



Chicken Production and Human Clinical *Escherichia coli* Isolates Differ in Their Carriage of Antimicrobial Resistance and Virulence Factors

Reed Woyda,^a Adelumola Oladeinde,^b Zaid Abdo^{a,c}

^aProgram of Cell and Molecular Biology, Colorado State University, Fort Collins, Colorado, USA

^bU.S. National Poultry Research Center, USDA-ARS, Athens, Georgia, USA

^cDepartment of Microbiology, Immunology, and Pathology, Colorado State University, Fort Collins, Colorado, USA

ABSTRACT Contamination of food animal products by *Escherichia coli* is a leading cause of foodborne disease outbreaks, hospitalizations, and deaths in humans. Chicken is the most consumed meat both in the United States and across the globe according to the U.S. Department of Agriculture. Although *E. coli* is a ubiquitous commensal bacterium of the guts of humans and animals, its ability to acquire antimicrobial resistance (AMR) genes and virulence factors (VFs) can lead to the emergence of pathogenic strains that are resistant to critically important antibiotics. Thus, it is important to identify the genetic factors that contribute to the virulence and AMR of *E. coli*. In this study, we performed in-depth genomic evaluation of AMR genes and VFs of *E. coli* genomes available through the National Antimicrobial Resistance Monitoring System GenomeTrackr database. Our objective was to determine the genetic relatedness of chicken production isolates and human clinical isolates. To achieve this aim, we first developed a massively parallel analytical pipeline (Reads2Resistome) to accurately characterize the resistome of each *E. coli* genome, including the AMR genes and VFs harbored. We used random forests and hierarchical clustering to show that AMR genes and VFs are sufficient to classify isolates into different pathogenic phylogroups and host origin. We found that the presence of key type III secretion system and AMR genes differentiated human clinical isolates from chicken production isolates. These results further improve our understanding of the interconnected role AMR genes and VFs play in shaping the evolution of pathogenic *E. coli* strains.

IMPORTANCE Pathogenic *Escherichia coli* causes disease in both humans and food-producing animals. *E. coli* pathogenesis is dependent on a repertoire of virulence factors and antimicrobial resistance genes. Food-borne outbreaks are highly associated with the consumption of undercooked and contaminated food products. This association highlights the need to understand the genetic factors that make *E. coli* virulent and pathogenic in humans and poultry. This research shows that *E. coli* isolates originating from human clinical settings and chicken production harbor different antimicrobial resistance genes and virulence factors that can be used to classify them into phylogroups and host origins. In addition, to aid in the repeatability and reproducibility of the results presented in this study, we have made a public repository of the Reads2Resistome pipeline and have provided the accession numbers associated with the *E. coli* genomes analyzed.

KEYWORDS *Escherichia coli*, antimicrobial resistance, poultry production, random forests, virulence factors

Escherichia coli are ubiquitous commensal bacteria in the gut of both humans and food-producing animals and rarely cause disease but may acquire antimicrobial resistance (AMR) genes and virulence factors (VFs) resulting in increased pathogenicity (1). Pathogenic *E. coli* has consistently ranked in the top five causative agents of

Editor Martha Vives, Universidad de los Andes

Copyright © 2023 Woyda et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Zaid Abdo, zaid.abdo@colostate.edu, Adelumola Oladeinde, ade.oladeinde@usda.gov, or Reed Woyda, reed.woyda@colostate.edu.

The authors declare no conflict of interest.

[This article was published on 18 January 2023 with a standard copyright line ("© 2023 American Society for Microbiology. All Rights Reserved."). The authors elected to pay for open access for the article after publication, necessitating replacement of the original copyright line, and this change was made on 30 January 2023.]

Received 13 July 2022

Accepted 17 December 2022

Published 18 January 2023

disease outbreaks, outbreak-associated illnesses, and hospitalizations in the United States (2–4) and is responsible for billions of dollars of annual health care associated costs in the United States (5, 6). Pathogenic *E. coli* infections in humans may result in acute to severe diarrhea or dysentery, urinary tract infections, and meningitis (7–9). *E. coli* pathogenesis is dependent on the VF repertoire, which enables the bacterium to evade host defenses, adhere to host surfaces, and successfully invade—and replicate in—host tissues. For example, VFs such as toxins, iron-acquisition systems, and fimbriae play integral roles in the pathogenicity of extraintestinal *E. coli* strains that cause urinary tract infections in humans (10). These uropathogenic *E. coli* strains are able to colonize human mucosal surfaces due to surface adhesion VFs, including the P, F, and type 1 fimbriae encoded by *pap*, *sfa*, and *fim* genes (11).

Outbreaks in human populations caused by pathogenic *E. coli* are attributed to the consumption of undercooked and contaminated foods, including meats and vegetables (12). *E. coli* infections in food-producing animals usually results in diarrhea but can include acute fatal septicemia, airsacculitis, pericarditis, and perihepatitis (13). In poultry, avian-pathogenic *E. coli* infections, termed avian colibacillosis, can result in economic loss due to the cost of treatment, as well as from culling of flocks (14). Colibacillosis is a leading cause of morbidity and mortality in poultry, as noted by decreases in the feed conversion ratio, egg production, and hatching rates (15–17). The severity of the disease depends on the VF repertoire of the strain, including genes encoding iron acquisition and transport systems (18).

Extraintestinal pathogenic *E. coli* strains that cause urinary tract infections, neonatal meningitis, and sepsis have VFs similar to those of *E. coli* isolates from meat and animal sources (19). Likewise, some avian-pathogenic *E. coli* and extraintestinal pathogenic *E. coli* isolates have been reported to share similar VFs and belong to the same multilocus sequence type (MLST) and phylogroup (20–23). Specifically, a microarray-based study of *E. coli* isolates from both human and animal sources in Denmark identified 66 to 87 genes, including both virulence factors and antimicrobial resistance genes, which were present both in human urinary tract infection isolates and in isolates obtained from poultry and pig products (24). Genes such as the iron acquisition genes *iutA* and *iroN*, as well as fimbrial *papA* encoding the Pap fimbrial major subunit and adhesion genes (e.g., *papGII* encoding Pap adhesion), were detected in both urinary tract infection and poultry-related isolates (24). *In vivo* and *in vitro* experimental studies have shown that extraintestinal pathogenic *E. coli* recovered from humans can cause disease in avian hosts and, similarly, avian-pathogenic *E. coli* isolates recovered from avian host can cause disease in mammalian models (25–29).

Infections with pathogenic *E. coli* are commonly treated with antibiotics; however, increasing levels of AMR impose difficulties in selecting effective treatment options (30). The spread and increased prevalence of AMR has been linked to both the overuse and the misuse of antibiotics in human clinical settings, as well as in food animal production (31). Consumer opinion has led to a reduction in antimicrobial use in food animal production to mitigate the spread of AMR. However, recent studies have indicated that AMR can still persist in food animal production even after the removal or stoppage of antibiotics (32, 33). The comobilization and coacquisition of AMR genes and VFs through horizontal gene transfer can result in highly pathogenic *E. coli* strains. Recent studies have demonstrated a correlation and close genetic linkage of AMR genes and VFs in pathogenic bacterial strains (34, 35). Pan et al. (35) performed a comprehensive analysis of over 9,000 bacterial genomes from multiple species and hosts and observed the coexistence of AMR genes and VFs from human-associated pathogens. Therefore, the aim of this study was to evaluate and compare the current distribution of AMR genes and VFs present in *E. coli* isolates from chicken production and human clinical settings in the United States.

To do so, we determined the AMR genes and VFs present in nearly 800 *E. coli* genomes collected from chicken production and the human clinical settings. Data for all isolates were obtained from the National Antimicrobial Resistance Monitoring System (NARMS). We used the most recent World Health Organization (WHO) classification of antimicrobials

important to human health to determine the extent to which resistance to these antimicrobials are prevalent in chicken production and human clinical settings. We hypothesized that resistance to antimicrobials of critical and high importance to human health would be present in both human- and chicken-derived isolates. In addition, we hypothesized that *E. coli* isolates from human clinical settings and chicken production will differ in their carriage of AMR genes and VFs. We found that *E. coli* isolates obtained from chicken production and human clinical settings harbored AMR genes predicted to confer resistance to many antimicrobials classified as highly and critically important to human health by the WHO. Furthermore, we were able to classify all *E. coli* isolates into their respective phylogroups and host origins using random forest classification and hierarchical clustering of the AMR genes and VFs found in their genomes.

RESULTS

***E. coli* isolate selection and genome assembly.** To evaluate and compare the current distribution of AMR genes and VFs present in *E. coli* isolates from chicken production and the human clinical settings in the United States, we took advantage of the U.S. National Antimicrobial Resistance Monitoring System (NARMS) GenomeTrackr database (available through the NCBI [<https://www.ncbi.nlm.nih.gov/pathogens>]). We chose NARMS because it monitors enteric bacteria and pathogens across all states to determine their resistance to antimicrobials used in veterinary and human medicine. NARMS is a collaboration between agencies within the U.S. Department of Health and Human Services (HHS; Food and Drug Administration [FDA] and Centers for Disease Control and Prevention [CDC]) and the U.S. Department of Agriculture (USDA; Food Safety and Inspection Service, Animal and Plant Health Inspection Service, and the Agricultural Research Service). Each USDA agency taking part in NARMS tests bacterial samples at various stages of food animal production. The NARMS database holds bacterial genome sequences that originate from 15 distinct human, animal, and food sources and are categorized into four NARMS samples sources: (i) human clinical isolates; (ii) food animal production isolates from cecal samples at slaughter; (iii) samples routinely collected at inspected establishments as part of FSIS verification testing; and (iv) raw meats (chicken, ground turkey, ground beef, and pork chops) collected at retail outlets in 20 states (23 sites). We selected 837 chicken production isolates and 874 human clinical isolates, with collection dates from 1 January 2018 until 30 March 2020, using “species,” “host,” and “location” search filters (see Materials and Methods for a detailed description). *E. coli* isolate data were either available as Illumina or Oxford Nanopore FASTQ reads or as FASTA assemblies.

A total of 130 FASTA assemblies were obtained for human clinical isolates, whereas all chicken production isolates and remaining human clinical isolates had available FASTQ reads. The FASTA assemblies were obtained through the isolate’s associated BioProject, and quality control was done using QUAST (36). The FASTQ reads for each isolate were first normalized to 30× coverage using Trinity’s *insilico_read_normalization.pl* script (<https://github.com/trinityrnaseq/trinityrnaseq/wiki>). After quality control, the average genome coverage was 35.8× and 31.7× for human clinical and chicken production isolates, respectively (Fig. 1A). Totals of 385 chicken production isolates and 535 human clinical isolates were excluded from further analysis because they did not meet our preset quality control metrics based on genome coverage $\geq 20\times$ and assembled genome quality ($N_{50} \geq 10,000$ bp; Fig. 1B). The remaining *E. coli* isolates from chicken production and human clinical settings were confirmed to be *E. coli* through MLST (see Table S1, <https://github.com/tseemann/mlst>). This resulted in our study population consisting of 452 (57%) chicken production isolates and 339 (43%) human clinical isolates (Fig. 2A and B). Isolates originated from a diverse set of sources (see Table S1 in the supplemental material) and locations within the United States (Fig. 2C).

Reads2Resistome enables high-throughput genome assembly and resistome characterization. We developed Reads2Resistome (R2R) (37) to streamline the quality control, assembly, annotation, and resistome characterization of the chicken production and human clinical genomes (see the supplemental material for detailed description of

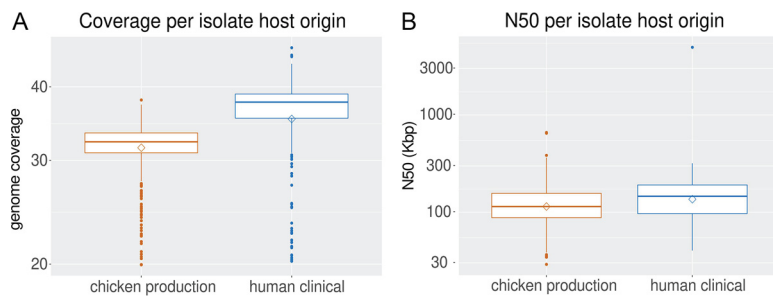


FIG 1 Average coverage, postnormalization, and average N_{50} , postassembly, of human clinical and poultry production isolates. (A) Isolate FASTQ reads were downloaded from the NCBI and normalized to $30\times$ using Trinity’s *insilico_read_normalization.pl* script. (B) Isolates with available FASTQ reads were assembled and annotated using Reads2Resistome, including N_{50} calculation by QUAST. FASTA assemblies were annotated using the additional script provided by Reads2Resistome for postassembly annotation. A diamond shape indicates a mean value, and the horizontal lines within a box indicate the mean. The upper and lower sides of the box correspond to the 25th and 75th percentiles. Plots were generated in R v4.0.4 using ggplot2 v3.3.3.

Reads2Resistome [Fig. S1]). Utilizing a singularity (38) container, R2R maintains software and database versions resulting in reproducible results that can be replicated using the same input data and command-line options. We first assessed the accuracy and performance of R2R using short (Illumina) and long FASTQ reads (PacBio and Oxford Nanopore) of *Salmonella enterica* serovar Heidelberg and *E. coli* isolates recovered from the ceca of broiler chickens (see Table S2). Short and long reads-only assemblies resulted in the shortest run-time with an average of 6 min per sample, while hybrid assemblies (short and long reads were combined) required ~ 1 h per sample (see Table S3). Genomes assembled with short reads only or a hybrid approach had better resistome characterization and gene annotation than genomes assembled using long reads only (see Tables S4 and S5). For this study, all *E. coli* isolates were assembled from short Illumina reads except for a single isolate with long-read data (SRR11174251).

AMR gene distribution not explained by isolation year or location. We conducted principal-component analysis (PCA) to determine whether the presence or absence of identified AMR genes corresponded to the state origin of the isolates. Overlapping 95% confidence intervals were observed for all isolates in all states (see Fig. S1). We also conducted PCA on the presence/absence matrix of these AMR genes to assess impact of time of sampling. Separation of isolates based on their associated isolation year was also not observed (see Fig. S2). Accordingly, year and location will not be included in further analysis.

Determining the phylogeny of *E. coli* isolate phylogroups. To determine whether there is an evolutionary based separation between chicken production and human clinical isolates, we performed a whole-genome alignment to the *E. coli* K12-MG1655 reference genome (accession U00096). Single-nucleotide polymorphisms (SNPs) with

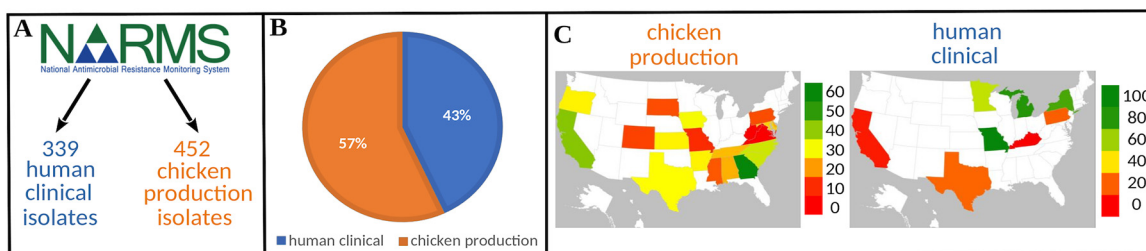


FIG 2 Graphical description of human clinical and poultry production isolates. (A and B) FASTQ reads or FASTA assemblies for each *E. coli* isolate was downloaded from the NARMS GenomeTrackr for chicken production and human clinical isolates. Isolates were selected from submission dates of 1 January 2018 to 30 March 2020 using the filtering criteria described in Materials and Methods. After quality control and filtering, 452 chicken production isolates (57%) and 399 human clinical isolates (43%) remained. (C) Isolate geographic locations across the United States. Maps were generated using the “Filled Map” chart type in Microsoft Excel from Microsoft Office 365 (Microsoft Corporation).

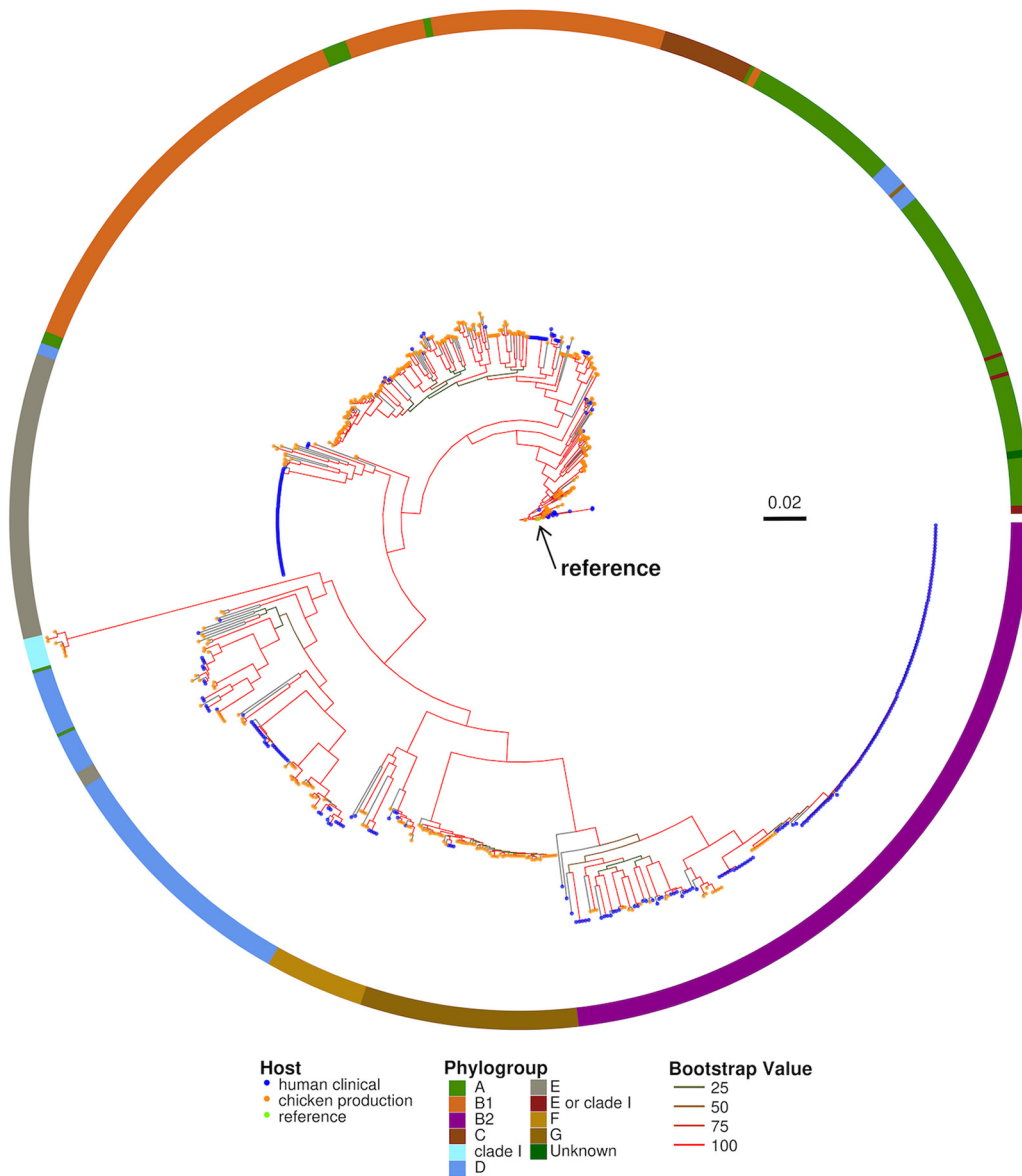


FIG 3 Whole-genome SNP-based maximum-likelihood phylogenetic tree. SNPs found between the *E. coli* K12-MG1655 (accession [U00096](#)) reference genome, and each of the 791 *E. coli* isolates was used for alignment using Snippy v4.6.0 (Snippy 2018). The resulting whole-genome SNP alignment was used to construct maximum-likelihood phylogenetic trees under a GTR+GAMMA model using RAxML with 100 bootstraps (RAxMLHPC-PTHREADS, v8.2.12). The strength of nodal support is indicated by the branch color: red (100), orange (50> to ≤ 75), dark orange (25> to ≤ 50), and green (0> to ≤ 25). (Ring) Phylogroups identified using ClermonTyping v1.4.0. Phylogroups are labeled by color: A (green), B1 (orange), B2 (magenta), C (brown), clade I (aquamarine), D (blue), E (gray), E or clade I (maroon), F (gold), G (dark gold), and unknown (dark green).

respect to the reference genome were used to generate a phylogenetic tree. In addition, we used ClermonTyping (39) to classify the isolates into phylogroups. *E. coli* isolates within the phylogenetic tree clustered into single clades and chicken production isolates were less divergent from the *E. coli* K-12 reference genome than human clinical isolates (Fig. 3). *E. coli* isolates were phylotyped into one of the phylogroups A, B1, B2, C, clade I, D, E, “E or clade I,” F, and G (Table 1). The majority of the *E. coli* genomes in our data set belonged to phylogroups A, B1, and B2. Phylogroups B1 and B2 were the most represented groups, with each accounting for ~23% of the total *E. coli* isolates. Phylogroup determination for one isolate was unsuccessful (identified as unknown), but the isolate clustered with phylogroup A isolates on the SNP-based tree.

Phylogroup classification is useful for the rapid and easy identification of potentially

TABLE 1 Distribution of phylogroups in chicken production and human clinical isolates^a

Phylogroup	No. of isolates (% total)		
	Total (n = 791)	Chicken production (n = 452)	Human clinical (n = 339)
A	132 (16.62)	109 (85.58)*	23 (17.42)*
B1	183 (23.05)	148 (80.87)*	35 (19.13)*
B2	184 (23.17)	28 (15.21)*	156 (84.78)*
C	23 (2.90)	18 (78.26)	5 (21.74)
Clade I	8 (1.01)	8 (100)	0 (0)
D	103 (12.97)	56 (54.37)	47 (45.63)
E	75 (9.45)	14 (18.67)*	61 (81.33)*
E or clade I	4 (0.05)	4 (100)	0 (0)
F	26 (3.27)	15 (57.69)	11 (42.31)
G	55 (6.93)	52 (94.55)*	3 (5.45)*
Unknown	1 (0.13)	0 (0)	1 (100)

^aThe phylogroup for each isolate was determined using ClermonTyping v1.4.0 with default settings for each isolate genome assembly. Phylogroups that are significantly associated with a specific host are indicated with an asterisk (*), as determined by maximum-likelihood ratio testing with a Benjamini-Hochberg adjusted *P* value of <0.05.

virulent and AMR isolates (39). Based on the Clermont quadruplex *in silico* method, our study isolates represented all seven main phylogroups (A, B1, B2, C, D, E, and F), cryptic clade I, and the newly identified phylogroup G. One isolate was classified as unknown. We found significant differences in the proportion of times phylogroups A, B1, B2, E, and G were seen with respect to their host origin. Phylogroups A, B1, and G were found at higher frequencies in chicken production isolates, while B2 and E were found at higher frequencies in human clinical isolates (Table 1). Eighty-one percent of chicken production isolates belonged to phylogroup B1, whereas 85% of human clinical isolates were classified as B2. The most divergent subclade within B2 consisted of closely related human clinical isolates, many of which originated from the same BioProject ([PRJNA489090](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA489090)) which included many isolates from the human infant gut. Although the B2 phylogroup was dominated by human clinical isolates, there were a few chicken isolates nested within the B2 clades. Phylogroup E was dominated by closely related human clinical isolates (81%) and was the largest cluster of human clinical isolates within the phylogenetic tree. Isolates identified as belonging to the *Escherichia* cryptic clade I phylogroup (NCBI genome accession numbers [SRR9984913](https://www.ncbi.nlm.nih.gov/nuccore/SRR9984913), [SRR10687720](https://www.ncbi.nlm.nih.gov/nuccore/SRR10687720), [SRR9852843](https://www.ncbi.nlm.nih.gov/nuccore/SRR9852843), [SRR9875260](https://www.ncbi.nlm.nih.gov/nuccore/SRR9875260), [SRR9875260](https://www.ncbi.nlm.nih.gov/nuccore/SRR9875260), [SRR10267809](https://www.ncbi.nlm.nih.gov/nuccore/SRR10267809), [SRR10687983](https://www.ncbi.nlm.nih.gov/nuccore/SRR10687983), and [SRR9852805](https://www.ncbi.nlm.nih.gov/nuccore/SRR9852805)) (40) clustered together within the phylogenetic tree and were all of chicken origin. No human clinical *E. coli* isolates were identified as belonging to clade I, even though clade I has been reported to have 2 to 3% prevalence in human *E. coli* isolates (41–43). Phylogroup G was dominated by chicken production isolates (95%).

The analysis presented above shows that in most cases the phylogroup classification supported the phylogenetic distribution within the SNP-based tree; however, we found several instances where human clinical and chicken production isolates were closely related. Likewise, we found cases where the SNP-based classification contradicted the phylogroup classification. For instance, human clinical strains mainly belong to phylogroup A (41–43); however, we identified 85.58% of our chicken production isolates belong to this phylogroup. Taken together, our isolate collection serves as a good representation and distribution of the currently described *E. coli* phylogroups, with a few exceptions.

Chicken production and human clinical isolates harbor a diverse set of AMR genes and mutations. We used the ResFinder (44) to identify acquired AMR genes in each *E. coli* genome. In addition, we used R2R and the Comprehensive Antimicrobial Resistance Database (CARD) via the Resistance Gene Identifier (RGI) (45) (see Materials and Methods for details) to identify all AMR genes, including those encoded on the bacterial chromosome and gene mutations that can result in resistant genotypes. We found 42 acquired AMR genes predicted to confer resistance to 21 drug classes (Table 2). Fifty-seven percent of the isolates carried at least 1 acquired AMR gene, while one human clinical isolate carried 21 acquired AMR genes. Chicken production isolates were more likely to harbor acquired AMR genes (65.7%) than human clinical isolates (46.3%) (adjusted *P* << 0). Chicken

TABLE 2 Acquired antimicrobial resistance genes identified within chicken production and human clinical isolate FASTA assemblies using ResFinder^a

Gene	Drug class	Gene	Drug class
<i>aac(3)-IIa</i>	Aminoglycoside antibiotic	<i>dfrA1</i>	Diaminopyrimidine antibiotic
<i>aac(3)-IIId</i>	Aminoglycoside antibiotic	<i>dfrA12</i>	Diaminopyrimidine antibiotic
<i>aac(3)-IV</i>	Aminoglycoside antibiotic	<i>dfrA14</i>	Diaminopyrimidine antibiotic
<i>aac(6')-Ib-cr</i>	Fluoroquinolone antibiotic, Aminoglycoside antibiotic	<i>dfrA17</i>	Diaminopyrimidine antibiotic
<i>aadA1</i>	Aminoglycoside antibiotic	<i>erm(B)</i>	Lincosamide antibiotic, macrolide antibiotic, streptogramin antibiotic
<i>aadA2</i>	Aminoglycoside antibiotic	<i>floR</i>	Phenicol antibiotic
<i>aadA5</i>	Aminoglycoside antibiotic	<i>formA</i>	Disinfecting agents
<i>aph(3'')-Ib</i>	Aminoglycoside antibiotic	<i>fosA7</i>	Fosfomycin
<i>aph(3')-Ia</i>	Aminoglycoside antibiotic	<i>mcr-9</i>	Peptide antibiotic
<i>aph(3')-IIa</i>	Aminoglycoside antibiotic	<i>mdf(A)</i>	Rhodamine, tetracycline antibiotic, benzalkonium chloride
<i>aph(3')-III</i>	Aminoglycoside antibiotic	<i>mph(A)</i>	Macrolide antibiotic
<i>aph(4)-Ia</i>	Aminoglycoside antibiotic	<i>qacE</i>	Disinfecting agents and intercalating dyes
<i>aph(6)-Ic</i>	Aminoglycoside antibiotic	<i>qnrB19</i>	Fluoroquinolone antibiotic
<i>aph(6)-Id</i>	Aminoglycoside antibiotic	<i>rmtB</i>	Aminoglycoside antibiotic
<i>bla_{CMY-2}</i>	Penam, carbapenem, cephalosporin, cephamycin	<i>sitABCD</i>	Disinfecting agents
<i>bla_{CTX-M-14}</i>	Cephalosporin	<i>sul1</i>	Sulfonamide antibiotic
<i>bla_{CTX-M-15}</i>	Penam, cephalosporin	<i>sul2</i>	Sulfonamide antibiotic
<i>bla_{OXA-1}</i>	Penam, carbapenem, cephalosporin	<i>sul3</i>	Sulfonamide antibiotic
<i>bla_{OXA-244}</i>	Penam, carbapenem, cephalosporin	<i>tet(A)</i>	Tetracycline antibiotic
<i>bla_{TEM-1B}</i>	Monobactam, penam, penem, cephalosporin	<i>tet(B)</i>	Tetracycline antibiotic
<i>catA1</i>	Phenicol antibiotic	<i>tet(C)</i>	Tetracycline antibiotic

^aResFinder output hits for each gene were filtered ($\geq 80\%$ identity to database reference query). The corresponding resistance-conferring drug class for each gene was identified using the Comprehensive Antibiotic Resistance Database (CARD).

production *E. coli* isolates carried more acquired AMR genes per isolate (range, 0 to 8; mean, 1.97 per isolate) compared to human clinical isolates (range, 0 to 21; mean, 1.87 per isolate). Fosfomycin (*fosA7*) and peptide antibiotics (*mcr-9*) were significantly higher in chicken production isolates, while aminoglycosides [*aac(6')-Ib-cr*] and fluoroquinolones (*qnrB19*) were significantly higher in human clinical isolates (Fig. 4A; see also Table S6).

Using the CARD database we found 208 AMR genes, including all but three genes [*aac(3')-IIa*, *sitABCD*, and *formA*] identified by ResFinder (see Table S1). A majority of AMR genes were shared between human clinical isolates and chicken production isolates (Fig. 5A). The AMR genes were predicted to confer resistance to 31 drug classes, including cephalosporins, fluoroquinolones, penams, and tetracyclines (Fig. 4B). We found AMR genes for 31 drug classes in human clinical isolates, while 28 of the 31 drug classes were found in chicken production isolates. AMR genes absent from chicken production isolates were predicted to confer resistance to streptogramin (*ermB* and *msrA*), oxazolidinone (*msrA*), and pleuromutilin (*msrA* and *taeA*). We found significant differences in antimicrobial drug classes between chicken production and human clinical isolates for 14 of the 31 drug classes (adjusted $P < 0.05$, Fig. 4B; see also Table S7). Genes predicted to confer resistance to diaminopyrimidine, phenicol, sulfonamide, penem, aminoglycoside, fluoroquinolone, macrolide, and peptide antibiotics were more prevalent within human clinical isolates (adjusted $P \ll 0$, Fig. 4B). Genes predicted to confer cephameycin resistance – a highly important antimicrobial group for human medicine (46) – were more prevalent in human clinical isolates (adjusted $P \ll 0$, Fig. 4B) compared to chicken isolates. Genes predicted to confer resistance to elfamycin and nucleoside antibiotics were more prevalent within the chicken production isolates compared to human isolates (adjusted $P \ll 0$, Fig. 4B). The frequency of genes predicted to confer resistance to tetracycline was similar (adjusted $P = 0.27$, Fig. 4B) between host origins.

It is important to mention that the majority of AMR genes identified by CARD are likely chromosomally encoded and may not be mobilizable. Many of the AMR genes harbored on the chromosome confer levels of resistance that are not medically relevant but may confer fitness advantages to these isolates in their native environments (47, 48). In contrast, most of the AMR genes identified by ResFinder are more likely to be mobilized within bacterial populations, via plasmids or other mobile genetic elements (MGEs), are therefore more likely to spread between/within *E. coli* strains originating from human and food animal populations. We hypothesized that a majority of

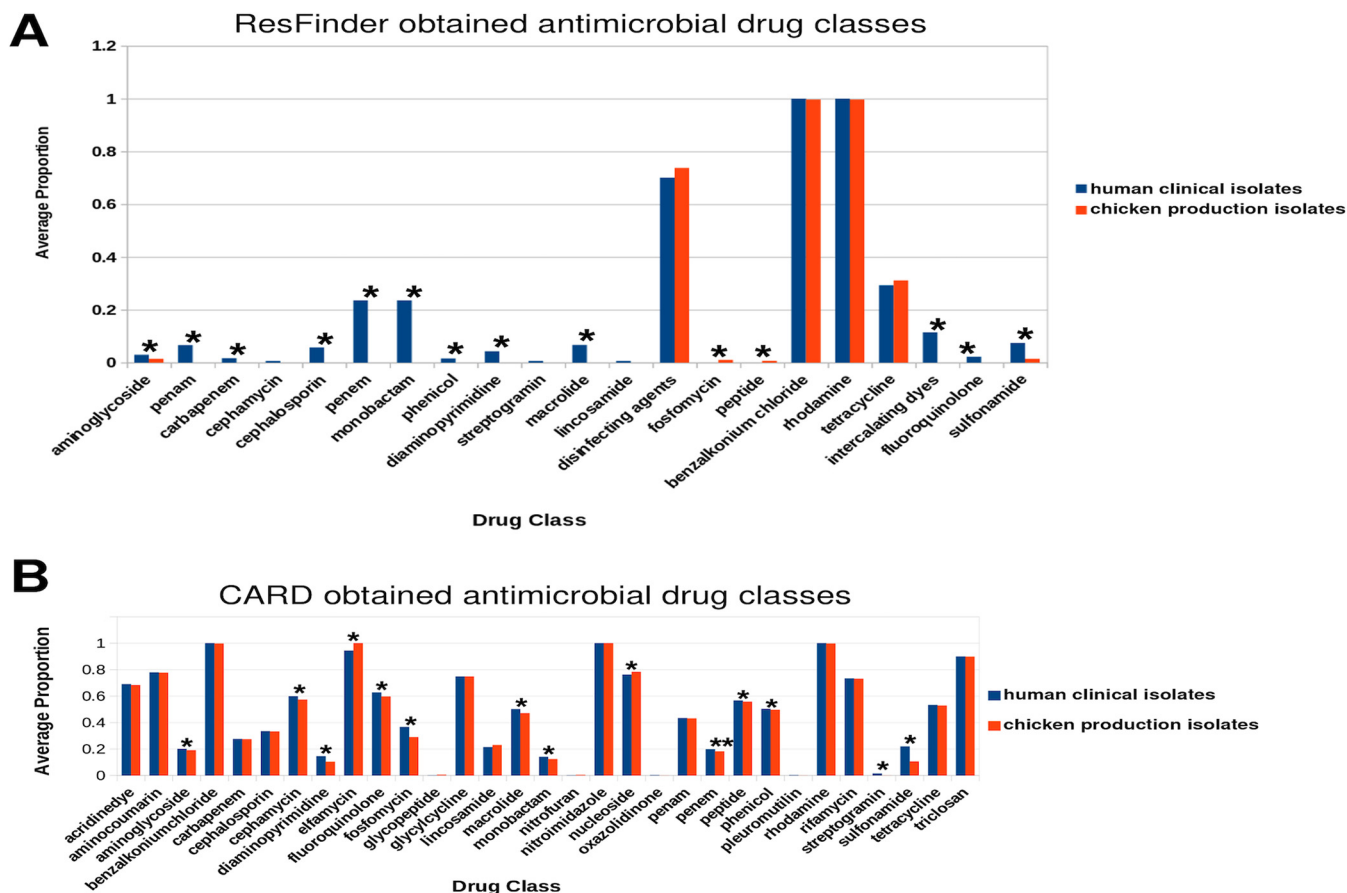


FIG 4 Average proportions of identified antimicrobial resistance genes in human clinical and poultry production isolates. (A) Acquired antimicrobial resistance genes identified within chicken production and human clinical isolate FASTA assemblies using ResFinder. ResFinder output hits for each gene were filtered ($\geq 80\%$ identity to database reference query). (B) Antimicrobial resistance drug classes identified within chicken production and human clinical isolate FASTA assemblies using Resistance Gene Identifier (RGI) and Abricate employed via Reads2Resistome. AMR database hits were filtered ($\geq 95\%$ identity to database reference query), and the highest hit from each database was retained. Genes conferring resistance to drug classes were enumerated for each isolate, and a proportion was calculated using the total number of genes in the study population conferring resistance to a given drug class. Drug classes significantly differing between chicken production isolates and human clinical isolates were determined by Wilcoxon rank sum test, and an asterisk (*) indicates a Benjamini-Hochberg adjusted P value of < 0.05 .

the AMR genes identified by ResFinder were located either on plasmid or chromosomal contigs and not located in close proximity to other MGEs such as genomic islands, insertion sequences, gene cassettes, etc. This hypothesis was proposed due to all but three ResFinder identified AMR genes being also identified by CARD which are majorly present on chromosomal contigs. To test this hypothesis, we submitted four isolates to VRProfile2 (49), an online bacterial mobile element detection pipeline capable of predicting AMR gene presence on plasmid or chromosomal contigs or those in close proximity to other MGEs. VRProfile2 results indicated AMR genes found by ResFinder were located on plasmid contigs, as was the case for *aph(3')-III*. No AMR genes were identified in close proximity to any other MGE types (see Table S8).

Overall, we observed a high prevalence of genes that can confer resistance to antimicrobials considered highly and critically important to human medicine in both chicken production and human clinical *E. coli* isolates. The prevalence of these genes in both human and chicken *E. coli* isolates indicates that these populations may serve as a possible reservoir of antimicrobial resistance.

The presence of type III secretion system genes classified the *E. coli* isolates into two separate clusters. To determine whether AMR genes and VFs (Fig. 5B) could separate the human clinical and chicken production isolates into distinct groups, we employed hierarchical clustering. Hierarchical clustering was based on the filtered presence/absence table of AMR genes and VFs (Fig. 6; see also Table S1) and resulted in two

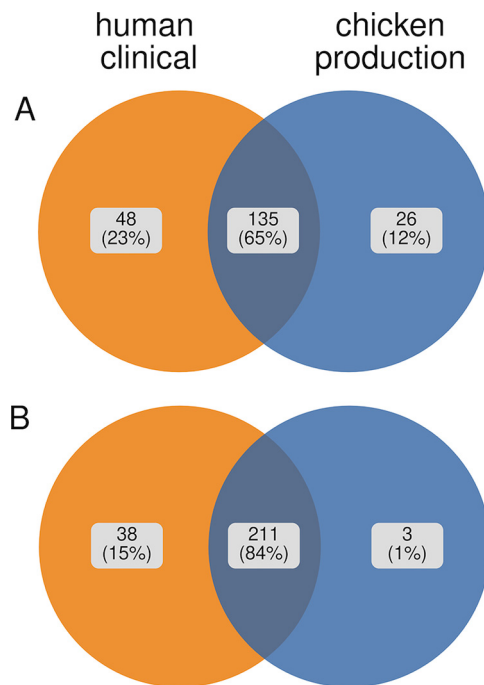


FIG 5 Venn diagrams depicting the percentage of antimicrobial resistance genes identified via Reads2Resistome (RGI) and The Comprehensive Antibiotic Resistance Database and ResFinder (A) and virulence factors identified via Reads2Resistome (Abricate) and the Virulence Factor Database (B).

distinct clusters (A and B) consisting of 661 and 130 isolates, respectively. These two clusters were significantly different from one another (adjusted $P = P \ll 0.05$) compared by isolate host origin. Cluster A comprised mostly of chicken production isolates (407 of 664 [61%]) while cluster B composed of human clinical isolates (85 of 130 [65%]).

To determine the genes driving the separation of these isolates into the two distinct clusters, we performed random forest classification on the filtered AMR gene and VF table (see Table S1). This reclassification resulted in 100% correct classification of all isolates to the two clusters (A and B) (see Table S9). Type 3 secretion system (T3SS) genes play a key role in the virulence of many Gram-negative bacterial pathogens (49) and were prevalent in all isolates from cluster B and completely absent from cluster A isolates. Twenty-five percent of all human clinical isolates ($n = 339$) carried genes for the T3SS operon (50), whereas only 9.29% of all chicken production isolates ($n = 352$) harbored the T3SS operon. Along with genes from the T3SS operon, the intimin gene (*eae*) that is required for intimate adherence and virulence in both humans and animals was a high contributor to the mean decreasing accuracy and was present in 100% of cluster B isolates but completely absent from cluster A isolates (Fig. 7A). Reducing the input AMR and virulence factors to only include the T3SS outer ring protein (*escD*) and the T3SS secretin (*escC*), two genes that encode oligomerizing proteins for the T3SS and that are essential for T3SS functionality (49), resulted in 100% correct classification of all isolates to either cluster A or B. Our results suggest that cluster B isolates that harbor T3SS may have the potential to deliver effector proteins that promote bacterial colonization, replication, and transmission in human host cell cytoplasm, while cluster A isolates may lack this capability. Nevertheless, we found both clusters contained isolates from human clinical and chicken production settings.

AMR genes and virulence factors classified *E. coli* isolates by their host origin.

To assess which AMR genes and VFs differentiate isolates based on their host origin we performed Random Forests classification (Fig. 7B). This classification resulted in 1.1% (5 isolates) error in classifying chicken isolates and a 5.6% (19 isolates) error in classifying human clinical isolates (see Table S10). All misclassified isolates belonged to the

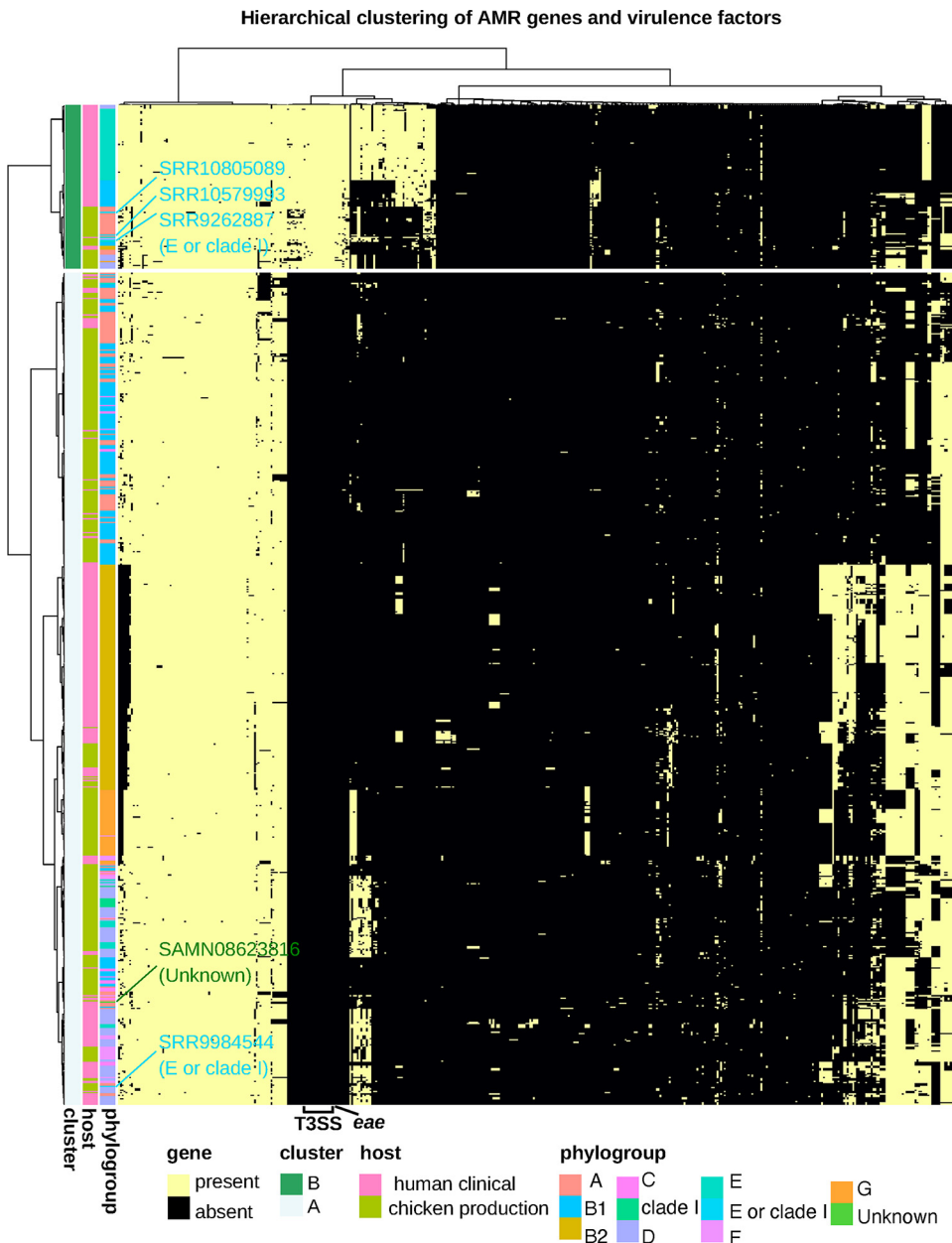


FIG 6 Heatmap of AMR genes and virulence factors across all isolates (see Table S1). Heatmap was generated in R v4.0.4 with pheatmap v1.0.12 (clustering_method = "average" [UPGMA], clustering_distance_cols = "binary," clustering_distance_rows = "Euclidean") using the filtered AMR gene and virulence factor table. Isolates from the phylogroups "unknown" and "E or clade I" are highlighted.

hierarchical cluster A (see Table S11) described earlier. The *E. coli* topoisomerase IV (*parC*) gene with the S80I mutation, was the highest contributor to the mean decreasing accuracy. The *parC* gene plays a critical role in DNA replication and confers reduced susceptibility to fluoroquinolones if the relevant *parC* and gyrase A (*gyrA*) mutations are present together (47, 51–54). A total of 185 isolates (1 chicken production and 184 human clinical) had *parC* mutations, and 215 isolates (25 chicken production and 190 human clinical) had *gyrA* mutations. Four human clinical isolates had *parC* mutations but no *gyrA* mutation, while 34 isolates (24 chicken production and 10 human clinical) had *gyrA* mutations but no *parC* mutation. One *E. coli* isolate, misclassified as a human isolate, harbored both *parC* (S80I) and *gyrA* (S83L, S84L, and D87Y) mutations (see Table S12). The remaining 572 isolates did not have a *parC* or a *gyrA* mutation. In

Variable Importance of Random Forests Classification

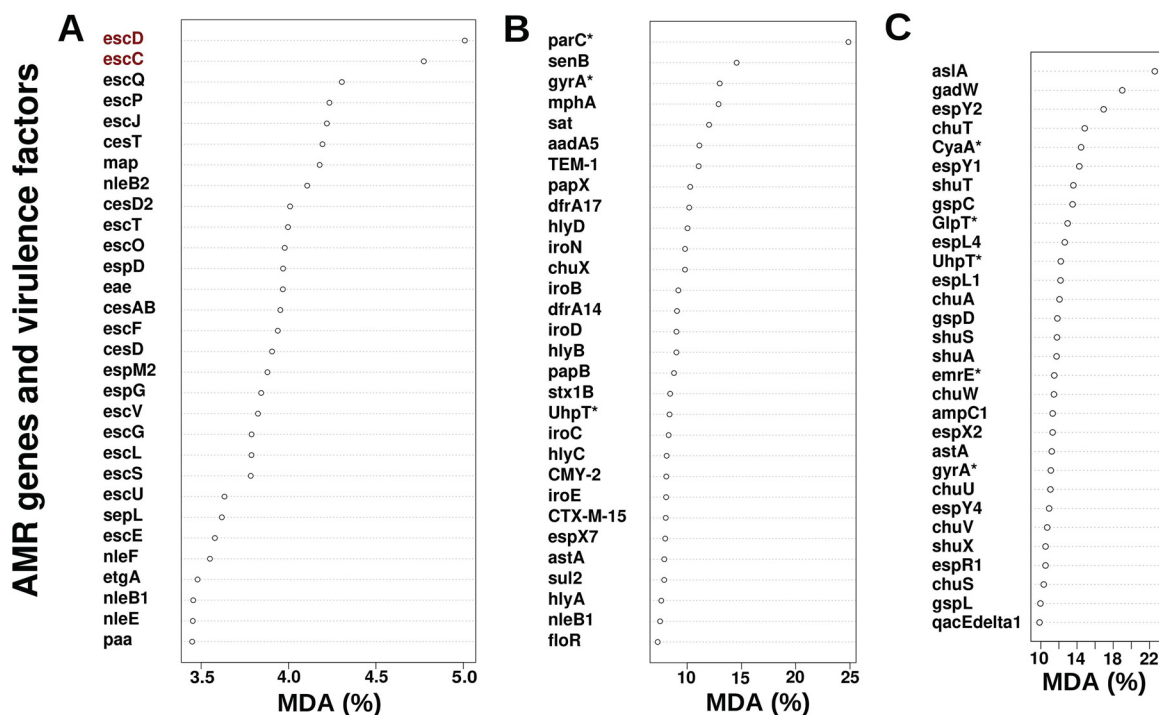


FIG 7 Mean decreasing accuracy (MDA) plot of variables in random forest classification models. MDA represents how much accuracy the model loses by exclusion of each variable (AMR gene or virulence factor). Random forest classification was performed on AMR gene and virulence factor table (see Materials and Methods for details). (A) Resulting MDA plot after classification of isolates into hierarchical clustering identified clusters A and B. Genes highlighted in red, on their own, result in 100% classification of isolates into hierarchical clustering identified clusters A and B. (B) Resulting MDA plot after classification of isolates into chicken production isolates and human clinical isolates. (C) Resulting MDA plot after classification of isolates into ClermonTyping identified phylogroups. Random forest classification in R was performed using the randomForest v4.6-14. Asterisks (*) indicate genes with mutations: *parC* (S80I), *gyrA*(S83L) (*n* = 30), *gyrA*(S84L) (*n* = 2), *gyrA*(D87Y) (*n* = 275), UtpT(E350Q), CyaA(S352T), and GlpT(E448K) (see Table S14).

summary, *parC* and *gyrA* mutations that confer resistance to fluoroquinolones were present in many of the human clinical isolates, and it was a major factor that separated the *E. coli* isolates into their respective host origins.

The iron acquisition genes (*iroN*, *iroD*, and *iroE*) were high contributors to the mean decreasing accuracy of random forest classification. The catechol siderophore uptake system (*iroBCDEN*) plays a critical role in virulence since iron is required for many cellular processes, but it is limited in host sites of extraintestinal infections (55). This gene cluster has been reported to be located in a pathogenicity island on the chromosome and has been found on plasmids (33, 56). However, this gene cluster was absent from all misclassified chicken isolates but was present in 56 and 7% of chicken production isolates and human clinical isolates, respectively (see Table S1). The enterotoxin TieB protein (*senB*) was the second most important factor and was absent from all chicken production isolates regardless of their random forest classification but present in 28% of human clinical isolates. Only one misclassified human clinical isolate harbored *senB*. These data suggest that the human and chicken isolates in our data set carry distinct AMR determinants and VFs which differentiate them from one another.

AMR genes and VF content can classify *E. coli* isolates by phylogroup. Next, we hypothesized that if the AMR genes and VF content is specific to each phylogroup, we should be able to classify the *E. coli* isolates correctly into their respective phylogroups based on their AMR and VF profiles. To evaluate this, we used random forests to classify the isolates into their identified phylogroups. Classification resulted in the following error rates: phylogroup A (11%), B1 (3%), B2 (0%), C (48%), clade I (0%), D (0%), E (5%), “E or clade I” (100%), F (4%), G (2%), and unknown (100%) (see Table S13). Unlike

host origin classification, reducing the input table for phylogroup classification resulted in an increase in the misclassification error. Interestingly, the top 30 factors, needed to classify isolates into their respective phylogroups, consisted of 8 AMR genes and 22 VFs (Fig. 7C), suggesting that VFs have a greater influence than AMR genes for phylogroup classification.

Of 11 VF functions found using the virulence factor database (see Table S14), 6 were present in all phylogroups, and the remaining functions were either entirely absent from individual phylogroups or present in some (see Table S15). Interestingly, virulence functions relating to adherence, secretion systems, and iron uptake were present in all phylogroups, but functions relating to immune evasion were present only in phylogroups A, B1, B2, D, and E. (see Table S15). Thus, both chicken production and human clinical isolates have the potential to serve as VF reservoirs. Albeit *E. coli* isolates of phylogroups C, "E or clade I," clade I, G and F, lacked many of the VF functions compared to A, B1, B2, D, and E.

We found that an isolate's AMR and VF profile was sufficient, in most cases, to determine its phylogroup. In addition, misclassified isolates were supported by their AMR and VF similarity as seen using the UPGMA (unweighted pair-group method with arithmetic averages) clustering approach (see Materials and Methods for details; Fig. 6). For instance, the *E. coli* isolate that was identified by ClermonTyping as unknown was classified into phylogroup A based on its AMR gene and VF profile similarity with other phylogroup A, B1, and C isolates (Fig. 6). The misclassification of "E or clade I" isolates (NCBI accession numbers [SRR10805089](#), [SRR10579993](#), and [SRR9262887](#)) as phylogroups D and E, respectively, were also supported by clustering in close proximity to other phylogroup D and E isolates (Fig. 6). An additional "E or clade I" isolate (NCBI accession number [SRR9984544](#)) was misclassified as D but contained a similar AMR and VF profile as other phylogroup D isolates. The misclassification was due to the absence of iron uptake genes *shuS*, *shuA*, and *shuX*, as well as the non-locus of enterocyte effacement (LEE)-encoded T3SS genes *espL4* and *espY1*. Although the AMR and VFs harbored by an *E. coli* isolate provided sufficient signal for correct classification of phylogroups, it could not correctly classify all phylogroups identified by ClermonTyping. However, the misclassification of "unknown" groups and E or clade I isolates was explained and supported by their clustering based on the similarity of their AMR gene and VF profiles (Fig. 6). Similarly, the location of the "unknown" isolate within the phylogenetic tree supports the random forests classification (Fig. 3), i.e., clustered with phylogroup A isolates.

DISCUSSION

In this study, we utilized *E. coli* isolates obtained from the NARMS surveillance program to study differences in phylotype, AMR, and VFs between chicken production and human clinical isolates. To do this, we first developed a massively parallel pipeline (Reads2Resistome) to streamline our genomic analysis and to ensure that our results are repeatable and reproducible. We found that *E. coli* can serve as a reservoir of AMR genes and VFs which supports previously reported findings (57–59). We determined that a majority of AMR genes and VFs identified were shared (65 and 84%, respectively) between human clinical isolates and chicken production isolates. Although there is a possibility of horizontal gene transfer in the environment, the *E. coli* isolates from this study were mostly host-specific, suggesting that strains from chicken production rarely colonize humans. We identified AMR genes that have been acquired, as well as those encoded on the chromosome that can acquire mutations that will result in reduced susceptibility to antibiotics. We confirmed the presence of 42 acquired AMR genes conferring resistance to 21 drug classes (Table 2, Fig. 4A, and Fig. 5A). The drug classes identified include those considered critically and highly important to human medicine. For example, genes conferring resistance to fosfomycin, considered a high priority and critically important antimicrobial by the WHO (46), were significantly more prevalent within the chicken production isolates. Utilizing the CARD database, a collection of all AMR genes, including those with mutations known to confer a level of resistance to known antimicrobials, we identified genes conferring resistance to 31 antibiotic drug

classes in our human clinical isolates, 28 of which were observed in the chicken production isolates (Fig. 4B). In addition, genes conferring resistance to these 31 drug classes were observed across a majority of identified phylogroups.

We determined that chicken production isolates carried a higher proportion of acquired AMR genes than human clinical isolates. Acquired AMR genes are those which have been obtained by a strain via mobile genetic elements such as plasmids, integrative conjugative elements, bacteriophages, insertion sequences, and gene capture elements such as a site-specific recombination system (59). Acquired AMR genes are important in the context of animal-to-human transmission since they are more readily shared between bacteria and provide a route for dissemination of AMR genes. The high proportion of benzalkonium chloride is not surprising due to its high usage, as a disinfectant, in the food and medical industries (60). The gene conferring resistance to benzalkonium chloride, *mdfA*, a multidrug efflux system, also confers resistance to tetracycline and rhodamine. When comparing the presence of AMR genes from CARD, which includes genes with mutations resulting in a resistance genotype to acquired resistance genes from ResFinder, we observed a drastic decline in the number of drug classes a given isolate may present resistance to. The prevalence of tetracycline in human clinical and chicken production isolates is expected due to its widespread usage in both agriculture and human clinical medicine and is of concern because of its high importance to human medicine.

Of particular interest is our finding of widespread potential resistance to fluoroquinolone antibiotics within the chicken production isolates since their usage was banned in poultry in 2005 (61), especially given that our data were obtained from samples between 2018 and 2020. Fluoroquinolone is a critically important antimicrobial drug class to human medicine due to its use in treatment of *Campylobacter* spp., invasive *Salmonella* spp., and multidrug-resistant *Shigella* spp. infections (46). Further investigation of the causes for fluoroquinolone resistance persistence in poultry is needed to ensure its efficacy in treating human infections. Taken together, we identified many antimicrobial drug classes which were highly present, but expected, in our study population. However, the presence of AMR genes conferring resistance to antibiotics banned for use in poultry, or food animal production in general, raises concern for their future effectiveness in treating human infections and diseases.

We found 11 virulence functions, conferred by 251 identified VFs (Fig. 5B; see also Table S14) to be present in both chicken and human clinical isolates. Based on hierarchical clustering of T3SS genes, the *E. coli* isolates were classified into two distinct clusters, A and B. The T3SS is associated with pathogens which can adhere to the epithelial surface of the host, many of which can cause diarrheal disease; the second leading cause of death in children globally (62, 63). Cluster A contained a higher representation of chicken production isolates, while cluster B comprised mainly of human clinical isolates. This result suggests that *E. coli* isolates from chicken production and human clinical isolates differ in their carriage of T3SS genes. The T3SS is essential for virulence and colonization of the human gut and is the genetic basis for enteropathogenic *E. coli* classification (63). In addition, the T3SS in poultry contributes to virulence in avian-pathogenic *E. coli* (64). This result suggests that the T3SS genes can be used as targets for a rapid on-farm and clinical diagnostic detection of virulent *E. coli* strains in chicken production and hospital settings. Identifying a lower proportion of chicken production isolates harboring the T3SS is a positive finding since the T3SS increases virulence in both humans and chickens.

Classification into host origin was successful with little error. Two of the most important factors for this separation were the *parC* gene (S80I) and *gyrA* (S83L [$n = 30$], S84L [$n = 2$], and D87Y [$n = 275$]), which are both linked to fluoroquinolone resistance. It has been documented that these resistance mutants are commonly found in environmental *E. coli* even in the absence of fluoroquinolone selective pressures (52). We speculate that these genes, and their respective mutations, which relate to DNA replication and transcription, may confer a fitness advantage. The secreted enterotoxin TieB, encoded by *senB*, was the second most important factor in separating human and chicken isolates. *senB* was only identified in human clinical isolates, and it was expected since *senB* is typically associated with

enteroinvasive and uropathogenic *E. coli* infection in humans (65). Similarly, *TEM-1*, a broad-spectrum beta-lactamase conferring resistance to penicillins and first-generation cephalosporins, was also a high contributor to the separation of hosts and was only present in human-linked isolates. While *TEM-1* has been identified in poultry isolates (66), cephalosporin usage is more common in beef production (67).

This study reemphasizes the utility of phylogroup classification as a convenient way to identify pathogenic strains. The most recent version of ClermonTyper (39) relies on five genes (*arpA*, *chuA*, *yjaA*, *TspE4.C2*, and *ybgD*) for phylogroup classification, only one of the quadruplex genes, *chuA* (outer membrane protein responsible for heme uptake), was present in our presence/absence table of identified AMR and VFs. Even though *yjaA* (gene that differentiates phylogroups B2 and D isolates) was absent from the AMR gene and VF table, we were able to reclassify B2 and D isolates with 100% accuracy. Similarly, *TspE4.C2* gene differentiates phylogroup A from phylogroup B1, and its absence from our AMR gene and VF table resulted in an 11% classification error for A isolates and a 3% error for B1 isolates. The absence of four of the five ClermonTyping genes might have contributed to the 48% error in classifying phylogroup C. The ability to separate phylogroups F and G from B2 was recently accomplished with the addition of the genes *cfbB* (CFA/I fimbrial subunit B), specific to phylogroup G strains, and *ybgD* (fimbrial-like adhesin protein), specific to phylogroup F strains (39, 68). The absence of these genes in our table only resulted in one phylogroup F (4% error rate) and one phylogroup G misclassified isolates (2% error rate). Our results show that the phylotype classification based on AMR and VFs could be achieved even when the majority of the genes used for Clermont *E. coli* phylotyping are missing.

The use of whole-genome sequencing (WGS) is an effective tool to predict the AMR and virulence potential of a bacterium; however, WGS still has its limitations. The depth of coverage is crucial for genome assembly and gene annotation, both of which can be significantly influenced by a coverage of less than 30 \times . Our study excluded many isolates from both chicken production and the human clinical settings due to low sequencing coverage. Due to the nature of the database chosen for isolate selection, i.e., a surveillance system for foodborne and other enteric bacteria, we expected an over representation of isolates harboring AMR genes and VFs, and an under-representation of susceptible commensal and environmental *E. coli* isolates in our data sets.

In addition, the choice of AMR and virulence databases is important as not all databases have the same entries and methods of detection. Here, we implemented Reads2Resistome which allowed us to incorporate the results from multiple databases. Our final AMR gene and VFs were comprised of the highest percent hit based on similarity to the reference gene (cutoff for gene coverage >80%) from each database. Another limitation of this study is the lack of AMR or virulence phenotype data for each isolate. Therefore, the presence an AMR gene or VFs in a genome is not a confirmation that the isolate is resistant to the antibiotic/drug class predicted or that the isolate can cause diseases.

MATERIALS AND METHODS

Study isolates, normalization, and assembly. A total of 1,711 *Escherichia coli* isolates, including 874 human clinical and 837 chicken production isolates, were selected from the USDA's GenomeTrackr network via NCBI's Pathogen Detection Network. Agricultural poultry isolates were selected based on the following criteria: species, "*E. coli* and *Shigella*"; location, "USA"; target creation, "01/01/2018 to 03/30/2020"; host, "Poultry," "poultry," "*Gallus gallus*," and "*Gallus gallus domesticus*." Human clinical isolates were selected based on the following criteria: species, "*E. coli* and *Shigella*"; location, "USA"; target creation, "01/01/2018 to 03/30/2020"; host, "*Homo sapiens*," "*Homosapiens*," "*homo sapiens*," "human," "human (infant)." Only isolates with a "strain" denoted as "*E. coli*" were retained for downstream analysis; isolates identified as "*Shigella*" were not retained. If available, Sequence Read Archive (SRA) reads, in FASTQ format, were downloaded based on the BioSample ID for both agricultural and human isolates. In the event the SRA reads were not available, FASTA assemblies were downloaded from the associated BioProject submission. To ensure accurate and unbiased antimicrobial resistance identification, SRA reads were further filtered based on coverage $\geq 20\times$. To guarantee inclusion of quality genome assemblies, FASTA assemblies were filtered based on having an N_{50} greater than 10,000 bp. Three isolates in total were identified to be identical and were removed from downstream statistical analysis.

To compare genetic structure and contents, specifically AMR genes, virulence factors, and plasmid replicons across our study isolates, we assembled the SRA FASTQ reads into genomes, annotated these

genomes, and annotated the downloaded FASTA assemblies. To reduce bias in gene identification and assembly quality across isolates, SRA reads were normalized to 30× coverage using the *insilico_read_normalization.pl* script from Trinity RNA-Seq v2.11.0 (Trinity's GitHub). SRA reads were assembled using Reads2Resistome v0.0.2 (options: slidingwindow = 4:20, assembly = nonhybrid, threads 60). Resulting SRA assemblies were further filtered based on the N_{50} (>10,000 bp) and final coverage ($\geq 20\times$) like the FASTA assemblies described earlier. The final study data set consisted of R2R assembled isolate genomes and downloaded FASTA assemblies totaling 791 *E. coli* isolates (452 chicken production and 339 human clinical; Fig. 2A).

Phylogenetic and sequence type analysis. To evaluate the relatedness of the isolates with respect to their assembled genomes, we constructed a phylogenetic tree utilizing SNPs. These SNPs were identified by comparing the *E. coli* K12-MG1655 (accession U00096) reference genome to each of the 791 *E. coli* isolates using Snippy v4.6.0 (69). The resulting whole-genome SNP alignment was then used to construct a maximum-likelihood phylogenetic tree using the GTR+GAMMA model of evolution and utilizing RAxML (RAxMLHPC-PTHREADS) v8.2.12 (70, 71). In addition, we performed 100 bootstraps to assess nodal support. Only 100 bootstraps were performed due to time and computational limitations; it took 75.49 days utilizing 64 compute cores and 504 Gb memory to construct the phylogenetic tree and perform the 100 bootstraps. To determine which of the seven main phylogroups the isolates belong to, we used ClermonTyping v1.4.0 (39), utilizing the default settings, to determine the phylogroup designation (A, B1, B2, C, D, E, and F) for each isolate. Tree visualization, along with phylogroup and metadata overlay, was done in R v4.0.4 using ggtree v2.4.1 (72), factoextra v1.0.7 (73), and ggnewscale v0.4.5 (74). The sequence type of each isolate was determined using ECTyper v1.0.0 (75) with default settings.

Antimicrobial resistance gene and virulence-factor characterization using Reads2Resistome. Reads2Resistome (R2R; see the supplemental Materials and Methods) was used to characterize the antimicrobial resistance genes, virulence factors, and prophage regions. Reads2Resistome utilizes ABRICATE v0.5 (76) to screen the assembled contigs against the following databases via BLAST and reports a percentage identity to the reference sequence. Databases used by ABRICATE, and provided by R2R, were as follows: the ARG-ANNOT antibiotic resistance gene database (77), the Comprehensive Antibiotic Resistance Database (CARD) (45), the MEGARes Antimicrobial Database for High-Throughput Sequencing (78), NCBI AMRFinderPlus (79), PlasmidFinder (80), and the VirulenceFinder database (81). Downloaded FASTA assemblies' resistome characterization was performed by the modified Reads2Resistome (R2R-0.0.1-Fasta-QC-Ann-Only.nf). In addition, resistance gene identification was performed using the Resistance Gene Identifier v5.1.1 (45) (RGI), and acquired resistance genes were identified using ResFinder (82). AMR genes and virulence factors with the highest percent identity across all databases were selected for further analysis. The heatmap of AMR genes and virulence factors, using the table (described below; see also Table S1) was generated using pheatmap v1.0.12 in R v4.0.4 with the "average" (unweighted pair group method with arithmetic mean [UPGMA]) clustering method (see Fig. S1) (83). (See the supplemental Materials and Methods for further description of R2R). Mobile genetic elements were determined for selected isolates using VRProfile2 (updated most recently on 2 October 2021) (49).

Hierarchical clustering and statistical analysis of resistome and virulence factors. To evaluate the relationships between poultry and human isolates with respect to their AMR and virulence factors, we employed hierarchical clustering and ordination analysis. AMR and virulence database hits from R2R and RGI were filtered utilizing $\geq 95\%$ and $\geq 80\%$ sequence identities to the reference database query, respectively. ResFinder results were filtered utilizing $\geq 85\%$ identity to the reference database query. A table was then generated based on the presence/absence of identified VFs and AMR genes from all isolates (see Table S1). A distance matrix was generated using the tanimoto metric via the Distance() (84) function from the IntClust v0.1.0 (85) package. hclust() from the stats v3.6.2 package (86) was then utilized to perform the hierarchical clustering under the "average" (UPGMA) method using the determined optimal number of clusters. We identified the optimal number of clusters which separate the 791 isolates using the silhouette method implemented by the fviz_nbclust() function from the factoextra v1.0.7 package (73). Drug classes associated with each identified AMR gene were determined using the output from RGI, in combination with the CARD database (45), and virulence factor-associated functions were determined using the comparative tables from the Virulence Factor Database (81). Genes conferring resistance to drug classes were enumerated for each isolate, and a proportion was calculated using the total number of genes in the study population conferring resistance to a given drug class. Drug classes significantly differing between hosts were determined by using a Wilcoxon rank sum test with the wilcox.test() function from the stats v3.6.2 package (86). *P* value adjustment for multiple comparisons was conducted using the p.adjust() function, according to the Benjamini and Hochberg ("BH") method, from the stats v3.6.2 package (86). PCA was conducted using the prcomp() function from the stats v3.6.2 package (86), and plots were generated using ggplot2 (87). All analyses were done in R v4.0.4 utilizing RStudio v1.2.1106 (88).

Random forest classification. To determine the most influential AMR and virulence factors driving the separation of the isolates into (i) clusters identified via hierarchical clustering, (ii) isolate host origin, and (iii) identified phylogroups, we employed random forest classification (89). The described presence/absence table of AMR genes and virulence factors was used as input (see Table S1). TuneRF() (90) was used to determine the optimal mtry parameter value, with respect to out-of-bag error estimate, for RandomForest(). RandomForest() v4.6-14 (90) was used for classification with ntree = 200 based on the filtered AMR and virulence-factor presence/absence table. All plots were generated using ggplot2 v3.3.3 (87). All analyses were done in R v4.0.4 utilizing RStudio v1.2.1106 (88).

Data availability. All raw isolate data, both FASTQ and FASTA files, were obtained from and are available through the NCBI using the NCBI accession numbers in Table S1 in the supplemental material. All other data are available upon request.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.9 MB.

SUPPLEMENTAL FILE 2, XLSX file, 0.9 MB.

SUPPLEMENTAL FILE 3, XLSX file, 0.01 MB.

ACKNOWLEDGMENTS

We are grateful to Jodie Plumblee Lawrence and Denice Cudnik for logistical and technical assistance.

This study was supported by the U.S. Department of Agriculture (USDA) Agricultural Research Service (project 6040-32000-012-000D). R.W. was supported by the National Institutes of Health (grants 5T32GM132057-04 and 5T32GM132057-03). This research was partially supported in part by an appointment to the Agricultural Research Service (ARS) Research Participation Program administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and the USDA (CRIS project 60-6040-6-009). ORISE is managed by ORAU under DOE contract DE-SC0014664. All opinions expressed here are the authors' and do not necessarily reflect the policies and views of the USDA, the DOE, or ORAU/ORISE. Any mention of products or trade names does not constitute recommendation for use. We declare no competing commercial interests in relation to the submitted work. The USDA is an equal opportunity provider and employer.

REFERENCES

- Boyle DP, Zembower TR. 2015. Epidemiology and management of emerging drug-resistant Gram-negative bacteria: extended-spectrum β -lactamases and beyond. *Urol Clin North Am* 42:493–505. <https://doi.org/10.1016/j.ucl.2015.05.005>.
- Lynch M, Painter J, Woodruff R, Braden C, Centers for Disease Control and Prevention. 2006. Surveillance for foodborne-disease outbreaks—United States, 1998–2002. *MMWR Surveill Summ* 55:1–42.
- Dewey-Mattia D, Manikonda K, Hall AJ, Wise ME, Crowe SJ. 2018. Surveillance for foodborne disease outbreaks—United States, 2009–2015. *MMWR Surveill Summ* 67:1–11. <https://doi.org/10.15585/mmwr.ss6710a1>.
- Centers for Disease Control and Prevention. 2019. Surveillance for foodborne disease outbreaks United States, 2014: annual report, p 1–16. Centers for Disease Control and Prevention, Atlanta, GA. <http://www.cdc.gov/foodsafety/fdoss/>. Accessed 3 May 2022.
- Russo TA, Johnson JR. 2003. Medical and economic impact of extraintestinal infections due to *Escherichia coli*: focus on an increasingly important endemic problem. *Microbes Infect* 5:449–456. [https://doi.org/10.1016/S1286-4579\(03\)00049-2](https://doi.org/10.1016/S1286-4579(03)00049-2).
- Nyachuba DG. 2010. Foodborne illness: is it on the rise? *Nutr Rev* 68:257–269. <https://doi.org/10.1111/j.1753-4887.2010.00286.x>.
- Ramos S, Silva V, Dapkevicius MdLE, Caniça M, Tejedor-Junco MT, Igrejas G, Poeta P. 2020. *Escherichia coli* as commensal and pathogenic bacteria among food-producing animals: health implications of extended spectrum β -lactamase (ESBL) production. *Animals (Basel)* 10:2239. <https://doi.org/10.3390/ani10122239>.
- Croxen MA, Law RJ, Scholz R, Keeney KM, Wlodarska M, Finlay BB. 2013. Recent advances in understanding enteric pathogenic *Escherichia coli*. *Clin Microbiol Rev* 26:822–880. <https://doi.org/10.1128/CMR.00022-13>.
- Kaper JB, Nataro JP, Mobley HLT. 2004. Pathogenic *Escherichia coli*. *Nat Rev Microbiol* 2:123–140. <https://doi.org/10.1038/nrmicro818>.
- Mainil J. 2013. *Escherichia coli* virulence factors. *Vet Immunol Immunopathol* 152:2–12. <https://doi.org/10.1016/j.vetimm.2012.09.032>.
- Khairy RM, Mohamed ES, Ghany HMA, Abdelrahim SS. 2019. Phylogenetic classification and virulence genes profiles of uropathogenic *Escherichia coli* and diarrheagenic *E. coli* strains isolated from community acquired infections. *PLoS One* 14:e0222441. <https://doi.org/10.1371/journal.pone.0222441>.
- Yang SC, Lin CH, Aljuffali IA, Fang JY. 2017. Current pathogenic *Escherichia coli* foodborne outbreak cases and therapy development. *Arch Microbiol* 199:811–825. <https://doi.org/10.1007/s00203-017-1393-y>.
- Panth Y. 2019. Colibacillosis in poultry: a review. *J Agric Nat Res* 2:301–311. <https://doi.org/10.3126/janr.v2i1.26094>.
- Ewers C, Antão EM, Diehl I, Philipp HC, Wieler LH. 2009. Intestine and environment of the chicken as reservoirs for extraintestinal pathogenic *Escherichia coli* strains with zoonotic potential. *Appl Environ Microbiol* 75:184–192. <https://doi.org/10.1128/AEM.01324-08>.
- Dho-Moulin M, Morris Fairbrother J. 1999. Avian pathogenic *Escherichia coli* (APEC). *Vet Res* 30(2-3):299–366.
- Guabiraba R, Schouler C. 2015. Avian colibacillosis: still many black holes. *FEMS Microbiol Lett* 362:fnv118. <https://doi.org/10.1093/femsle/fnv118>.
- Mellata M. 2013. Human and avian extraintestinal pathogenic *Escherichia coli*: infections, zoonotic risks, and antibiotic resistance trends. *Foodborne Pathog Dis* 10:916–932. <https://doi.org/10.1089/fpd.2013.1533>.
- Johnson TJ, Siek KE, Johnson SJ, Nolan LK. 2006. DNA sequence of a ColV plasmid and prevalence of selected plasmid-encoded virulence genes among avian *Escherichia coli* strains. *J Bacteriol* 188:745–758. <https://doi.org/10.1128/JB.188.2.745-758.2006>.
- Mitchell NM, Johnson JR, Johnston B, Curtiss R, Mellata M. 2015. Zoonotic potential of *Escherichia coli* isolates from retail chicken meat products and eggs. *Appl Environ Microbiol* 81:1177–1187. <https://doi.org/10.1128/AEM.03524-14>.
- Johnson TJ, Wannemuehler Y, Johnson SJ, Stell AL, Doetkott C, Johnson JR, Kim KS, Spanjaard L, Nolan LK. 2008. Comparison of extraintestinal pathogenic *Escherichia coli* strains from human and avian sources reveals a mixed subset representing potential zoonotic pathogens. *Appl Environ Microbiol* 74:7043–7050. <https://doi.org/10.1128/AEM.01395-08>.
- Johnson TJ, Kariyawasam S, Wannemuehler Y, Mangiamela P, Johnson SJ, Doetkott C, Skyberg JA, Lynne AM, Johnson JR, Nolan LK. 2007. The genome sequence of avian pathogenic *Escherichia coli* strain O1:K1:H7 shares strong similarities with human extraintestinal pathogenic *E. coli* genomes. *J Bacteriol* 189:3228–3236. <https://doi.org/10.1128/JB.01726-06>.
- Moulin-Schouleur M, Schouler C, Tailliez P, Kao M-R, Brée A, Germon P, Oswald E, Mainil J, Blanco M, Blanco J. 2006. Common virulence factors and genetic relationships between O18:K1:H7 *Escherichia coli* isolates of

- human and avian origin. *J Clin Microbiol* 44:3484–3492. <https://doi.org/10.1128/JCM.00548-06>.
23. Ewers C, Li G, Wilking H, Kiessling S, Alt K, Antão E-M, Laturnus C, Diehl I, Glodde S, Homeier T, Böhnke U, Steinrück H, Philipp H-C, Wieler LH. 2007. Avian pathogenic, uropathogenic, and newborn meningitis-causing *Escherichia coli*: how closely related are they? *Int J Med Microbiol* 297:163–176. <https://doi.org/10.1016/j.ijmm.2007.01.003>.
 24. Jakobsen L, Garneau P, Kurbasic A, Bruant G, Stegger M, Harel J, Jensen KS, Brousseau R, Hammerum AM, Frimodt-Møller N. 2011. Microarray-based detection of extended virulence and antimicrobial resistance gene profiles in phylogroup B2 *Escherichia coli* of human, meat and animal origin. *J Med Microbiol* 60:1502–1511. <https://doi.org/10.1099/jmm.0.033993-0>.
 25. Tivendale KA, Logue CM, Kariyawasam S, Jordan D, Hussein A, Li G, Wannemuehler Y, Nolan LK. 2010. Avian-pathogenic *Escherichia coli* strains are similar to neonatal meningitis *E. coli* strains and are able to cause meningitis in the rat model of human disease. *Infect Immun* 78:3412–3419. <https://doi.org/10.1128/IAI.00347-10>.
 26. Comery R, Thanabalasuriar A, Garneau P, Portt A, Boerlin P, Reid-Smith RJ, Harel J, Manges AR, Gruenheid S. 2013. Identification of potentially diarrheagenic atypical enteropathogenic *Escherichia coli* strains present in Canadian food animals at slaughter and in retail meats. *Appl Environ Microbiol* 79:3892–3896. <https://doi.org/10.1128/AEM.00182-13>.
 27. Jakobsen L, Garneau P, Bruant G, Harel J, Olsen SS, Porsbo LJ, Hammerum AM, Frimodt-Møller N. 2012. Is *Escherichia coli* urinary tract infection a zoonosis? Proof of direct link with production animals and meat. *Eur J Clin Microbiol Infect Dis* 31:1121–1129. <https://doi.org/10.1007/s10096-011-1417-5>.
 28. Li G, Cai W, Hussein A, Wannemuehler YM, Logue CM, Nolan LK. 2012. Proteome response of an extraintestinal pathogenic *Escherichia coli* strain with zoonotic potential to human and chicken sera. *J Proteomics* 75:4853–4862. <https://doi.org/10.1016/j.jprot.2012.05.044>.
 29. Skyberg JA, Johnson TJ, Johnson JR, Clabots C, Logue CM, Nolan LK. 2006. Acquisition of avian pathogenic *Escherichia coli* plasmids by a commensal *E. coli* isolate enhances its abilities to kill chicken embryos, grow in human urine, and colonize the murine kidney. *Infect Immun* 74:6287–6292. <https://doi.org/10.1128/IAI.00363-06>.
 30. Caniça M, Manageiro V, Abriouel H, Moran-Gilad J, Franz CMAP. 2019. Antibiotic resistance in foodborne bacteria. *Trends Food Sci Technol* 84:41–44. <https://doi.org/10.1016/j.tifs.2018.08.001>.
 31. Ferri M, Ranucci E, Romagnoli P, Giaccone V. 2017. Antimicrobial resistance: a global emerging threat to public health systems. *Crit Rev Food Sci Nutr* 57:2857–2876. <https://doi.org/10.1080/10408398.2015.1077192>.
 32. Fenske GJ, Scaria J. 2021. Analysis of 56,348 genomes identifies the relationship between antibiotic and metal resistance and the spread of multi-drug-resistant non-typhoidal *Salmonella*. *Microorganisms* 9:1468. <https://doi.org/10.3390/microorganisms9071468>.
 33. Oladeinde A, Abdo Z, Zwirzitz B, Woyda R, Lakin SM, Press MO, Cox NA, Thomas JC, Looft T, Rothrock MJ, Zock G, Plumlee Lawrence J, Cudnik D, Ritz C, Aggrey SE, Liachko I, Grove JR, Wiersma C. 2022. Litter commensal bacteria can limit the horizontal gene transfer of antimicrobial resistance to *Salmonella* in chickens. *Appl Environ Microbiol* 88:e02517-22. <https://doi.org/10.1128/aem.02517-21>.
 34. Biswas K, Sharma P, Joshi SR. 2019. Co-occurrence of antimicrobial resistance and virulence determinants in enterococci isolated from traditionally fermented fish products. *J Glob Antimicrob Resist* 17:79–83. <https://doi.org/10.1016/j.jgar.2018.11.012>.
 35. Pan Y, Zeng J, Li L, Yang J, Tang Z, Xiong W, Li Y, Chen S, Zeng Z. 2020. Coexistence of antibiotic resistance genes and virulence factors deciphered by large-scale complete genome analysis. *mSystems* 5:e00821-19. <https://doi.org/10.1128/mSystems.00821-19>.
 36. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
 37. Woyda R, Oladeinde A, Abdo Z. 2020. Reads2Resistome: an adaptable and high-throughput whole-genome sequencing pipeline for bacterial resistome characterization. *bioRxiv*. <https://www.biorxiv.org/content/10.1101/2020.05.18.102715v1>.
 38. Kurtzer GM, Sochat V, Bauer MW. 2017. Singularity: scientific containers for mobility of compute. *PLoS One* 12:e0177459. <https://doi.org/10.1371/journal.pone.0177459>.
 39. Clermont O, Dixit OVA, Vangchhia B, Condamine B, Dion S, Bridier-Nahmias A, Denamur E, Gordon D. 2019. Characterization and rapid identification of phylogroup G in *Escherichia coli*, a lineage with high virulence and antibiotic resistance potential. *Environ Microbiol* 21:3107–3117. <https://doi.org/10.1111/1462-2920.14713>.
 40. Clermont O, Christenson JK, Denamur E, Gordon DM. 2013. The Clermont *Escherichia coli* phylotyping method revisited: improvement of specificity and detection of new phylogroups. *Environ Microbiol Rep* 5:58–65. <https://doi.org/10.1111/1758-2229.12019>.
 41. Ingle DJ, Clermont O, Skurnik D, Denamur E, Walk ST, Gordon DM. 2011. Biofilm formation by and thermal niche and virulence characteristics of *Escherichia* spp. *Appl Environ Microbiol* 77:2695–2700. <https://doi.org/10.1128/AEM.02401-10>.
 42. Smati M, Clermont O, Le Gal F, Schichmanoff O, Jauréguy F, Eddi A, Denamur E, Picard B, Colville Group. 2013. Real-time PCR for quantitative analysis of human commensal *Escherichia coli* populations reveals a high frequency of subdominant phylogroups. *Appl Environ Microbiol* 79:5005–5012. <https://doi.org/10.1128/AEM.01423-13>.
 43. Ahumada-Santos YP, Báez-Flores ME, Díaz-Camacho SP, Uribe-Beltrán MdJ, Eslava-Campos CA, Parra-Unda JR, Delgado-Vargas F. 2020. Association of phylogenetic distribution and presence of integrons with multi-drug resistance in *Escherichia coli* clinical isolates from children with diarrhoea. *J Infect Public Health* 13:767–772. <https://doi.org/10.1016/j.jiph.2019.11.019>.
 44. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 67:2640–2644. <https://doi.org/10.1093/jac/dks261>.
 45. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen A-LV, Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran H-K, Werfalli RE, Nasir JA, Oloni M, Speicher DJ, Florescu A, Singh B, Faltyn M, Hernandez-Koutoucheva A, Sharma AN, Bordeleau E, Pawlowski AC, Zubyk HL, Dooley D, Griffiths E, Maguire F, Winsor GL, Beiko RG, Brinkman FSL, Hsiao WWL, Domselaar GV, McArthur AG. 2020. CARD 2020: antibiotic resistance surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res* 48:D517–D525. <https://doi.org/10.1093/nar/gkz935>.
 46. WHO. 2019. WHO list of critically important antimicrobials (CIA). World Health Organization, Geneva, Switzerland.
 47. Bagel S, Hüllen V, Wiedemann B, Heisig P. 1999. Impact of *gyrA* and *parC* mutations on quinolone resistance, doubling time, and supercoiling degree of *Escherichia coli*. *Antimicrob Agents Chemother* 43:868–875. <https://doi.org/10.1128/AAC.43.4.868>.
 48. Girgis HS, Hottes AK, Tavazoie S. 2009. Genetic architecture of intrinsic antibiotic susceptibility. *PLoS One* 4:e5629. <https://doi.org/10.1371/journal.pone.0005629>.
 49. Wang M, Goh Y-X, Tai C, Wang H, Deng Z, Ou H-Y. 2022. VRprofile2: detection of antibiotic resistance-associated mobilome in bacterial pathogens. *Nucleic Acids Res* 50:W768–W773. <https://doi.org/10.1093/nar/gkac321>.
 50. Tseytin I, Dagan A, Oren S, Sal-Man N. 2018. The role of EscD in supporting EscC polymerization in the type III secretion system of enteropathogenic *Escherichia coli*. *Biochim Biophys Acta Biomembr* 1860:384–395. <https://doi.org/10.1016/j.bbmem.2017.10.001>.
 51. Slater SL, Sägfors AM, Pollard DJ, Ruano-Gallego D, Frankel G. 2018. The type III secretion system of pathogenic *Escherichia coli*. *Curr Top Microbiol Immunol* 416:51–72. https://doi.org/10.1007/82_2018_116.
 52. Tavio MM, Vila J, Ruiz J, Ruiz J, Martín-Sánchez AM, Jiménez de Anta MT. 1999. Mechanisms involved in the development of resistance to fluoroquinolones in *Escherichia coli* isolates. *J Antimicrob Chemother* 44:735–742. <https://doi.org/10.1093/jac/44.6.735>.
 53. Johnning A, Kristiansson E, Fick J, Weijdegård B, Larsson DGJ. 2015. Resistance mutations in *gyrA* and *parC* are common in *Escherichia* communities of both fluoroquinolone-polluted and uncontaminated aquatic environments. *Front Microbiol* 6:1355. <https://doi.org/10.3389/fmicb.2015.01355>.
 54. Sáenz Y, Zarazaga M, Briñas L, Ruiz-Larrea F, Torres C. 2003. Mutations in *gyrA* and *parC* genes in nalidixic acid-resistant *Escherichia coli* strains from food products, humans and animals. *J Antimicrob Chemother* 51:1001–1005. <https://doi.org/10.1093/jac/dkg168>.
 55. Gao Q, Wang X, Xu H, Xu Y, Ling J, Zhang D, Gao S, Liu X. 2012. Roles of iron acquisition systems in virulence of extraintestinal pathogenic *Escherichia coli*: salmochelin and aerobactin contribute more to virulence than heme in a chicken infection model. *BMC Microbiol* 12:143. <https://doi.org/10.1186/1471-2180-12-143>.
 56. Sorsa LJ, Dufke S, Heesemann J, Schubert S. 2003. Characterization of an iroBCDEN gene cluster on a transmissible plasmid of uropathogenic *Escherichia coli*: evidence for horizontal transfer of a chromosomal virulence factor. *Infect Immun* 71:3285–3293. <https://doi.org/10.1128/IAI.71.6.3285-3293.2003>.

57. Katouli M. 2010. Population structure of gut *Escherichia coli* and its role in development of extra-intestinal infections. *Iran J Microbiol* 2:59–72.
58. Cortés P, Blanc V, Mora A, Dahbi G, Blanco JE, Blanco M, López C, Andreu A, Navarro F, Alonso MP, Bou G, Blanco J, Llagostera M. 2010. Isolation and characterization of potentially pathogenic antimicrobial-resistant *Escherichia coli* strains from chicken and pig farms in Spain. *Appl Environ Microbiol* 76: 2799–2805. <https://doi.org/10.1128/AEM.02421-09>.
59. Van Hoek AHAM, Mevius D, Guerra B, Mullany P, Roberts AP, Aarts HJM. 2011. Acquired antibiotic resistance genes: an overview. *Front Microbiol* 2:203. <https://doi.org/10.3389/fmicb.2011.00203>.
60. García MR, Cabo ML. 2018. Optimization of *Escherichia coli* inactivation by benzalkonium chloride reveals the importance of quantifying the inoculum effect on chemical disinfection. *Front Microbiol* 9:1259. <https://doi.org/10.3389/fmicb.2018.01259>.
61. Price LB, Lackey LG, Vailes R, Silbergeld E. 2007. The persistence of fluoroquinolone-resistant campylobacter in poultry production. *Environ Health Perspect* 115:1035–1039. <https://doi.org/10.1289/ehp.10050>.
62. Black RE, Cousens S, Johnson HL, Lawn JE, Rudan I, Bassani DG, Jha P, Campbell H, Walker CF, Cibulskis R, Eisele T, Liu L, Mathers C, Child Health Epidemiology Reference Group of WHO and UNICEF. 2010. Global, regional, and national causes of child mortality in 2008: a systematic analysis. *Lancet* 375:1969–1987. [https://doi.org/10.1016/S0140-6736\(10\)60549-1](https://doi.org/10.1016/S0140-6736(10)60549-1).
63. Deborah Chen H, Frankel G. 2005. Enteropathogenic *Escherichia coli*: unraveling pathogenesis. *FEMS Microbiol Rev* 29:83–98. <https://doi.org/10.1016/j.femsre.2004.07.002>.
64. Wang S, Liu X, Xu X, Yang D, Wang D, Han X, Shi Y, Tian M, Ding C, Peng D, Yu S. 2016. *Escherichia coli* type III secretion system 2 ATPase EivC is involved in the motility and virulence of avian pathogenic *Escherichia coli*. *Front Microbiol* 7:1387. <https://doi.org/10.3389/fmicb.2016.01387>.
65. Mao B-H, Chang Y-F, Scaria J, Chang C-C, Chou L-W, Tien N, Wu J-J, Tseng C-C, Wang M-C, Chang C-C, Hsu Y-M, Teng C-H. 2012. Identification of *Escherichia coli* genes associated with urinary tract infections. *J Clin Microbiol* 50:449–456. <https://doi.org/10.1128/JCM.00640-11>.
66. Seo KW, Lee YJ. 2018. Prevalence and characterization of β -lactamase genes and class 1 integrons in multidrug-resistant *Escherichia coli* isolates from chicken meat in Korea. *Microb Drug Resist* 24:1599–1606. <https://doi.org/10.1089/mdr.2018.0019>.
67. FDA. 2018. Summary report on antimicrobials sold or distributed for use in food-producing animals. *Cent Vet Med* 2018:1–14. <https://www.fda.gov/media/144427/download>.
68. Lu S, Jin D, Wu S, Yang J, Lan R, Bai X, Liu S, Meng Q, Yuan X, Zhou J, Pu J, Chen Q, Dai H, Hu Y, Xiong Y, Ye C, Xu J. 2016. Insights into the evolution of pathogenicity of *Escherichia coli* from genomic analysis of intestinal *E. coli* of *Marmota himalayana* in Qinghai–Tibet plateau of China. *Emerg Microbes Infect* 5:e122. <https://doi.org/10.1038/emi.2016.122>.
69. Seemann T. 2015. Snippy, rapid haploid variant calling and core genome alignment. GitHub Repository <https://github.com/tseemann/snippy>.
70. Lutteropp S, Kozlov AM, Stamatakis A. 2019. A fast and memory-efficient implementation of the transfer bootstrap. *bioRxiv*. <https://www.biorxiv.org/content/10.1101/734848v2>.
71. Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690. <https://doi.org/10.1093/bioinformatics/btl446>.
72. Yu G. 2020. Using ggtree to visualize data on tree-like structures. *Curr Protoc Bioinform* 69:e96. <https://doi.org/10.1002/cpbi.96>.
73. Kassambara A, Mundt F. 2020. Factoextra: extract and visualize the results of multivariate data analyses. R Package Version 1.0.7. <https://CRAN.R-project.org/package=factoextra>.
74. Campitelli E. 2022. ggnewscale: multiple fill and colour scales in ggplot2. <https://doi.org/10.5281/zenodo.2543762>.
75. Bessonov K, Laing C, Robertson J, Yong I, Ziebell K, Gannon VPJ, Nichani A, Arya G, Nash JHE, Christianson S. 2021. ECTyper: *in silico* *Escherichia coli* serotype and species prediction from raw and assembled whole-genome sequence data. *Microb Genom* 7:000728. <https://doi.org/10.1099/mgen.0.000728>.
76. Seemann T. 2020. Abricate. GitHub <https://github.com/tseemann/abricate>.
77. Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, Rolain J-M. 2014. ARG-annot, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob Agents Chemother* 58: 212–220. <https://doi.org/10.1128/AAC.01310-13>.
78. Lakin SM, Dean C, Noyes NR, Dettenwanger A, Ross AS, Doster E, Rovira P, Abdo Z, Jones KL, Ruiz J, Belk KE, Morley PS, Boucher C. 2017. MEGARes: an antimicrobial resistance database for high throughput sequencing. *Nucleic Acids Res* 45:D574–D580. <https://doi.org/10.1093/nar/gkw1009>.
79. Feldgarden M, Brover V, Haft DH, Prasad AJ, Slotta DJ, et al. 2019. Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrob Agents Chemother* 63:e00483-19. <https://doi.org/10.1128/AAC.00483-19>.
80. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, Møller Aarestrup F, Hasman H. 2014. *In silico* detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 58:3895–3903. <https://doi.org/10.1128/AAC.02412-14>.
81. Liu B, Zheng D, Jin Q, Chen L, Yang J. 2019. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res* 47:D687–D692. <https://doi.org/10.1093/nar/gky1080>.
82. Bortolonia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, Philippon A, Allesoe RL, Rebelo AR, Florensa AF, Fagelhauer L, Chakraborty T, Neumann B, Werner G, Bender JK, Stingl K, Nguyen M, Coppens J, Xavier BB, Malhotra-Kumar S, Westh H, Pinholt M, Anjum MF, Duggett NA, Kempf I, Nykäsenoja S, Oikola S, Wiczorek K, Amaro A, Clemente L, Mossong J, Losch S, Ragimbeau C, Lund O, Aarestrup FM. 2020. ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother* 75:3491–3500. <https://doi.org/10.1093/jac/dkaa345>.
83. Sokal RR. 1958. A statistical method for evaluating systematic relationships. *Univ Kansas Sci Bull* 38:1409–1438. <https://doi.org/10.24517/00055549>.
84. Miller DL, Rexstad E, Thomas L, Marshall L, Laake JL. 2019. Distance sampling in R. *J Statist Softw* 89:1–28. <https://doi.org/10.18637/jss.v089.i01>.
85. Moerbeck MV. 2018. IntClust: integration of multiple data sets with clustering techniques. Version 0.1.0. <https://CRAN.R-project.org/package=IntClust>.
86. R Core Team. 2015. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org>.
87. Wickham H. 2016. ggplot2: elegant graphics for data analysis. Springer-Verlag, New York, NY. ISBN 978-3-319-24277-4. <https://ggplot2.tidyverse.org>.
88. RStudio, Inc. 2016. Integrated development environment for R. RStudio, Inc, Boston, MA.
89. Breiman L. 2001. Random forests. *Machine Learning* 45:5–32. <https://doi.org/10.1023/A:1010933404324>.
90. Liaw A, Wiener M. 2002. Classification and regression by randomForest. *R News* 2:18–22. <https://CRAN.R-project.org/doc/Rnews/>.