Enterobacterales-associated plasmid sharing amongst human bloodstream infections, livestock, wastewater, and waterway niches: a genomic surveillance study in Oxfordshire, UK

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

Background

Plasmids enable the dissemination of antimicrobial resistance (AMR) in common *Enterobacterales* pathogens, representing a major public health challenge. However, the extent of plasmid sharing between *Enterobacterales* causing human infections and those from other niches remains unclear. Studies to date have been small, with dispersed sampling frames, restricted to drug-resistant isolates only, and using incomplete plasmid sequence reconstruction.

Methods

We established a geographically and temporally restricted collection of human bloodstream infection (BSI)-associated, livestock-associated (cattle, pig, poultry, and sheep faeces, farm soils) and wastewater treatment work (WwTW)-associated (influent, effluent, waterways upstream/downstream of effluent outlets) *Enterobacterales*. Isolates were collected between 2008-2020 from sites <60km apart in Oxfordshire, UK, and sequenced using short- (Illumina) and long-read (PacBio, Nanopore) approaches to fully reconstruct bacterial genomes.

Findings

We analysed 1,458 complete *Enterobacterales* genomes, including 3,697 circularised plasmids, one-third of which represented novel diversity. Plasmid alignment-free clustering identified 122/247 (49%) closely related clusters containing plasmids found in human BSIs and ≥1 other niche (73/247 [30%] such clusters across human BSI and livestock-associated niches). Seventeen groups of near-identical plasmids (*n*=84 plasmids) were seen across human BSIs and ≥1 other sampling niche (eight groups across human BSI and livestock-associated niches), including six AMR-gene associated groups. Pangenome-style analyses of the 69 most prolific clusters (*n*=1,832/3,697 plasmids) revealed shared, core, “backbone”

gene sets. Core-gene plasmid phylogenies reflected an intertwined ecology where well-conserved plasmid backbones carried diverse accessory functions, potentially linked to niche adaptation. Closely related plasmids were found across distantly related bacterial lineages and species.

Interpretation

We identified significant plasmid diversity, highlighting the challenges in adequately sampling natural plasmid populations. Many genetically related plasmids were seen across species and niches, raising the possibility that plasmid movement between these followed by rapid accessory gene change may be relatively common. Dense, unselected sampling is highly relevant to developing our understanding of plasmid epidemiology and designing appropriate interventions to limit the dissemination of plasmid-associated AMR.

Funding

This study was funded by the Antimicrobial Resistance Cross-council Initiative supported by the seven research councils and the NIHR, UK.

Research in context

Evidence before this study

We searched PubMed using the search terms (*Enterobacterales* AND plasmid*) AND (Human OR Animal OR wastewater OR river OR sewage OR environment*) and (*Enterobacterales* AND plasmid*) AND ("One Health") to 04/Jul/2022 with no restrictions on start date or language. We identified 221 studies, of which 167 (76%) were investigating *Enterobacterales* plasmid epidemiology to a lesser or greater extent. 108 (65%) studies analysed >1 *Enterobacterales* species, 49 (29%) studies investigated >50 isolates, 38 (23%) studies completely reconstructed all genomes being evaluated (20 studies a subset of genomes), 23 (14%) studies sampled across multiple niches, and only seven (4%) studies

considered drug- and non-drug-resistant isolates. We identified no large-scale study which considered plasmid epidemiology in several species of drug-resistant and non-resistant *Enterobacterales* across multiple human and non-human niches.

Added value of this study

For the first time, we demonstrate that *Enterobacterales* plasmid transfer and evolution between geographically proximate niches, including human bloodstream infection and livestock-associated niches, is not uncommon. This includes closely related plasmids associated with AMR, observed across *Enterobacterales* species and lineages, underscoring the relevance of our findings for "One Health" for AMR management approaches. Our study also highlights that plasmid diversity in *Enterobacterales* is substantial and undersampled to date.

Implications of all the available evidence

Our understanding of *Enterobacterales* plasmid and AMR epidemiology has been limited by the size and scope of available studies. In the largest systematic study to date, we demonstrate that whilst in some cases niche-/host-associated plasmid structuring exists, multiple *Enterobacterales* plasmid clusters, including those associated with AMR genes, can disseminate widely.

Introduction

Enterobacterales are found both in human niches (e.g., hospital patients[1,2] and wastewater[3]) and non-human niches (e.g., livestock-associated[4,5] and waterways[6]). In recent decades, widespread carriage of antimicrobial resistance (AMR) genes has complicated the treatment of *Enterobacterales* infections[7,8]. The dissemination of AMR genes between *Enterobacterales* occurs in a 'Russian-doll'-style hierarchy of nested, mobilisable genetic structures[9]: genes not only move between bacterial hosts on mobilisable or conjugative plasmids but can also be transferred within and between plasmids and chromosomes by smaller mobile genetic elements (MGEs) such as insertion sequences[10,11]. Despite gene gain/loss events, many plasmids have been shown to have a persistent structure encoding replication and transfer machinery[12,13].

Many plasmids can transfer between species and are seen across different niches[14] but the extent to which they are shared between human and non-human niches remains poorly understood. Previous studies investigating this topic have often been limited in size given the genetic diversity in these niches[15,16], and/or restricted to single species[17] or drug-resistant isolates[18]. Further, fragmented genome assemblies in many cases make recovering complete plasmids, and other MGEs, impossible[19]. There are clearly multiple instances where AMR genes are believed to have emerged from non-human niches and subsequently become major clinical problems in human *Enterobacterales* infections, highlighting the relevance of inter-niche transfer in AMR gene dissemination (e.g., *mcr-1*[20] and $bla_{\text{NDM-1}}$[21]).

To explore *Enterobacterales* plasmid diversity and sharing across niches in a geographically and temporally restricted context, we studied hybrid assemblies (i.e., using both long and short reads) of large *Enterobacterales* isolate collections in Oxfordshire, UK, from (i) human

bloodstream infections (BSI; 2008-2018), (ii) livestock-associated sources (faeces from cattle, pigs, poultry, sheep; surrounding environmental soils; [all 2017 except poultry 2019-2020], and (iii) wastewater treatment work (WwTW)-associated sources (influent, effluent, waterways upstream/downstream of effluent outlets; Oxfordshire, 2017).

Methods

Livestock-associated isolates

n=247 *Enterobacterales* isolates from farm-proximate soils and poultry faeces (*n*=19 farms; *n*=5 cattle, *n*=4 pig, *n*=5 poultry, *n*=5 sheep) were collected and sequenced for this study in 2017-2020. DNA extraction and sequencing was performed as in Shaw *et al.*, 2021[11]. Genomes were hybrid assemblies reconstructed using Unicycler[22] (v. 0.4.4; default hybrid assembly parameters except min_component_size 500 and --min_dead_end_size 500). Only complete assemblies (plasmids and chromosomes) were considered (*n*=162/247).

BSI isolates

Sequenced Human BSI Enterobacterales isolates from patients presenting to *n*=4 hospitals within Oxfordshire, UK, September 2008-December 2018, as described in Lipworth *et al.*, 2021[23] were also included. Although all patients were sampled in Oxfordshire, a total of *n*=505/738 patients resided in Oxfordshire, *n*=133/738 in surrounding counties, and *n*=100/738 had location information omitted. Only complete assemblies (*n*=738/953 total assembled) were considered.

Other livestock-associated and WwTW-associated isolates

Enterobacterales isolates from faeces from the *n*=14 non-poultry farms and wastewater influent, effluent, and waterways upstream/downstream of effluent outlets surrounding *n*=5 WwTWs, across 3 seasonal timepoints in 2017 (as in [11]) were included. Only complete assemblies (*n*=558/827 total assembled) were considered.

Statistical analysis and bioinformatics

Chromosome sequence types (STs) were determined with mlst (v. 2.19.0; see Supplementary Methods). To generate accumulation curves (ACs), new plasmid diversity was recorded for each isolate sampled randomly, without replacement. A bootstrapped average of *b*=1,000 ACs was used to estimate a Heap's parameter (γ) by fitting a linear regression to log-log transformed data see (Supplementary Methods). We adopted three approaches to plasmid classification, using COPLA to classify plasmids into broad plasmid taxonomic units (PTUs), and also grouping and clustering plasmids into smaller clusters using alignment-free distances (see Supplementary Methods). Within plasmid clusters, we identified core genes with Panaroo (v. 1.2.9), aligned them with mafft (v7.407) and produced trees with IQ-tree (v. 2.0.6). Plots were primally produced using the R library ggplot2, with additional graphics in BioRender. More information can be found in the Supplementary Methods.

Results

Our dataset of *n*=3,697 plasmids from *n*=1,458 isolates (Fig. 1a, Table 1) contained bacteria from human bloodstream infections (BSI; *n*=1,880 plasmids from *n*=738 isolates), livestock-associated sources (cattle, pig, poultry, and sheep faeces, soils surrounding livestock farms; *n*=1,155 plasmids from *n*=512 isolates), and from wastewater treatment works (WwTW)-associated sources (influent, effluent, waterways upstream/downstream of effluent outlets; *n*=662 plasmids from *n*=208 isolates). All sampling sites were <60km apart (Fig. 1b) and timeframes overlapped (2008-2020; Fig. 1c). Isolates had a median 2 plasmids (IQR=1-4, range=0-16). Major *Enterobacterales* genera represented included: *n*=1,044 *Escherichia*, *n*=211 *Klebsiella*, *n*=125 *Citrobacter*, and *n*=63 *Enterobacter*.

Table 1. Isolate niche breakdown.

Niche	**Sample type(s)**	**No. isolates**	**No. plasmids**
Bloodstream infections (BSI)	Community, nosocomial, other healthcare associated infections	738	1,880
Livestock-associated	Cattle faeces	133	215
	Sheep faeces	113	286
	Pig faeces	104	352
	Poultry faeces	34	112
	Soil surrounding livestock farms	128	190
Wastewater treatment work (WwTW)-associated	Influent	88	313
	Upstream waterways	25	60
	Effluent/downstream waterways	95	289
Total		**1,458**	**3,697**

167

Table 2. Isolate genus breakdown.

Niche	Isolate genus					Total
	Citrobacter	***Enterobacter***	***Escherichia***	***Klebsiella***	**Other**	
Bloodstream infections (BSI)	6	11	547	161	13	**738**
Livestock-associated	54	10	433	14	1	**512**
Wastewater treatment work (WwTW)-associated	65	42	64	37	0	**208**
Total	**125**	**63**	**1,044**	**212**	**14**	**1,458**

Sampling niche was strongly associated with isolate genus (Fisher's test, *p*-value<5e-4; Table 2). *Klebsiella* isolates were disproportionately derived from BSI versus other niches (76% [161/212] versus 51% [738/1,458]). *Citrobacter* and *Enterobacter* were disproportionately derived from WwTW-associated versus other niches (52% [65/125] and 67% [42/63] versus 14% [208/1,458]). Chromosomal Mash trees (see Supplementary Methods) for the two most common species in the dataset, *E. coli* (72% [1,044/1,458]; see Fig. S1) and *K. pneumoniae* (11% [163/1,458]; Fig. S2) demonstrated intermixing of human and non-human isolates within clades, consistent with species-lineages not being structured by niche.

We contextualised our plasmids within known plasmid diversity using 'plasmid taxonomic units' (PTUs; using COPLA, see Supplementary Methods), designed to be equivalent to a plasmid 'species'. We found 32% (1,193/3,697) of plasmids were unclassified, highlighting the substantial plasmid diversity within this geographically restricted dataset. In total, we

found *n*=67 known PTUs, containing a median 9 plasmids (IQR=4-30, range=1-556), with the largest PTU-F_E (556/2,504), corresponding to F-type *Escherichia* plasmids.

Near-identical plasmid sharing observed between human and livestock-associated *Enterobacterales*

We screened for near-identical plasmids shared across isolates by grouping those with a low Mash distance (d<0.0001) and highly similar lengths (longest plasmid ≤1% longer than shorter plasmids; see Supplementary Methods). We found *n*=225 near-identical groups of ≥2 members, recruiting 19% (712/3,697) plasmids. Bootstrapping accumulation curves for near-identical plasmid groups and singletons per the number of isolates (ACs; see Supplementary Methods), we revealed a highly 'open' accumulation (Heap's parameter γ=0.97, Fig. S3) suggesting further isolate sampling would detect more unique plasmids approximately linearly. Restricted to BSI/livestock-associated isolates alone, we found similar curves for both niches (BSI γ=0.98, livestock-associated γ=0.94), suggesting they had similar levels of plasmid diversity.

Near-identical pairs of plasmids were most common, representing 71% (159/225) of groups (group size IQR=2-3, range=2-32). Plasmid members of near-identical groups represented multiple bacterial host STs (25% [56/225]), species (4% [9/225]), and genera (4% [9/225]), consistent with plasmids capable of inter-lineage/species/genus transfer. Further, 8% (17/225) of near-identical groups contained plasmids found across human BSIs and ≥1 other sampling niche (livestock-associated/WwTW-associated), suggesting inter-niche transfer (i.e., 'cross-niche groups'; Fig. 2a). Within cross-niche groups, *n*=3/17 contained plasmids from multiple bacterial species (Fig. 2b), and most consisted of conjugative plasmids (*n*=5/17 conjugative, *n*=9/17 mobilisable, *n*=3/17 non-mobilisable; Fig. 2c). AMR genes were carried by plasmids

in *n*=6/17 cross-niche groups (Fig. 2d), with *n*=5/6 of these groups containing ≥1 beta-lactamase protein.

Sharing between BSI and livestock-associated isolates was supported by 8/17 cross-niche groups (*n*=45 plasmids). Of these, *n*=2/8 contained non-mobilisable Col-type plasmids (one group contained BSI/pig/poultry/influent plasmids, and one group contained BSI/poultry plasmids); *n*=4/8 contained mobilisable Col-type plasmids (two groups contained BSI/pig plasmids, one group contained BSI/sheep plasmids, and one group contained BSI/cattle/pig/poultry/sheep/influent plasmids), of which one group contained BSI/pig plasmids carrying the AMR genes *aph(3'')-Ib*, *aph(6)-Id*, *dfrA14*, and *sul2* (see Supplementary Methods). The remaining 2/8 groups contained conjugative FIB-type BSI/sheep plasmids. One group contained plasmids, carrying the AMR genes *aph(3'')-Ib*, *aph(6)-Id*, bla_{TEM-1}, *dfrA5*, *sul2*, and the other group contained plasmids carrying the MDR efflux pump protein *robA*.

Plasmid clustering reveals a diverse but intertwined population structure across niches

Near-identical plasmids shared across niches are a likely signature of recent transfer events, but we also wanted to examine the wider plasmid population structure. We therefore agnostically clustered all plasmids based on alignment-free sequence similarity (clusters were groups of *n*≥3 plasmids; see Supplementary Methods and Figs. S5-6). We defined *n*=247 plasmid clusters with median 5 members (IQR=3-10, range=3-123) recruiting 71% (2,627/3,697) of the plasmids. The remainder were either singletons (i.e., single, unconnected plasmids; 19% [718/3,697]) or doubletons (i.e., pairs of connected plasmids; 10% [352/3,697]). By bootstrapping *b*=1,000 ACs for plasmid clusters, doubletons, and singletons found against number of isolates sampled (Fig. S7; see Supplementary Methods), we

232 estimated that the rarefaction curve had a Heap's parameter γ=0.75, suggesting further isolate
233 sampling would likely detect more plasmid diversity and clusters.
234
235 Of the plasmid clusters, *n*=69/247 (28%) plasmid clusters had ≥10 members, representing
236 50% (1,832/3,697) of all plasmids (Fig. 3a). 122/247 (49%) clusters contained BSI plasmids
237 and plasmids from ≥1 other niche. This included 73/247 (30%) of clusters with both BSI and
238 livestock-associated plasmids, representing *n*=38 unique plasmid replicon haplotypes (i.e.,
239 combinations of replication proteins) of which only 24% (9/38) were Col-type plasmids,
240 which are often well-conserved and carry few genes[24]. 72/247 (29%) of clusters contained
241 both BSI, and influent/effluent/downstream plasmids, reflecting a route of *Enterobacterales*
242 dissemination into waterways. In contrast, only 18/247 (7%) of clusters contained both BSI
243 and upstream waterway plasmids, of which most (13/18 [72%]) also contained
244 influent/effluent/downstream plasmids.
245
246 Overall, plasmid clusters scored high homogeneity (*h*) but low completeness (*c*) with respect
247 to biological and ecological characteristics (non-putative PTUs [*h*=0.99, *c*=0.66]; replicon
248 haplotype [*h*=0.92, *c*=0.69]; bacterial host ST [*h*=0.84, *c*=0.14] in Fig. 3b; predicted mobility
249 [*h*=0.93, *c*=0.20] in Fig 3c). This indicated that clustered plasmids often had similar
250 characteristics, but the same characteristics were often observed in multiple clusters. The
251 imperfect homogeneity is to be anticipated as replicon haplotypes and mobilities can vary
252 within plasmid families, and plasmid families can have diverse host ranges[14].
253
254 Plasmids carrying AMR genes were found in 21% (52/247) of the plasmid clusters (i.e.,
255 'AMR-carrying clusters'), representing *n*=550 plasmids (Fig. 3d). Of the AMR-carrying
256 clusters, 92% (48/52) contained at least one beta-lactamase-carrying plasmid (*n*=437

plasmids in total). AMR genes were present in a median proportion 67% of AMR-carrying cluster members (IQR=28-100%, range=3-100%). This highlights that AMR genes are not necessarily widespread on genetically similar plasmids and can be potentially acquired multiple different times through the activity of smaller MGEs (e.g. transposons) or recombination. For example, cluster 12 was a group of *n*=42 conjugative, PTU-F_E plasmids found in BSI, wastewater, and waterways. Of these, 31% (13/42) carried the AMR gene bla_{TEM-1}, and in a range of genetic contexts: *n*=9/13 bla_{TEM-1} genes were found within Tn*3* and *n*=4/13 were carried without a transposase, of which *n*=2/4 were found with the additional AMR genes *aph(6)-Id*, *aph(3'')-Ib*, and *sul2*. AMR genes were disproportionately carried by F-type plasmids (61% [337/550] AMR-carrying cluster plasmids versus 34% [891/2627] of the total clustered plasmids), further underlining the known role of F-type plasmids in AMR gene dissemination[13].

An intertwined ecology of plasmids across human and livestock-associated niches

Plasmids can change their genetic content, particularly when subject to new selective pressures[25,26]. Many plasmids have a structure with a 'backbone' of conserved core genes and a 'cargo' of variable accessory genes[12,13]. We wanted to explore evidence for cross-niche plasmids with minimal mutational evolution in a shared backbone (compatible with ~years of evolutionary separation) but variable accessory gene repertoires.

We first conducted a pangenome-style analysis (see Supplementary Methods) on the *n*=69/247 plasmid clusters with ≥10 members. For each cluster, we determined "core" (genes found in ≥95% of plasmids) and "accessory" gene repertoires (found in <95% of plasmids). Within clusters, we found median 9 core genes (IQR=4-53, range=0-219), and median 9 accessory genes (IQR=3-145, range=0-801) (Fig. 3e). Core genes comprised a median

proportion 42.2% of the total pangenome sizes (IQR=20.9-66.7%). At an individual plasmid level, core genes shared by a cluster comprised a median proportion 62.5% of each plasmid's gene repertoire (IQR=37.4-83.3%; Fig. 3e). Putatively conjugative plasmids carried a significantly higher proportion of accessory genes in their repertoires than mobilisable/non-mobilisable plasmids (Kruskal-Wallis test [$H(2)=193.01$, p-value<2.2e-16] followed by Dunn's test).

Using multiple sequence alignments of the core genes within each cluster, we produced maximum likelihood phylogenies (see Supp. File 1 and Supplementary Methods). For this step, we only considered the n=62/69 clusters where each plasmid had ≥1 core gene. With the n=27/62 clusters that contained both BSI and livestock-associated plasmids, we measured the phylogenetic signal for plasmid sampling niche using Fritz and Purvis' D (see Table S1 and Supplementary Methods). The analysis indicated that the evolutionary history of plasmid clusters is neither strictly segregated by sampling niche nor completely intermixed, but something intermediate.

Alongside the core gene phylogenies, we generated gene repertoire heatmaps (example cluster 2 in Fig. 4a-b; all clusters and heatmaps in Supp. File 1). By visualising the genes in a consensus synteny order (see Supplementary Methods), the putative backbone within each plasmid cluster is shown alongside its accessory gene and transposase repertoire. This highlights how plasmids might gain/lose accessory functions within a persistent backbone. Log-transformed linear regression revealed a significant relationship between Jaccard distance of accessory genes presence against core gene cophenetic distance ($y=0.080\log(x)+0.978$, $R^2=0.47$, $F(1,52988)=4.75e4$, p-value< 2.2e-16; see Fig. S8 and Supplementary Methods).

307

308 **Plasmid evolution between human and livestock-associated niches is not structured by**
309 **bacterial host**

310 Alongside vertical inheritance, conjugative and mobilisable plasmids are capable of inter-host
311 transference, crossing between bacterial lineages, species, up to phyla[14]. However, bacterial
312 surveillance often only tracks clonally evolving lineages[27], which might not account for
313 clinically relevant AMR genes mobilised on plasmids. Phylogenetic analysis can determine
314 whether plasmid evolution between BSI and livestock-associated niches is driven by host
315 clonal expansion or other means.

316

317 As a detailed example, we evaluated the largest plasmid cluster containing both human and
318 livestock-associated plasmids (cluster 2, n=100 members). All plasmids carried at least one
319 F-type replicon and were all putatively conjugative, with 75% (75/100) and 25% (25/100)
320 assigned PTU-F_E and a putative PTU, respectively. Further, 48% (48/100) plasmids carried
321 bla_{TEM-1}, and 51% (51/100) carried >1 AMR gene. All host chromosomes were *E. coli* except
322 OX-BSI-481_2 (*S. enterica* ST 2998; hereon omitted from the analysis). The n=99 *E. coli*
323 isolates represented six phylogroups: A (5/99), B1 (18/99), B2 (52/99), C (14/99), D (7/99),
324 and G (3/99; see Supplementary Methods).

325

326 Figure 4b-c shows the plasmid core gene phylogeny (T_{plasmid}) and the *E. coli* host core gene
327 phylogeny ($T_{\text{chromosome}}$). The *E. coli* phylogeny was structured by six clades corresponding to
328 the six phylogroups (see Supplementary Methods). We found low congruence between the
329 plasmid core-gene phylogeny and the chromosomal core-gene phylogeny as seen in the
330 central 'tanglegram' (i.e., lines connecting pairs of plasmid and chromosome tips from the
331 same isolate). Additionally, we calculated a Robinson-Foulds distance $RF(T_{\text{plasmid}},$

$T_{chromosome}$)=162, reflecting a high number of structural differences between the phylogenies (see Supplementary Methods). There was some evidence of plasmid structuring by niche (Fritz and Purvis' *D*=0.24; see Supplementary Methods).

Within the plasmid phylogeny, there was a clade of *n*=44 plasmids (support 100%; circled in grey in Figure 4b) containing both BSI and livestock-associated plasmids, which were within median 4 core gene SNPs of each other (IQR=2-8, range=0-59). Estimating plasmid evolution at an approximate rate of one SNP per year (see Supplementary Methods) would give a median time to most recent common ancestor of the backbone at approximately 4 years prior to sampling, consistent with recent movement between human and livestock-associated niches. This plasmid clade was mainly present in phylogroup B2 (20/44), but also A (3/44), B1 (9/44), C (8/44), and D (4/44), suggesting plasmid movement. Further, 77% (34/44) of plasmids within the clade carried bla_{TEM-1} (BSI: 25/34, Livestock-associated: 8/34, WwTW-associated: 1/34), and 82% (36/44) carried ≥1 AMR gene, highlighting the role of plasmids in cross-niche dissemination of AMR.

To examine the evolution of entire plasmid sequences within the clade, we represented all *n*=44 plasmids as a 'pangraph' (Figure 4d; see Supplementary Methods). Briefly, pangraph converts input sequences into a consensus graph, where each sequence is a path along a set of homologous sequence alignments i.e., 'blocks', which in series form 'pancontigs'. Filtering for 'core blocks' (i.e., those found in ≥95% plasmids), we found 4 pancontigs (40 blocks total), with the longest 98,269bp (total length 125,369bp), indicating a putative plasmid backbone (Fig. 4e). Then, filtering for 'accessory blocks' (i.e., those found in <95% plasmids), we found 18 pancontigs (39 blocks total), with median length 2,380bp (total length

63,753bp), forming the accessory gene repertoire (Fig. 4f). This points to a persistent plasmid backbone structure with loss/gain events at particular 'hotspots' as well as rearrangements.

Discussion

By analysing a dataset of n=3,697 systematically collected *Enterobacterales* plasmids sampled from human BSI, livestock- and WwTW-associated sources in a geographically and temporally restricted context, we find evidence of plasmid dissemination across niches, including those carrying clinically relevant AMR genes. We found 225 instances of shared, near-identical plasmid groups, 25% of which were found across multiple bacterial STs, 4% across multiple bacterial species, and 8% in both human BSI and ≥1 non-BSI niche. Beyond this near-identical sharing, we analysed 'clusters' of plasmids and found that that 73/247 clusters contained plasmids seen in both human BSIs and other contexts. Over one fifth (52/247) of plasmid clusters contained plasmids carrying AMR genes (*n*=550 plasmids). Our results suggest the need for broad, unselected, and detailed sampling frames to fully understand plasmid diversity and evolution.

Whilst some plasmid clusters are strongly structured by host phylogeny and isolate source, some plasmids from human BSIs are highly genetically related to those in other niches, including livestock. However, recovering these similarities is a sampling challenge. Accumulation curve analyses suggested increasing the size of our dataset would have led to further near-identical matches at an approximately linear rate, meaning even a dataset of this size captures only a small fraction of the true extent of plasmid sharing between human clinical and other non-human/clinical niches. This presents a challenge for designing appropriately powered studies. Had we only sampled *n*=100 livestock-associated isolates

(i.e., around 20% of our actual sample), there was only a 39% chance that we would have detected ≥5 matches with BSI plasmids (Fig. S4).

Given that plasmids observed in BSI isolates represent a restricted and small proportion of human *Enterobacterales* diversity, many more sharing events may occur in the human gut[28] which we only sampled incompletely using wastewater influent as a proxy. The human colon contains around 10^{14} bacteria[29], with large ranges of *Enterobacteriaceae* abundance. Further, even small numbers of across-niche sharing events, such as transfer events of important AMR genes from species-to-species or niche-to-niche, may have significant clinical implications, as has been seen with several important AMR genes globally (e.g., *mcr*-1, $bla_{\text{NDM-1}}$). Future studies need to carefully consider the limitations of sampling frames in detecting any genetic overlap, given both substantial diversity and the effects of niches and geography[11,16].

By examining plasmid relatedness compared to bacterial host relatedness, we demonstrated that cross-niche plasmid spread is not driven by clonal lineages. Using a pangenome-style analysis, we showed that plasmids can share sets of near-identical core genes alongside diverse accessory gene repertoires. While plasmids with more distantly related core genes tended to have dissimilar accessory gene content, plasmids with more closely related core genes shared a wide range of accessory gene content. This would be consistent with a hypothesis of persistent 'backbone' structures gaining and losing accessory functions as they move between hosts and niches. We suggest that this mode of transfer might be worth considering. Evolutionary models for plasmids which can accommodate well-conserved backbone evolution alongside accessory structural changes and gain/loss events are urgently needed. Estimating plasmid evolutionary rates remains a challenge, with little known about

404 appropriate values for mutation rates in plasmids, and even less for non-mutational processes
405 such as gene gain/loss.
406
407 Our study had several limitations. Our non-BSI isolates were not as temporally varied as the
408 BSI isolates, meaning we could not fully explore temporal evolution. Isolate-based
409 methodologies are limited in evaluating the true diversity of the niches sampled; composite
410 approaches including metagenomics might shed additional insight in future studies. Further,
411 the exact source of an isolate is poorly defined for wastewater/waterway isolates as they act
412 as a confluence of multiple sources, although they represent important niches in their own
413 right. We only analysed plasmids from complete genomes i.e., where the chromosome and all
414 plasmids were circularised, meaning we disregarded ~23% and ~33% of BSI and non-BSI
415 assemblies, respectively. We only focused on plasmids as horizontally transmissible elements
416 here; detailed study of other smaller mobile genetic elements across-niches would represent
417 interesting future work. We have also investigated a limited subset of *Enterobacterales*:
418 plasmid sharing likely extends to other bacterial hosts not investigated here. Lastly, our
419 isolate culture methods for livestock-associated samples may not have been as sensitive for
420 the identification of *Klebsiella* spp. as for other *Enterobacterales* such as *Escherichia,* as we
421 did not use enrichment and selective culture on Simmons citrate agar with inositol[30].
422
423 In conclusion, this study presents to our knowledge the largest evaluation of systematically
424 collected *Enterobacterales* plasmids across human and non-human niches within a
425 geographically and temporally restricted context. Plasmids can clearly disseminate between
426 niches, although this dynamic likely varies by cluster; the overall number of near-identical
427 plasmid groups identified across niches consistent with recent transfer events was 8%
428 (17/225) and influenced by sample size. We demonstrate a likely intertwined ecology of

plasmids across human and non-human niches, where different plasmid clusters are variably but incompletely structured and putative 'backbone' plasmid structures can rapidly gain and lose accessory genes following cross-niche spread. Future "One Health" studies require dense and unselected sampling, and complete/near-complete plasmid reconstruction, to appropriately understand plasmid epidemiology across niches.

Data availability

Study metadata is provided in Table S2. Accessions for poultry and environmental soil isolate reads are given in Table S3, and assemblies will shortly be made available on NCBI. Accessions for existing BSI and REHAB reads and assemblies can be found in Lipworth *et al.*, 2021[23] and Shaw *et al.*, 2021[11], respectively.

Code availability

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

REHAB Consortium.

Manal AbuOun[2], Muna F. Anjum[2], Mark J. Bailey[3], Brett H[8], Mike J. Bowes[3], Kevin K. Chau[1], Derrick W. Crook[1,6,7], Nicola de Maio[1], Nicholas Duggett[2], Daniel J. Wilson[1,9], Daniel Gilson[2], H. Soon Gweon[3,4], Alasdair Hubbard[10], Sarah J. Hoosdally[1], William Matlock[1], James Kavanagh[1], Hannah Jones[2], Timothy E. A. Peto[1,6,7], Daniel S. Read[3], Robert Sebra[5], Liam P. Shaw[1], Anna E. Sheppard[1,6], Richard P. Smith[2], Emma Stubberfield[2], Nicole Stoesser[1,6,7], Jeremy Swann[1], A. Sarah Walker[1,6,7], Neil Woodford[11]

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

[2] Animal and Plant Health Agency, Weybridge, Addlestone, UK

[3] UK Centre for Ecology & Hydrology, Wallingford, UK

[4] University of Reading, Reading, UK

[5] Icahn Institute of Data Science and Genomic Technology, Mt Sinai, NY, USA

[6] NIHR HPRU in healthcare-associated infection and antimicrobial resistance, University of Oxford, Oxford, UK

[7] NIHR Oxford Biomedical Research Centre, University of Oxford, Oxford, UK

[8]Thames Water Utilities, Clearwater Court, Vastern Road, Reading, UK

[9]Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, UK

[10]Department of Tropical Disease Biology, Liverpool School of Tropical Medicine, Liverpool, UK

[11]Antimicrobial Resistance and Healthcare Associated Infections (AMRHAI) Reference Unit, National Infection Service, Public Health England, London, United Kingdom

Declarations of interests. The authors declare no interests.

Role of the funding source.

This work was funded by the Antimicrobial Resistance Cross-council Initiative supported by the seven research councils [grant NE/N019989/1]. The UKCEH component of the REHAB consortium was supported by the Natural Environment Research Council (NERC) [grant NE/N019660/1]. DWC, SG, TEAP and NS are supported by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Healthcare-Associated Infections and Antimicrobial Resistance at the University of Oxford in partnership with Public Health England (PHE) [grant HPRU-2012–10041 and NIHR200915]. DWC and TEAP are also supported by the NIHR Oxford Biomedical Research Centre. The computational aspects of this research were funded from the NIHR Oxford BRC with additional support from a Wellcome Trust Core Award Grant [grant 203141/Z/16/Z]. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health or Public Health England. WM and KKC are supported by a

479 scholarship from the Medical Research Foundation National PhD Training Programme in
480 Antimicrobial Resistance Research (MRF-145-0004-TPG-AVISO). NS is an Oxford Martin
481 Fellow and a Senior NIHR BRC Oxford Fellow. LPS is a Sir Henry Wellcome Postdoctoral
482 Fellow funded by Wellcome (grant 220422/Z/20/Z).

483

Bibliography

1. Linh, T. D. *et al.* Expansion of KPC-producing Enterobacterales in four large hospitals in Hanoi, Vietnam. *Journal of Global Antimicrobial Resistance* **27**, (2021).
2. Kraftova, L. *et al.* Evidence of an epidemic spread of KPC-producing Enterobacterales in Czech hospitals. *Scientific Reports* **11**, (2021).
3. Cahill, N. *et al.* Hospital effluent: A reservoir for carbapenemase-producing Enterobacterales? *Science of the Total Environment* **672**, (2019).
4. Subramanya, S. H. *et al.* Detection and characterization of ESBL-producing Enterobacteriaceae from the gut of subsistence farmers, their livestock, and the surrounding environment in rural Nepal. *Scientific Reports* **11**, (2021).
5. Abuoun, M. *et al.* A genomic epidemiological study shows that prevalence of antimicrobial resistance in enterobacterales is associated with the livestock host, as well as antimicrobial usage. *Microbial Genomics* **7**, (2021).
6. Díaz-Gavidia, C. *et al.* Isolation of Ciprofloxacin and Ceftazidime-Resistant Enterobacterales From Vegetables and River Water Is Strongly Associated With the Season and the Sample Type. *Frontiers in Microbiology* **12**, (2021).
7. Buchy, P. *et al.* Impact of vaccines on antimicrobial resistance. *International Journal of Infectious Diseases* vol. 90 Preprint at https://doi.org/10.1016/j.ijid.2019.10.005 (2020).
8. Ruppé, E. *et al.* From genotype to antibiotic susceptibility phenotype in the order Enterobacterales: a clinical perspective. *Clinical Microbiology and Infection* **26**, (2020).
9. Sheppard, A. E. *et al.* Nested Russian doll-like genetic mobility drives rapid dissemination of the carbapenem resistance gene blakpc. *Antimicrobial Agents and Chemotherapy* **60**, (2016).

509 10. Che, Y. *et al.* Conjugative plasmids interact with insertion sequences to shape the
510 horizontal transfer of antimicrobial resistance genes. *Proc Natl Acad Sci U S A* **118**,
511 (2021).
512 11. Shaw, L. P. *et al.* Niche and local geography shape the pangenome of wastewater-and
513 livestock-associated Enterobacteriaceae. *Science Advances* **7**, (2021).
514 12. Orlek, A. *et al.* Plasmid classification in an era of whole-genome sequencing:
515 Application in studies of antibiotic resistance epidemiology. *Frontiers in Microbiology*
516 vol. 8 Preprint at https://doi.org/10.3389/fmicb.2017.00182 (2017).
517 13. Matlock, W. *et al.* Genomic network analysis of environmental and livestock F-type
518 plasmid populations. *ISME Journal* **15**, (2021).
519 14. Redondo-Salvo, S. *et al.* Pathways for horizontal gene transfer in bacteria revealed by
520 a global map of their plasmids. *Nature Communications* **11**, (2020).
521 15. Mounsey, O. *et al.* Limited phylogenetic overlap between fluoroquinolone-resistant
522 Escherichia coli isolated on dairy farms and those causing bacteriuria in humans living
523 in the same geographical region. *Journal of Antimicrobial Chemotherapy* **76**, (2021).
524 16. Hanage, W. P. Two health or not two health? That is the question. *MBio* **10**, e00550-19
525 (2019).
526 17. Ludden, C. *et al.* One health genomic surveillance of escherichia coli demonstrates
527 distinct lineages and mobile genetic elements in isolates from humans versus livestock.
528 *mBio* **10**, (2019).
529 18. Shen, C. *et al.* Dynamics of mcr-1 prevalence and mcr-1-positive Escherichia coli after
530 the cessation of colistin use as a feed additive for animals in China: a prospective
531 cross-sectional and whole genome sequencing-based molecular epidemiological study.
532 *The Lancet Microbe* **1**, e34–e43 (2020).

19. Hilpert, C., Bricheux, G. & Debroas, D. Reconstruction of plasmids by shotgun sequencing from environmental DNA: Which bioinformatic workflow? *Briefings in Bioinformatics* **22**, (2021).

20. Wang, R. *et al.* The global distribution and spread of the mobilized colistin resistance gene mcr-1. *Nat Commun* **9**, 1–9 (2018).

21. Sekizuka, T. *et al.* Complete sequencing of the bla NDM-1-positive IncA/C plasmid from Escherichia coli ST38 isolate suggests a possible origin from plant pathogens. *PLoS ONE* **6**, (2011).

22. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Computational Biology* **13**, (2017).

23. Lipworth, S. *et al.* Ten-year longitudinal molecular epidemiology study of Escherichia coli and Klebsiella species bloodstream infections in Oxfordshire, UK. *Genome Medicine* **13**, (2021).

24. Rozwandowicz, M. *et al.* Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. *Journal of Antimicrobial Chemotherapy* **73**, (2018).

25. Rodríguez-Beltrán, J., DelaFuente, J., León-Sampedro, R., MacLean, R. C. & San Millán, Á. Beyond horizontal gene transfer: the role of plasmids in bacterial evolution. *Nature Reviews Microbiology* vol. 19 Preprint at https://doi.org/10.1038/s41579-020-00497-1 (2021).

26. Pesesky, M. W., Tilley, R. & Beck, D. A. C. Mosaic plasmids are abundant and unevenly distributed across prokaryotic taxa. *Plasmid* **102**, (2019).

27. David, S. *et al.* Integrated chromosomal and plasmid sequence analyses reveal diverse modes of carbapenemase gene spread among Klebsiella pneumoniae. *Proc Natl Acad Sci U S A* **117**, (2020).

558 28. Forster, S. C. *et al.* A human gut bacterial genome and culture collection for improved
559 metagenomic analyses. *Nature Biotechnology* **37**, (2019).
560 29. Sender, R., Fuchs, S. & Milo, R. Revised Estimates for the Number of Human and
561 Bacteria Cells in the Body. *PLoS Biology* **14**, (2016).
562 30. Rodrigues, C. *et al.* High Prevalence of Klebsiella pneumoniae in European Food
563 Products: a Multicentric Study Comparing Culture and Molecular Detection Methods.
564 *Microbiology Spectrum* **10**, (2022).

565

566

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

(a) Number of chromosomes and plasmids by niche, stratified by isolate genus. **(b)** Map of approximate, relative distances between sampling sites, coloured by niche (human bloodstream infection [BSI], livestock-associated (cattle, pig, poultry, and sheep faeces, soils nearby livestock sites), and wastewater treatment work (WwTW)-associated sources (influent, effluent, waterways upstream/downstream of effluent outlets). Number in circles indicates how many of the *n*=1,458 isolates are from that location. **(c)** Sampling timeframe for BSI and REHAB (non-BSI) isolates.

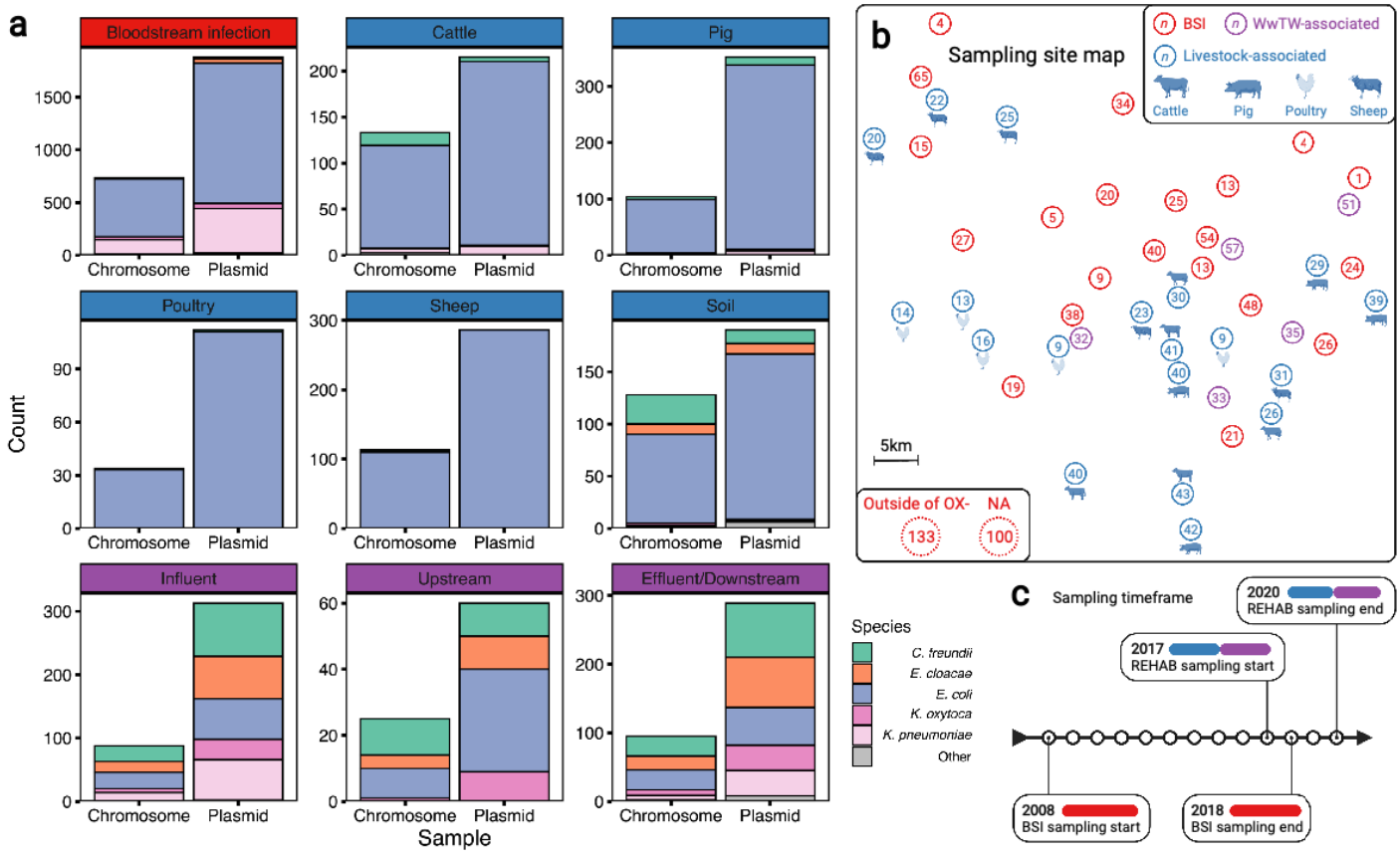


Fig. 2 Cross-niche, near-identical plasmids.

(a) Size of cross-niche, near-identical plasmid groups, coloured by niche (total *n*=84 plasmids). Median length (bp) of plasmids within groups increases from left to right. **(b)** Proportion of plasmid host species by group. **(c)** Predicted mobility of plasmid. **(d)** AMR gene carriage in plasmid.

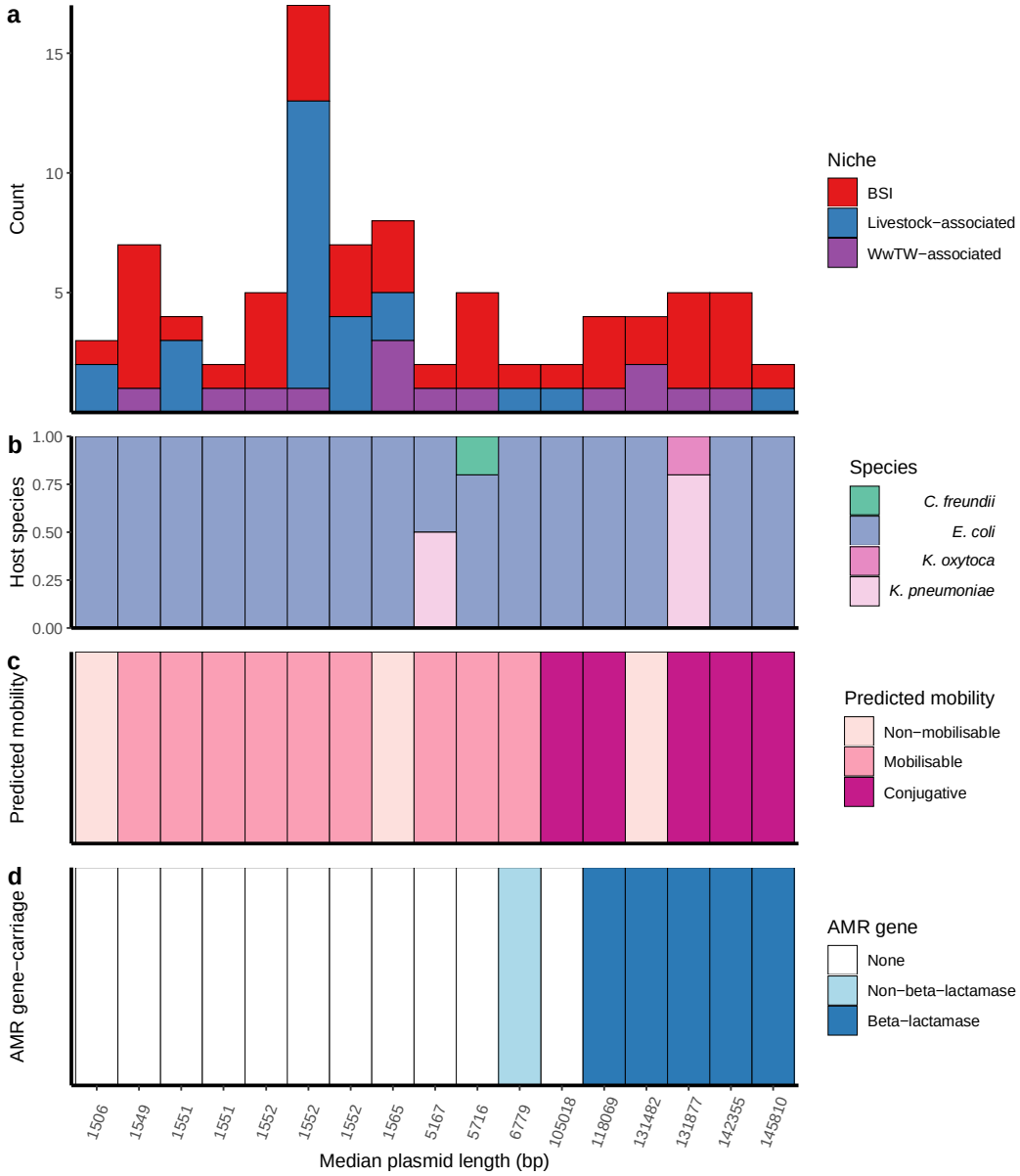
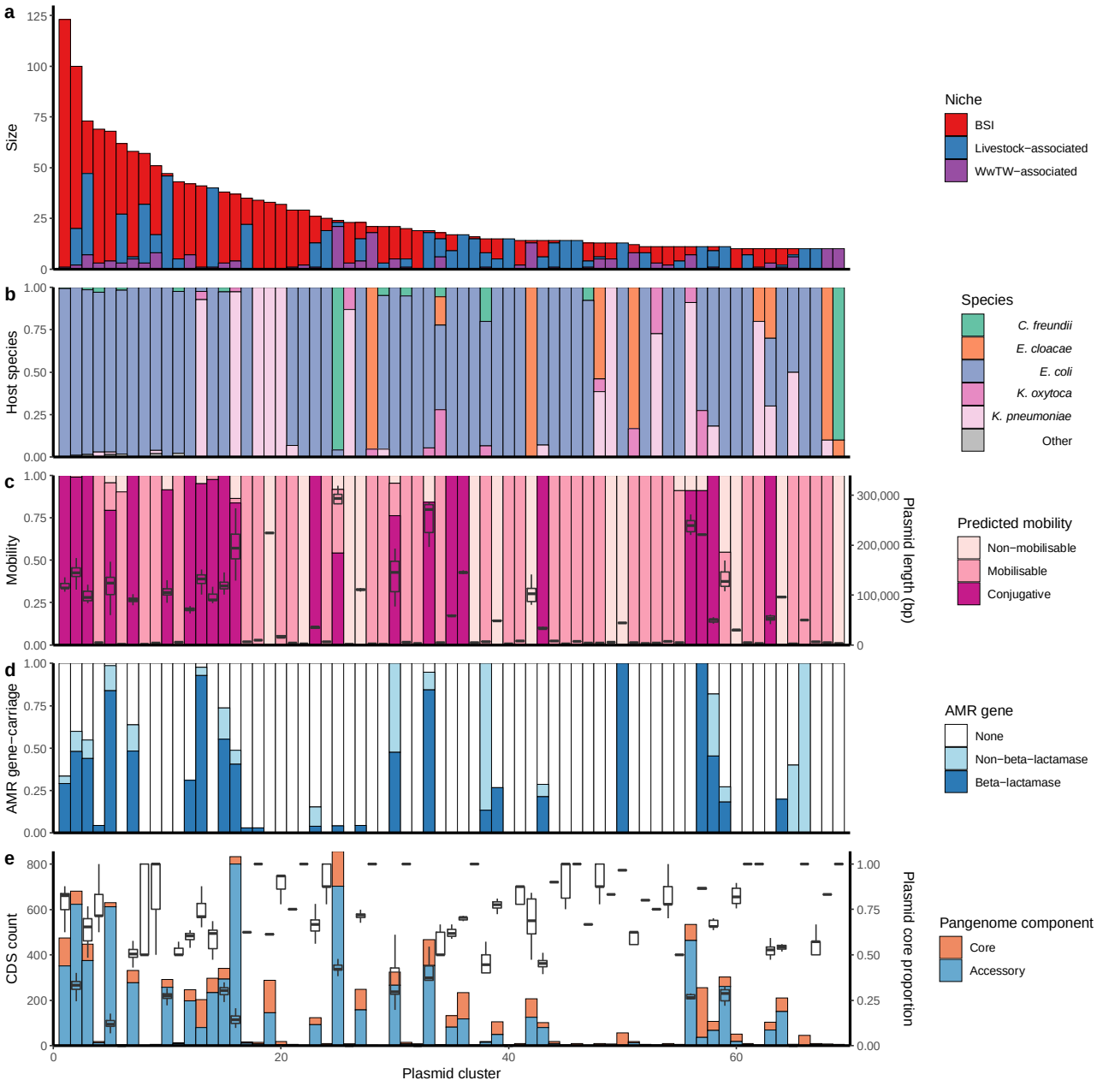


Fig. 3. Genetically similar plasmids share between niches

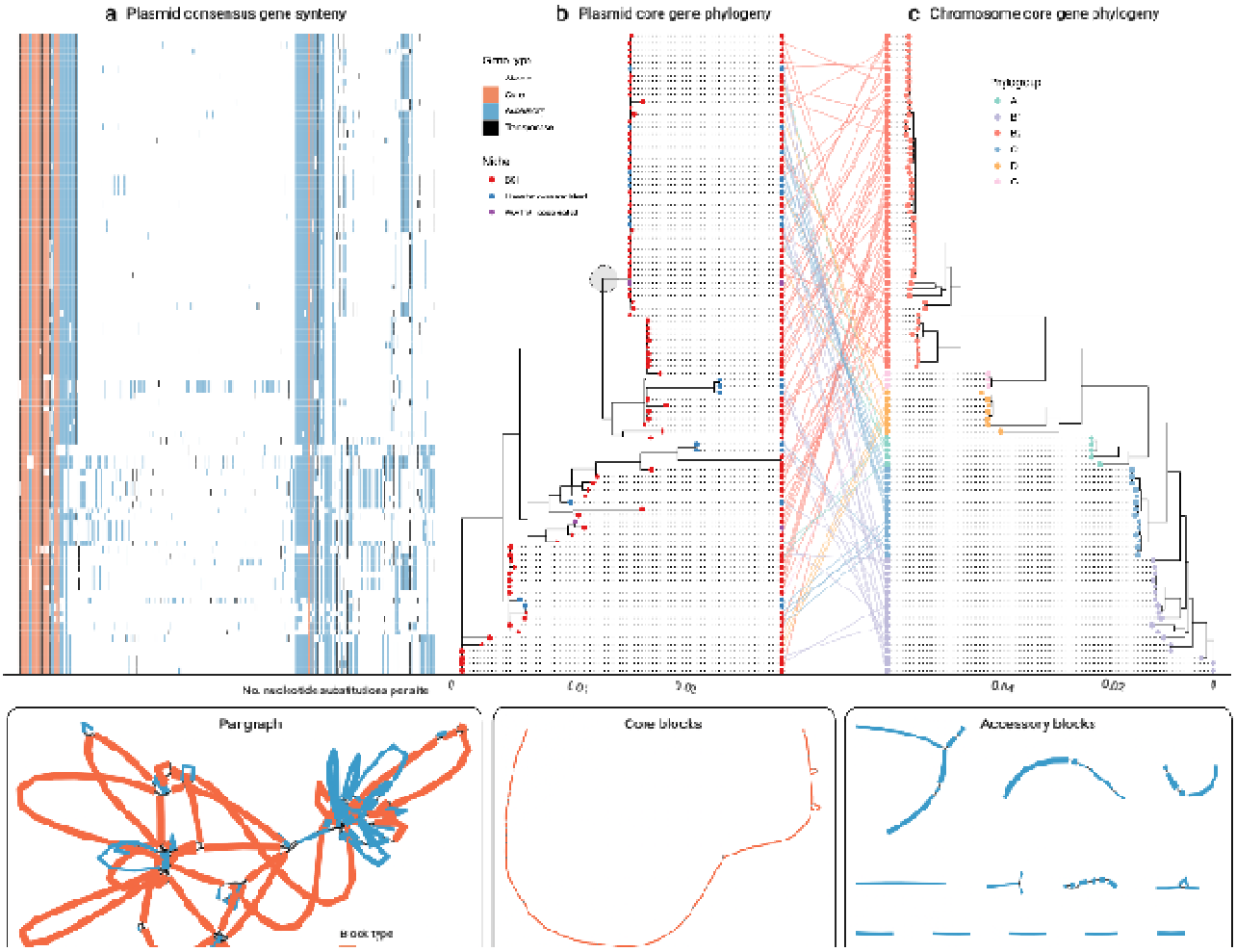
(a) Size of plasmid clusters with at least 10 members, coloured by niche. Size of clusters decreases from left to right. (**b)** Proportion of plasmid host species by cluster. **(c)** Plasmid mobility class and size: Left hand axis shows proportion of plasmids with a predicted mobility class by cluster. 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)

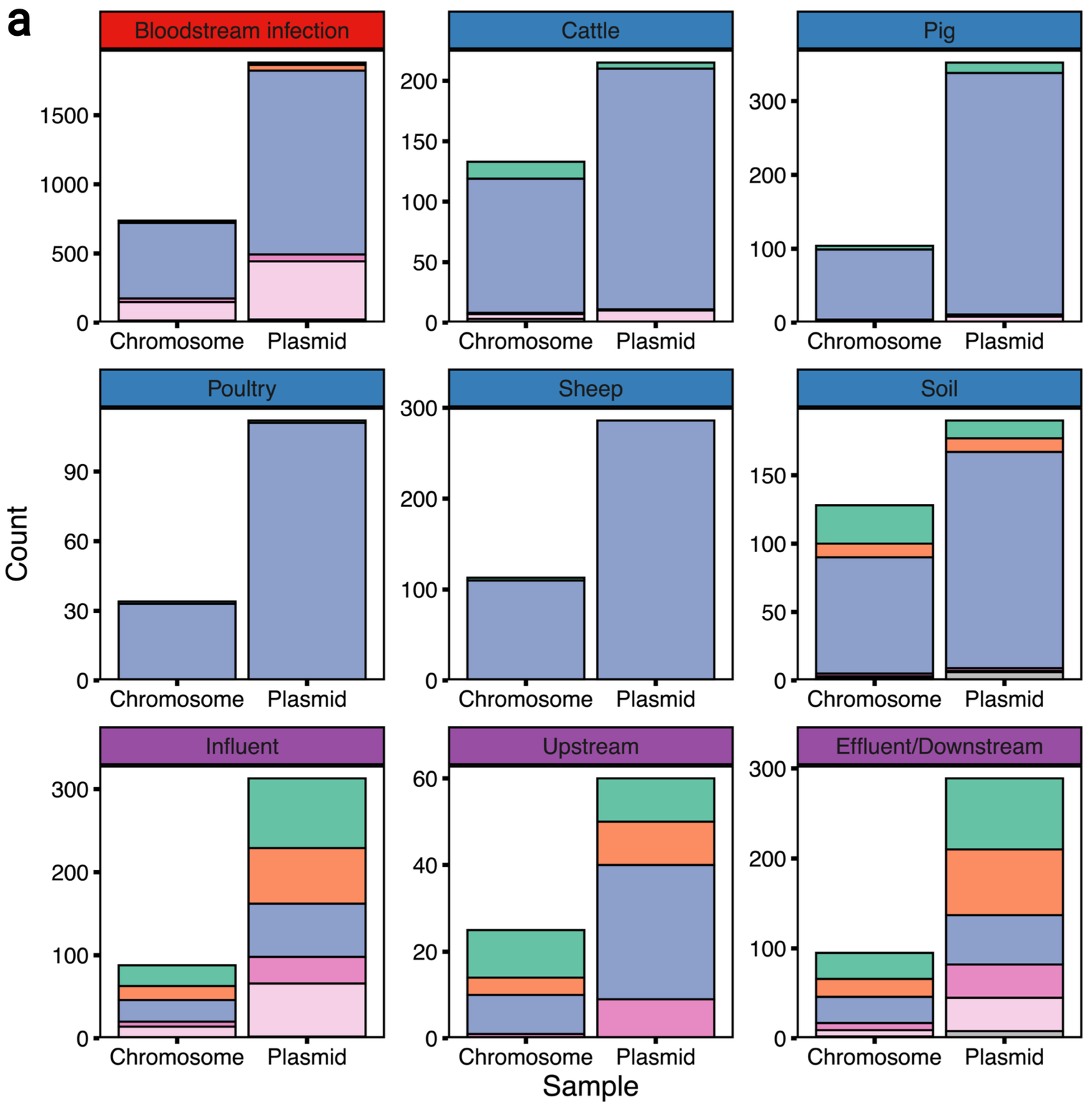


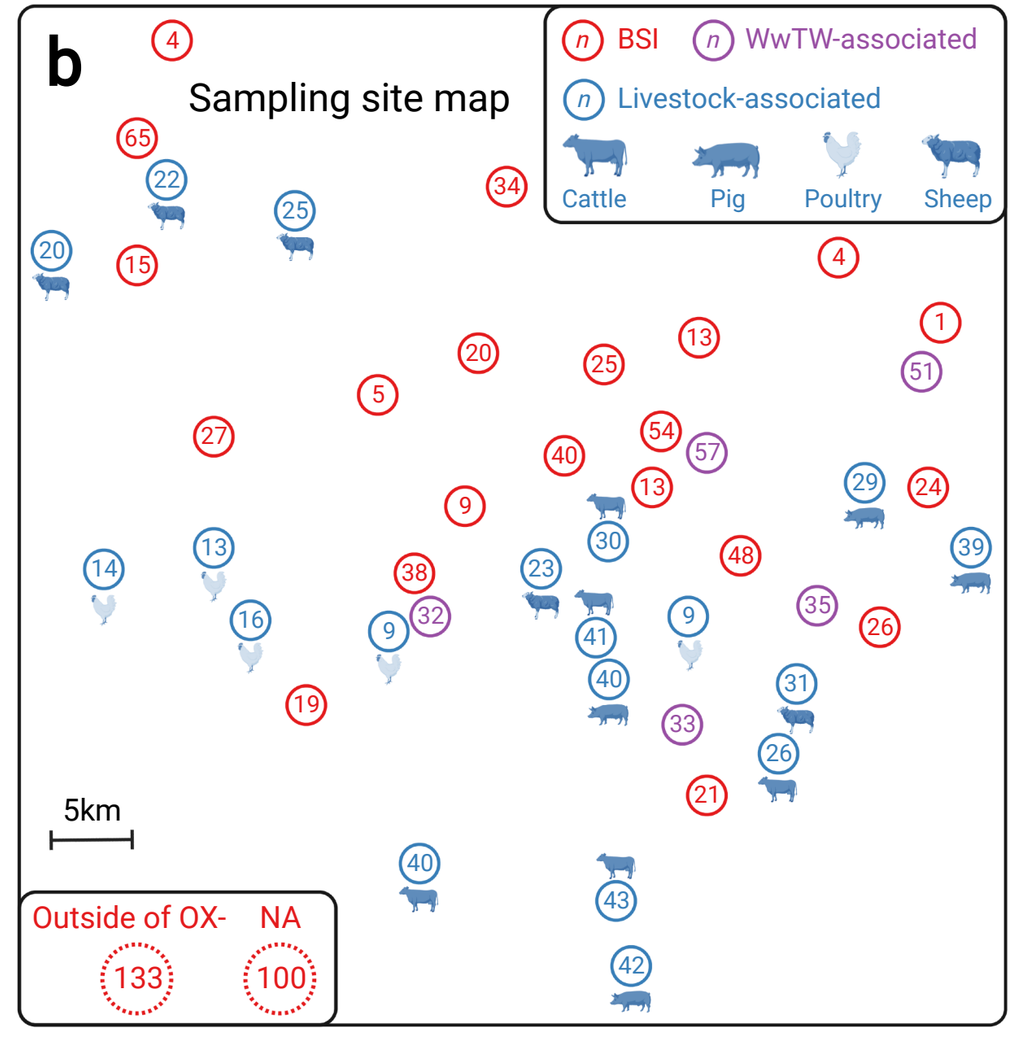
588 boxplots by cluster.

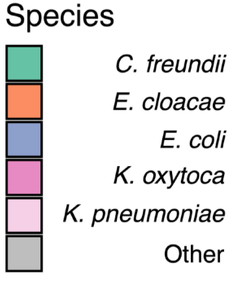
Fig. 4. Cluster 2 plasmid and host evolution

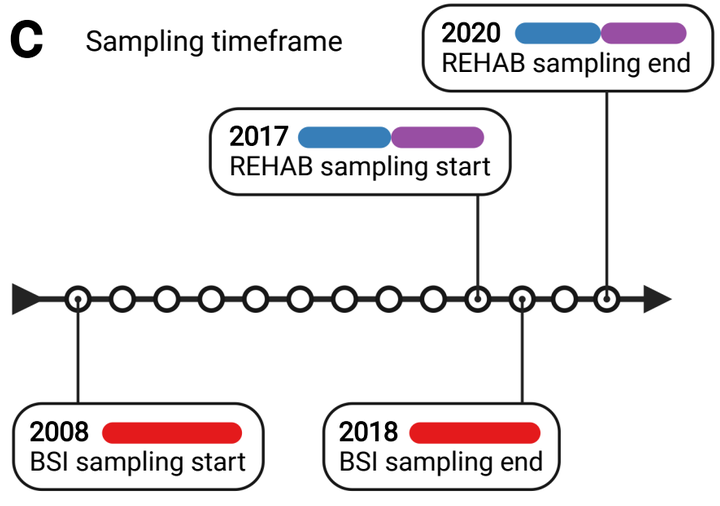
(a) Consensus gene ordering for plasmid cluster 2, coloured by gene type (total *n*=99 plasmids; *n*=1 *S. enterica* isolate omitted). Genes are coloured by core, accessory, or transposase. **(b)** Plasmid core gene phylogeny with tips coloured by sampling niche. The grey circle highlights the clade of *n*=44 plasmids which were further analysed. **(c)** Plasmid host chromosome core gene phylogeny with tips coloured by sampling niche. 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).

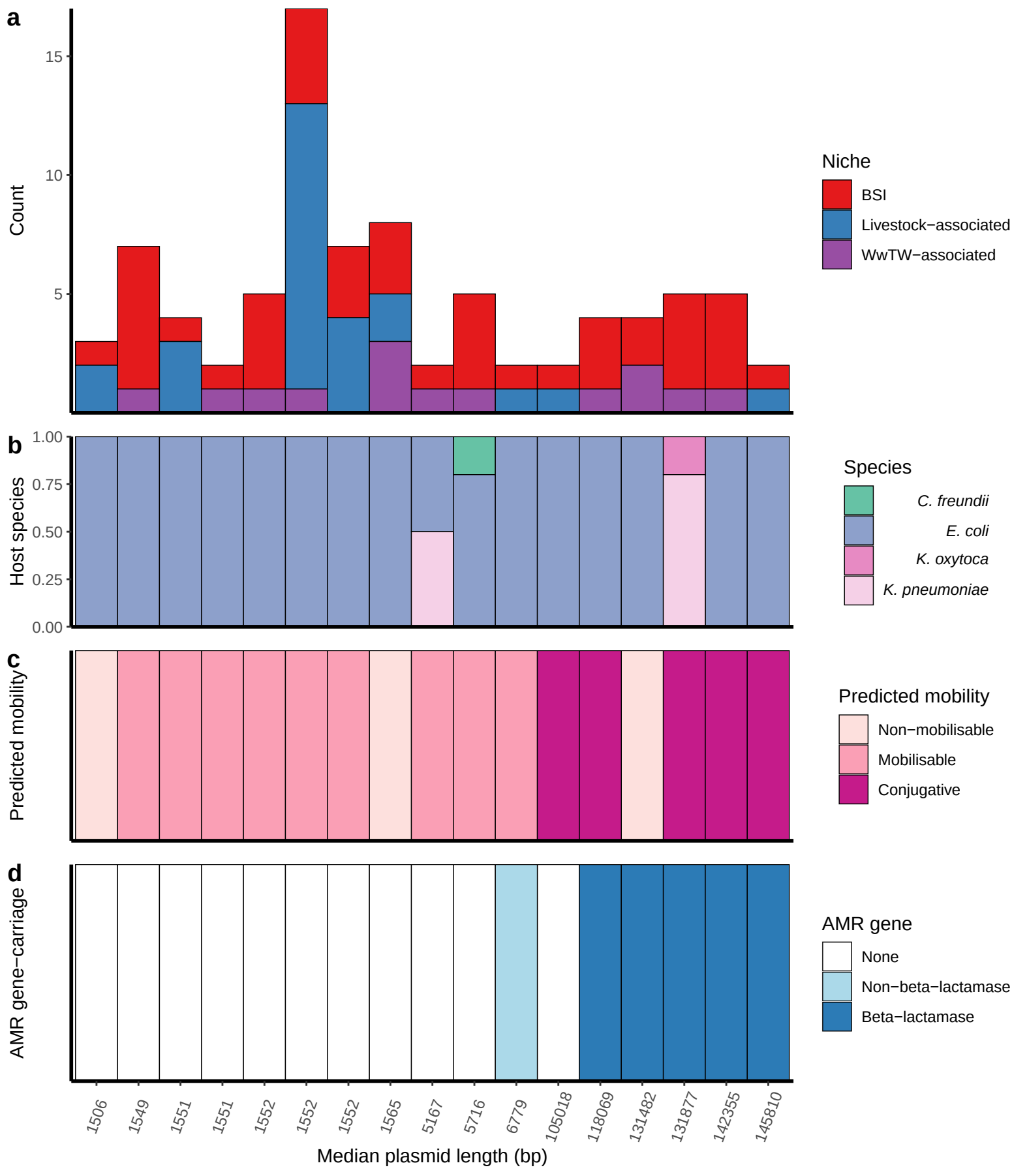


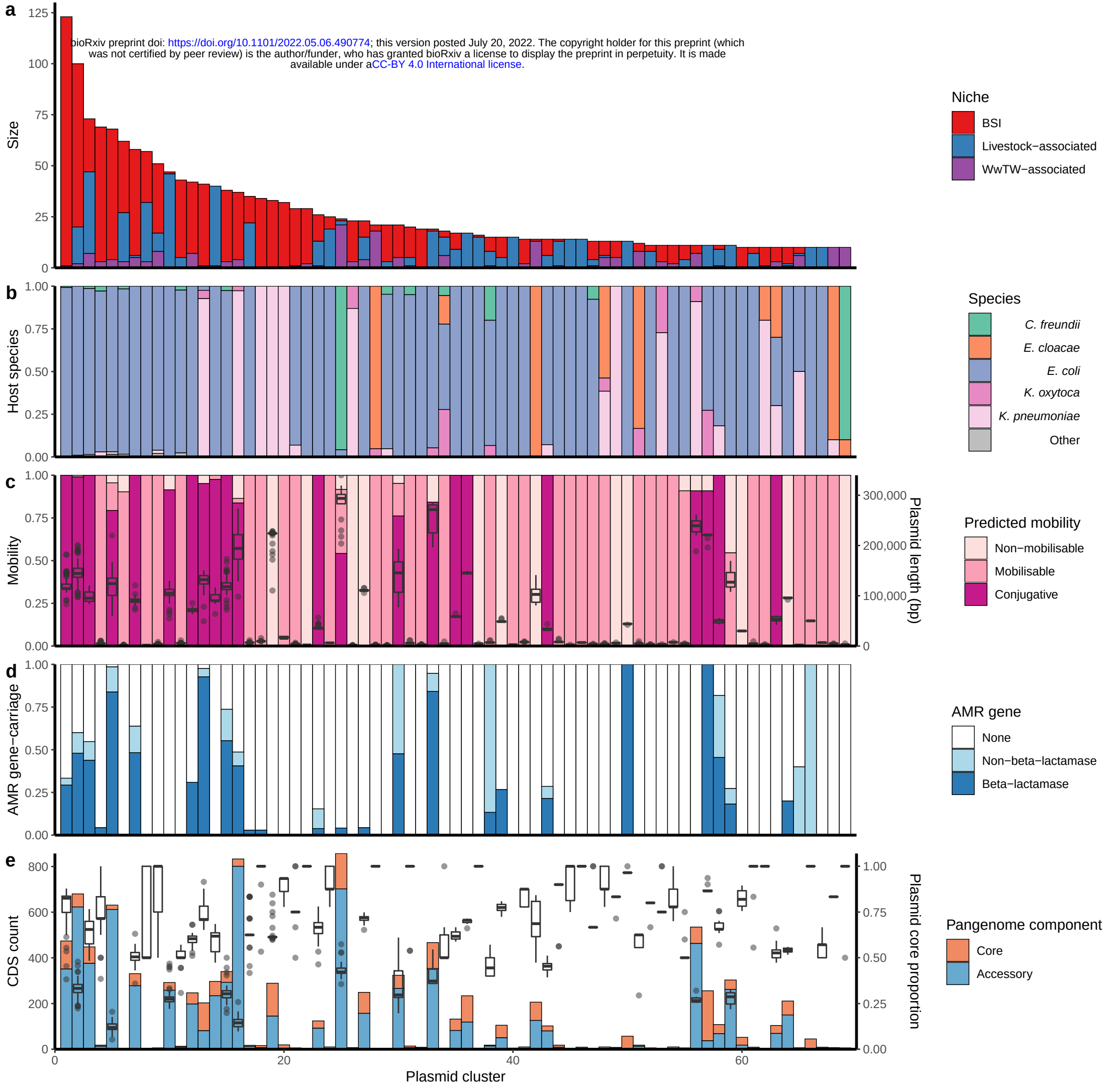

a
Bloodstream infection
Cattle
Pig
Poultry
Sheep
Soil
Influent
Upstream
Effluent/Downstream
Chromosome
Plasmid
Count
Sample

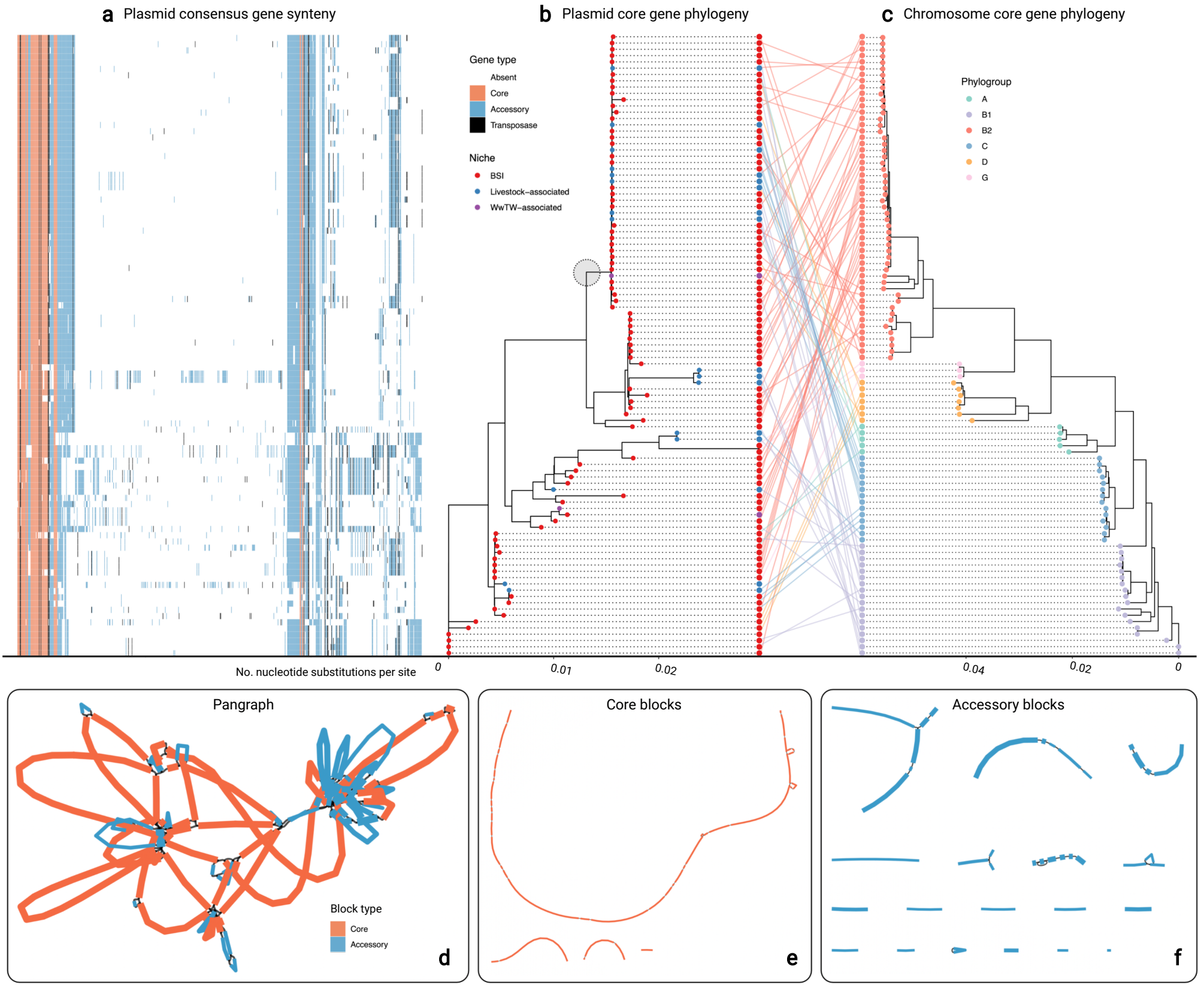

b
Sampling site map
n BSI
n WwTW-associated
n Livestock-associated
Cattle
Pig
Poultry
Sheep
5km
Outside of OX-
133
NA
100


Species
C. freundii
E. cloacae
E. coli
K. oxytoca
K. pneumoniae
Other


c
Sampling timeframe
2020
REHAB sampling end
2017
REHAB sampling start
2008
BSI sampling start
2018
BSI sampling end


a
Count
15
10
5
Niche
BSI
Livestock−associated
WwTW−associated
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
105018
118069
131482
131877
142355
145810
Median plasmid length (bp)




a Plasmid consensus gene synteny
b Plasmid core gene phylogeny
c Chromosome core gene phylogeny
Gene type
Absent
Core
Accessory
Transposase
Niche
BSI
Livestock–associated
WwTW–associated
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