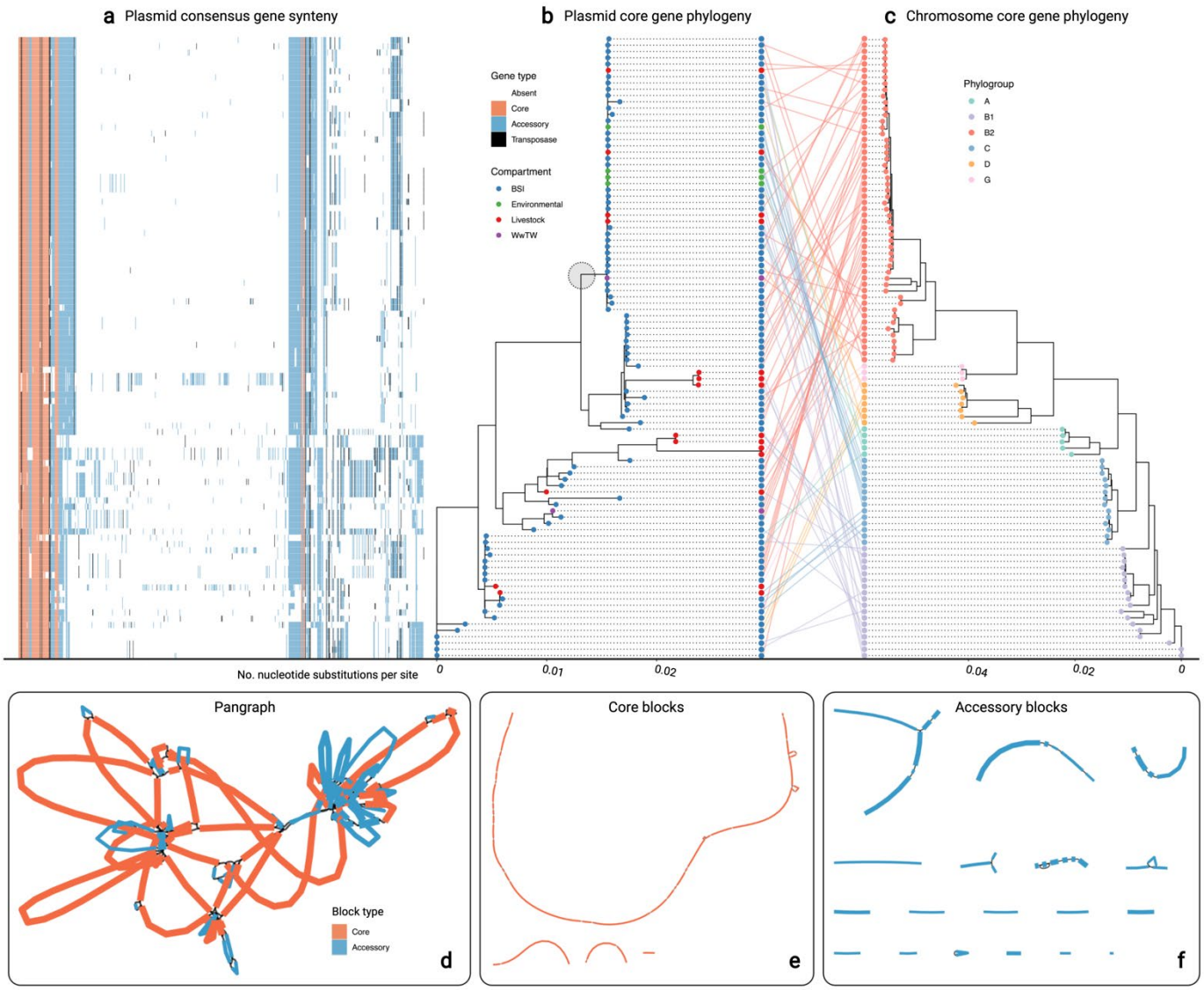
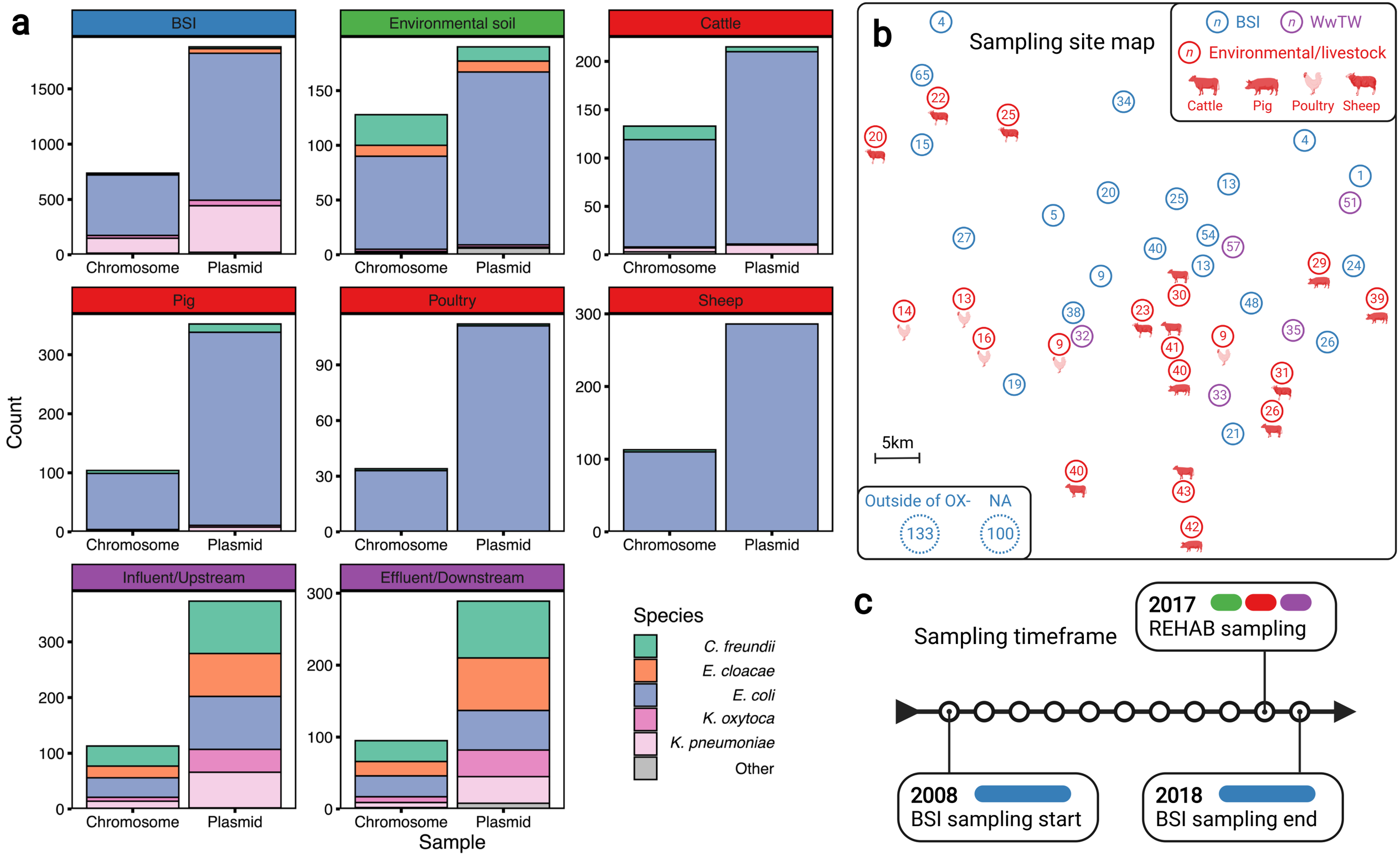
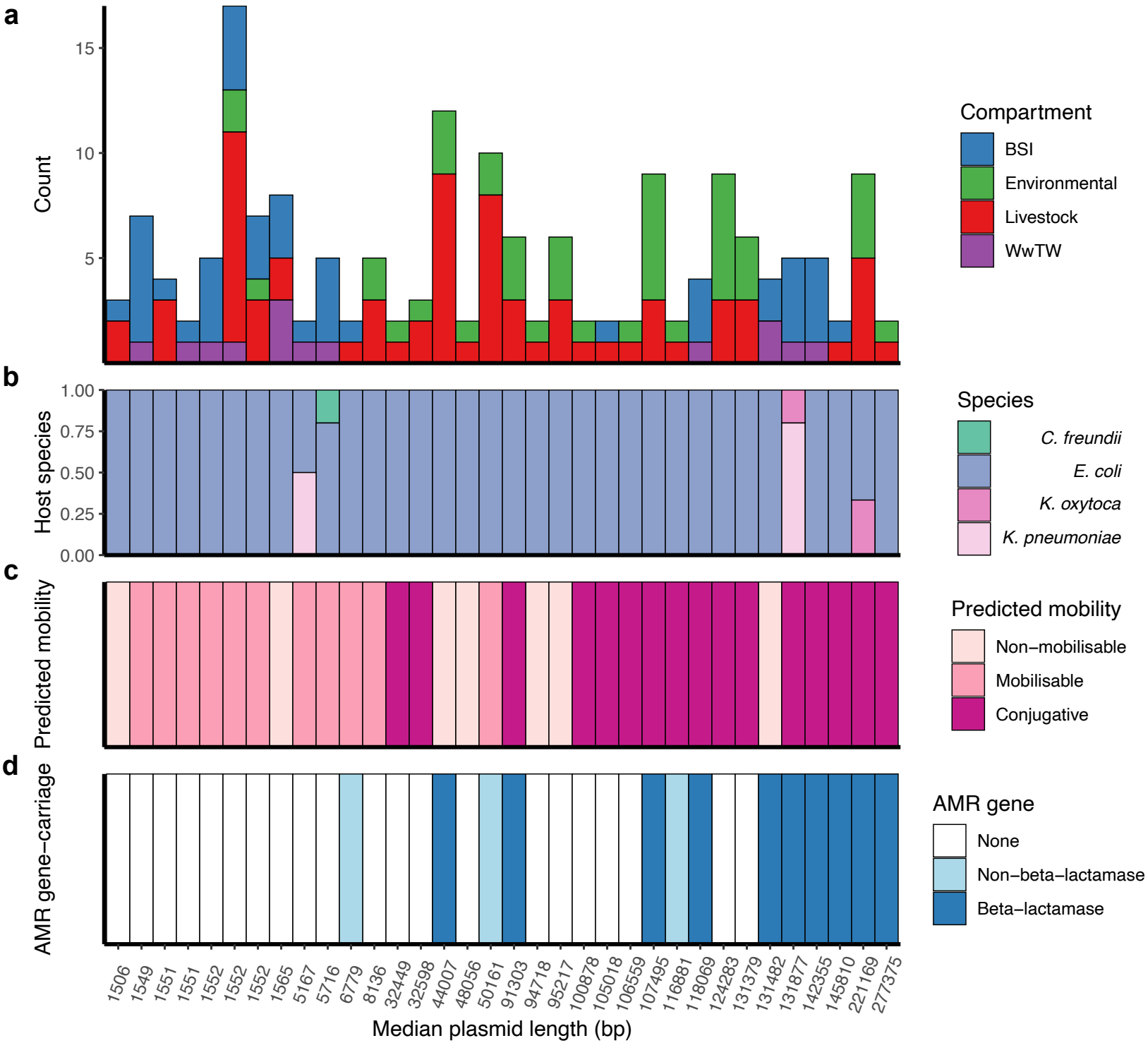


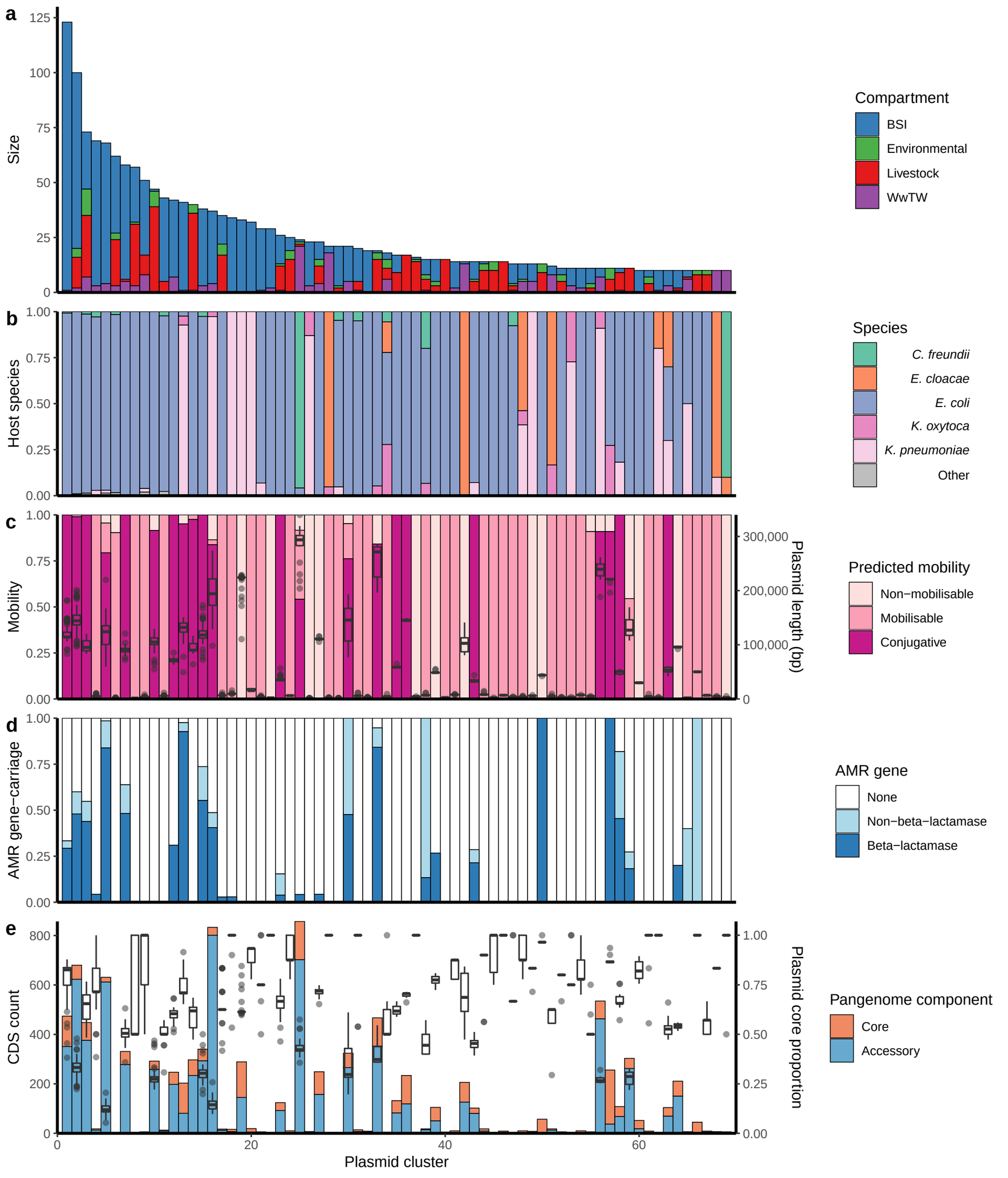
Fig. 4. Cluster 2 plasmid and host evolution

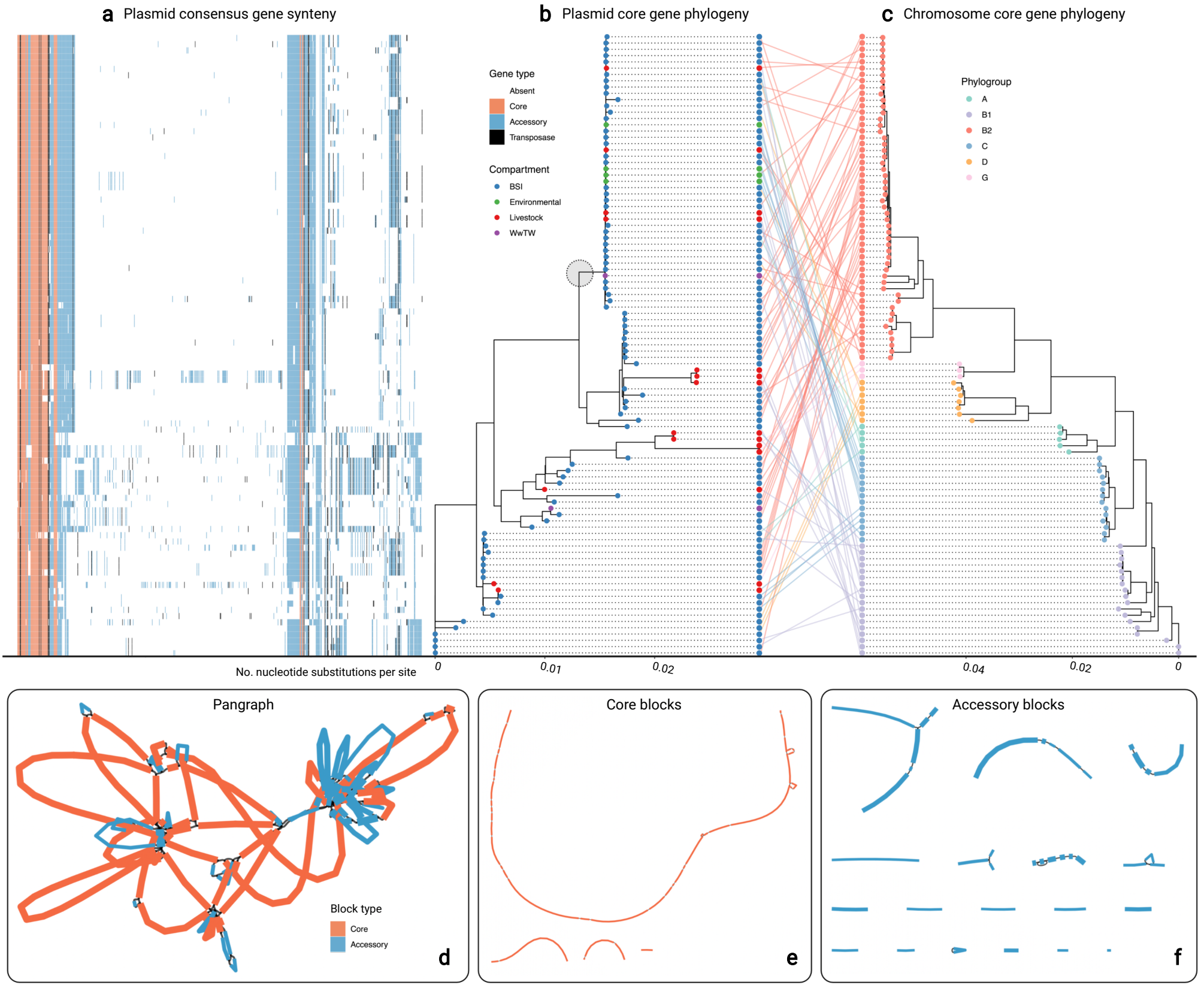
(a) Consensus gene ordering for plasmid cluster 2, coloured by gene type (total *n*=99 plasmids; *n*=1 *S. enterica* isolate omitted). Genes are coloured by core, accessory, or transposase. **(b)** Plasmid core gene phylogeny with tips coloured by sampling compartment. The grey circle highlights the clade of *n*=44 plasmids which were further analysed. **(c)** Plasmid host chromosome core gene phylogeny with tips coloured by sampling compartment. Plasmid and host phylogeny tips are connected in a 'tanglegram' which connects pairs of plasmids and chromosomes from the same isolate. **(d)** Visualisation of the pangraph for *n*=44 plasmids in the grey-circled clade in (b). Blocks are coloured by presence in plasmids. **(e)** Core blocks (found in at least 95% of the *n*=44 plasmids). **(f)** Accessory blocks (found in less than 95% of the *n*=44 plasmids).




a
BSI
Environmental soil
Cattle
Pig
Poultry
Sheep
Influent/Upstream
Effluent/Downstream
Count
Chromosome
Plasmid
Sample
Species
C. freundii
E. cloacae
E. coli
K. oxytoca
K. pneumoniae
Other
b
Sampling site map
BSI
WwTW
Environmental/livestock
Cattle
Pig
Poultry
Sheep
5km
Outside of OX-
NA
133
100
c
Sampling timeframe
2017
REHAB sampling
2008
BSI sampling start
2018
BSI sampling end


a
Count
15
10
5
Compartment
BSI
Environmental
Livestock
WwTW
b
Host species
1.00
0.75
0.50
0.25
0.00
Species
C. freundii
E. coli
K. oxytoca
K. pneumoniae
c
Predicted mobility
Predicted mobility
Non−mobilisable
Mobilisable
Conjugative
d
AMR gene−carriage
AMR gene
None
Non−beta−lactamase
Beta−lactamase
1506
1549
1551
1551
1552
1552
1552
1565
5167
5716
6779
8136
32449
32598
44007
48056
50161
91303
94718
95217
100878
105018
106559
107495
116881
118069
124283
131379
131482
131877
142355
145810
221169
277375
Median plasmid length (bp)


a
Size
125
100
75
50
25
0
Compartment
BSI
Environmental
Livestock
WwTW
b
Host species
1.00
0.75
0.50
0.25
0.00
Species
C. freundii
E. cloacae
E. coli
K. oxytoca
K. pneumoniae
Other
c
Mobility
1.00
0.75
0.50
0.25
0.00
Plasmid length (bp)
300,000
200,000
100,000
0
Predicted mobility
Non−mobilisable
Mobilisable
Conjugative
d
AMR gene−carriage
1.00
0.75
0.50
0.25
0.00
AMR gene
None
Non−beta−lactamase
Beta−lactamase
e
CDS count
800
600
400
200
0
Plasmid core proportion
1.00
0.75
0.50
0.25
0.00
Pangenome component
Core
Accessory
0
20
40
60
Plasmid cluster


a Plasmid consensus gene synteny
b Plasmid core gene phylogeny
c Chromosome core gene phylogeny
Gene type
Absent
Core
Accessory
Transposase
Compartment
BSI
Environmental
Livestock
WwTW
Phylogroup
A
B1
B2
C
D
G
No. nucleotide substitutions per site
0
0.01
0.02
0.04
0.02
0
Pangraph
Block type
Core
Accessory
d
Core blocks
e
Accessory blocks
f