

1 **Genomic diversity of *Escherichia coli* isolates from backyard chickens and guinea fowl** 2 **in the Gambia**

3 Ebenezer Foster-Nyarko^{1,2}, Nabil-Fareed Alikhan¹, Anuradha Ravi¹, Nicholas M. Thomson¹,
4 Sheikh Jarju², Brenda Anna Kwambana-Adams^{2,5}, Arss Secka³, Justin O’Grady¹, Martin
5 Antonio^{2,4}, Mark J. Pallen^{1,6#}

6

7 ¹ Quadram Institute Bioscience, Norwich Research Park, Norwich, Norfolk, United Kingdom

8 ² Medical Research Council Unit The Gambia at the London School of Hygiene and Tropical
9 Medicine, Atlantic Boulevard Road, Fajara, the Gambia

10 ³ West Africa Livestock Innovation Centre (WALIC), MB 14, Banjul, the Gambia

11 ⁴ Warwick Medical School, University of Warwick, Coventry, United Kingdom

12 ⁵ NIHR Global Health Research Unit on Mucosal Pathogens, Division of Infection and
13 Immunity, University College London, London, United Kingdom

14 ⁶ School of Veterinary Medicine, University of Surrey, Guildford, Surrey, United Kingdom.

15

16

17 **#Corresponding author:** Professor Mark J. Pallen, Quadram Institute Bioscience, Norwich
18 Research Park, Norwich, Norfolk, United Kingdom

19 Email: M.Pallen@uea.ac.uk

20

21

22

23 **Abstract**

24 Chickens and guinea fowl are commonly reared in Gambian homes as affordable sources of
 25 protein. Using standard microbiological techniques, we obtained 68 caecal isolates of
 26 *Escherichia coli* from ten chickens and nine guinea fowl in rural Gambia. After Illumina
 27 whole-genome sequencing, 28 sequence types were detected in the isolates (four of them
 28 novel), of which ST155 was the most common (22/68, 32%). These strains span four of the
 29 eight main phylogroups of *E. coli*, with phylogroups B1 and A being most prevalent. Nearly
 30 a third of the isolates harboured at least one antimicrobial resistance gene, while most of the
 31 ST155 isolates (14/22, 64%) encoded resistance to ≥ 3 classes of clinically relevant
 32 antibiotics, as well as putative virulence factors, suggesting pathogenic potential in humans.
 33 Furthermore, hierarchical clustering revealed that several Gambian poultry strains were
 34 closely related to isolates from humans. Although the ST155 lineage is common in poultry
 35 from Africa and South America, the Gambian ST155 isolates belong to a unique cgMLST
 36 cluster comprised of closely related (38-39 alleles differences) isolates from poultry and
 37 livestock from sub-Saharan Africa—suggesting that strains can be exchanged between
 38 poultry and livestock in this setting. Continued surveillance of *E. coli* and other potential
 39 pathogens in rural backyard poultry from sub-Saharan Africa is warranted.

40

41 **Keywords**

42 *Escherichia coli*, genomic diversity, ST155, backyard poultry, chickens, guinea fowl.

43

44

45 **Author notes**

46 All supporting data and protocols have been provided within the article or as supplementary
47 data files. Eleven supplementary figures and eight supplementary files are available with the
48 online version of this article.

49

50 **Abbreviations**

51 APEC, Avian Pathogenic *E. coli*; ExPEC, Extraintestinal pathogenic *E. coli*; ST, Sequence
52 type; AMR, Antimicrobial resistance; MDR, Multiple-drug resistance; MLST, Multi-locus
53 sequence typing; cgMLST, core-genome multi-locus sequence typing; ARIBA,
54 Antimicrobial resistance identification by assembly;; VFDB, Virulence factors database;
55 SNP, single nucleotide polymorphism; STGG, Skimmed milk tryptone glucose glycerol;
56 SDS, Sodium dodecyl-sulphate; EDTA, Ethylenediaminetetraacetic acid; ESBL, Expanded-
57 spectrum beta-lactamase; MIC, minimum inhibitory concentrations.

58

59 **Data summary**

60 The genomic assemblies for the isolates reported here are available for download from
61 EnteroBase (<http://enterobase.warwick.ac.uk/species/index/ecoli>) and the EnteroBase
62 assembly barcodes are provided in File S2.

63 Sequences have been deposited in the NCBI SRA, under the BioProject ID:

64 PRJNA616250 and accession numbers SAMN14485281 to SAMN14485348 (File S2).

65 Assemblies have been deposited in GenBank under the BioProject ID: PRJNA616250 and

66 accession numbers CP053258 and CP053259.

67

68 **Impact statement**

69 Domestic birds play a crucial role in human society, in particular contributing to food
70 security in low-income countries. Many households in Sub-Saharan Africa rear free-range
71 chickens and guinea fowl, which are often left to scavenge for feed in and around the family
72 compound, where they are frequently exposed to humans, other animals and the environment.
73 Such proximity between backyard poultry and humans is likely to facilitate transmission of
74 pathogens such as *Escherichia coli* or antimicrobial resistance between the two host species.
75 Little is known about the population structure of *E. coli* in rural chickens and guinea fowl,
76 although this information is needed to contextualise the potential risks of transmission of
77 bacterial strains between humans and rural backyard poultry. Thus, we sought to investigate
78 the genomic diversity of *E. coli* in backyard poultry from rural Gambia.
79

80 **Introduction**

81 The domestic chicken (*Gallus gallus domesticus*) is the most numerous bird on the planet,
82 with an estimated population of over 22.7 billion—ten times more than any other bird [1].
83 Since their domestication from the red jungle fowl in Asia between 6, 000 and 8, 000 years
84 ago [2, 3], chickens have been found almost everywhere humans live. Other poultry, such as
85 turkeys, guinea fowl, pheasants, duck and geese, are derived from subsequent domestication
86 events across Africa, Europe and the Americas [4]. For example, the helmeted guinea fowl
87 (*Numida meleagris*) originated in West Africa, although domesticated forms of this bird are
88 now found in many parts of the tropics.

89 Poultry are reared for meat, eggs and feathers [5]. Poultry production is classified into four
90 sectors, based on the marketing of poultry products and the level of biosecurity [6]. Intensive
91 poultry farming falls under sectors 1 to 3, characterised by moderate to high levels of
92 biosecurity, while sector 4 pertains to the “backyard”, “village” or “family” poultry system,
93 with little or no biosecurity measures.

94 In rural backyard farming—prevalent in low- to middle-income countries such as the
95 Gambia—a small flock of birds (between one and fifty) usually from indigenous breeds are
96 allowed to scavenge for feed over a wide area during the daytime, with minimal
97 supplementation, occasional provision of water and natural hatching of chicks. The poultry
98 may be confined at night in rudimentary shelters to minimise predation, or birds may roost in
99 owners’ kitchens, family dwellings, tree-tops, or nest in the bush [7]. Urban and peri-urban
100 backyard poultry farming—for example, in Australia, New Zealand (North Island), the US
101 and in the UK—differs from rural backyard farming in that the birds are kept on an enclosed
102 residential lot [8-10].

103 Backyard poultry fulfils important social, economic and cultural roles in many societies.
104 Seventy percent of poultry production in low-income countries comes from backyard poultry

[11]. The sale of birds and eggs generates income, while occasional consumption of poultry meat provides a source of protein in the diet. In traditional societies, domestic poultry meat is considered tastier than commercial broiler meat and, as it is perceived to be tougher in texture, is preferred for preparing dishes that require prolonged cooking [12]. It is estimated that meat and eggs from backyard poultry contribute about 30% of the total animal protein supply of households in low-income countries [13, 14]. In rural Gambia, backyard poultry can be offered as gifts for newlyweds or sold to solve family needs such as paying school fees, buying new clothes or other household needs [7]. Chickens may also be used as offering to a traditional healer, consumed when there is a guest, or during ceremonies. Urban and peri-urban poultry are kept mostly for home consumption of their eggs or meat, but also as pets or used for pest control [8, 15-17]. The proximity between backyard poultry and humans may facilitate transmission of pathogens such as *Escherichia coli* between the two host species.

E. coli is a generalist bacterium that commonly colonises the gastrointestinal tract of mammals and avian species [18]. Based on their pathogenic potential, *E. coli* can be divided into three categories: commensals, diarrhoeagenic *E. coli* and extraintestinal pathogenic *E. coli* (ExPEC). ExPEC frequently colonise the gut asymptotically; however, they possess a wide range of unique virulence factors that enable them to colonise extraintestinal tissues in humans, pets and poultry [19, 20]. A sub-pathotype of ExPEC strains, known as Avian Pathogenic *E. coli* (APEC), cause colibacillosis—an extraintestinal disease in birds, with manifestations such as septicaemia, air sacculitis and cellulitis [21]. Avian colibacillosis results in high mortality and condemnation of birds, resulting in significant annual economic losses for the poultry industry [22]. As a result, antimicrobials are often used in intensive farming systems to prevent bacterial infections and treat sick birds—a practice that has been linked to the development of antimicrobial resistance (AMR) in poultry.

Although previous studies have focused on the detection of AMR and documented the emergence of multiple-drug resistance (MDR) in this niche [23-27], little is known about the population structure of *E. coli* in rural backyard poultry. Given the increased exposure to humans, the natural environment and other animals, the population of *E. coli* in birds raised under the backyard system may differ considerably from those reared in intensive systems. It is also possible that the lineages of *E. coli* within local genotypes of rural poultry might differ between geographical regions. The absence of biosecurity measures in backyard poultry farming increases the potential for zoonotic transmission of pathogenic and/or antimicrobial-resistant strains to humans.

In a recent study of commercial broiler chickens, multiple colony sampling revealed that a single broiler chicken could harbour up to nine sequence types of *E. coli* [28]. However, within-host diversity of *E. coli* in backyard poultry—particularly in guinea fowl—has not been well studied and so we do not know how many lineages of *E. coli* can co-colonise a single backyard bird. To address these gaps in our knowledge, we exploited whole-genome sequencing to investigate the genomic diversity and burden of AMR among *E. coli* isolates from backyard chickens and guinea fowl in rural Gambia, West Africa.

Materials and methods

Study population

The study population comprised ten local-breed chickens and nine guinea fowl from a village in Sibanor in the Western Region of the Gambia (Table 1). Sibanor covers an area of approximately 90 km² and is representative of rural areas in the Gambia [29]. It has a population of about 10,000. Most of the villagers are subsistence farmers growing peanuts, maize and millet. Households within this community comprise extended family units of up to

fifteen people, which make up the “compound”. All guinea fowl were of the pearl variety, characterised by purplish-grey feathers dotted with white.

Sample collection

The sampling was done in November 2016. Poultry birds were first observed in motion for the presence of any abnormalities. Healthy-looking birds were procured from eight contiguous households within 0.3-0.4 km of each other and transported to the Abuko Veterinary Station, the Gambia in an air-conditioned vehicle. A qualified veterinarian then euthanised the birds and removed their caeca under aseptic conditions. These were placed into sterile falcon tubes and flash-frozen on dry ice in a cooler box. The samples were transported to the Medical Research Council Unit The Gambia at the London School of Hygiene and Tropical Medicine labs in Fajara, where the caecal contents were aseptically emptied into new falcon tubes for storage at -80°C within 3 h. A peanut-sized aliquot was taken from each sample into a 1.8 ml Nunc tube containing 1 ml of Skimmed Milk Tryptone Glucose Glycerol (STGG) transport and storage medium (Oxoid, Basingstoke, UK), vortexed at 4200rpm for 2 min and frozen at -80°C. Figure 1 summarises the sample processing flow.

Microbiological processing

Isolation of *E. coli* was performed as follows. The caecal-STGG suspension was removed from -80 °C storage and allowed to thaw briefly on wet ice. A 100 µl aliquot was then taken into 900 µl of physiological saline (0.85%) and taken through four ten-fold serial dilutions. A100 µl aliquot each was then taken from the dilutions and uniformly streaked onto tryptone-bile-X-glucoronide agar plates using the spread plate technique. The inoculated plates were incubated at 37°C for 18–24 h under aerobic conditions. Following overnight incubation, colony counts were determined for raised, translucent and entire colonies that exhibited

bluish-green pigmentation typical of *E. coli*. Up to five candidate colonies were selected per sample and sub-cultured on MacConkey agar. These were incubated at 37°C in air for 18–24 h and stored in 20% glycerol broth at -80°C. The isolates from chickens were designated “C1-C10”, while those from guinea fowl were prefixed by “G1-G9”, followed by the respective colony number (1 up to 5).

Genomic DNA extraction

Genomic DNA was extracted from overnight broth cultures prepared from each single colony sub-culture using the 96-well plate lysate method as described previously [30]. The DNA was eluted in Tris-Cl (pH, 8.0) and quantified using the Qubit high sensitivity DNA assay kit (Invitrogen, MA, USA). DNA samples were kept at -20°C until Illumina sequencing library preparation. Broth cultures were spun at 3500rpm for 2 min and lysed using lysozyme, proteinase K, 10% SDS and RNase A in Tris EDTA buffer (pH 8.0).

Illumina sequencing

Whole-genome shotgun sequencing of the DNA extracts was performed on the Illumina NextSeq 500 instrument (Illumina, San Diego, CA) using a modified Illumina Nextera library preparation protocol as described previously [30]. We run the final pooled library at a concentration of 1.8 pM on a mid-output flow cell (NSQ® 500 Mid Output KT v2 300 cycles; Illumina Catalogue No. FC-404-2003) according to manufacturer’s instructions. Following sequencing, FASTQ files were downloaded from BaseSpace to a local server hosted at the Quadram Institute Bioscience.

Genome assembly and phylogenetic analysis

The raw sequences were initially analysed on the Cloud Infrastructure for Microbial Bioinformatics [31]. This included concatenating paired-end short reads, quality checks with FastQC v0.11.7 [32], trimming with Trimmomatic v0.39 [33] and assembly by Spades v3.13.2 [34]. The quality of the assemblies was checked using QUAST v5.0.0, de6973bb [35] and annotation of the draft genomes was carried out using Prokka v1.13.3 [36]. We used the mlst software (<https://github.com/tseemann/mlst>) to call multi-locus sequence types (MLSTs) using the Achtman scheme [37], based on the seven house-keeping genes, *adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA* and *recA*. We used Snippy v4.3.2 (<https://github.com/tseemann/snippy>) for variant calling and to generate a core-genome alignment, from which a maximum-likelihood phylogenetic tree was reconstructed using RAxML v8.2.4 [38], based on a general time-reversible nucleotide substitution model with 1,000 bootstrap replicates. We included representative reference genome sequences for the major phylogroups of *E. coli* and *Escherichia fergusonii* as an outgroup (File S1). Given that recombination is widespread in *E. coli* and tends to blur phylogenetic signals [37], we used Gubbins (Genealogies Unbiased By recomBinations In Nucleotide Sequences) [39] to detect and mask recombinant regions of the core-genome alignment prior to the phylogenetic reconstruction. We used the GrapeTree [40] to visualise and annotate phylogenetic trees. We calculated pair-wise single nucleotide polymorphism (SNP) distances between genomes from the core-genome alignment using snp-dists v0.6 (<https://github.com/tseemann/snp-dists>).

Subsequently, the short-read sequences were uploaded to EnteroBase [41], an online genome database and integrated software environment which currently hosts more than 138,164 *E. coli* genomes, sourced from all publicly available sequence databases and user uploads. EnteroBase routinely retrieves short-read *E. coli* sequences from the public domain, performs quality control and *de novo* assemblies of Illumina short-read sequences, annotates these and assigns seven-allele MLST (ST) and phylogroups from genome assemblies using

standardised pipelines. In addition, EnteroBase assigns unique core-genome MLST (cgMLST) numbers to each genome, based on the typing of 2,512 genes in *E. coli*.

Population structure analysis

We utilised the Hierarchical Clustering (HierCC) algorithm in EnteroBase to assign our poultry genomes to stable clusters designated as HC0 up to HC1100, based on pair-wise differences between genomes at cgMLST alleles. In *E. coli*, the HC400 cluster has been shown to correspond to strain endemicity, while HC1100 corresponds to the seven-allele MLST clonal complexes [41]. The HierCC algorithm therefore lends itself as a very useful tool for the analysis of bacterial population structures at multiple levels of resolution. In a recent study of the population structure of *Clostridioides difficile*, Frentrup et al [42] showed that HierCC allows closely-related neighbours to be detected at 89% consistency between cgMLST pair-wise allelic differences and SNPs. We determined the closest relatives to our study *E. coli* isolates using the HC1100 cluster and reconstructed neighbour-joining trees using NINJA [43]. In order to compare the strain distribution that we observed among our study isolates with what pertains in poultry *E. coli* isolates from elsewhere, we further retrieved genomic assemblies from all publicly-available poultry *E. coli* isolates, stratified by their source continent and reconstructed NINJA neighbour-joining trees depicting the prevalence of STs per continent.

Analysis of accessory gene content

We used ARIBA v2.12.1 [44] to detect virulence factors, antimicrobial resistance genes and plasmid replicons. Briefly, this tool scans the short-read sequences against the core Virulence Factors Database [45] (virulence factors), ResFinder (AMR) [46] and PlasmidFinder (plasmid-associated genes) [47] databases and generates customised outputs, based on a

percentage identity of $\geq 90\%$ and coverage of $\geq 70\%$. The VFDB-core, ResFinder and PlasmidFinder databases were downloaded on 29 October 2018. Virulence factors were visualised by overlaying them onto the phylogenetic tree using the ggtree, ggplot2 and phangorn packages in RStudio v3.5.1.

We determined the prevalence of AMR genes among poultry *E. coli* isolates from the rest of the world, for comparison with what we found in isolates from this study. To do this, we interrogated the downloaded continent-stratified genomes as above using ABRicate v0.9.8 (<https://github.com/tseemann/abrigate>) to predict AMR-associated genes by scanning against the ResFinder database (accessed 28 July 2019), based on a percentage identity threshold of $\geq 90\%$ and a coverage of $\geq 70\%$.

Antimicrobial susceptibility

Due to logistic constraints, a third of the study isolates (20/68, 29%) were randomly selected for phenotypic susceptibility testing by minimum inhibitory concentrations (MICs). MICs were performed by the agar dilution method [48], according to the European Committee on Antimicrobial Susceptibility Testing v. 9.0 (EUCAST, 2019) guidelines. Stock solutions of 1,000 mg l⁻¹ were initially prepared, from which the working solutions were made. For each antibiotic, duplicate two-fold serial dilutions (from 32mg/L to 0.03 mg l⁻¹) were done in molten Mueller-Hinton agar (Oxoid, Basingstoke, UK). The results were interpreted according to EUCAST breakpoint tables (<http://www.eucast.org>). Where EUCAST cut-off values were not available, the recommended cut-off values from the Clinical Laboratory Standards Institute (<https://www.clsi.org>) (Performance Standards for Antimicrobial Susceptibility Testing (28th Information Supplement, M100-S28) were used.

Oxford nanopore sequencing

Prior to sequencing, DNA fragments were assessed using the Agilent 2200 TapeStation (Agilent Catalogue No. 5067-5579) to determine the fragment lengths. Long-read sequencing was carried out using the rapid barcoding kit (Oxford Nanopore Catalogue No. SQK-RBK004). Libraries were prepared following the manufacturer's instructions. An input DNA concentration of 400 ng was used for the library preparation and a final concentration of 75 µl of the prepared library loaded onto an R9.4 MinION flow cell. The final concentration of the library pool was assessed using the Qubit high-sensitivity DNA assay (Invitrogen, MA, USA).

Hybrid assembly and analysis of plasmids and phages

The long reads were base-called with Guppy, the Oxford Nanopore Technologies' post-sequencing processing software (<https://nanoporetech.com/>). The base-called FASTQ files were then concatenated into a single file each and demultiplexed based on their respective barcodes, using the qcat python command-line tool v1.1.0 (<https://github.com/nanoporetech/qcat>). We performed hybrid assemblies of the Illumina and nanopore reads with Unicycler v0.4.8.0 [49]. The quality of the hybrid assemblies was assessed with QUAST v5.0.0, de6973bb [35]. The hybrid assemblies were analysed for the presence of plasmids and prophages using ABRicate PlasmidFinder and PHASTER [50] respectively. Annotations of the assemblies were carried out using Prokka v1.13.3 [36].

Ethical statement

The study was approved by the joint Medical Research Council Unit The Gambia-Gambian Government ethical review board.

Results

Study population

We analysed nineteen caecal samples obtained from ten chickens and nine guinea fowl. Fifteen out of the nineteen (79%) samples yielded growth of *E. coli* on culture, from which 68 colonies were recovered.

Sequence type and phylogroup distribution

We recovered 28 seven-allele sequence types (STs), of which ST155 was the most common (22/68, 32%). Four of the STs were novel—two from chickens and two from guinea fowl. Seventeen of the 28 STs have been previously isolated from humans and or other vertebrates, while five (ST942, ST2165, ST2461, ST4392 and ST5826) have not been seen in humans before (Table 2). The isolates were spread over four of the eight known phylogroups of *E. coli*, but most belonged to phylogroups B1 and A, which are home to strains associated with human intestinal infections and avian colibacillosis [51, 52] (Figure 2). Hierarchical clustering resolved the study strains into 22 cgMLST complexes, indicating a high level of genomic diversity (File S2).

We generated complete, circular genome assemblies of the two novel sequence types isolated from guinea fowl: ST10654 (GF3-3) and ST9286 (GF4-3). Although neither strain encoded AMR genes or plasmids, GF3-3 contained three prophages (two intact, one incomplete), while GF4-3 harboured four prophages (three intact, one incomplete) (File S3).

Within-host genomic diversity and transmission of strains

Several birds (12/19, 63%) were colonised by two or more STs; in most cases, the STs spanned more than two phylotypes (Table 1). In two chickens, all five colony picks belonged to distinct STs. We observed some genetic diversity among multiple colonies of the same ST recovered from the same host (Table 3A). Most of these involved variants that differed by 0-

4 SNPs, i.e. variation likely to have arisen due to within-host evolution. However, in one instance, pair-wise SNP differences (ranging from 4 to 255) suggested independent acquisition of distinct clones. Pair-wise SNP analysis also suggested transmission of strains—including MDR isolates—between chickens and between chicken and guinea fowl (Table 3B and 3C) from the same household (File S4).

Prevalence of AMR, virulence factors and plasmid replicons among the study isolates

Twenty isolates (20/68, 29%) harboured at least one AMR gene and sixteen (16/68, 24%) were MDR, i.e. positive for genes predicted to convey resistance to three or more classes of antibiotics (Figure 4; File S5). Fourteen of the sixteen MDR isolates belonged to ST155—representing 64% (14/22) of the ST155 isolates recovered in this study. Notable among the resistance genes detected was the class A broad-spectrum beta-lactamase resistance (18/68, 26%). Phenotypic resistance was confirmed in >50% of the isolates tested, with an MDR rate of 75% (15/20).

Interestingly, the MDR isolates also harboured more genes encoding putative virulence factors than did less-resistant isolates (Figure 2). Overall, 125 unique virulence-associated genes were detected from the study isolates (File S6). Notably, the virulence and AMR profiles of co-colonising STs tended to differ from each other.

One or more plasmid replicons were detected in 69% (47/68) of the study isolates—with seventeen plasmid types detected overall (File S7). IncF plasmids were the most common. A single isolate carried the col156 virulence plasmid. The multi-drug resistant isolates often co-carried large IncF plasmids (IncFIA_1, ~27kb; IncFIB(AP001918)_1, ~60kb; IncFIC(FII)_1, ~56kb). Scrutiny of annotated assemblies revealed that resistance genes were often co-located on the same contig as one of the IncF plasmids. In three birds (Guinea fowl 2, Guinea fowl 5

and Guinea fowl 7), co-colonising strains (belonging to different STs) shared the same plasmid profile.

Population dynamics of study strains

Hierarchical clustering analyses provided evidence of genomic relationships between strains from poultry and those from humans (Table 5). Significant among these were ST2772 and ST4392, which were separated from human isolates belonging to these STs by just 41 and 68 alleles in the core-genome MLST scheme respectively (Figures 4 and 5). Similarly, ST86, ST6186 and ST602 were closest to isolates from livestock (Figures S9-11), suggesting exchange of strains between livestock species.

By contrast, three of the novel STs from this study (ST10654, ST9285, ST9286) were genetically distinct from anything else in the public domain. These belonged to unique HC1100 clusters in the cgMLST scheme and did not have any relatives in the seven-allele MLST scheme, even after allowing for two mismatches. Two of these (ST10654 from Guinea fowl 3 and ST9286 from Guinea fowl 4) now have complete genomic assemblies.

The global prevalence of strains and AMR among avian *E. coli* isolates

Phylogenomic analyses of 4,846 poultry *E. coli* isolates from all over the world revealed that ST155 is common among poultry isolates from Africa and South America (Figures S1 and S2). In contrast, ST117 is prevalent among poultry isolates from Europe and North America (Figures S3-S4), with ST156 and ST254 being the most common *E. coli* STs found in poultry from Asia and Oceania respectively (Figures S5 and S6).

Our phylogenetic analyses revealed that ST155 strains from Africa were dispersed among other ST155 isolates from the rest of the world; however, the majority of ST155 strains from this study belonged to a tight genomic cluster, comprised of isolates from poultry and

livestock from sub-Saharan Africa (separated by 38-39 alleles), except for a single isolate sourced from poultry in the US. In the cgMLST scheme, all the study ST155 isolates fell into four HC100 sub-clusters (100 alleles difference) (Figure S7). The largest sub-cluster (sub-cluster 1, HC100_43137) comprised ST155 isolates from this study and isolates from Uganda and Kenya; while sub-clusters 2 (HC100_73903), 3 (HC100_73905) and 4 (HC100_93719) occurred in the Gambia only—although distantly related to isolates from humans and a companion animal (Figure S8).

Antimicrobial resistance was high across the continents, with the highest prevalence of MDR in South America (100/131, 77%), followed by Asia (175/249, 70%), then Africa (392/591, 66%) (Table 4; File S8). Of note, the highest percentages of resistance globally were that of broad-spectrum beta-lactamases, while the lowest percentages of resistance were to colistin (File S6). Interestingly, the prevalence of colistin resistance was highest in Europe but did not occur in Oceania and North America.

Discussion

Here, we have described genomic diversity of *E. coli* from backyard chickens and guinea fowl reared in households in rural Gambia, West Africa. Backyard poultry from this rural setting harbour a remarkably diverse population of *E. coli* strains that encode antimicrobial-resistance genes and virulence factors important for infections in humans. Furthermore, we provide evidence of sharing of strains (including MDR strains) from poultry to poultry and between poultry, livestock and humans, with potential implications for public health.

Our results reflect the rich diversity that exists within the *E. coli* population from backyard poultry. Although our sample size was small (19 birds), we recovered as many as 28 sequence types of *E. coli*, four of which have not been seen before—even though more than quarter of a million of *E. coli* strains had been sequence-typed by March 2020. Three of our

novel STs differed by >945 alleles from their nearest relative—two of these now have complete assemblies. Also, some of the strains from this study were found in unique cgMLST HierCC clusters containing strains only from this study.

Our results confirm previous reports that phylogroups B1 and A are dominant phylogroups among *E. coli* isolates from both intensive and backyard poultry [54-57]. Hierarchical clustering analysis suggested that ST155 is common in African poultry. However, most of our ST155 strains belong to a unique cgMLST cluster containing closely related (38-39 alleles differences—and so presumably recently diverged) isolates from poultry and livestock from sub-Saharan Africa, suggesting that strains can be exchanged between livestock and poultry in this setting.

Rural backyard poultry can act as a source of transmission of infections to humans, due to the absence of biosecurity and daily contact with humans [58]. Indirect contact might occur through food or through contact with faeces, for example by children who are often left to play on the ground [59].

We observed a high prevalence of AMR genes among *E. coli* isolates sourced from African poultry. Similarly, high rates of genotypic MDR were detected among poultry *E. coli* isolates from the rest of the world, with ESBL (various types) being the most significant resistant gene detected. Poultry-associated ESBL genes have also been found among human clinical isolates [60]. Strikingly, most of our ST155 isolates encoded resistance to ≥ 3 classes of clinically relevant antibiotics, with the highest percentages to *bla*-TEM-1 beta-lactamase and tetracycline. This is worrying, as beta-lactamase-positive isolates are often resistant to several other classes of antibiotics [61, 62].

Our results are consistent with previous studies that reported ST155 isolates to be commonly associated with MDR [63, 64], but differ from other studies that have reported a low prevalence of AMR in backyard poultry. For example, in a study that compared the

prevalence of ESBL genes in backyard poultry and commercial flocks from West Bengal, India, none of the 272 *E. coli* isolates from backyard birds harboured any ESBL gene [65], while 30% of commercial birds carried ESBL genes. The absence of resistance in that study was attributed to a lack of exposure to antimicrobials. Similarly, *E. coli* from organic poultry in Finland were reported to be highly susceptible to most of the antimicrobials studied and no ESBL resistance was detected [66].

Although tetracycline is commonly used in poultry farming for therapeutic purposes [67], resistance to this antibiotic is known to be prevalent in poultry, even in the absence of the administration of this antibiotic [68]. Our results also suggest that IncF plasmids may play a role in the dissemination of AMR in our study population.

Many sub-Saharan countries lack clear guidelines on the administration of antibiotics in agriculture, although an increasing trend in the veterinary use of antimicrobials has been documented [69]. The usage of antimicrobials in developing countries is likely to increase because of increasingly intensive farming practices [70]. Europe has banned the use of antimicrobials as growth promoters since 2006 [71] and the use of all essential antimicrobials for prophylaxis in animal production since 2011 [72]. However, AMR may be less well controlled in other parts of the world.

Although APEC strains span several phylogroups (A, B1, B2 and D) and serogroups [52], the majority of APEC strains encode virulence genes associated with intestinal or extra-intestinal disease in humans. These include adhesion factors, toxins, iron-acquisition genes, and genes associated with serum resistance—such as *fyuA*, *iucD*, *iroN*, *iss*, *irp2*, *hlyF*, *vat*, *kpsM* and *ompT*. Although APEC isolates present different combinations of virulence factors, each retains the capability to cause colibacillosis [21, 73]. We did not detect haemolysin or serum survival genes in our study isolates; however, we recovered some of the known markers of intestinal and extraintestinal virulence in some study isolates—such the

enteroaggregative *E. coli* heat-stable enterotoxin and the vacuolating autotransporter toxin (*vat*, *astA*), invasion and evasion factors (*kpsM*, *kpsD*, *pla*) and adherence factors (*fim* and *pap* genes) that are associated with intestinal and extraintestinal infections in humans. Thus, these strains could cause disease in humans, should they gain access to the appropriate tissues.

Several birds were colonised with two or more STs and at least two phylotypes of *E. coli*. This level of diversity is probably a consequence of the frequent exposure of backyard poultry to the environment, livestock and humans. Co-colonisation of single hosts with multiple strains may facilitate the spread of AMR- and virulence-associated genes from resistant strains to other bacteria via both horizontal and vertical gene transfer [74]. A high co-colonisation rate of *E. coli* has been described in humans [75, 76] and in non-human primates [30]—involving pathogenic strains of *E. coli*. Recently, Li *et al* reported three to nine sequence types of colistin-resistant *E. coli* to co-exist within a single broiler chicken [28]. Here, we report co-colonisation with different lineages of *E. coli* in backyard chickens and guinea fowl. Unsurprisingly, co-colonising strains often had different AMR and virulence patterns.

An obvious limitation of our study is the small sample size. However, the inclusion of publicly-available sequences strengthens our analysis and inference of the population of *E. coli* in this setting. We also could not perform phenotypic susceptibility testing on all isolates. We acknowledge that a minor percentage of genotypic resistance predictions fail to correspond with phenotypic resistance [77].

Taken together, our results indicate a rich diversity of *E. coli* within backyard poultry from the Gambia, characterised by strains with a high prevalence of AMR and the potential to contribute to infections in humans. This, coupled with the potential for the exchange of

strains between poultry and livestock within this setting, might have important implications for human health and warrants continued surveillance.

Acknowledgements

We thank Dr Duto Fofana and Dr Ousman Ceesay of the Department of Livestock Services; Veterinary Services; Ministry of Agriculture, the Gambia for kindly providing access to their veterinary station at Abuko. We gratefully acknowledge the Research Support Office at the Medical Research Council Unit The Gambia at the London School of Hygiene and Tropical Medicine, in particular Mrs Elizabeth Stanley-Batchilly and Ms Isatou Cham, for project management support. We also thank Mr Steven Rudder and Mr David Baker for their assistance with the Illumina and Oxford nanopore sequencing.

Author contributions

Conceptualization, MA, MP; data curation, MP, NFA; formal analysis, EFN; funding, MP and MA; sample collection, EFN, SJ, AS; laboratory experiments, EFN, supervision, BKA, AR, NFA, JO, MP, MA; manuscript preparation – original draft, EFN; review and editing, NT, NFA, MP; review of final manuscript, all authors.

Funding information

MP, EFN, NT, AR and JO were supported by the BBSRC Institute Strategic Programme Microbes in the Food Chain BB/R012504/1 and its constituent projects 44414000A and 4408000A. NFA was supported by the Quadram Institute Bioscience BBSRC funded Core Capability Grant (project number BB/ CCG1860/1). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Bennett, C.E., et al., *The broiler chicken as a signal of a human reconfigured biosphere*. Royal Society Open Science. **5**(12): p. 180325.
2. Storey, A.A., et al., *Investigating the global dispersal of chickens in prehistory using ancient mitochondrial DNA signatures*. Plos One, 2012. **7**(7): p. e39171.
3. Miao, Y.W., et al., *Chicken domestication: an updated perspective based on mitochondrial genomes*. Heredity (Edinburgh), 2013. **110**(3): p. 277-82.
4. Alders, R. and R. Pym, *Village poultry: Still important to millions, eight thousand years after domestication*. World's Poultry Science Journal, 2009. **65**: p. 181-190.
5. Alders, R.G., et al., *Family poultry: Multiple roles, systems, challenges and options for sustainable contributions to household nutrition security through a planetary health lens*. Maternal and Child Nutrition, 2018. **14**(3): p. e12668.
6. Food and Agriculture Organization of the United Nations (FAO), *Recommendations on the prevention, control and eradication of highly pathogenic avian influenza in Asia*. FAO position paper, September 2004. Rome, Italy.
7. Olaniyan, O.F. and S. Camara, *Rural household chicken management and challenges in the Upper River Region of the Gambia*. Tropical Animal Health and Production, 2018. **50**(8): p. 1921-1928.
8. Alders, R., et al., *Smallholder poultry: Contributions to food and nutrition security*, in Encyclopedia of Food Security and Sustainability, 2019. Elsevier: Oxford: p. 292-298.

- 525 9. Kilonzo-Nthenge, A., et al., *Prevalence and antimicrobial resistance of pathogenic*
526 *bacteria in chicken and guinea fowl*. Poultry Science, 2008. **87**(9): p. 1841-8.
- 527 10. Samanta, I., S.N. Joardar and P.K. Das, *Biosecurity Strategies for Backyard Poultry:*
528 *A Controlled Way for Safe Food Production*. Food Control and Biosecurity, 2018. p.
529 481-517.
- 530 11. Alders, R. *Poultry for profit and pleasure. Diversification booklet number 3*.
531 Agricultural Support Systems Division, Food and Agriculture Organization of the
532 United Nations, 2004. Rome, Italy.
- 533 12. Food and Agriculture Organization of the United Nations (FAO), *Small-scale poultry*
534 *production technical guide*. 2004. Rome, Italy.
- 535 13. Alam J 1997: *Impact of smallholder livestock development project in some selected*
536 *areas of rural Bangladesh*. Livestock Research for Rural Development. Volume 9,
537 Article number 25. Retrieved May 4, 2020, from
538 <http://www.lrrd.org/lrrd9/3/bang932.htm>.
- 539 14. Branckaert, R.D.S., Guèye, E.F., 1999. *FAO's programme for support to family*
540 *poultry production*, in *Poultry as a tool in poverty eradication and promotion of*
541 *gender equality*. Proceedings of a Workshop, Tune Landboskole, Denmark, March
542 22-26, 1999. p. 244-256. Retrieved May 4, 2020, from
543 <http://www.fao.org/3/AC154E/AC154E00.htm>.
- 544 15. Burns, T.E., et al., *Preliminary investigation of bird and human movements and*
545 *disease-management practices in noncommercial poultry flocks in southwestern*
546 *British Columbia*. Avian Diseases, 2011. **55**(3): p. 350-7.
- 547 16. Karabozhilova, I., et al., *Backyard chicken keeping in the Greater London Urban*
548 *Area: welfare status, biosecurity and disease control issues*. British Poultry Science,
549 2012. **53**(4): p. 421-30.

- 550 17. Beam, A., et al., *Salmonella awareness and related management practices in U.S.*
551 *urban backyard chicken flocks*. Preventive Veterinary Medicine, 2013. **110**(3-4): p.
552 481-8.
- 553 18. Nataro, J.P. and J.B. Kaper, *Diarrheagenic Escherichia coli*. Clinical Microbiology
554 Reviews, 1998. **11**(1): p. 142-201.
- 555 19. Picard, B., et al., *The link between phylogeny and virulence in Escherichia coli*
556 *extraintestinal infection*. Infection and Immunity, 1999. **67**(2): p. 546-53.
- 557 20. Escobar-Paramo, P., et al., *Identification of forces shaping the commensal*
558 *Escherichia coli genetic structure by comparing animal and human isolates*.
559 Environmental Microbiology, 2006. **8**(11): p. 1975-84.
- 560 21. Rodriguez-Siek, K.E., et al., *Characterizing the APEC pathotype*. Veterinary
561 Research, 2005. **36**(2): p. 241-56.
- 562 22. Barnes HJ, N.L., Vaillancourt J. *Colibacillosis*, in Diseases of Poultry. Iowa: Iowa
563 State University Press. 2008. **12**: p. 691–738.
- 564 23. Hedman, H.D., et al., *Impacts of small-scale chicken farming activity on*
565 *antimicrobial-resistant Escherichia coli carriage in backyard chickens and children*
566 *in rural Ecuador*. One Health, 2019. **8**: p. 100112.
- 567 24. Sarba, E.J., et al., *Identification and antimicrobial susceptibility profile of Escherichia*
568 *coli isolated from backyard chicken in and around ambo, Central Ethiopia*. BMC
569 Veterinary Research, 2019. **15**(1): p. 85.
- 570 25. Langata, L.M., et al., *Antimicrobial resistance genes in Salmonella and Escherichia*
571 *coli isolates from chicken droppings in Nairobi, Kenya*. BMC Research Notes, 2019.
572 **12**(1): p. 22.
- 573 26. Adzitey, F., P. Assoah-Peprah and G.A. Teye, *Whole-genome sequencing of*
574 *Escherichia coli isolated from contaminated meat samples collected from the*

575 *Northern Region of Ghana reveals the presence of multiple antimicrobial resistance*
576 *genes*. Journal of Global Antimicrobial Resistance, 2019. **18**: p. 179-182.

577 27. Borzi, M.M., et al., *Characterization of avian pathogenic Escherichia coli isolated*
578 *from free-range helmeted guineafowl*. Brazillian Journal of Microbiology, 2018.
579 **49**(1): p. 107-112.

580 28. Li, X.-P., et al., *Within-host heterogeneity and flexibility of mcr-1 transmission in*
581 *chicken gut*. International Journal of Antimicrobial Agents, 2020. **55**(1): p. 105806.

582 29. Roca, A., et al., *Effects of community-wide vaccination with PCV-7 on pneumococcal*
583 *nasopharyngeal carriage in The Gambia: A Cluster-randomized trial*. Plos Medicine,
584 2011. **8**(10): p. e1001107.

585 30. Foster-Nyarko, E., et al., *Genomic diversity of Escherichia coli isolates from non-*
586 *human primates in the Gambia*. bioRxiv, 2020: p. 2020.02.29.971309.

587 31. Connor, T.R., et al., *CLIMB (the Cloud Infrastructure for Microbial Bioinformatics):*
588 *an online resource for the medical microbiology community*. Microbial Genomics,
589 2016. **2**(9): p. e000086-e000086.

590 32. Wingett, S.W. and S. Andrews, *FastQ Screen: A tool for multi-genome mapping and*
591 *quality control*. F1000Research, 2018. **7**: p. 1338-1338.

592 33. Bolger, A.M., M. Lohse and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina*
593 *sequence data*. Bioinformatics (Oxford, England), 2014. **30**(15): p. 2114-2120.

594 34. Bankevich, A., et al., *SPAdes: a new genome assembly algorithm and its applications*
595 *to single-cell sequencing*. Journal of Computational Biology : a Journal of
596 Computational Molecular Cell Biology, 2012. **19**(5): p. 455-477.

597 35. Gurevich, A., et al., *QUAST: quality assessment tool for genome assemblies*.
598 Bioinformatics, 2013. **29**(8): p. 1072-1075.

- 599 36. Seemann, T., *Prokka: rapid prokaryotic genome annotation*. Bioinformatics, 2014.
600 **30**(14): p. 2068-9.
- 601 37. Wirth, T., et al., *Sex and virulence in Escherichia coli: an evolutionary perspective*.
602 Molecular Microbiology, 2006. **60**(5): p. 1136-1151.
- 603 38. Stamatakis, A., *RAxML version 8: a tool for phylogenetic analysis and post-analysis*
604 *of large phylogenies*. Bioinformatics (Oxford, England), 2014. **30**(9): p. 1312-1313.
- 605 39. Croucher, N.J., et al., *Rapid phylogenetic analysis of large samples of recombinant*
606 *bacterial whole genome sequences using Gubbins*. Nucleic Acids Research, 2015.
607 **43**(3): p. e15-e15.
- 608 40. Zhou, Z., et al., *GrapeTree: visualization of core genomic relationships among*
609 *100,000 bacterial pathogens*. Genome Research, 2018. **28**(9): p. 1395-1404.
- 610 41. Zhou, Z., et al., *The Enterobase user's guide, with case studies on Salmonella*
611 *transmissions, Yersinia pestis phylogeny and Escherichia core genomic diversity*.
612 Genome Research, 2020. **30**(1): p. 138-152.
- 613 42. Frentrop, M., et al., *Global genomic population structure of Clostridioides difficile*.
614 BioRxiv, 2019: p. 727230.
- 615 43. Wheeler, T.J. *Large-Scale Neighbor-Joining with NINJA*. in *Algorithms in*
616 *Bioinformatics*. 2009. Berlin, Heidelberg: Springer Berlin Heidelberg.
- 617 44. Hunt, M., et al., *ARIBA: rapid antimicrobial resistance genotyping directly from*
618 *sequencing reads*. Microbial Genomics, 2017. **3**(10): p. e000131-e000131.
- 619 45. Liu, B., et al., *VFDB 2019: a comparative pathogenomic platform with an interactive*
620 *web interface*. Nucleic Acids Research, 2019. **47**(D1): p. D687-d692.
- 621 46. Zankari, E., et al., *Identification of acquired antimicrobial resistance genes*. Journal
622 of Antimicrobial Chemotherapy, 2012. **67**(11): p. 2640-4.

- 623 47. Carattoli, A., et al., *In silico detection and typing of plasmids using PlasmidFinder*
624 *and plasmid multilocus sequence typing*. Antimicrobial Agents and Chemotherapy,
625 2014. **58**(7): p. 3895-903.
- 626 48. Wiegand, I., K. Hilpert and R.E. Hancock, *Agar and broth dilution methods to*
627 *determine the minimal inhibitory concentration (MIC) of antimicrobial substances*.
628 Nature Protocols, 2008. **3**(2): p. 163-75.
- 629 49. Wick, R.R., et al., *Unicycler: Resolving bacterial genome assemblies from short and*
630 *long sequencing reads*. Plos Computational Biology, 2017. **13**(6): p. e1005595.
- 631 50. Arndt, D., et al., *PHASTER: a better, faster version of the PHAST phage search tool*.
632 Nucleic Acids Research, 2016. **44**(W1): p. W16-21.
- 633 51. Clermont, O., et al., *Characterization and rapid identification of phylogroup G in*
634 *Escherichia coli, a lineage with high virulence and antibiotic resistance potential*.
635 Environmental Microbiology, 2019. **21**(8): p. 3107-3117.
- 636 52. Mellata, M., *Human and avian extraintestinal pathogenic Escherichia coli: infections,*
637 *zoonotic risks and antibiotic resistance trends*. Foodborne Pathogens and Disease,
638 2013. **10**(11): p. 916-932.
- 639 53. Escobar-Paramo, P., et al., *A specific genetic background is required for acquisition*
640 *and expression of virulence factors in Escherichia coli*. Molecular Biology and
641 Evolution, 2004. **21**(6): p. 1085-94.
- 642 54. Coura, F.M., et al., *Phylogenetic group determination of Escherichia coli isolated*
643 *from animals samples*. Scientific World Journal, 2015. **2015**: p. 258424.
- 644 55. Asadi, A., et al., *ECOR phylotyping and determination of virulence genes in*
645 *Escherichia coli isolates from pathological conditions of broiler chickens in poultry*
646 *slaughter-houses of southeast of Iran*. Veterinary Research Forum : an international
647 quarterly journal, 2018. **9**(3): p. 211-216.

- 648 56. Messaili, C., Y. Messai and R. Bakour, *Virulence gene profiles, antimicrobial*
649 *resistance and phylogenetic groups of fecal Escherichia coli strains isolated from*
650 *broiler chickens in Algeria*. Veterinaria Italiana, 2019. **55**(1): p. 35-46.
- 651 57. Vounba, P., et al., *Molecular characterization of Escherichia coli isolated from*
652 *chickens with colibacillosis in Senegal*. Foodborne Pathogens and Disease, 2018.
653 **15**(8): p. 517-525.
- 654 58. Patyk, K.A., et al., *An epidemiologic simulation model of the spread and control of*
655 *highly pathogenic avian influenza (H5N1) among commercial and backyard poultry*
656 *flocks in South Carolina, United States*. Preventive Veterinary Medicine, 2013.
657 **110**(3): p. 510-524.
- 658 59. Dione, M.M., et al., *Clonal differences between Non-Typhoidal Salmonella (NTS)*
659 *recovered from children and animals living in close contact in The Gambia*. Plos
660 Neglected Tropical Diseases, 2011. **5**(5): p. e1148.
- 661 60. Castellanos, L.R., et al., *High Heterogeneity of Escherichia coli Sequence Types*
662 *Harbouring ESBL/AmpC Genes on IncII Plasmids in the Colombian Poultry Chain*.
663 PloS one, 2017. **12**(1): p. e0170777-e0170777.
- 664 61. Mathers, A.J., G. Peirano and J.D. Pitout, *The role of epidemic resistance plasmids*
665 *and international high-risk clones in the spread of multidrug-resistant*
666 *Enterobacteriaceae*. Clinical Microbiology Reviews, 2015. **28**(3): p. 565-91.
- 667 62. Reich, F., V. Atanassova and G. Klein, *Extended-spectrum beta-lactamase- and*
668 *AmpC-producing enterobacteria in healthy broiler chickens, Germany*. Emerging
669 Infectious Diseases, 2013. **19**(8): p. 1253-9.
- 670 63. Dominguez, J.E., et al., *Characterization of Escherichia coli carrying mcr-1-plasmids*
671 *recovered from food animals from Argentina*. Frontiers in Cellular and Infection
672 Microbiology, 2019. **9**: p. 00041

- 673 64. van Hoek, A., et al., *Longitudinal study of ESBL Escherichia coli carriage on an*
674 *organic broiler farm*. Journal of Antimicrobial Chemotherapy, 2018. **73**(12): p. 3298-
675 3304.
- 676 65. Samanta, I., et al., *Comparative possession of Shiga toxin, intimin, enterohaemolysin*
677 *and major extended spectrum beta lactamase (ESBL) genes in Escherichia coli*
678 *isolated from backyard and farmed poultry*. Iranian Journal of Veterinary Research,
679 2015. **16**(1): p. 90-3.
- 680 66. Pohjola, L., et al., *Zoonotic public health hazards in backyard chickens*. Zoonoses and
681 Public Health, 2016. **63**(5): p. 420-430.
- 682 67. Fairchild, A.S., et al., *Effects of orally administered tetracycline on the intestinal*
683 *community structure of chickens and on tet determinant carriage by commensal*
684 *bacteria and Campylobacter jejuni*. Applied and Environmental Microbiology, 2005.
685 **71**(10): p. 5865-5872.
- 686 68. van den Bogaard, A.E. and E.E. Stobberingh, *Epidemiology of resistance to*
687 *antibiotics. Links between animals and humans*. International Journal of
688 Antimicrobial Agents, 2000. **14**(4): p. 327-35.
- 689 69. Alonso, C.A., et al., *Antibiotic resistance in Escherichia coli in husbandry animals:*
690 *the African perspective*. Letters in Applied Microbiology, 2017. **64**(5): p. 318-334.
- 691 70. Van Boeckel, T.P., et al., *Global trends in antimicrobial use in food animals*.
692 Proceedings of the National Academy of Sciences, 2015. **112**(18): p. 5649.
- 693 71. Castanon, J.I.R., *History of the use of antibiotic as growth promoters in european*
694 *poultry feeds*. Poultry Science, 2007. **86**(11): p. 2466-2471.
- 695 72. Maron, D.F., T.J.S. Smith and K.E. Nachman, *Restrictions on antimicrobial use in*
696 *food animal production: an international regulatory and economic survey*.
697 Globalization and Health, 2013. **9**: p. 48-48.

698 73. Schouler, C., et al., *Diagnostic strategy for identifying avian pathogenic Escherichia*
699 *coli based on four patterns of virulence genes*. Journal of Clinical Microbiology,
700 2012. **50**(5): p. 1673-8.

701 74. Laxminarayan, R., et al., *Antibiotic resistance-the need for global solutions*. Lancet
702 Infectious Diseases, 2013. **13**(12): p. 1057-98.

703 75. Lidin-Janson, G., et al., *The homogeneity of the faecal coliform flora of normal*
704 *school-girls, characterized by serological and biochemical properties*. Medical
705 Microbiology and Immunology, 1978. **164**(4): p. 247-53.

706 76. Schlager, T.A., et al., *Clonal diversity of Escherichia coli colonizing stools and*
707 *urinary tracts of young girls*. Infection and Immunity, 2002. **70**(3): p. 1225-9.

708 77. Doyle, R.M., et al., *Discordant bioinformatic predictions of antimicrobial resistance*
709 *from whole-genome sequencing data of bacterial isolates: an inter-laboratory study*.
710 Microbial Genomics, 2020. **6**(2): p. e000335.

711
712
713

Table 1: Characteristics of the study population

| Sample ID | Poultry species | Gender | Household | Colony picks | Recovered Sequence Types (No. of colonies per ST) | Phylogroup distribution (STs per phylogroup) |
|-----------|-----------------|---------|-----------|----------------------------|---|--|
| C1 | Chicken | Rooster | 1 | No <i>E. coli</i> isolated | | |
| C2 | Chicken | Hen | 3 | 1 | 155 (1) | B1 (155) |
| C3 | Chicken | Rooster | 2 | 5 | 155 (1), 48 (1), 746 (1) 2461 (1), 542 (1) | A (48, 746, 2461, 542), B1 (155) |
| C4 | Chicken | Rooster | 2 | 5 | 1423 (1), 337 (1), 9285* (1), 540 (1), 58 (1) | A (540), B1 (1423, 337, 9285*, 58) |
| C5 | Chicken | Hen | 2 | 2 | 155 (2) | B1 (155) |
| C6 | Chicken | Rooster | 2 | 5 | 155 (3), 9284* (2) | B1 (155), E (9284*) |
| C7 | Chicken | Rooster | 3 | 5 | 155 (4), 602 (1) | B1 (155, 602) |
| C8 | Chicken | Rooster | 4 | 5 | 5286 (1), 2772 (2), 6186 (1), 2165 (1) | A (5286), B1 (2772, 6186, 2165) |
| C9 | Chicken | Hen | 5 | No <i>E. coli</i> isolated | | |
| C10 | Chicken | Rooster | 5 | No <i>E. coli</i> isolated | | |
| GF1 | Guinea fowl | Rooster | 1 | 5 | 540 (5) | A (540) |
| GF2 | Guinea fowl | Rooster | 1 | 5 | 155 (4), 540 (1) | A (540), B1 (155) |
| GF3 | Guinea fowl | Rooster | 3 | 5 | 540 (2), 443 (1), 6025 (1), 10654* (1) | A (540), B1 (443), D (6025), E (10654) |
| GF4 | Guinea fowl | Rooster | 6 | 5 | 155 (4), 9286* (1) | B1 (155, 9286) |
| GF5 | Guinea fowl | Hen | 6 | 5 | 155 (2), 4392 (1), 86 (1), 942 (1) | B1 (155, 4392, 86, 942) |
| GF6 | Guinea fowl | Hen | 1 | 5 | 540 (1), 2067 (4) | A (540), B1 (2067) |
| GF7 | Guinea fowl | Rooster | 2 | 5 | 212 (4), 155 (1) | B1 (155, 212) |
| GF8 | Guinea fowl | Rooster | 7 | No <i>E. coli</i> isolated | | |
| GF9 | Guinea fowl | Rooster | 8 | 5 | 2614 (2), 295 (1) 196 (1), 2067 (1) | B1 (2614, 295, 196) |
| Total | | | | 68 | | |

Novel sequence types are designated by an asterisk (*).

Table 2: Prevalence of the study sequence types in Enterobase

| ST | Source | Phylotype | Prevalence in Enterobase |
|-------|----------------------|-----------|--|
| 48 | Chicken | A | Human, livestock, <i>Celebes</i> ape |
| 58 | Chicken | B1 | Human, livestock, poultry |
| 86 | Guinea fowl | B1 | Human, livestock, companion animal, poultry |
| 155 | Chicken, Guinea fowl | B1 | Human, poultry, mink, livestock |
| 196 | Guinea fowl | B1 | Human, livestock, companion animal, environment |
| 212 | Guinea fowl | B1 | Human, poultry, deer, companion animal |
| 295 | Guinea fowl | B1 | Human, poultry, livestock, companion animal, environment, food, |
| 337 | Chicken | B1 | Human, rhinoceros, poultry, environment (soil and water) |
| 443 | Guinea fowl | B1 | Human, environment, livestock |
| 540 | Chicken, Guinea fowl | A | Human, environment (water and sewage), livestock, poultry, gull, rabbit, plant, oyster, fish |
| 542 | Chicken | A | Human, livestock, poultry |
| 602 | Chicken | B1 | Human, poultry, livestock, bird, fish, reptile |
| 746 | Chicken | A | Human, poultry, fish, livestock, environment (water) |
| 942 | Guinea fowl | B1 | Environment, food, companion animal, livestock |
| 1423 | Chicken | B1 | Human, reptile, livestock |
| 2067 | Guinea fowl | B1 | Human, environment |
| 2165 | Chicken | B1 | Livestock, companion animal, reptile, bird |
| 2461 | Chicken | A | Sheep, poultry |
| 2614 | Guinea fowl | B1 | Human |
| 2772 | Chicken | B1 | Human, livestock, environment |
| 4392 | Guinea fowl | B1 | Livestock, wild animal, companion animal |
| 5826 | Chicken | A | Poultry |
| 6025 | Guinea fowl | D | Unknown source§ |
| 6186 | Chicken | B1 | Livestock, environment |
| 9284 | Chicken | E | Novel |
| 9285 | Chicken | B1 | Novel |
| 9286 | Guinea fowl | B1 | Novel |
| 10654 | Guinea fowl | D | Novel |

§ ST6025 occurred only one other isolate in Enterobase, beside the study ST6025 strain. However, the source of isolation of this other isolate was not available.

Table 3A: Within-host single nucleotide polymorphism diversity between multiple genomes of the same ST recovered from the same bird

| Sample ID | Sequence type (ST) | Colonies per ST | Pair-wise SNP distances between multiple colonies of the same ST |
|-----------|--------------------|-----------------|--|
| C5 | 155 | 2 | 0 |
| C6 | 155 | 3 | 0 |
| C6 | 9284 | 2 | 4 |
| C7 | 155 | 4 | 0 |
| C8 | 2772 | 2 | 4 |
| GF1 | 540 | 5 | 0-3 |
| GF2 | 155 | 4 | 0 |
| GF3 | 540 | 2 | 2 |
| GF4 | 155 | 4 | 0-4 |
| GF5 | 155 | 2 | 0 |
| GF6 | 2067 | 4 | 0 |
| GF7 | 212 | 4 | 4-255 |
| GF9 | 2614 | 2 | 0 |

“C” denotes chickens and “GF” denotes guinea fowl.

Table 3B: Single nucleotide polymorphism differences between isolates recovered from Chicken 3, Chicken 5, Chicken 6 and Guinea fowl 7

| | C3-5 | C5-1 | C5-2 | GF7-2 | C6-1 | C6-2 | C6-3 |
|-------|------|------|------|-------|------|------|------|
| C3-5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C5-1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C5-2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GF7-2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C6-1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C6-2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C6-3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Chicken samples are denoted by “C” and guinea fowl samples by “GF”. All the isolates in this transmission network encoded resistance to ≥ 3 classes of antimicrobials.

Table 3C: Single nucleotide diversity differences between isolates recovered from guinea fowls 1, 2 and 6.

| | GF1-1 | GF1-2 | GF1-3 | GF1-4 | GF1-5 | GF2-3 | GF6-1 |
|-------|-------|-------|-------|-------|-------|-------|-------|
| GF1-1 | 0 | 2 | 3 | 1 | 1 | 2 | 3 |
| GF1-2 | 2 | 0 | 3 | 1 | 1 | 2 | 3 |
| GF1-3 | 3 | 3 | 0 | 2 | 2 | 3 | 2 |
| GF1-4 | 1 | 1 | 2 | 0 | 0 | 1 | 2 |
| GF1-5 | 1 | 1 | 2 | 0 | 0 | 1 | 2 |
| GF2-3 | 2 | 2 | 3 | 1 | 1 | 0 | 3 |
| GF6-1 | 3 | 3 | 2 | 2 | 2 | 3 | 0 |

Table 4: Global prevalence of AMR genes

| | Europe | Africa | South America | North America | Asia | Oceania |
|-----------------|--------------|--------------|---------------|----------------|--------------|--------------|
| Tetracycline | 564/752, 75% | 559/591, 95% | 108/131, 83% | 2480/2975, 83% | 228/249, 92% | 132/148, 90% |
| Aminoglycoside | 303/752, 40% | 378/591, 64% | 94/131, 72% | 1497/2975, 50% | 172/249, 69% | 56/148, 38% |
| Beta-lactamase | 303/752, 40% | 246/591, 42% | 127/131, 98% | 933/2975, 31% | 157/249, 63% | 61/148, 41% |
| Sulphonamide | 338/752, 45% | 377/591, 64% | 84/131, 65% | 1174/2975, 39% | 167/249, 67% | 52/148, 35% |
| Trimethoprim | 192/752, 25% | 353/591, 52% | 58/131, 45% | 176/2975, 6% | 143/249, 57% | 66/148, 45% |
| Chloramphenicol | 303/752, 40% | 69/591, 13% | 36/131, 28% | 69/2975, 2% | 131/249, 53% | 0/148, 0% |
| Quinolone | 51/752, 7% | 144/591, 24% | 24/131, 18% | 17/2975, 1% | 74/249, 30% | 0/148, 0% |
| Lincosamide | 57/752, 8% | 0/591, 0% | 12/131, 9% | 0/2975, 0% | 14/249, 6% | 1/148, 1% |
| Macrolide | 20/752, 3% | 79/591, 13% | 3/131, 2% | 30/2975, 1% | 92/249, 37% | 0/148, 0% |
| Fosfomycin | 8/752, 1% | 4/591, 1% | 31/131, 24% | 19/2975, 1% | 71/249, 29% | 0/148, 0% |
| Streptogramin | 0/752, 0% | 0/591, 0% | 23/131, 18% | 0/2975, 0% | 0/249, 0% | 0/148, 0% |
| Colistin | 29/752, 4% | 0/591, 0% | 9/131, 7% | 0/2975, 0% | 119/249, 48% | 0/148, 0% |
| MDR | 406/752, 54% | 392/591, 66% | 100/131, 77% | 1236/2975, 42% | 175/249, 70% | 56/148, 44% |

The full list of resistance genes that were detected is presented in File S6.

Table 5: Closest relatives to the Gambian poultry strains

| 7-gene ST | cgST HC100 sub-cluster | Study poultry host | Neighbour host | Neighbour's country of isolation | Allelic distance |
|-----------|------------------------|-------------------------|-------------------------|----------------------------------|------------------|
| ST9286 | NA | Guinea fowl | Chicken | Gambia (this study) | 945 |
| ST9285 | NA | Chicken | Guinea fowl | Gambia (this study) | 945 |
| ST10654 | NA | Guinea fowl | Unknown avian source | Kenya | 1324 |
| ST155 | 43137 | Chicken and Guinea fowl | Poultry | US | 32-34 |
| ST2772 | NA | Chicken | Human | Kenya | 41 |
| ST6186 | NA | Chicken | Livestock | US | 58 |
| ST540 | 10207 | Guinea fowl | Human | UK | 59 |
| ST58 | 25133 | Chicken | Unknown | Unknown | 59 |
| ST2461 | 93699 | Chicken | Human | Kenya | 64 |
| ST2165 | 12281 | Chicken | Food | Kenya | 66 |
| ST4392 | NA | Guinea fowl | Human | UK | 68 |
| ST602 | NA | Chicken | Livestock | US | 70 |
| ST540 | 70056 | Chicken | Food | UK | 72 |
| ST540 | 1320 | Guinea fowl | Poultry | US | 73 |
| ST942 | NA | Guinea fowl | Environment (tap water) | Australia | 76 |
| ST212 | NA | Guinea fowl | Seagull | Australia | 81 |
| ST5826 | NA | Chicken | Water | UK | 91 |
| ST1423 | 27957 | Chicken | Reptile | US | 96 |
| ST337 | 73054 | Chicken | Reptile | US | 96 |
| ST196 | NA | Guinea fowl | Human | Kenya | 102 |
| ST155 | 93719 | Chicken | Tanzania | Human | 106 |
| ST86 | NA | Guinea fowl | US | Livestock | 131 |
| ST155 | 73905 | Guinea fowl | Companion animal | US | 137 |
| ST542 | 93732 | Chicken | Poultry | US | 148 |
| ST746 | NA | Chicken | Poultry | US | 148 |
| ST295 | NA | Guinea fowl | Human | Mexico | 162 |
| ST48 | 93724 | Chicken | Unknown | UK | 163 |
| ST542 | 93697 | Chicken | Environment (soil/dust) | US | 194 |
| ST155 | 73903 | Guinea fowl | Nepal | Human | 195 |
| ST443 | 93721 | Guinea fowl | Unknown | Unknown | 224 |
| ST6025 | NA | Guinea fowl | Unknown | US | 245 |
| ST2614 | NA | Guinea fowl | Human | China | 284 |
| ST9284 | NA | Chicken | Environment (soil/dust) | North America | 293 |
| ST2067 | NA | Guinea fowl | Human | Gambia | 458 |

Figure Legends

Figure 1. Study sample-processing flow diagram. TBX, Tryptone-Bile-X-glucuronide agar; MLST, multi-locus sequence typing; cgMLST, core genome multi-locus sequence typing.

Figure 2. A maximum-likelihood phylogeny of the study isolates reconstructed with RAxML, based on non-repetitive, non-recombinant core SNPs, using a general time-reversible nucleotide substitution model with 1,000 bootstrap replicates. The tip labels indicate the sample names, with the respective Achtman sequence types (ST) and HC1100 (cgST complexes) indicated next to the sample names. The colour codes indicate the respective phylogroups to which the isolates belong. The outgroup and the other *E. coli* reference genomes denoting the major *E. coli* phylogroups are in black. Asterisks (*) are used to indicate novel STs. Overlaid on the tree are the predicted virulence factors for each isolate. The virulence genes are grouped according to their function. Chicken isolates are denoted “C” and guinea fowl samples by “GF”, with the suffix indicating the colony pick. We have not shown multiple colonies of the same Achtman ST recovered from a single individual—in such instances, only one representative isolate is shown. Nor have we shown virulence factors that were detected only in the reference genomes. The red box highlights multi-drug resistant isolates which concurrently harbour putative fitness and colonisation factors that are important for invasion of host tissues and evasion of host immune defences. The full names of virulence factors and their known functions are provided in File S8.

Figure 3. A: A bar graph showing the prevalence of resistance genes found among the study isolates, using the core Virulence Factors Database (reference 45) (virulence factors), ResFinder (AMR) (reference 46) and PlasmidFinder (plasmid-associated genes) (reference

47) databases, with a cut-off percentage identity of $\geq 90\%$ and coverage of $\geq 70\%$. The full list of the resistance genes that were detected is presented in File S5.

B: A bar graph depicting the prevalence of phenotypic antimicrobial resistance in 20 isolates.

The results were interpreted using the recommended breakpoint tables from EUCAST

(<http://www.eucast.org>) or the Clinical Laboratory Standards Institute (<https://www.clsi.org>)

(Performance Standards for Antimicrobial Susceptibility Testing (28th Information

Supplement, M100-S28) where EUCAST cut-off values were not available.

Figure 4. A NINJA neighbour-joining tree showing the phylogenetic relationship between our study ST2772 (Achtman) strain and all other publicly available genomes that fell within the same HC1100 cluster (cgST complex). The locations of the isolates are displayed, with the genome counts displayed in square brackets. The branch lengths are annotated with the allelic distances separating the genomes. Strains from this study are highlighted in red. The sub-tree (B) shows the closest relatives to the study strains, with the allelic distance separating them displayed with the arrow (41 alleles).

Figure 5: A NINJA neighbour-joining tree showing the phylogenetic relationship between the avian ST4392 (Achtman) strain from this study and all other publicly available genomes that cluster together at HC1100 level (cgST complex). The legend shows the continent of isolation of the isolates, with genome counts displayed in square brackets. Gambian poultry strains are highlighted in red. The study ST strain is separated from a human ST4392 isolate by 68 alleles, as shown in the subtree (B).

Supplementary Figures

Figure S1. A NINJA neighbour-joining tree of all publicly available *E. coli* poultry isolates from Africa, showing the prevalence of Achtman sequence types (STs). The dominant ST is highlighted with a red box. The legend displays the top 27 STs, with the respective genome counts displayed in square brackets.

Figure S2. A NINJA neighbour-joining tree of all publicly available *E. coli* poultry isolates from South America, depicting the prevalence of Achtman sequence types (STs). The most common ST found among *E. coli* isolates from this continent is ST155 (highlighted with a red box), similar to Africa (see Figure S1). The top 20 STs are displayed in the legend, with the respective genome counts displayed [in square brackets].

Figure S3. A NINJA neighbour-joining tree of all publicly available *E. coli* poultry isolates from Europe, depicting the prevalence of Achtman sequence types (STs). The top 20 STs are displayed in the legend, with the most common ST among poultry isolates from this continent (ST117) highlighted with a red box. The respective genome count per ST is also displayed.

Figure S4. A NINJA neighbour-joining tree of all publicly available *E. coli* poultry isolates from North America, showing the prevalence of Achtman sequence types (STs). The most common ST among poultry isolates from this continent is ST117 (highlighted with a red box). The legend displays the top 23 STs, with the respective genome counts displayed next to the STs.

Figure S5. A NINJA neighbour-joining tree of all publicly available *E. coli* poultry isolates from Asia, showing the prevalence of Achtman sequence types (STs). The most common ST

among poultry isolates from this continent is ST156 (highlighted with a red box). The legend displays the top 25 STs, with the respective genome counts displayed next to the STs.

Figure S6. A NINJA neighbour-joining tree of all publicly available *E. coli* poultry isolates from Oceania, depicting the prevalence of Achtman sequence types (STs). The most common ST found among *E. coli* isolates from this continent is ST354 (highlighted with a red box). The first 18 STs are displayed in the legend, with the respective genome counts displayed.

Figure S7. A phylogenetic tree showing the global distribution of *E. coli* ST155 isolates. The study ST155 isolates are highlighted in red. Hierarchical clustering resolved four sub-clusters, encompassing the Gambian ST155 strains, displayed in red boxes. The legend displays the locations of the isolates, with the genome counts depicted in square brackets.

Figure S8. NINJA phylogenetic trees showing the sub-clusters for the study ST155 population within the cgMLST hierarchical clustering scheme. The largest sub-cluster (HC100_43137) (**A**) encompassed most of the study ST155 isolates (13/22, 59%), which were closely related to isolates from poultry and livestock in sub-Saharan Africa (separated by 38-39 alleles). Sub-clusters 2 (HC100_73903) (**B**), 3 (HC100_73905) (**C**) and 4 (HC100_93719) (**D**) were unique to the Gambia, although distantly related to isolates from humans and a companion animal. The red highlights indicate the study ST155 isolates. The locations of the isolates are displayed in the respective legends, with the genome counts indicated in square brackets.

Figure S9. NINJA phylogenetic trees showing the closest neighbours to avian ST86 isolates from this study. The nearest relatives occurred in livestock, depicted with the arrow. The

legend indicates the location of isolation, with the genome count displayed in square brackets.

Figure S10. A NINJA phylogenetic tree showing the closest neighbours to avian ST6186 isolates from this study (A). The nearest relatives occurred in livestock from Kenya (B), separated by 58 alleles (depicted with the arrow). The branch lengths display the allelic distance between the genomes. The legend indicates the location of isolation, with the genome count displayed in square brackets.

Figure S11. A NINJA phylogenetic tree showing the closest neighbours to avian ST602 isolates from this study. The nearest relatives were isolated from livestock from the US (B), separated by 70 alleles (depicted with the arrow). The branch lengths display the allelic distance between the genomes. The legend indicates the location of isolation, with the genome count displayed in square brackets.

Supplementary Files

File S1. Reference strains that were included in this study.

File S2. A summary of the sequencing statistics of the study isolates derived from this study.

File S3. A. A summary of the sequencing statistics of two novel sequence types derived from guinea fowl. **B.** Prophage types detected from long-read sequences using PHASTER (reference 50).

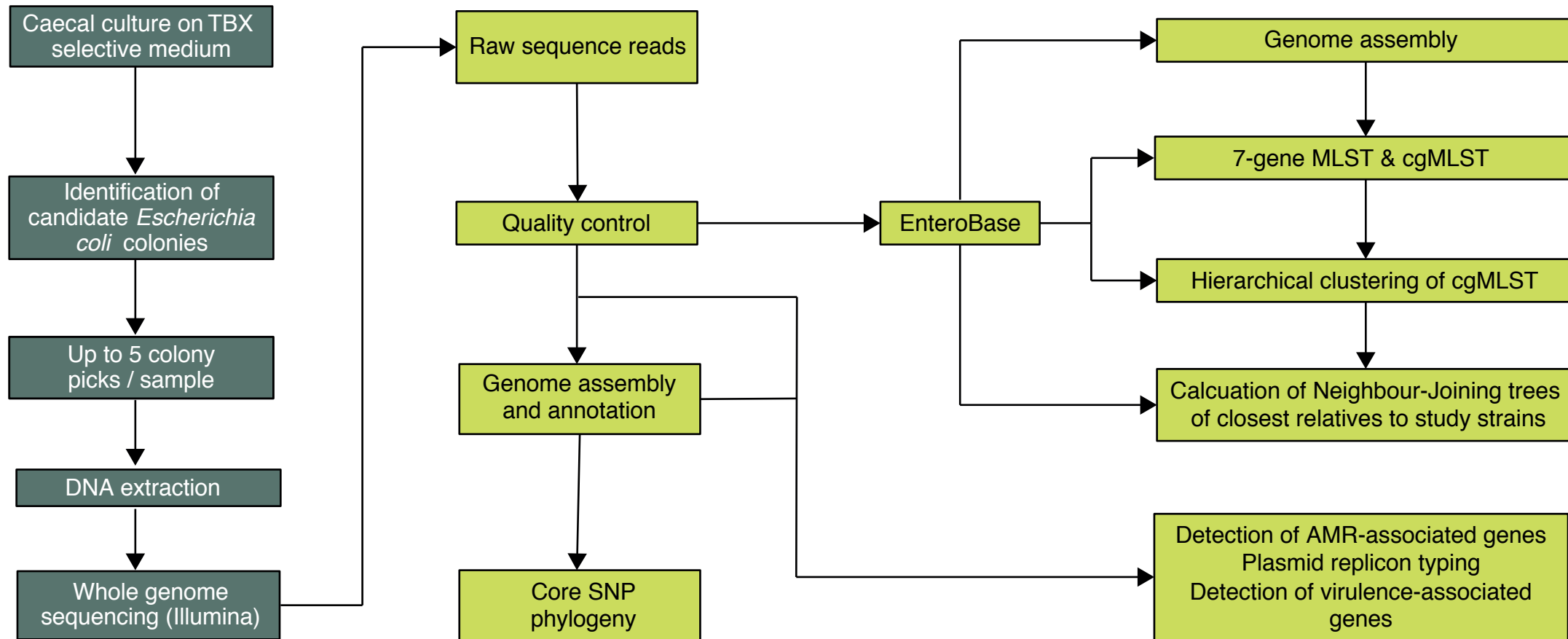
File S4. A pair-wise single nucleotide polymorphism matrix depicting the SNP distances between the study genomes calculated from the core genome alignment using snp-dists v0.6 (<https://github.com/tseemann/snp-dists>).

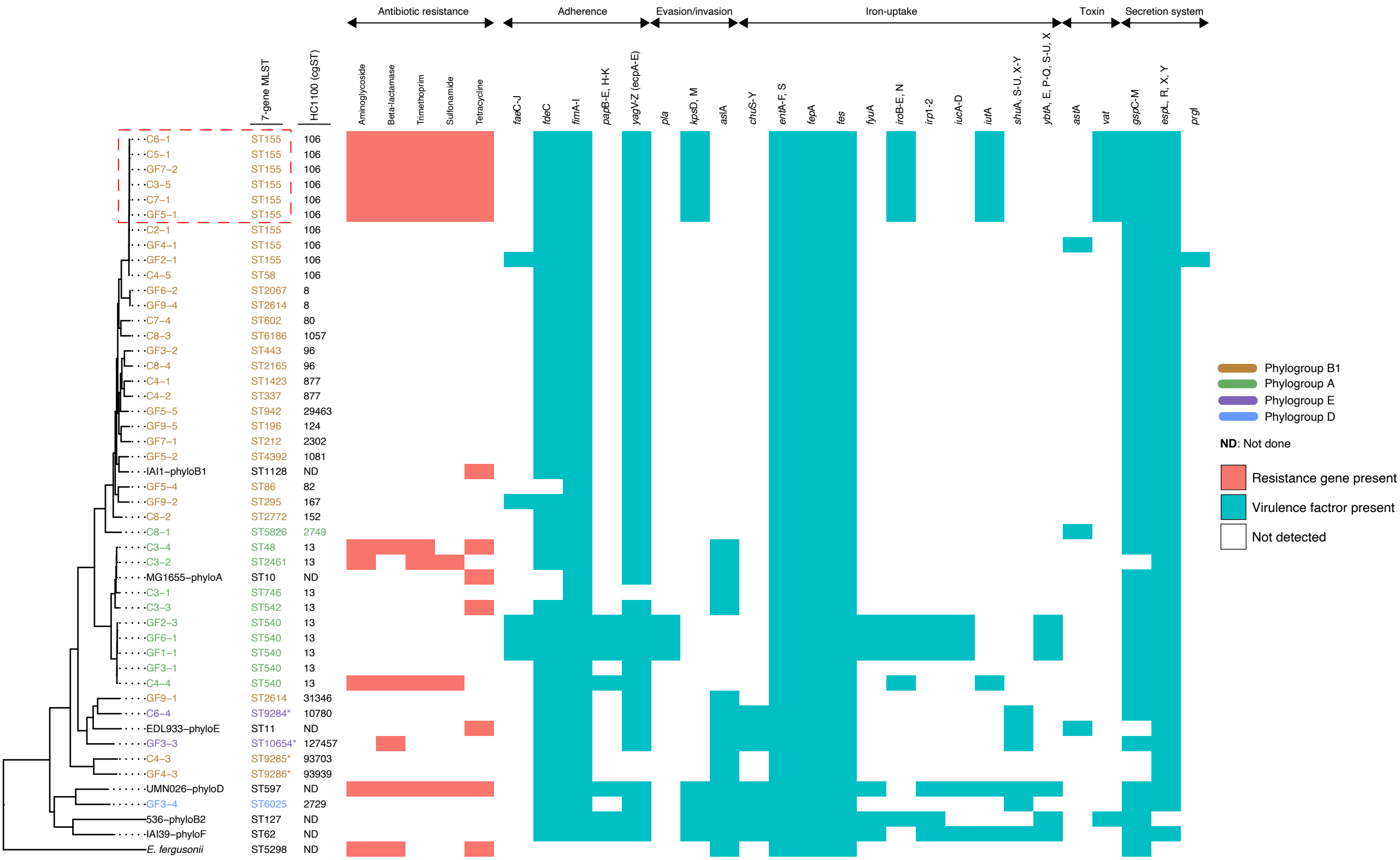
File S5. A summary of the predicted resistance genes for the study isolates, using ResFinder (Reference 24).

File S6. A list of the virulence factors detected among the study isolates using ARIBA VFDB (References 22 and 23) and their known functions.

File S7. Predicted plasmid-replicons from the study isolates.

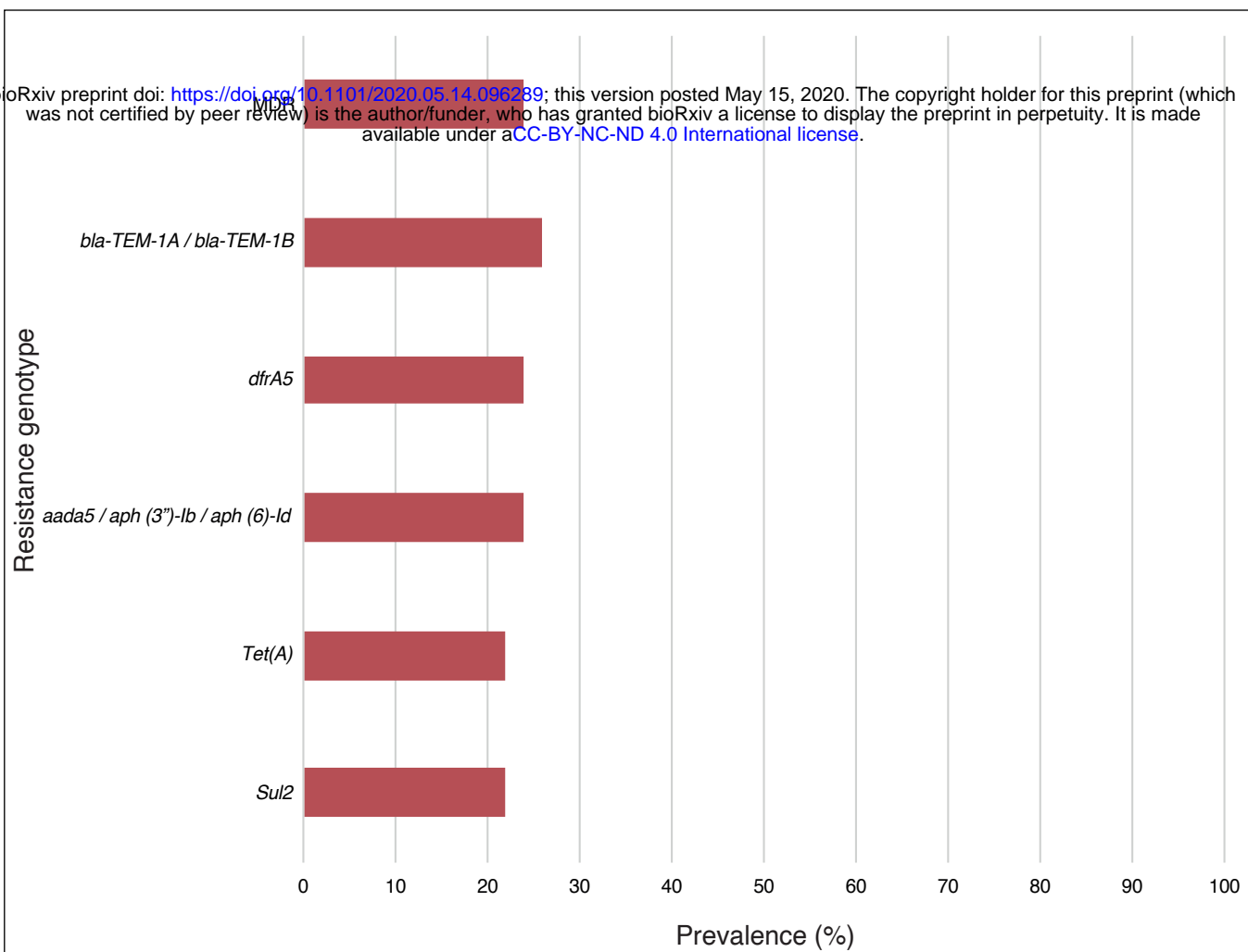
File S8. The global prevalence of antimicrobial resistance genes predicted from whole-genome sequences of all publicly-available poultry *E. coli* isolates.



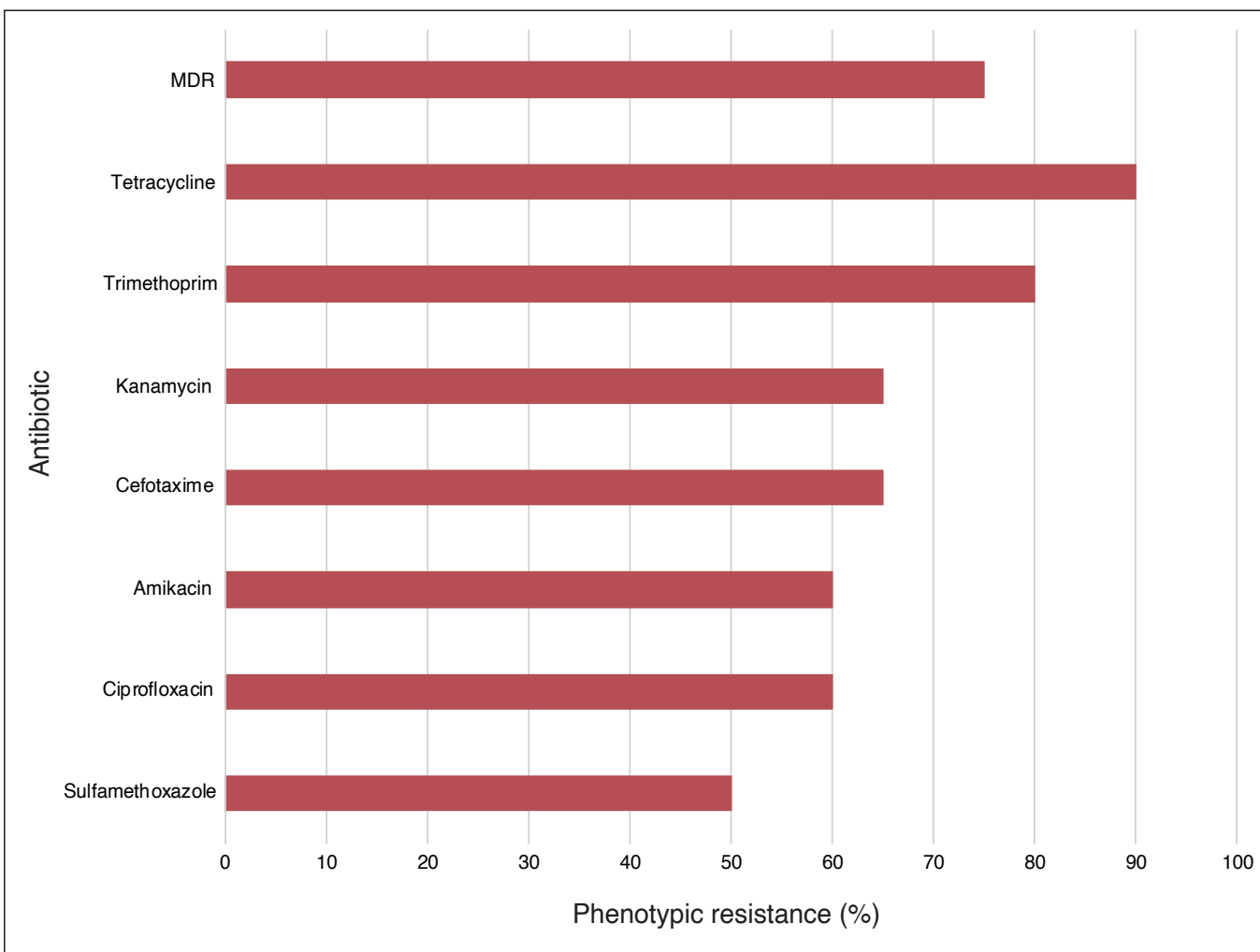


A

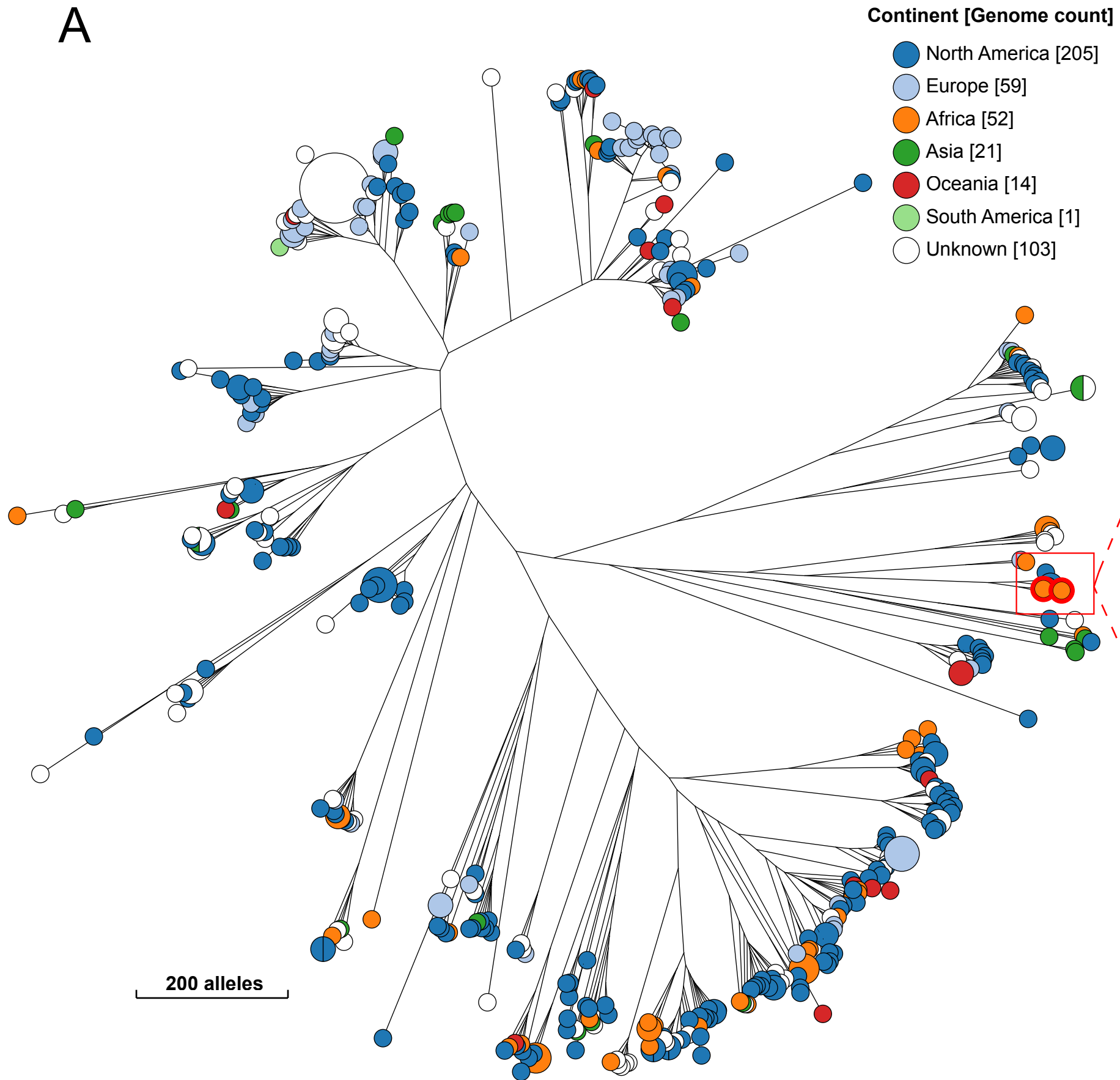
bioRxiv preprint doi: <https://doi.org/10.1101/2020.05.14.096289>; this version posted May 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



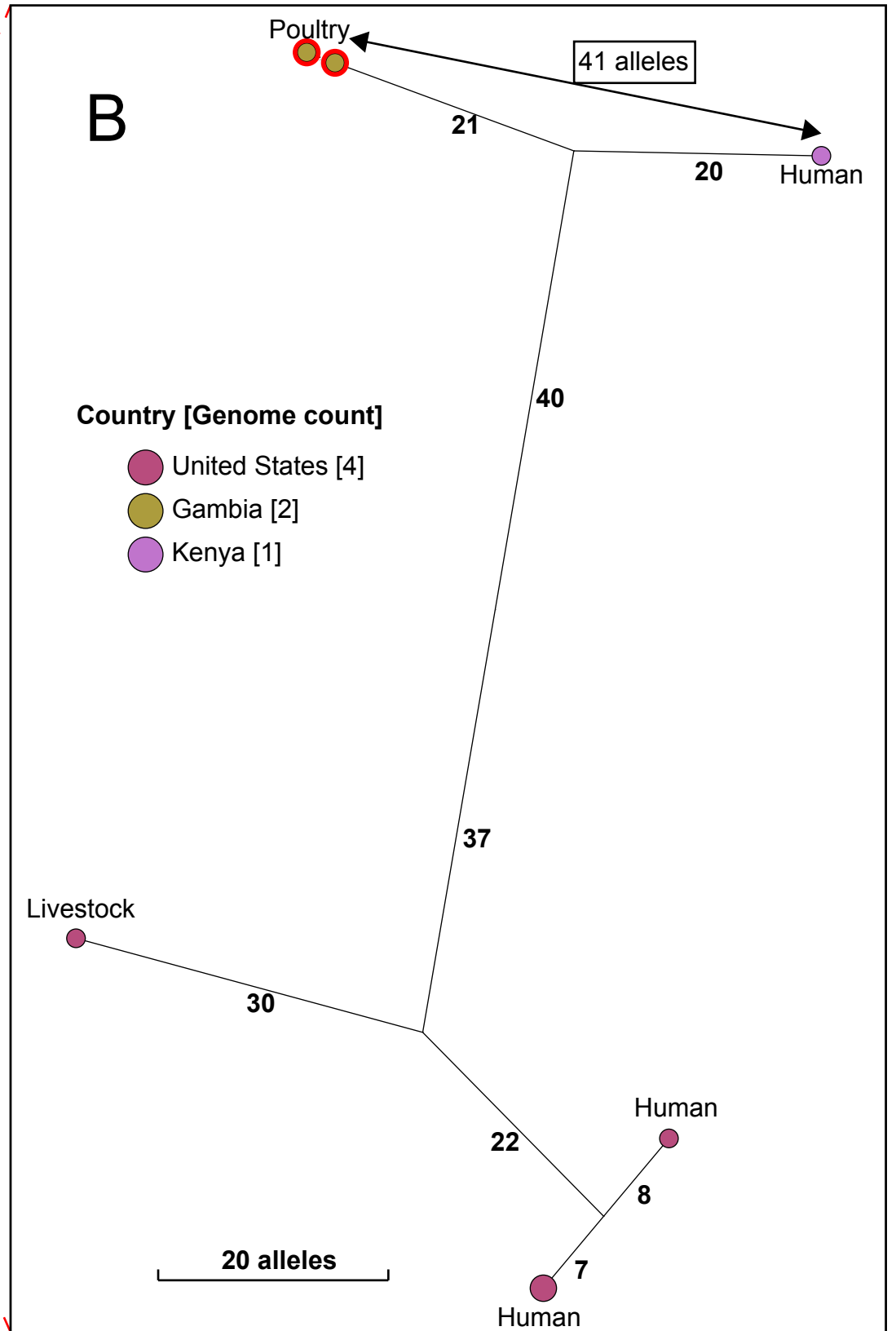
B



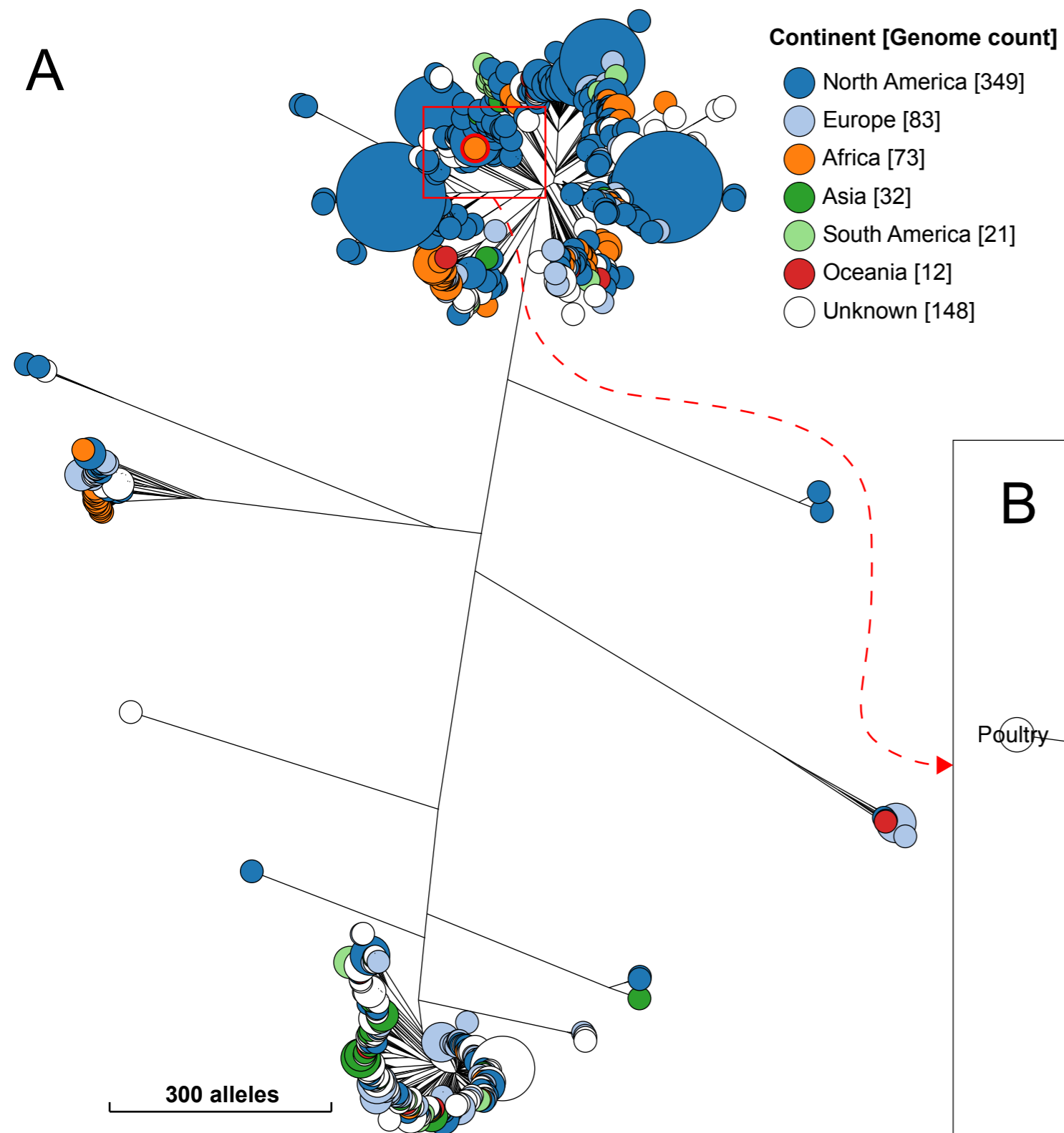
A



B



A



B

