Landscape genomics of Escherichia coli in livestock-keeping households across a rapidly developing urban city

Dishon Muloi
International Livestock Research Institute https://orcid.org/0000-0002-6236-2280
Bryan Wee
University of Edinburgh https://orcid.org/0000-0001-6135-5903
Deirdre McClean
University of Edinburgh https://orcid.org/0000-0002-5857-9081
Melissa Ward
Centre for Immunity, Infection and Evolution, University of Edinburgh
Louise Pankhurst
University of Oxford
Hang Phan
Nuffield Department of Clinical Medicine, University of Oxford
Alasdair Ivens
Institute of Infection and Immunology Research and Centre for Immunity, Infection and Evolution, School of Biological Sciences, University of Edinburgh
Velma Kivali
International Livestock Research Institute
Alice Kiyonga
International Livestock Research Institute
Christine Ndinda
International Livestock Research Institute
Nduhiu Gitahi
University of Nairobi
Tom Ouko
Kenya Medical Research Institute
James Hassell
International Livestock Research Institute
Titus Imboma
National Museums of Kenya
James Akoko
International Livestock Research Institute

Maurice Karani
International Livestock Research Institute
Samuel Njoroge
Centre for Microbiology Research, Kenya Medical Research Institute https://orcid.org/0000-0001-6965-3681
Patrick Muinde
International Livestock Research Institute
Yukiko Nakamura
Research Center for Zoonosis Control, Hokkaido University https://orcid.org/0000-0002-3379-0526
Lorren Alumasa
International Livestock Research Institute
Erin Furmaga
Department of Epidemiology, Columbia University
Titus Kaitho
Veterinary Services Department, Kenya Wildlife Service
Elin Öhgren
Uppsala University
Fredrick Amanya
International Livestock Research Institute
Allan Ogendo
International Livestock Research Institute
Daniel Wilson
Nuffield Department of Population Health, University of Oxford
Judy Bettridge
Natural Resources Institute, University of Greenwich
John Kiiru
Centre for Microbiology Research, Kenya Medical Research Institute
Catherine Kyobutungi
APHRC
Cecilia Tacoli
International Institute for Environment and Development
Erastus Kang'ethe
University of Nairobi
Julio Davila
University College London
Samuel Kariuki
samkariuki2@gmail.com https://orcid.org/0000-0003-3209-9503
Timothy Robinson
FAO

Jonathan Rushton
Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool
Mark Woolhouse
University of Edinburgh https://orcid.org/0000-0003-3765-8167
Eric Fèvre (Eric.Fevre@liverpool.ac.uk)
International Livestock Research Institute

Article

Keywords: Zoonoses, Antimicrobial Resistance, Epidemiological Interface, Bacterial Strain Sharing, Resistome Similarity

Posted Date: February 26th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-172737/v1

Version of Record: A version of this preprint was published at Nature Microbiology on March 14th, 2022. See the published version at https://doi.org/10.1038/s41564-022-01079-y.

Landscape genomics of *Escherichia coli* in livestock-keeping households across a rapidly developing urban city.

Dishon M. Muloi[1,2,3+], Bryan A. Wee[1+], Deirdre M. H. McClean[1], Melissa E. Ward[3,4], Louise Pankhurst[4], Hang Phan[4], Alasdair Ivens[3], Velma Kivali[2], Alice Kiyonga[2], Christine Ndinda[2], Nduhiu Gitahi[5], Tom Ouko[6], James Hassell[2,7], Titus Imboma[8], James Akoko[2], Maurice Karani[2], Samuel Njoroge[6], Patrick Muinde[2], Yukiko Nakamura[9], Lorren Alumasa[2], Erin Furmaga[10], Titus Kaitho[11], Elin Öhgren[12], Fredrick Amanya[2], Allan Ogendo[2], Daniel Wilson[13], Judy Bettridge [14], John Kiiru[6], Catherine Kyobutungi[15], Cecila Tacoli[16], Erastus Kang'ethe[5], Julio Davila[17], Sam Kariuki[6], Timothy Robinson[18], Jonathan Rushton[7], Mark E. J. Woolhouse[1*], Eric Fèvre[2,7*]

1. Usher Institute, University of Edinburgh, Edinburgh, United Kingdom
2. International Livestock Research Institute, Nairobi, Kenya
3. Centre for Immunity, Infection and Evolution, University of Edinburgh, Edinburgh, United Kingdom
4. Nuffield Department of Clinical Medicine, University of Oxford, John Radcliffe Hospital, Oxford, United Kingdom
5. University of Nairobi, Nairobi, Kenya
6. Kenya Medical Research Institute, Nairobi, Kenya
7. Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Neston, United Kingdom
8. National Museums of Kenya, Nairobi, Kenya
9. Research Center for Zoonosis Control, Hokkaido University, Japan
10. Department of Epidemiology, Columbia University, New York, United States
11. Veterinary Services Department, Kenya Wildlife Service, Kenya
12. Uppsala University, Uppsala, Sweden
13. Big Data Institute, Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom
14. Natural Resources Institute, University of Greenwich, Chatham Maritime, United Kingdom
15. African Population Health Research Centre, Nairobi, Kenya
16. International Institute for Environment and Development, London, United Kingdom
17. The Bartlett Development Planning Unit, Faculty of the Built Environment, University College London, London, United Kingdom
18. Food and Agriculture Organization of the United Nations, Rome, Italy

[+] These authors contributed equally

* Corresponding authors

Abstract

The keeping of livestock has been posited as a risk factor for the emergence of zoonoses and the spread of antimicrobial resistance. However, quantitative evidence regarding the major sources of pathogenic and drug-resistant bacteria and transmission routes between hosts remains lacking. In the largest epidemiological study of this nature to date, we sampled *Escherichia coli* from humans, livestock, food, wildlife and the environment of 99 households across Nairobi, Kenya to gain a deeper understanding of sharing of bacteria among hosts and potential reservoirs. By analysing whole genome sequencing data from 1,338 *E. coli* isolates, we reconstruct sharing patterns for the sampled *E. coli* and its antimicrobial resistance determinants. We find that the diversity and sharing patterns of *E. coli* is heavily structured by household, which is the primary epidemiological interface for bacterial strain sharing. Strain sharing within households was strongly shaped by host type. We also find evidence for inter-household and inter-host sharing, and importantly, between humans and animals, although this occurs much less frequently. We find similar strain sharing patterns for the *E. coli* accessory genome, suggesting that it is shaped by recent evolutionary history and is strongly associated with the core genome. Resistome similarity, however, were quite differently distributed across host and household, consistent with their being driven by shared exposure to antimicrobials. Our results indicate that there is potential for the exchange of bacteria between humans, livestock and wildlife in the same household in a tropical urban setting, with wider mixing occurring over a period of months or years, but this does not drive the distribution of antimicrobial resistance.

Introduction

The spread of bacterial pathogens and antimicrobial resistance (AMR) across human and animal populations present a significant and growing threat to global health and economic development. Identifying risk factors for emergence and spread is one of epidemiology's most important challenges. Many recent pandemics and newly emergent infectious diseases have animal origins (*1*, *2*) and are associated with rapidly urbanising environments (*3*, *4*). The dynamic interfaces between humans, domestic livestock and wild animals act as conduits by which humans can be exposed to zoonotic pathogens and AMR in an environment with inadequate sanitation infrastructure, limited access to appropriate and effective drugs, and unregulated antimicrobial usage (*5*–*8*).

The importance of livestock to the transmission of bacteria and AMR remains unclear(*9*). The practice of keeping livestock, particularly in urban settings, has been described as a risk factor for the emergence and spread of zoonoses (*10*, *11*). Antimicrobial agents used in human medicine are also used for growth promotion, disease prevention and treatment in livestock, enhancing selection pressures on bacterial pathogens for AMR emergence and spread.

Wild birds and mammals have also been documented to carry and exchange drug-resistant bacteria with livestock and humans (*6*, *12*, *13*). The rapid expansion of urban environments into previously pristine or sparsely populated natural landscapes also increases the potential for greater contact between wildlife, humans and livestock which can provide conduits for microbiome sharing (*14*).

Fundamental to whole genome sequencing studies is the availability of systematically sampled bacterial isolates obtained from humans, livestock and wildlife across overlapping geographical regions and time-frames, yet data are lacking (*15*). In this study, we sampled the bacterium *Escherichia coli* from humans, livestock and peri-domestic wildlife of 99 households and their environs across 33 sublocations in Nairobi, Kenya, in an epidemiologically structured study. The rapid development of Nairobi's urban landscape is comparable to that of many other cities in the developing world, making it an ideal system in which to explore how people's interactions and co-existence with animals

78 influences pathogen transmission across species. (*16*, *17*). As a common commensal and pathogen of
79 both human and animal populations, as well as ease of culture and wealth of available genetic
80 information, *E.coli* is an ideal organism for this study. Here, we report a genomic investigation of 1,338
81 *E. coli* isolates sourced from humans, livestock and wildlife across Nairobi to elucidate patterns of
82 bacterial strain sharing as a proxy for transmission potential. We test the hypothesis that the
83 distributions of bacterial strains and their genetic pools are limited to particular defined ecological
84 niches (households and hosts) versus an alternative that they display a cosmopolitan distribution – in
85 essence, recapitulating the famous tenet, "Everything is everywhere, but, the environment selects"
86 (*18*). Our study aims to identify risk factors to help inform surveillance strategies that target potential
87 hotspots for strain sharing and AMR transmission between populations in an urban setting, and more
88 broadly, to understand risks associated with transmission of multi-host pathogens in urban settings.

Results

E. coli from humans and animals in Nairobi originate from both global and local lineages.

A total of 1,338 *Escherichia coli* isolates were sequenced as part of this study (Table S1). 311 genomes were obtained from human isolates. 421 genomes isolated from 63 wildlife species, primarily comprised of wild birds (n=245), rodents and bats (n=130). 606 genomes were attained from 13 species of livestock that can be grouped into poultry (n=324), goat and sheep (n=109), cattle (n=61), pig (n=49) and rabbit (n=38) isolates. The isolates were distributed across 33 geographic sublocations spanning the entire urban area of Nairobi with a range of 20 to 63 isolates per sublocation (see supplementary methods). A large fraction of isolates in each sublocation were obtained from a household with livestock (minimum 75%). The sampling protocol also ensured that there was at least one and up to ten isolates from a household without livestock in each sublocation.

The genomes represent all major lineages of the *E. coli sensu stricto* phylogroup in addition to members of the cryptic clade I. The isolates belong to Clermont phylogroups B1 (45%), A (38%), B2 (6%), D (4%), E (2%), and to a lesser extent clades C, F, G and clade I (>1%). Phylogroup A was strongly associated with humans (41% of human isolates) compared to the other host categories. In livestock mammal, wild bird and wild mammal categories, isolates from phylogroup B1 was the most frequently isolated.

A total of 537 unique sequence types (STs), based on the 7-gene Achtman scheme, were represented, with the three most common being ST10 (n= 93, 7%), ST48 (n=64, 5%) and ST155 (n=54, 4%) (Table S2). 139 STs, representing 14% (184/1338) of isolates, have only been found in African countries (Kenya, Madagascar, South Africa, Uganda), based on the genomes that were present in Enterobase at the time this study was carried out. 133 of the 139 Africa-specific STs in this collection, representing 13% (173/1338) isolates, were unique to Kenya and reported for the first time in this study. Most of these novel and unique STs were isolated from livestock (52%, 96/184), and wildlife (34%, 63/184).

A core genome alignment comprising 80,722 nucleotide positions conserved across all 1,338 isolates was used to infer the overall phylogenetic relationship between isolates (Figure 1). Associations between the clustering of isolates at the tips of this phylogeny and host types or sublocations were investigated. We found two clusters that were associated with Kitisuru (n=5) and Karen (n=6) sublocations suggesting localised transmission clusters (Supplementary Figure S1). The cluster in Kitisuru involved two wild birds and three poultry isolates from the same household (KTS089), while the cluster in Karen was made up of cattle and primate isolates from one household (KAN007) and wild bird isolates sampled in other households in the same sublocation (KAN008 and KAN009). Both Kitisuru and Karen represent sublocations of the highest wealth category consisting of detached dwellings with relatively large surrounding compounds.

Three clusters of isolates were found to be significantly associated with goats (n=18), rats (n=14) and rabbits (n=5), respectively. The longer phylogenetic distances of well over 10 core SNPs separating isolates in these hosts suggest transmission over a longer timescale, which is supported by the distribution of these isolates across multiple households and sublocations (Figure 1 and Supplementary Figure S2). The largest cluster associated with a single host type is the ST297/ST9433 caprine cluster that are found in multiple sublocations (Figure S2a). The second largest cluster, belonging to the ST9441 lineage, is so far unique to this dataset and Nairobi. This cluster is also found across at least nine sublocations and significantly associated with wild rodents.

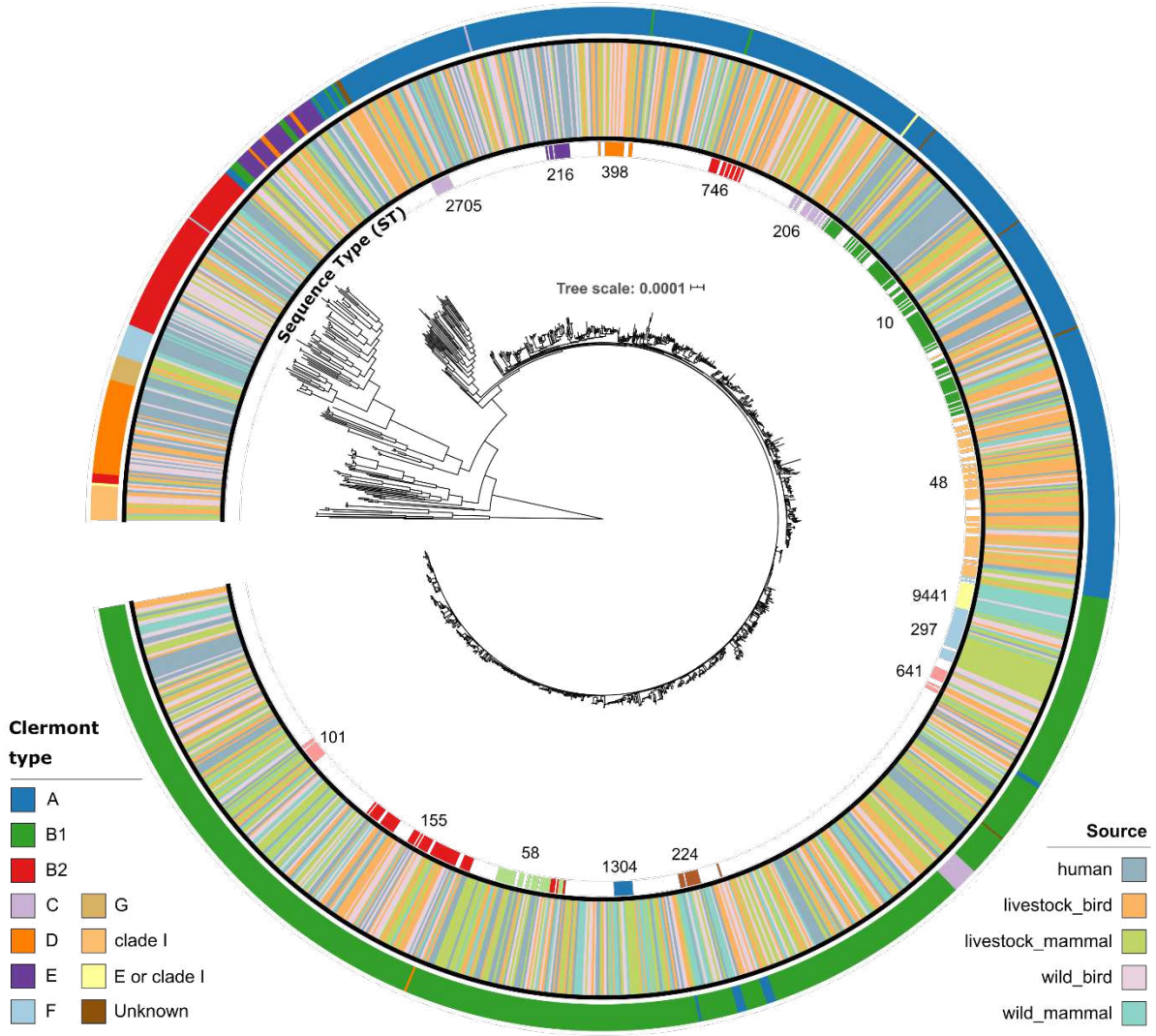

Figure 1. Core genome phylogeny of 1338 E. coli isolates showing (from inner to outer ring): i) Sequence Types (only STs with a minimum 10 isolates are shown), ii) source type of isolate, and iii) Clermont phylotype classifications. The tree is rooted on the Clade I group.

We employed core genome multi locus sequence typing (cgMLST) – a high resolution typing method, which is more reproducible and comparable across larger datasets (*19*). First, we used cgMLST to compare the global diversity of sequenced *E. coli* with the isolates in Nairobi by performing an all-vs-all pairwise comparison of cgMLST distances between the 1,338 Nairobi genomes and 28,382 publicly available whole genome sequences that were annotated with source type, place and year of isolation from Enterobase (downloaded June 2018). We found that the closest related isolates to those found in Nairobi differed by at least 9 cgMLST loci (Supplementary Figure S3). Two ST69 isolates from humans (INT005715 and INT005465) from different sublocations (Kahawa West and South C), differed by 9

and 10 cgMLST loci from two wastewater isolates from Arizona in the United States, respectively (Enterobase Uberstrain accessions: ESC_CA6087AA, ESC_CA6031AA). The number of pairwise connections to isolates in the public database increases considerably between the 25 to 60 cgMLST range due to isolates belonging to the globally dispersed multi-drug resistant ST131 lineage (Supplementary Figure S4a). The intensive sequencing effort of strain ST131, a major cause of multidrug resistant infections in humans, is concentrated in Europe and North America. This explains the disproportionately higher number of links between Nairobi and this part of the world, which is thus a feature of global sampling effort rather than underlying epidemiology (Supplementary Figure S4a). Within this range of 25-60 cgMLST differences between pairs, a large fraction (39%) of pairs also occur between isolates from Nairobi (Supplementary figure 4b) that originate from the two goat (ST297 and ST9433) and rodent-associated (ST9441) lineages, described previously.

Clonal strain sharing of isolates is primarily shaped by household structure.

When the frequency distribution of pairs of isolates differing by less than 100 cgMLST loci is plotted, we find a total of 150 pairs of isolates that differ by 10 or fewer cgMLST alleles to other isolates in our collection (Figure 2). These pairs comprise 187 (14%) isolates, with some isolates involved in multiple pairs. Data on household and host type for these 150 pairs revealed that the majority occur between hosts from the same household (n=101, 67%) and 33% (n=49) involving hosts from different households. Given the low genetic distances and epidemiological context, we refer to these pairs of ≤10 cgMLST as sharing pairs. These sharing pairs are inferred to be evidence of recent strain sharing either by direct transmission or acquisition from a common source.

WGS studies of *E. coli* outbreaks in humans showed that epidemiologically-linked isolates differed by up to four core genome SNPs when isolated within 30 days of each other and when separated by between five to ten core SNPs, this timeframe increases to an average of 8 months (*20*). Although the cgMLST genetic distance used in this analysis is not directly comparable to core single nucleotide polymorphisms (SNP) distances, 96% of the sharing pairs (n=144) were separated by four or fewer

core genome SNPs and almost all pairs (99%, n=149) by a maximum of 10 core SNPs. Therefore, the genetic diversity of isolates within the same household agrees with other examples of epidemiologically-linked *E. coli* in other settings and we estimate that length of evolutionary time separating two isolates from within the same household is within the range of several months to several years.

65% (n=97) of the pairs were between isolates from the same host category (57 [38%] within livestock, 26 [17%] within wildlife, 14 [9%] within humans) and the remaining 36% (n=53) found between host category (38 [25%] between wildlife and livestock [W-L], 10 [6%] between human and livestock [H-L], 5 [3%] between human and wildlife [H-L]). Further details on the breakdown of these sharing pairs are provided in Table S1. No correlation was evident between sharing pairs and particular *E. coli* lineages, as sharing pairs were distributed across the phylogeny for all six (H-H, L-H, L-L, W-H, W-L and W-W) categories of sharing (Supplementary Figure S5). However, in seven cases, wildlife isolates that were implicated in sharing pairs were found in the same cluster as isolates involved in sharing pairs with other host categories (Supplementary Figure S5). Similarly, there was no significant correlation between host-type sharing and the geographical distance between households ($X^2 = 8.83$, $p = 0.64$, Kruskal Wallis) (Supplementary Figure S6).

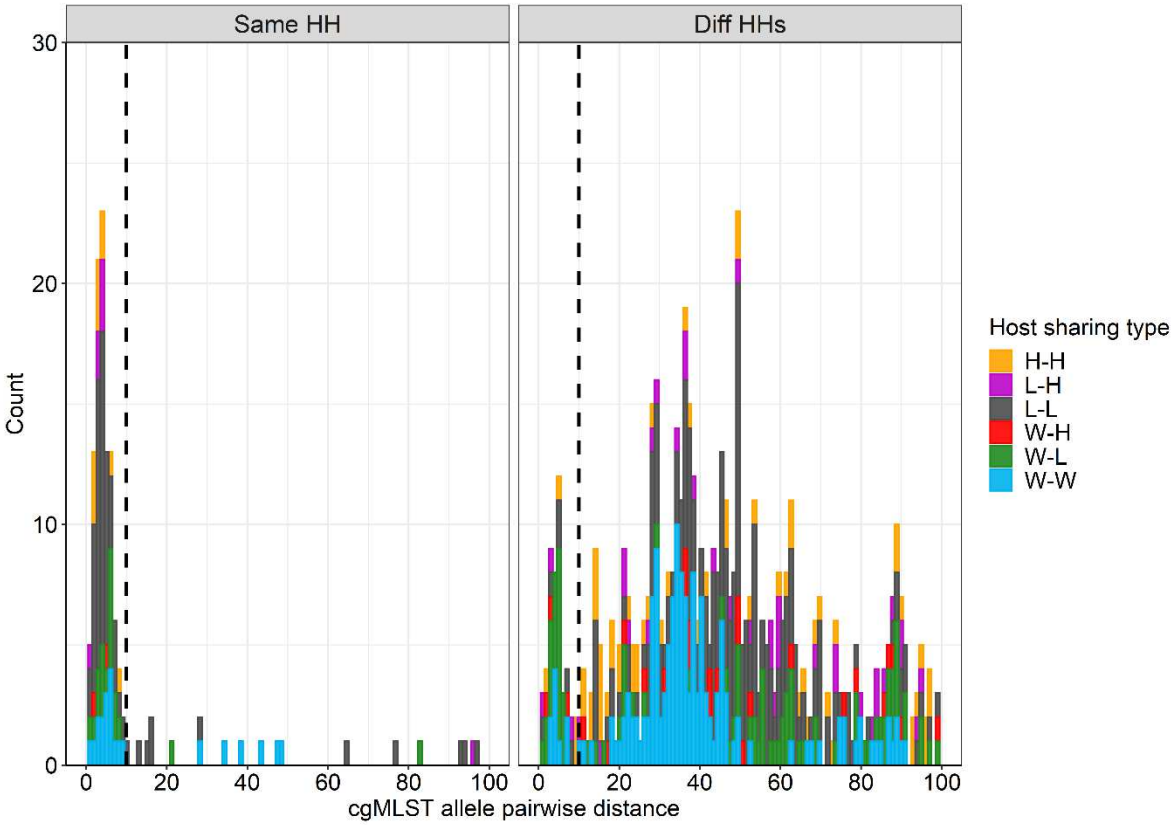

Figure 2. Frequency distribution of pairwise distances between isolates from the same households (left) and different (right) households. The sources of isolates in each pair is indicated by the colour. Only pairs that are closer than 100 cgMLST loci apart are shown. H = human, L = livestock and W= wildlife

To understand the contribution of intra-host diversity on the numbers of detected sharing pairs we obtained multiple isolates per host for a subset of six households (KHW050, KIG019, KIG020, KOR058, UTH029 and VIW002). Ten isolates per host were sequenced from two adult humans, a chicken and a goat from the same household. Comparing these isolates in the context of the larger dataset increased the number of human-livestock sharing pairs within households by only 7 pairs. These pairs were due to 2 clusters of clonal isolates found in either poultry or ruminants in 2 households (UTH029 and VIW002). Between households, we found only two more sharing pairs; involving a human and another human, and between a goat and a chicken. This lack of magnitudinous change in the distribution of sharing pairs provides confidence in our single sample approach. Our single isolate per host approach to sampling, maximises the ability to detect sharing potential among humans and livestock, while minimising the cost and effort of additional sampling and culturing.

Husbandry is a risk factor for *E. coli* sharing between humans and animals

We identified ten sharing pairs involving human and livestock isolates belonging to STs that are not host-restricted and have been associated with a variety of sources and host species (Table 1).

All sharing pairs involved human males (p-value = 0.003, Fisher's Exact test). Six of the ten sharing pairs involved humans and livestock in the same household while four humans (not keeping livestock) shared bacteria with livestock from other households. Six of seven persons (we lacked data for three people) had direct contact with livestock through collecting eggs, slaughter, milking or handling but one person had no history of livestock contact (Table 1).

Table 1 Details of humans involved in bacterial sharing with livestock (<10 cgMLST)

Sharing pair	Livestock host	cgMLST distance	ST	Household	Livestock keeping status	Human livestock handling status	Gender
1	Chicken	1	10	Different	Yes	Yes	Male
2	Goose	1	538	Same	Yes	Yes	Male
3	Chicken	3	23	Different	No	*	Male
4	Cattle	3	6178	Same	Yes	Yes	Male
5	Duck	3	58	Same	Yes	Yes	Male
6	Rabbit	4	9454	Same	Yes	Yes	Male
7	Turkey	4	9454	Same	Yes	Yes	Male
8	Chicken	4	206	Same	Yes	*	Male
9	Turkey	8	1237	Different	Yes	*	Male
10	Chicken	10	48	Different	No	None	Male

* Information not collected

Sharing of *E. coli* core genome, accessory genome and resistome is shaped by host and households

While the sharing threshold for the core genome of ≤10 cgMLST distance, sharing for the pangenome and resistome was based on a Jaccard similarity Index (JI), where a cut-off threshold for sharing was defined in the same way as the core genome. For the pangenome/accessory genome this was determined to be JI ≤ 0.98 (Fig 3c, d). Resistome sharing was defined as JI = 1 (Fig 3e, f), meaning that to be considered a sharing pair two isolates needed to have an identical antibiotic resistance gene profile, with a minimum of two AMR genes in each isolate. Denominator values were based on the

number of pairs of isolates in each category, assuming an equal probability of sharing among isolates. We resampled the observed values to generate expected distributions of events based on the frequencies of these expected values (see Methods for details). From this we were able to assess whether our observed number of sharing pairs fell above, below, or within the range we may expect given the sampling effort.

Household and host category strongly influenced the distribution of sharing of *E. coli* isolates in both the core genome and the pangenome in Nairobi (Figure 3; a-d). Within households, sharing of *E. coli* isolates was consistently higher than expected within the same host category (Figure 3; a, c). No strong pattern was observed between households where the observed shared *E. coli* isolates fell largely within the expected range (Figure 3; b, d). Resistome sharing was predominantly low between different hosts, but high between poultry isolates, irrespective of household structure (Figure 3; e, f). Sharing among poultry in the same household was particularly high across all three definitions of sharing and similarity, i.e. the core, pangenome and resistome (LB-LB in Figure 3).

To further investigate resistome similarity between hosts we performed the same analysis with sharing classed as two isolates sharing resistance genes that confer drug resistance to a given class of antibiotics. We compared 8 classes of antibiotic whose resistance genes were found in the population (Supplementary Figure 7) and found that between households, poultry-poultry sharing continued to be much greater than the expected range (Supplementary Figure 7). Resistome sharing among poultry does not therefore appear to be driven by resistance to a single or few antibiotic classes. Human-human sharing between households was also higher than expected, suggesting similar antibiotic selection pressures on human isolates across the board.

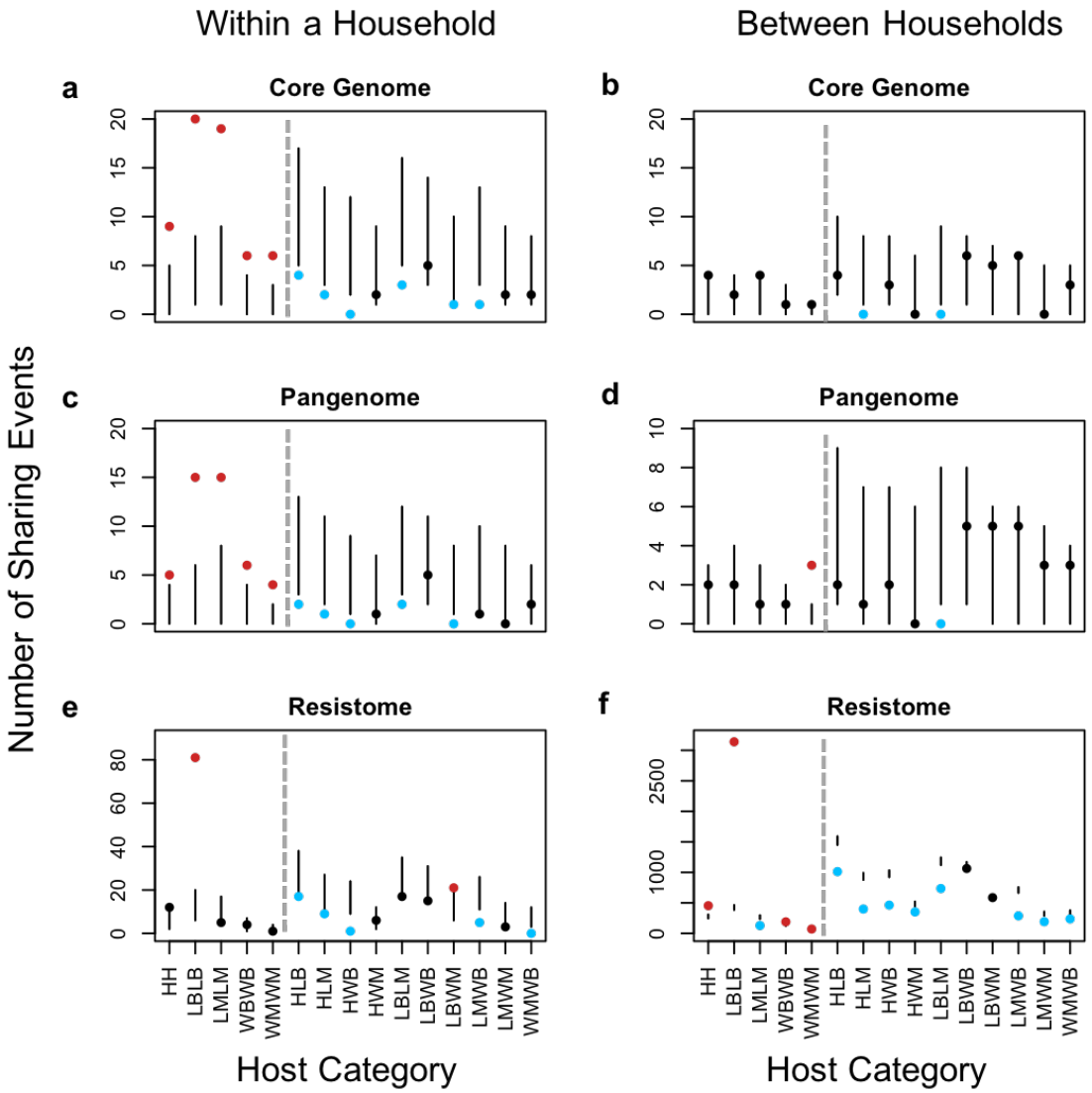

Figure 3. Number of sharing pairs across 15 host category types, within (a, c, e), and between (b, d, f) households for each of the core genome, pangenome, and resistome. Panels show the 95% confidence intervals of the calculated expected distribution using a resampling approach, points depict the observed number of sharing pairs in each category coloured according to whether they fall above (red), below (blue) or within (black) the expected distribution. Hosts in the same category (e.g. human-human) and different categories (e.g. Human-Livestock Bird) are separated by grey dashed lines. Source type of isolate pairs are indicated on the x-axis with either Human (H), Livestock birds (LB), Livestock mammals (LM), Wildlife birds (WB), Wildlife mammals (WM). In each plot, within-category connections are on the left of the grey dotted line and between-category connections are on the right.

Discussion

Our population genomic analysis, explicitly embedded within an epidemiologically structured sampling framework, provides a comprehensive overview of the genomic landscape of *E. coli* in humans, livestock and peri-domestic wildlife in a rapidly developing city. Our findings have implications for understanding the baseline level of bacterial diversity in settings where there is a potential for interaction between humans and animals. Our results reveal extensive strain sharing within households and a lower but detectable level of connectivity among human and animal populations across the urban environment beyond the household.

Isolates from Africa make up less than 3% (n=3626) of the publicly available *E. coli* genome sequences in Enterobase. Our study provides a substantial contribution to the record of *E. coli* diversity in this part of the world with 133 unique and novel STs, in addition to a significant detailed footprint at a city-wide scale. Previous work on the population structure of *E. coli* isolated from human, livestock and wildlife in other both rural and urban settings showed varying degrees of overlap in the genotypes between these populations, driven by frequent contact and close proximity (*13*, *14*, *21*). The wide range of genotyping methods used in these studies, each with varying levels of resolution, makes it difficult to make direct comparisons between studies. Earlier genotyping methods have lower resolution and are less robust (*22*). Other studies measure similarity in microbiome community composition but are less reliable at resolving strain differences between samples (*23*). Our approach combines high-resolution whole genome sequencing with a structured sampling design which captures more accurately the extent of strain sharing in this location.

In our study, we found that household stratification drives clonal strain sharing. Previous studies have shown an important role of the household as a driver for sharing similar microbiomes or bacteria in humans and companion animals (*24–27*). Our findings show that strain sharing can involve humans, livestock and wildlife found in the same household or area.

The use of contemporary isolates in our sampling increased our ability of finding clonal isolates that overlap between hosts, households and sublocations. Previous work using whole genomes either found no overlap or isolates that were separated by more than ten core SNPs, which does not provide strong evidence for a recent sharing event (*28*, *29*). While challenging in practice, we have demonstrated the importance of large-scale structured sampling to understand strain sharing at the population level.

Our comparison of the isolates in our study to isolates from a global collection (Enterobase) revealed that globally dispersed lineages differed by at least 9 cgMLST loci. At this level of divergence or higher, isolates in Nairobi could not be differentiated from globally circulating clones found in other parts of the world. The genetic diversity in the two largest clonal lineages in the dataset (ST297 and ST9441) circulating in rodents and goats was similar to the diversity of ST131 isolates circulating globally. This shows that beyond the 10 cgMLST threshold, epidemiological links from strain sharing or transmission events becomes obscured by the bacterial diversity present in the environment.

Genotype similarity of the core and accessory genome within households is posited to be driven by direct and social contact among individual hosts (*30*, *31*). Consistent with expectation, host type was also demonstrated to be a strong driver in *E. coli* isolate sharing within households (Fig 3). Members of the same host category, particularly in the same household, are more likely to have direct and/or indirect contact within shared environments, creating increased opportunity for bacterial sharing (*14*, *23*, *24*, *30*–*32*).

Eight of the ten observed human-livestock sharing pairs involved poultry. Inhalation and ingestion of faecal dust from poultry has previously been identified as a significant risk in the spread of bacteria from one host to another, both within the poultry populations and with humans working in close contact with them (*33*). Furthermore, it has been previously hypothesised that poultry is likely to be a reservoir of the global epidemic strain, ST131 (*34*, *35*). Humans in direct contact with livestock were more prone to sharing *E. coli* isolates, likely through direct contact with meat and faecal matter.

296 Though the sample size is small, this result is consistent with previous work postulating direct contact
297 as a risk for bacterial sharing events (*26*, *36*). We note that the strong host type signal for *E. coli* sharing
298 within a household (Figure 3a) does not hold true when examining pairs between households (Figure
299 3b). This could be due to a higher diversity of *E. coli* in the wider population, leading to a lower
300 probability of detecting closely related strains.

301 Our resistome sharing analysis also suggests disproportionately higher rates of resistome similarity
302 among poultry irrespective of the household compared to the other host groups. As poultry isolates
303 are phylogenetically diverse, the presence of a common selection pressure could explain this
304 observation. Across Nairobi poultry are routinely exposed to a set regimen of antimicrobial agents (for
305 therapeutic or prophylactic purposes) and such recipes vary minimally geographically from one
306 location to another (*37*). Conversely, a wider range of combinations of antimicrobials are available for
307 use in ruminants and monogastrics, including an array of injectable formulations, and these greatly
308 vary from one farm to another. We also find resistome similarity to be high among human and wildlife
309 isolates, both mammals and birds. The similar availability and usage patterns of antibiotics in the
310 human population across the city could explain the similarity seen in humans, suggesting resistome
311 similarity occurs from prevailing selective pressures rather than spread from a common source. The
312 presence of manure, rubbish and human waste – all contaminated with potentially similar kinds of
313 AMR pathogens and antimicrobials – across the urban landscape of Nairobi provide a conduit for
314 acquisition and/or selection of similar resistome in wildlife, which act as a sink population for AMR
315 (*12*).

316 We observed higher than expected level of accessory genome sharing between wild mammals (bats
317 and rodents), between households, apparently involving divergent lineages as we did not see the same
318 pattern at the core genome level. Other types of wildlife, for example, wild birds around the world
319 have been shown to carry and transmit *E. coli* and should be considered a public health risk (*38*–*40*).
320 Our findings suggest that the role of rodents and bats should also be considered.

Our study design focuses on the breadth of sampling over depth, and as a single isolate is sampled from each host our approach does not account for intra-host diversity. Previous studies on the intra-host diversity of *E. coli* strains were found to be variable across host populations and taking single isolates has the potential to underestimate the number of sharing pairs (*41*). However, we showed that for a subset of six households in our study, increased sampling by ten times had a minimal effect on the number of inter-household and inter-host sharing pairs that were detected. Higher intra-host diversity in different host populations for example, between wildlife and domesticated animals, may reduce the probability of finding sharing pairs in hosts with higher bacterial diversity. Future studies should therefore take into account both inter and intra-host diversity to expand on our findings.

Conclusions

Employing an epidemiologically structured sampling framework and using highly discriminatory whole genome sequencing, our study provides detailed insight into the strain diversity of *E. coli* across a fast-growing African city where livestock-keeping within households is commonplace. To our knowledge, this is one of the largest and most comprehensive survey of the bacterial genomic landscape in an urban environment to date, and serves as a model for epidemiologically structured, targeted sampling and whole genome sequencing of human and animal-borne bacteria. We found evidence of recent clonal sharing between humans and livestock and show that the *E. coli* population structure in humans, livestock and wildlife in this environment is primarily shaped by household and host type, but not by animal husbandry. We also found similarities in the resistome of the isolates that did not match the patterns of shared genomes and presumably reflects common antibiotic usage practices, particularly in poultry. These findings provide empirical support for the hypothesis 'everything is everywhere' (frequent sharing of bacteria and AMR genes between households) but `environment selects' (different households and hosts have different bacterial and resistome persistence). Further work, guided by the finding of where clonal sharing is most likely to be found, will be required to quantify spillover risk associated with the main routes of inter-host transmission.

Methods

Study Site

A cross-sectional study targeting synanthropic wildlife and sympatric human and livestock populations in Nairobi, Kenya was carried out from August 2015 to October 2016 as part of the Urban Zoo Project. Faecal samples (n=2081) from 75 wildlife species (birds and mammals, n=794), 13 livestock species (n=677) and humans (n=333) were collected from households across Nairobi that were participating in the UrbanZoo 99-household project. Our study design is described in detail in the supplementary methods. Briefly, Nairobi was split into administrative units, and 33 were chosen based on a socioeconomic stratification which was weighted by population, such that the larger proportion of low-income households was oversampled which ensuing representation of all other socio-economic groups. Three households were randomly selected in each sublocation to obtain two livestock-keeping and one non-livestock-keeping household (a total of 99 households), with the aim of maximising the spatial distribution and diversity of livestock-keeping practices captured within the sampling frame (Supplementary Figure S8). Households in each sublocation had to meet strict inclusion criteria of keeping small mammals (rabbits) or poultry, large mammals (cattle, goats and sheep) or pigs, or no livestock within the household perimeter. Wildlife samples were obtained by a range of taxon-specific trapping methods, which are described in the supplementary methods and elsewhere.(*12*, *42*)

Sample collection and Microbiological testing

Questionnaires detailing household composition and socioeconomic data, and livestock ownership and management, were administered at each household. Human, animal and wildlife faecal samples were collected and transported on ice to one of two laboratories (University of Nairobi or Kenya Medical Research Institute) within 5 h of collection. Samples were enriched in buffered peptone water for 24 h, and thereafter plated on to eosin methylene blue agar (EMBA) and incubated for 24 h at

37 °C. Subsequently, five colonies were selected and subcultured on EMBA, before being further subcultured on Müller-Hinton agar. A single colony was picked at random from the plate for each original sample (hereafter referred to as an 'isolate') and a 10-parameter biochemical test was used (triple sugar iron agar=4, Simmon's citrate agar=1, and motility-indole-lysine media=3, urease production from urea media =1, oxidase from tetra-methyl- p-phenylenediamine dihydrochloride = *1*) were used for presumptive identification of *E coli.*

Whole genome sequencing

DNA was extracted from bacterial isolates using commercial kits (Purelink® Genomic DNA Mini Kit, Invitrogen, Life Technologies, Carlsbad, California) at the International Livestock Research Institute, Nairobi, Kenya and transported under licence to The Wellcome Trust Centre for Human Genetics, Oxford, UK. Whole genome sequencing was carried out at the Wellcome Trust Centre for Human Genetics on the Illumina HiSeq 2500 platform. Whole genome sequences used in this study are available under the accession number PRJEB32607 and PRJEB41827.

Sequence analysis

Sequenced reads were filtered for quality and trimmed for adaptors with BBDuk (v38.46), k=19 mink=11 hdist=1 ktrim=r minoverlap=12 qtrim=rl trimq=15. The following sequencing quality thresholds were used based on Quast: i) at least 3 Mb aligned to EC958, ii) A maximum assembly length of 6.5 Mb iii) GC content of between 50% and 51%, iv) Assembly N50 of >30kb or a maximum of 100 cgMLST missing loci. A total of 1642 genomes were sequenced that passed this quality threshold.

Genomes were assembled using Spades v3.13.0 with the `--careful` option. Clermont phylotype of the isolates was determined using the ClermonTyping tool (*43*) and the multi-locus sequence type was determined and assigned by Enterobase (*44*).

393 The pangenome was estimated using Roary v3.12.0 with the following options: -s -i 95 -g 100000.
394 Acquired antibiotic resistance genes were identified from the assemblies using starAMR (v0.4.0)
395 (https://github.com/phac-nml/staramr), with a cutoff of 95% sequence identity and a minimum of
396 60% alignment to the query sequence, against the ResFinder database downloaded 25 September
397 2019 (*45*). Antibiotic class of each resistant gene was assigned using the ResFinder classification.

398 **Phylogenetic analyses**

399 A core genome alignment was generated using Snippy v4.6.0 (with default settings) using EC958 as a
400 reference genome (GCA_000285655.3). A phylogenetic analysis of the core genome alignment was
401 performed using IQTREE (v1.6.12) -m TVM+G4 -bb 1000 -safe. The tree and metadata were visualised
402 in iToLv4.3 (itol.embl.de). Due to the species-level diversity of the isolate collection, positions in the
403 alignment in recombinant region of the genome were not removed.

404 *Ad hoc* core genome multi Locus sequence typing (cgMLST) was performed on genome assemblies
405 using chewBBACA (v. 2.0.11) with the 2513 gene cgMLST profile from Enterobase (Downloaded
406 October 2018).

407 The association between metadata (sublocation, host category) and phylogeny was tested using
408 Phylotype (*46*). A minimum of 5 isolates were required to define a cluster, with a maximum of 200
409 core SNP maximum internal cluster distance.

410 **Identification of putative bacterial sharing**

411 A genetic distance matrix was calculated from all pairwise allelic profile comparisons using the library
412 “ape” in R (Paradis et al., 2004). The cgMLST cutoff of 11 allelles to define putative *E. coli* (defined here
413 as a sharing pair) transmission clusters was based on the observed bimodal distributions of inter- and
414 intrahousehold allele differences (Supplementary Figure S9). The R package “cutpointR” was used to
415 validate this cutoff as the optimal value to differentiate pairs that occur within and between
416 households (*47*).

417

418 **Epidemiological analysis of sharing**

419 We established epidemiological links between every possible pair of *E. coli* isolates through a
420 systematic comparison. Household level sharing was categorised as: within household, if a sharing pair
421 involved isolates/hosts from the same household; between household, if a sharing pair involved
422 isolates in different household. Wildlife isolates that could not be attributed to a specific household
423 were omitted from the sharing analysis (Table S2).

424 We condensed our host types into five broad categories (Tables S1, S2); (i) Humans, (ii) Livestock birds;
425 poultry dominated by chickens, (iii) Livestock mammals consisting of ruminants and monogastric
426 livestock, (iv) Wild birds; predominantly seed eating birds such as house sparrows, and (v) Wild
427 mammals; predominantly rodents, along with bats. Primates were omitted from the sharing analysis
428 as they were only associated with two households, along with some samples derived from populations
429 of bats and wild birds which could be attributed to sublocation but not household.

430 While the sharing threshold for the core genome was of ≤ 10 MLST distance, sharing for the
431 pangenome and resistome was based on a Jaccard similarity Index (JI; between 0 and 1, where 1 is
432 identical), where a cut-off threshold was defined, similar to the core genome. For the pangenome/
433 accessory genome this was determined to be JI ≤ 0.98 (Fig 3c, d). Resistome sharing was defined as JI
434 = 1 (Fig 3e, f), with each isolate having a minimum of two AMR genes. In practice, this means that two
435 isolates must share an identical set of AMR genes of length equal to or more than two.

436 We used the sharing thresholds for each facet of the *E. coli* genome and applied a correction to the
437 number of pairs counted when multiple connections involved the same isolate(s). For example, if 4
438 isolates formed a cluster that were all below the 10 cgMLST threshold with each other, the maximum
439 number of pairs/connections that can be drawn between these 4 isolates is 6 or $n(n\text{-}1)/2$, where n is
440 the number of isolates that form a cluster fewer than 10 cgMLST loci apart. However, when the

441 correction is applied, we count only 3 connections, or *n*-1. This avoids the overestimation of sharing
442 events between larger clusters of clonal isolates and provides a more realistic estimate of sharing.

443 Having defined the set of observed sharing events among each of our host categories within and
444 between households, we then sought to detect whether these observed events fell above or below
445 what might be expected given the sampling effort. These denominator values were based on the
446 number of pairs of isolates in each category, assuming an equal probability of sharing among isolates.
447 Within households this was calculated using the formula n(n-1)/2, where n = number of samples of a
448 given host type. Between household sharing was calculated as (n_1) * (n_2). Where n_1 = number of
449 samples of a given host in household 1, and n_2 = number of samples of a given host in household 2.
450 This approach generated a list of all possible paired (expected) sharing events for each category type.
451 From this we calculated the expected frequencies of each type of category sharing within and between
452 households. We then used a resampling approach of the observed values (1000 times) to generate
453 expected distributions (± 95% confidence intervals) of events based on the frequencies of these
454 expected values. From this we were able to assess whether our observed sharing events fell above,
455 below, or within the range we may expect given the sampling effort. The same approach was applied
456 to all aspects of genome sharing (Figure 3 a- f).

Acknowledgements

We are grateful to the residents of our 99 sampling locations in Nairobi for taking part in the study. We thank: Pablo Alarcon, Paula Dominguez-Salas, Laure Madé, Sohel Ahmed and Catherine Pfeifer for assistance with questionnaire design; Absolomon Kihara and Jason Rogena for support with implementing the data collection tool and database design; students on the FELTP project for assistance with fieldwork. We thank the City of Nairobi authorities, the Department of Veterinary Services and the Department of Health Services for their cooperation in this study. This study was supported by the UK Medical Research Council, Biotechnology and Biological Science Research Council (UK), the Economic and Social Research Council (UK), the Natural Environment Research Council (UK), through the Environmental & Social Ecology of Human Infectious Diseases Initiative (ESEI), Grant Reference: G1100783/1. This work was also supported in part by the CGIAR Research Program on Agriculture for Nutrition and Health (A4NH), led by the International Food Policy Research Institute (IFPRI); we acknowledge the CGIAR Fund Donors (https://www.cgiar.org/funders/). The High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics in Oxford (funded by Wellcome Trust grant reference 090532/Z/09/Z) generated the sequencing data. We thank the Modernizing Medical Microbiology Informatics Group for bioinformatics assistance; Tim Davies, Tim Peto, Sarah Walker and Derrick Crook for helpful discussions. D.J.W. is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and the Royal Society (Grant 101237/Z/13/B), and is supported by a Big Data Institute Robertson Fellowship. D.M. was supported by the Darwin Trust of Edinburgh and Centre for Immunology, Infection and Evolution (CIIE). M.J.W. was supported by a Sir Henry Wellcome Postdoctoral Fellowship from the Wellcome Trust (grant reference number WT103953MA) and by a Junior Research Fellowship as part of a Wellcome Trust Strategic grant to the Centre for Immunity, Infection and Evolution at the University of Edinburgh (grant reference number 095831). The views expressed are those of the authors and are not necessarily those of CGIAR, Government of Kenya, FAO or other institutions to which the authors are affiliated.

Ethical approval

The collection of data adhered to the legal requirements of the Government of Kenya. The International Livestock Research Institute Institutional Research Ethics Committee is registered and accredited by the National Commission for Science, Technology and Innovation in Kenya, and approved by the Federal wide Assurance for the Protection of Human Subjects in the USA. Ethical approval for human sampling and data collection was obtained from the ILRI Institutional Research Ethics Committee (ILRI-IACUC2015/09). Livestock samples were obtained under the approval of the ILRI Institutional Animal Care and Use Committee (Reference ILR-IACUC2015/18) and permits obtained from the Directorate of Veterinary Services. Wildlife were trapped under approval of an ILRI Institutional Animal Care and Use Protocol (IACUC2015/12), and permits were obtained from the National Museums of Kenya and Kenya Wildlife Service.

References

1. W. B. Karesh, A. Dobson, J. O. Lloyd-Smith, J. Lubroth, M. A. Dixon, M. Bennett, S. Aldrich, T. Harrington, P. Formenty, E. H. Loh, C. C. Machalaba, M. J. Thomas, D. L. Heymann, Ecology of zoonoses: natural and unnatural histories. *The Lancet*. **380**, 1936–1945 (2012).

2. N. D. Wolfe, C. P. Dunavan, J. Diamond, Origins of major human infectious diseases. *Nature*. **447**, 279–283 (2007).

3. T. Allen, K. A. Murray, C. Zambrana-Torrelio, S. S. Morse, C. Rondinini, M. Di Marco, N. Breit, K. J. Olival, P. Daszak, Global hotspots and correlates of emerging zoonotic diseases. *Nat. Commun.* **8**, 1124 (2017).

4. F. Zhang, M. Chase-Topping, C.-G. Guo, B. A. D. van Bunnik, L. Brierley, M. E. J. Woolhouse, Global discovery of human-infective RNA viruses: A modelling analysis. *PLoS Pathog.* (2020), doi:10.1371/journal.ppat.1009079.

5. N. L. Gottdenker, D. G. Streicker, C. L. Faust, C. R. Carroll, Anthropogenic Land Use Change and Infectious Diseases: A Review of the Evidence. *EcoHealth*. **11**, 619–632 (2014).

6. D. Muloi, J. Kiiru, M. J. Ward, J. M. Hassell, J. M. Bettridge, T. R. Robinson, B. A. D. van Bunnik, M. Chase-Topping, G. Robertson, A. B. Pedersen, E. M. Fevre, M. E. J. Woolhouse, E. K. Kang'ethe, S. Kariuki, Epidemiology of antimicrobial-resistant *Escherichia coli* carriage in sympatric humans and livestock in a rapidly urbanizing city. *Int. J. Antimicrob. Agents*. **54**, 531–537 (2019).

7. D. Santiago-Alarcon, I. MacGregor-Fors, Cities and pandemics: urban areas are ground zero for the transmission of emerging human infectious diseases. *J. Urban Ecol.* **6** (2020), doi:10.1093/jue/juaa012.

8. M. Woolhouse, M. Ward, B. van Bunnik, J. Farrar, Antimicrobial resistance in humans, livestock and the wider environment. *Philos. Trans. R. Soc. B Biol. Sci.* **370**, 20140083 (2015).

9. W. P. Hanage, Two Health or Not Two Health? That Is the Question. *mBio*. **10** (2019), doi:10.1128/mBio.00550-19.

10. J. M. Hassell, M. Begon, M. J. Ward, E. M. Fèvre, Urbanization and Disease Emergence: Dynamics at the Wildlife-Livestock-Human Interface. *Trends Ecol. Evol.* **32**, 55–67 (2017).

11. C.-J. Neiderud, How urbanization affects the epidemiology of emerging infectious diseases. *Infect. Ecol. Epidemiol.* **5** (2015), doi:10.3402/iee.v5.27060.

12. J. M. Hassell, M. J. Ward, D. Muloi, J. M. Bettridge, T. P. Robinson, S. Kariuki, A. Ogendo, J. Kiiru, T. Imboma, E. K. Kang'ethe, E. M. Öghren, N. J. Williams, M. Begon, M. E. J. Woolhouse, E. M. Fèvre, Clinically relevant antimicrobial resistance at the wildlife–livestock–human interface in Nairobi: an epidemiological study. *Lancet Planet. Health*. **3**, e259–e269 (2019).

13. M. Vittecoq, S. Godreuil, F. Prugnolle, P. Durand, L. Brazier, N. Renaud, A. Arnal, S. Aberkane, H. Jean-Pierre, M. Gauthier-Clerc, F. Thomas, F. Renaud, Antimicrobial resistance in wildlife. *J. Appl. Ecol.* **53**, 519–529 (2016).

14. I. B. Rwego, G. Isabirye-Basuta, T. R. Gillespie, T. L. Goldberg, Gastrointestinal Bacterial Transmission among Humans, Mountain Gorillas, and Livestock in Bwindi Impenetrable National Park, Uganda. *Conserv. Biol.* **22**, 1600–1607 (2008).

15. B. A. Wee, D. M. Muloi, B. A. D. van Bunnik, Quantifying the transmission of antimicrobial resistance at the human and livestock interface with genomics. *Clin. Microbiol. Infect.* (2020), doi:10.1016/j.cmi.2020.09.019.

16. F. Jaureguy, L. Landraud, V. Passet, L. Diancourt, E. Frapy, G. Guigon, E. Carbonnelle, O. Lortholary, O. Clermont, E. Denamur, B. Picard, X. Nassif, S. Brisse, Phylogenetic and genomic diversity of human bacteremic Escherichia coli strains. *BMC Genomics*. **9**, 560 (2008).

17. O. Tenaillon, D. Skurnik, B. Picard, E. Denamur, The population genetics of commensal Escherichia coli. *Nat. Rev. Microbiol.* **8**, 207–217 (2010).

18. L. B. Becking, *Geobiologie of inleiding tot de milieukunde* (WP Van Stockum & Zoon, 1934).

19. A. C. Schürch, S. Arredondo-Alonso, R. J. L. Willems, R. V. Goering, Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene–based approaches. *Clin. Microbiol. Infect.* **24**, 350–354 (2018).

20. T. J. Dallman, L. Byrne, P. M. Ashton, L. A. Cowley, N. T. Perry, G. Adak, L. Petrovska, R. J. Ellis, R. Elson, A. Underwood, J. Green, W. P. Hanage, C. Jenkins, K. Grant, J. Wain, Whole-Genome Sequencing for National Surveillance of Shiga Toxin–Producing Escherichia coli O157. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **61**, 305–312 (2015).

21. J. S. Iramiot, H. Kajumbula, J. Bazira, E. P. de Villiers, B. B. Asiimwe, Whole genome sequences of multi-drug resistant Escherichia coli isolated in a Pastoralist Community of Western Uganda: Phylogenomic changes, virulence and resistant genes. *PLOS ONE*. **15**, e0231852 (2020).

22. T. K. S. Richter, T. H. Hazen, D. Lam, C. L. Coles, J. C. Seidman, Y. You, E. K. Silbergeld, C. M. Fraser, D. A. Rasko, Temporal Variability of Escherichia coli Diversity in the Gastrointestinal Tracts of Tanzanian Children with and without Exposure to Antibiotics. *mSphere*. **3** (2018), doi:10.1128/mSphere.00558-18.

23. N. de C. Stoppe, J. S. Silva, C. Carlos, M. I. Z. Sato, A. M. Saraiva, L. M. M. Ottoboni, T. T. Torres, Worldwide Phylogenetic Group Patterns of Escherichia coli from Commensal Human and Wastewater Treatment Plant Isolates. *Front. Microbiol.* **8** (2017), doi:10.3389/fmicb.2017.02512.

24. T. L. Goldberg, T. R. Gillespie, I. B. Rwego, E. Wheeler, E. L. Estoff, C. A. Chapman, Patterns of gastrointestinal bacterial exchange between chimpanzees and humans involved in research and tourism in western Uganda. *Biol. Conserv.* **135**, 511–517 (2007).

25. S. J. Salipante, D. J. SenGupta, L. A. Cummings, T. A. Land, D. R. Hoogestraat, B. T. Cookson, Application of Whole-Genome Sequencing for Bacterial Strain Typing in Molecular Epidemiology. *J. Clin. Microbiol.* **53**, 1072–1079 (2015).

26. I. L. Brito, T. Gurry, S. Zhao, K. Huang, S. K. Young, T. P. Shea, W. Naisilisili, A. P. Jenkins, S. D. Jupiter, D. Gevers, E. J. Alm, Transmission of human-associated microbiota along family and social networks. *Nat. Microbiol.* **4**, 964–971 (2019).

27. S. Lax, D. P. Smith, J. Hampton-Marcell, S. M. Owens, K. M. Handley, N. M. Scott, S. M. Gibbons, P. Larsen, B. D. Shogan, S. Weiss, J. L. Metcalf, L. K. Ursell, Y. Vázquez-Baeza, W. V. Treuren, N. A. Hasan, M. K. Gibson, R. Colwell, G. Dantas, R. Knight, J. A. Gilbert, Longitudinal analysis of microbial interaction between humans and the indoor environment. *Science*. **345**, 1048–1052 (2014).

28. M. Mohamed, C. Clabots, S. B. Porter, T. Bender, P. Thuras, J. R. Johnson, Large Fecal Reservoir of Escherichia coli Sequence Type 131-H30 Subclone Strains That Are Shared Within Households and Resemble Clinical ST131-H30 Isolates. *J. Infect. Dis.* **221**, 1659–1668 (2020).

29. E. Mosites, M. Sammons, E. Otiang, A. Eng, C. Noecker, O. Manor, S. Hilton, S. M. Thumbi, C. Onyango, G. Garland-Lewis, D. R. Call, M. K. Njenga, J. N. Wasserheit, J. A. Zambriski, J. L. Walson, G. H. Palmer, J. Montgomery, E. Borenstein, R. Omore, P. M. Rabinowitz, Microbiome sharing between children, livestock and household surfaces in western Kenya. *PLOS ONE*. **12**, e0171017 (2017).

30. Z. Naziri, A. Derakhshandeh, R. Firouzi, M. Motamedifar, A. Shojaee Tabrizi, DNA fingerprinting approaches to trace Escherichia coli sharing between dogs and owners. *J. Appl. Microbiol.* **120**, 460–468 (2016).

31. L. Falgenhauer, C. Imirzalioglu, H. Ghosh, K. Gwozdzinski, J. Schmiedel, K. Gentil, R. Bauerfeind, P. Kämpfer, H. Seifert, G. B. Michael, S. Schwarz, Y. Pfeifer, G. Werner, M. Pietsch, U. Roesler, B. Guerra, J. Fischer, H. Sharp, A. Käsbohrer, A. Goesmann, K. Hille, L. Kreienbrock, T. Chakraborty, Circulation of clonal populations of fluoroquinolone-resistant CTX-M-15-producing Escherichia coli ST410 in humans and animals in Germany. *Int. J. Antimicrob. Agents*. **47**, 457–465 (2016).

32. C. Ludden, K. E. Raven, D. Jamrozy, T. Gouliouris, B. Blane, F. Coll, M. de Goffau, P. Naydenova, C. Horner, J. Hernandez-Garcia, P. Wood, N. Hadjirin, M. Radakovic, N. M. Brown, M. Holmes, J. Parkhill, S. J. Peacock, One Health Genomic Surveillance of *Escherichia coli* Demonstrates Distinct Lineages and Mobile Genetic Elements in Isolates from Humans versus Livestock. *mBio*. **10** (2019), doi:10.1128/mBio.02693-18.

33. A. McNally, Y. Oren, D. Kelly, B. Pascoe, S. Dunn, T. Sreecharan, M. Vehkala, N. Välimäki, M. B. Prentice, A. Ashour, O. Avram, T. Pupko, U. Dobrindt, I. Literak, S. Guenther, K. Schaufler, L. H. Wieler, Z. Zhiyong, S. K. Sheppard, J. O. McInerney, J. Corander, Combined Analysis of Variation in Core, Accessory and Regulatory Genome Regions Provides a Super-Resolution View into the Evolution of Bacterial Populations. *PLoS Genet.* **12** (2016), doi:10.1371/journal.pgen.1006280.

34. A. H. Moeller, S. Foerster, M. L. Wilson, A. E. Pusey, B. H. Hahn, H. Ochman, Social behavior shapes the chimpanzee pan-microbiome. *Sci. Adv.* **2**, e1500997 (2016).

35. J. Tung, L. B. Barreiro, M. B. Burns, J.-C. Grenier, J. Lynch, L. E. Grieneisen, J. Altmann, S. C. Alberts, R. Blekhman, E. A. Archie, Social networks predict gut microbiome composition in wild baboons. *eLife*. **4**, e05224 (2015).

36. A. Springer, A. Mellmann, C. Fichtel, P. M. Kappeler, Social structure and Escherichia coli sharing in a group-living wild primate, Verreaux's sifaka. *BMC Ecol.* **16**, 6 (2016).

37. M. R. Tobin, J. L. Goldshear, L. B. Price, J. P. Graham, J. H. Leibler, A Framework to Reduce Infectious Disease Risk from Urban Poultry in the United States. *Public Health Rep.* **130**, 380–391 (2015).

610 38. C. M. Liu, M. Stegger, M. Aziz, T. J. Johnson, K. Waits, L. Nordstrom, L. Gauld, B. Weaver, D.
611 Rolland, S. Statham, J. Horwinski, S. Sariya, G. S. Davis, E. Sokurenko, P. Keim, J. R. Johnson, L. B.
612 Price, *Escherichia coli* ST131-H22 as a Foodborne Uropathogen. *mBio*. **9** (2018),
613 doi:10.1128/mBio.00470-18.

614 39. L. Roer, S. Overballe-Petersen, F. Hansen, T. B. Johannesen, M. Stegger, V. Bortolaia, P.
615 Leekitcharoenphon, H. B. Korsgaard, A. M. Seyfarth, J. Mossong, P. Wattiau, C. Boland, D. S.
616 Hansen, H. Hasman, A. M. Hammerum, R. S. Hendriksen, ST131 fimH22 Escherichia coli isolate
617 with a blaCMY-2/IncI1/ST12 plasmid obtained from a patient with bloodstream infection: highly
618 similar to E. coli isolates of broiler origin. *J. Antimicrob. Chemother.* **74**, 557–560 (2019).

619 40. S. J. Song, C. Lauber, E. K. Costello, C. A. Lozupone, G. Humphrey, D. Berg-Lyons, J. G. Caporaso,
620 D. Knights, J. C. Clemente, S. Nakielny, J. I. Gordon, N. Fierer, R. Knight, Cohabiting family
621 members share microbiota with one another and with their dogs. *eLife*. **2**, e00458 (2013).

622 41. D. Muloi, E. M. Fèvre, J. Bettridge, R. Rono, D. Ong'are, J. M. Hassell, M. K. Karani, P. Muinde, B.
623 van Bunnik, A. Street, M. Chase-Topping, A. B. Pedersen, M. J. Ward, M. Woolhouse, A cross-
624 sectional survey of practices and knowledge among antibiotic retailers in Nairobi, Kenya. *J. Glob.*
625 *Health*. **9**, 010412–010412 (2019).

626 42. J. Bonnedahl, M. Drobni, M. Gauthier-Clerc, J. Hernandez, S. Granholm, Y. Kayser, Å. Melhus, G.
627 Kahlmeter, J. Waldenström, A. Johansson, B. Olsen, Dissemination of Escherichia coli with CTX-
628 M Type ESBL between Humans and Yellow-Legged Gulls in the South of France. *PLOS ONE*. **4**,
629 e5958 (2009).

630 43. M. D. Kauffman, J. LeJeune, European Starlings (Sturnus vulgaris) challenged with Escherichia
631 coli O157 can carry and transmit the human pathogen to cattle. *Lett. Appl. Microbiol.* **53**, 596–
632 601 (2011).

633 44. S. Mukerji, M. Stegger, A. V. Truswell, T. Laird, D. Jordan, R. J. Abraham, A. Harb, M. Barton, M.
634 O'Dea, S. Abraham, Resistance to critically important antimicrobials in Australian silver gulls
635 (Chroicocephalus novaehollandiae) and evidence of anthropogenic origins. *J. Antimicrob.*
636 *Chemother.* **74**, 2566–2574 (2019).

637 45. N. Stoesser, A. E. Sheppard, C. E. Moore, T. Golubchik, C. M. Parry, P. Nget, M. Saroeun, N. P. J.
638 Day, A. Giess, J. R. Johnson, T. E. A. Peto, D. W. Crook, A. S. Walker, Extensive Within-Host
639 Diversity in Fecally Carried Extended-Spectrum-Beta-Lactamase-Producing Escherichia coli
640 Isolates: Implications for Transmission Analyses. *J. Clin. Microbiol.* **53**, 2122–2131 (2015).

641 46. J. M. Hassell, M. J. Ward, D. Muloi, J. M. Bettridge, H. Phan, T. P. Robinson, A. Ogendo, T. Imboma,
642 J. Kiiru, S. Kariuki, M. Begon, E. K. Kang'ethe, M. E. J. Woolhouse, E. M. Fèvre, Deterministic
643 processes structure bacterial genetic communities across an urban landscape. *Nat. Commun.* **10**,
644 2643 (2019).

645 47. J. Beghain, A. Bridier-Nahmias, H. Le Nagard, E. Denamur, O. Clermont, ClermonTyping: an easy-
646 to-use and accurate in silico method for Escherichia genus strain phylotyping. *Microb. Genomics*.
647 **4**, e000192 (2018).

648 48. Z. Zhou, N.-F. Alikhan, K. Mohamed, Y. Fan, A. S. Group, M. Achtman, *Genome Res.*, in press,
649 doi:10.1101/gr.251678.119.

650 49. E. Zankari, H. Hasman, S. Cosentino, M. Vestergaard, S. Rasmussen, O. Lund, F. M. Aarestrup, M.
651 V. Larsen, Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.*
652 **67**, 2640–2644 (2012).

653 50. F. Chevenet, M. Jung, M. Peeters, T. de Oliveira, O. Gascuel, Searching for virus phylotypes.
654 *Bioinforma. Oxf. Engl.* **29**, 561–570 (2013).

655 51. C. Thiele, G. Hirschfeld, cutpointr: Improved Estimation and Validation of Optimal Cutpoints in
656 R. *ArXiv200209209 Stat* (2020) (available at http://arxiv.org/abs/2002.09209).

657

Figures

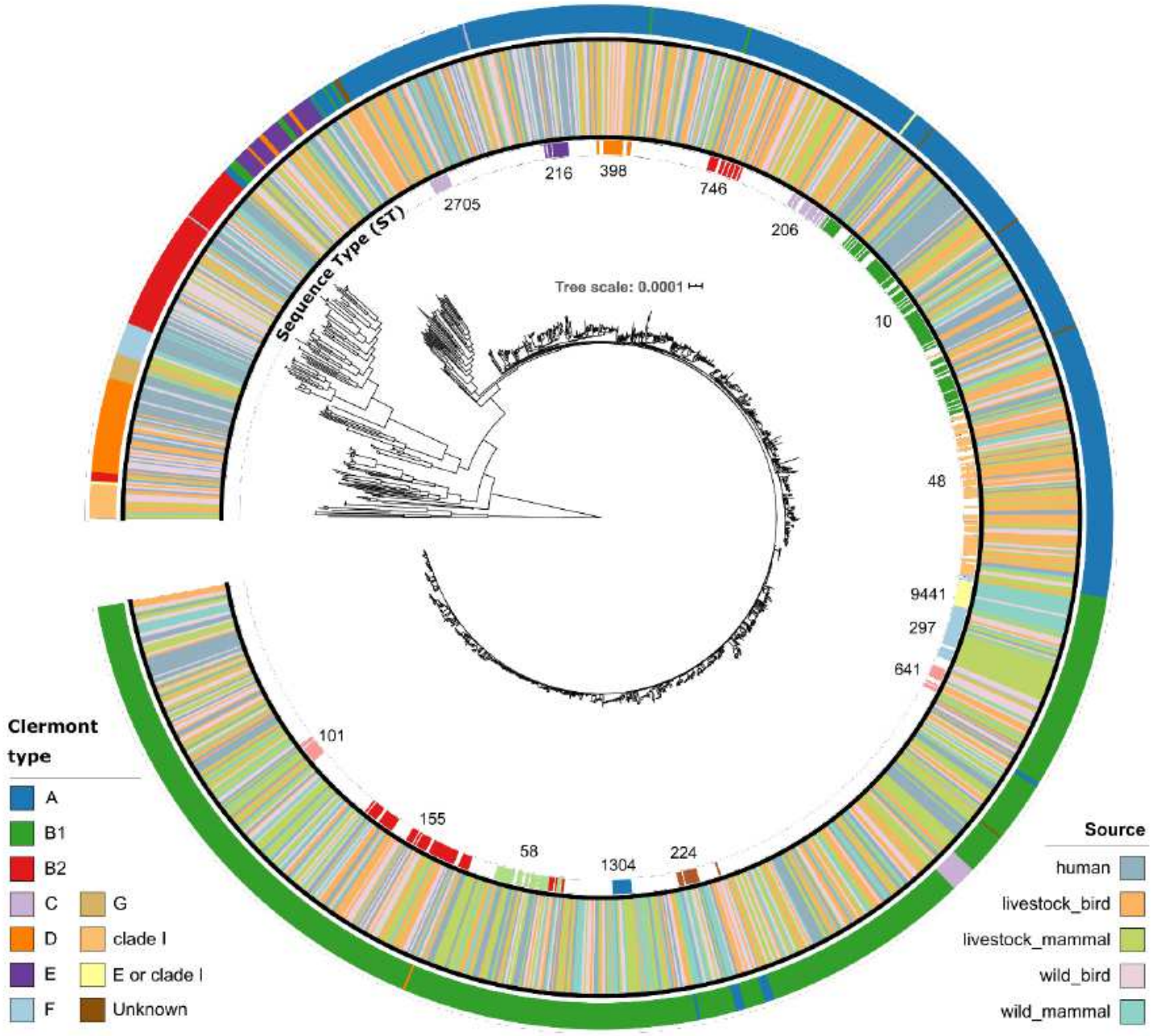

Figure 1

Core genome phylogeny of 1338 E. coli isolates showing (from inner to outer ring): i) Sequence Types (only STs with a minimum 10 isolates are shown), ii) source type of isolate, and iii) Clermont phylotype classifications. The tree is rooted on the Clade I group.

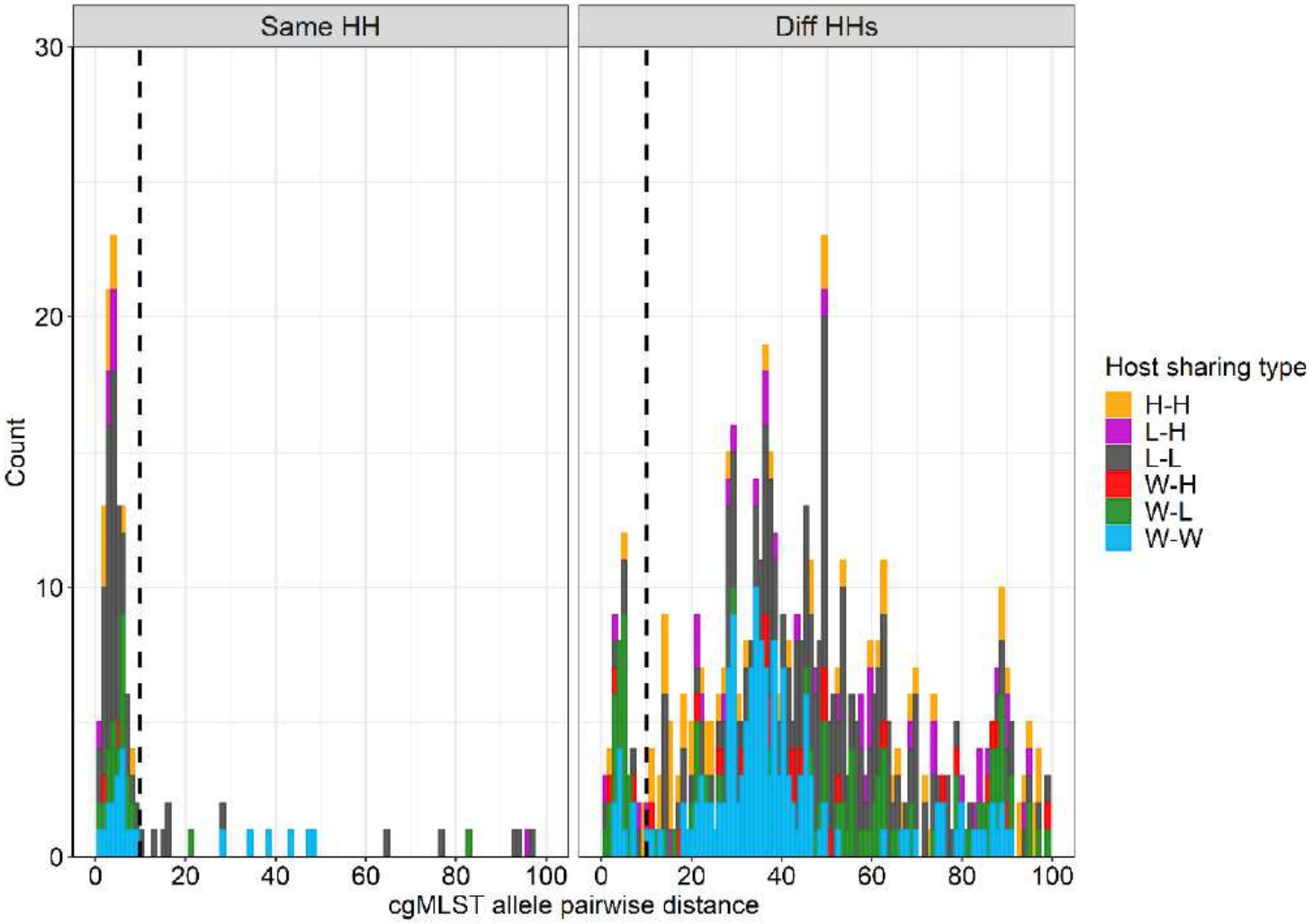

Figure 2

Frequency distribution of pairwise distances between isolates from the same households (left) and different (right) households. The sources of isolates in each pair is indicated by the colour. Only pairs that are closer than 100 cgMLST loci apart are shown. H = human, L = livestock and W= wildlife

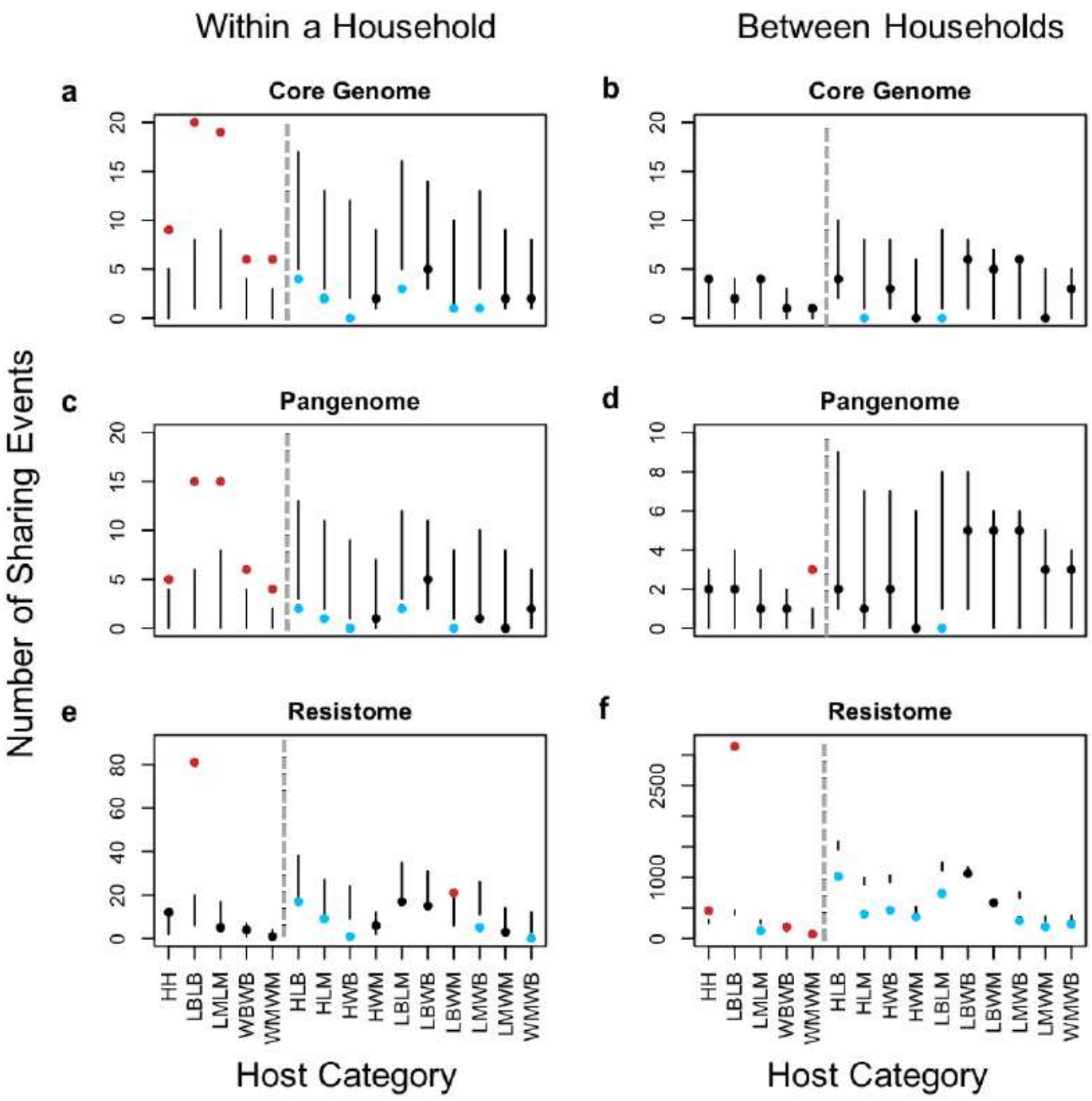

Figure 3

Number of sharing pairs across 15 host category types, within (a, c, e), and between (b, d, f) households for each of the core genome, pangenome, and resistome. Panels show the 95% confidence intervals of the calculated expected distribution using a resampling approach, points depict the observed number of sharing pairs in each category coloured according to whether they fall above (red), below (blue) or within (black) the expected distribution. Hosts in the same category (e.g. human-human) and different categories (e.g. Human-Livestock Bird) are separated by grey dashed lines. Source type of isolate pairs

are indicated on the x-axis with either Human (H), Livestock birds (LB), Livestock mammals (LM), Wildlife birds (WB), Wildlife mammals (WM). In each plot, within-category connections are on the left of the grey dotted line and between category connections are on the right.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryinformation.pdf
- TableS1.v2.20201030.csv
- UrbanZoopapers.docx
- FlowDiagramv1.pptx
- 99HHhouseholdselectionprotocolfinal.docx