



OPEN

Genomic evolution of antimicrobial resistance in *Escherichia coli*

Pimlapas Leekitcharoenphon^{1✉}, Markus Hans Kristofer Johansson¹, Patrick Munk¹, Burkhard Malorny², Magdalena Skarżyńska³, Katharina Wadepohl⁴, Gabriel Moyano⁵, Ayla Hesp⁶, Kees T. Veldman⁶, Alex Bossers^{7,6}, EFFORT Consortium*, Magdalena Zajac³, Dariusz Wasyl³, Pascal Sanders⁸, Bruno Gonzalez-Zorn⁵, Michael S. M. Brouwer⁶, Jaap A. Wagenaar⁹, Dick J. J. Heederik⁷, Dik Mevius⁶ & Frank M. Aarestrup¹

The emergence of antimicrobial resistance (AMR) is one of the biggest health threats globally. In addition, the use of antimicrobial drugs in humans and livestock is considered an important driver of antimicrobial resistance. The commensal microbiota, and especially the intestinal microbiota, has been shown to have an important role in the emergence of AMR. Mobile genetic elements (MGEs) also play a central role in facilitating the acquisition and spread of AMR genes. We isolated *Escherichia coli* (n = 627) from fecal samples in respectively 25 poultry, 28 swine, and 15 veal calf herds from 6 European countries to investigate the phylogeny of *E. coli* at country, animal host and farm levels. Furthermore, we examine the evolution of AMR in *E. coli* genomes including an association with virulence genes, plasmids and MGEs. We compared the abundance metrics retrieved from metagenomic sequencing and whole genome sequenced of *E. coli* isolates from the same fecal samples and farms. The *E. coli* isolates in this study indicated no clonality or clustering based on country of origin and genetic markers; AMR, and MGEs. Nonetheless, mobile genetic elements play a role in the acquisition of AMR and virulence genes. Additionally, an abundance of AMR was agreeable between metagenomic and whole genome sequencing analysis for several AMR classes in poultry fecal samples suggesting that metagenomics could be used as an indicator for surveillance of AMR in *E. coli* isolates and vice versa.

Escherichia coli, is an inhabitant of the gastrointestinal tract and feces of warm-blooded animals and reptiles^{1,2}. *Escherichia coli* has the interesting characteristic of being both a widespread gut commensal bacteria in vertebrates and a versatile pathogen, resulting in killing more than 2 million humans per year through both intra- and extraintestinal diseases^{3,4}. The genetic structure of commensal *E. coli* is shaped by multiple host and environmental factors, and genetic determinants involved in the virulence of the bacteria reflect adaptation to commensal habitats⁵. Selective pressure in the habitats of commensal strains may coincidentally promote the emergence of virulence factors and antimicrobial resistance (AMR), rendering commensal *E. coli* strain reservoirs of virulence and AMR determinants⁵. Emergence of antimicrobial resistance is one of the biggest global threat to public health⁶. In addition to the use of antimicrobials in humans, the use of antimicrobials in livestock is considered an important driver of AMR, subsequently compromising human health⁷.

Besides AMR in zoonotic pathogens, AMR in commensal bacteria is perturbing due to its ability to spread horizontally to pathogens⁸. The commensal microbiota, and especially the intestinal microbiota, has shown to play an important role in the emergence of antimicrobial resistance⁹. The extensive use of antimicrobials in both human and veterinary medicine, trigger the selection of antimicrobial resistance in the commensal microorganisms⁵. Antimicrobial resistance genes are frequently mobilized and disseminated through the bacterial population through the interplay of various mobile genetic elements (MGEs) of different types¹⁰.

¹National Food Institute, Technical University of Denmark, Kemitorvet, Building 204, 2800 Kgs. Lyngby, Denmark. ²Department Biological Safety, German Federal Institute for Risk Assessment, Berlin, Germany. ³National Veterinary Research Institute, Puławy, Poland. ⁴University of Veterinary Medicine Hannover, Bakum, Germany. ⁵Department of Animal Health and Health Surveillance Center (VISAVET), Complutense University of Madrid, Madrid, Spain. ⁶Wageningen Bioveterinary Research, Lelystad, The Netherlands. ⁷Faculty of Veterinary Medicine, Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands. ⁸Fougères Laboratory, French Agency for Food, Environmental and Occupational Health and Safety, Fougères, France. ⁹Department of Infectious Diseases and Immunology, Utrecht University, Utrecht, The Netherlands. ¹⁰A list of authors and their affiliations appears at the end of the paper. ✉email: pile@food.dtu.dk

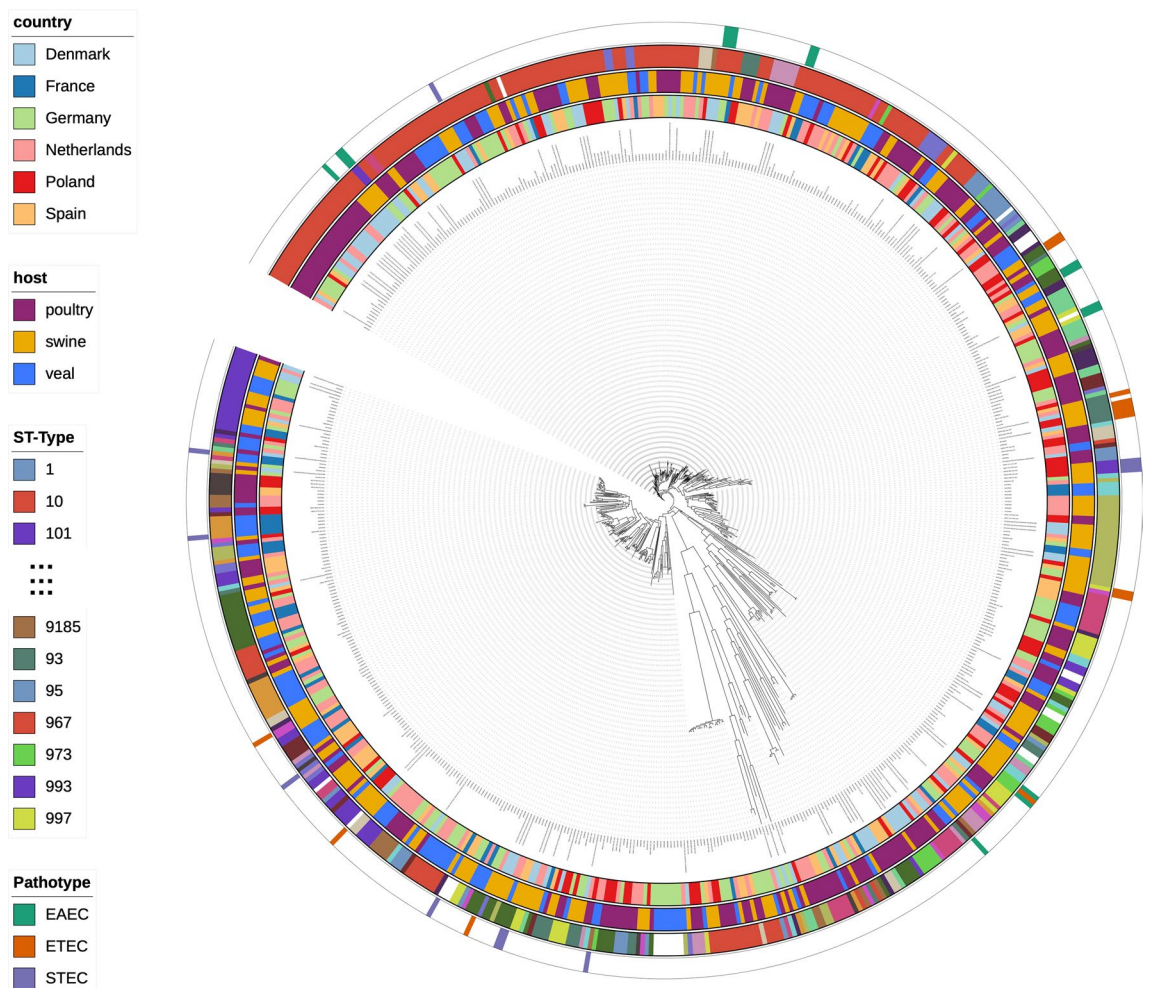


Figure 1. SNP tree of 627 *E. coli* isolates. The tree was constructed from 17,213 SNPs. The legends from inner circle to outer circle are country, host, ST type and pathotype. There are 182 different ST types with 24 isolates characterized as unknown ST type. Only a number of ST type can be showed in figure legend. EAEC Enterotoxigenic *E. coli*, ETEC Enterotoxigenic *E. coli* and STEC Shiga toxin-producing *E. coli*.

In this study, we isolated *E. coli* from fecal samples in poultry flocks, swine, and veal calf herds from 6 European countries to investigate phylogenetic background and clustering of *E. coli* at country, animal source and farm levels. Furthermore, this study provided an insight into the evolution of AMR in *E. coli* genomes including an association with virulence genes, plasmid and mobile genetic elements (MGEs). In addition, the correlation of AMR abundance between metagenomics and WGS analyses was determined.

Results

***Escherichia coli* phylogeny and genetic diversity.** The phylogenetic SNP tree exhibited diverse strains of *E. coli* with 17,213 total SNPs (Fig. 1). There was no clear clustering either by country of origin or animal host. In addition, within a small cluster from the same animal host, it contained strains from more than one country. Most of *E. coli* isolates were non-pathogenic strains. There were 39 isolates (6%) identified as EAEC/Enterotoxigenic *E. coli* ($n = 15$, 2%), ETEC/Enterotoxigenic *E. coli* ($n = 13$, 2%) and STEC/Shiga toxin-producing *E. coli* ($n = 11$, 2%). Furthermore, the SNP tree did not show any country and animal reservoir clusters of AMR genes nor chromosomal point mutation conferring to antimicrobial resistance (Figs. S2–S4). Looking at SNP tree in a country level (Figs. S5–S10), *E. coli* isolates within the same country could be related to strains from the same or different farms. The 627 *E. coli* isolates comprised of 182 different sequence types (ST) types with 24 isolates being of unknown ST type, indicating high diversity in the *E. coli* isolates. The major identified ST type was ST10 ($n = 143$, 23%) and *E. coli* isolates seems to be clustered based on ST type rather than geographical location, animal reservoir, *E. coli* pathotype or AMR (Fig. 1). Similar tree topology of the *E. coli* isolates was observed in core genome MLST tree (Fig. S11).

Average nucleotide identity (ANI) from comparative genomics of the *E. coli* isolates was summarised on farm and animal host levels (Fig. 2). Mean of overall ANI was 98.3% (red horizontal line). Five out of 10 *E. coli* isolates from Danish and Spanish farms and five of 15 isolates from Netherlands had ANI lower than average (98.3%). Meanwhile Germany, France and Poland had three out of 17, one out of five and one out of 10 farms with ANI lower than average respectively. Most of the lower average ANI farms were poultry and swine. Figure 2B showed

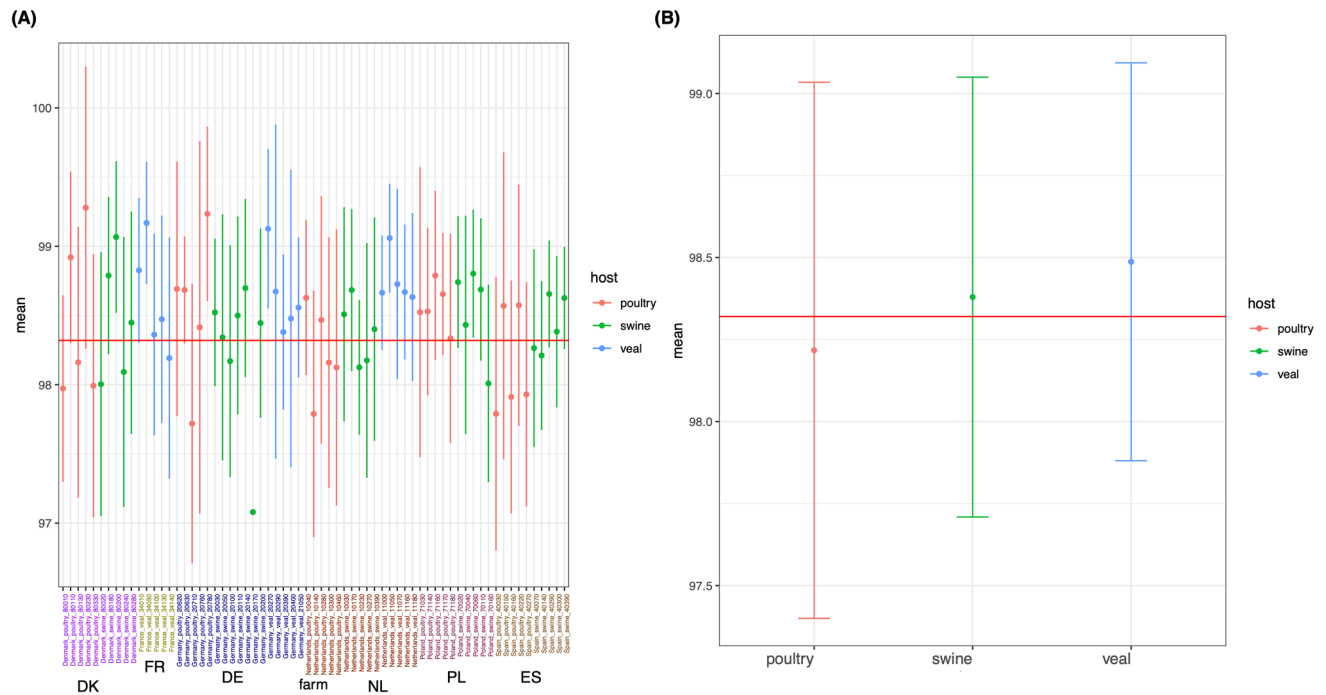


Figure 2. Average nucleotide identity in mean percent identity of *E. coli* isolates from the same farm (A) and from the same animal host (B). DK Denmark, FR France, DE Germany, NL The Netherlands, PL Poland and ES Spain. Vertical line represents points within 1st/3rd + 1.5 IQR. A red horizontal line indicates mean of overall nucleotide identity (98.34%).

that *E. coli* isolates from poultry had the lowest mean ANI and broadest range of ANI compared to the ones from swine and veal, reflecting the high strain diversity.

Antimicrobial resistance genes. Prevalence and distribution of resistant *E. coli* strains per country and host were illustrated in Figs. 3 and 4. Denmark had the lowest abundance of AMR strains, meanwhile Spain had the highest abundance. All countries had a similar distribution of AMR with the exception of some genes that were more abundant in Poland, Spain and France. Poland had higher proportion of fluoroquinolone resistant *E. coli* (Figs. 3, 4) which were caused by *qnrB19* and *qnrS1* genes (Fig. S12). Spain and France had higher proportion of sulphonamide resistant *E. coli* (Figs. 3, 4) associated with *sul1*, *sul2* and *sul3* genes (Fig. S12 and Table S1 sub-sheet “Percentage of resistant *E. coli*”). Regarding the animal reservoir, isolates from veal calves had higher proportion of AMR resistant *E. coli* strains (Fig. 3), while the proportion of fluoroquinolone resistant *E. coli* was higher in poultry and swine. Additionally, the proportion of beta-lactam resistant *E. coli* was higher in poultry which was caused mainly by *bla_{TEM-1B}* (Fig. S12 and Table S1 sub-sheet “Percentage of resistant *E. coli*”).

The prevalence of AMR was ranked based on percentage of *E. coli* carrying AMR genes in country of origin and animal host. The most prevalent five AMR genes were revealed (Table S1 sub-sheet “Percentage of resistant *E. coli*”). They were *tet(A)*, *bla_{TEM-1B}*, *sul2*, *aadA1* and *aph(6)-Id* coding for resistance against tetracyclines, aminopenicillins, sulphonamides and aminoglycosides. These five AMR genes were also the top five AMR genes in the strains from France and Poland. Three to Four of them were among the top five AMR genes in other countries in this study. In addition, these five AMR genes were most prevalent AMR genes in the *E. coli* strains from swine and veal calves. Four of them were among the top five AMR genes in poultry (Table S1 sub-sheet “Percentage of resistant *E. coli*”).

The most prevalent five AMR genes were analysed for conserved regions on 1 kb upstream and 1 kb downstream of the gene (Fig. S13). Although, upstream and downstream regions from almost all five AMR genes showed variable regions and not any large conserved regions, the upstream region from *aph(6)-Id* exhibited ~600 bp conserved region among the *E. coli* isolates (Fig. S13, grey colour) suggesting that among the top 5 AMR genes, only *aph(6)-Id* gene seems to be specifically conserved in *E. coli* isolates.

Mobile genetic element linked with plasmid, AMR and virulence genes. There was little difference in both total number of MGEs and in the average number of MGEs carried by an *E. coli* between the different animal hosts (Fig. 5A). However, the composition of MGEs carried by *E. coli* differed between the hosts. More unity transposon (Tn) were predicted in isolates from poultry and swine than in isolates from veal (Fig. S14). Some insertion sequences (IS) were very prevalent, for instance *ISEc1* and *IS609* were found in 87.9% and 79.9% of all isolates (Table S2). The predicted IS existed in multiple copies with varying degree of alignment completeness and level of truncation compared to the reference sequence. For all of the predicted IS a majority of the identified copies were highly truncated but had an intact inverted repeat region which could help form composite transposons (ComTn) (Tables S2 and S3). The most frequent transposon was Tn2 (Fig. 5B) which was

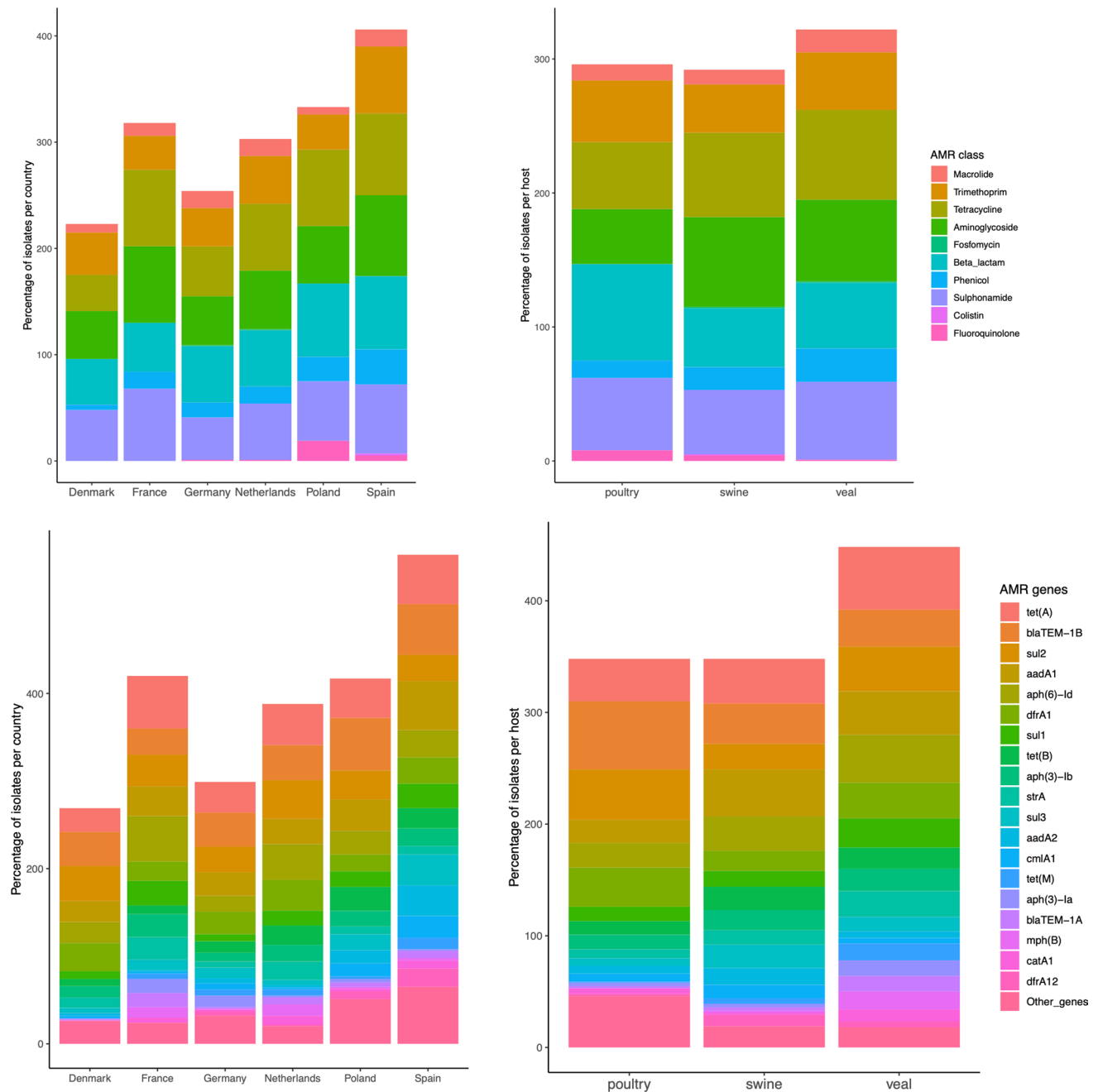


Figure 3. Bar plot showing percentage of resistant *E. coli* isolates in AMR classes (top) and genes (bottom) stratified by country (left) and host (right).

frequently found on IncX1, Inc11-I(gamma) and IncFIC(FII) plasmids (Fig. 5B). Tn2 is known that *E. coli* unit-transposon carries *bla*TEM-1 gene¹¹.

Escherichia coli isolates were clustered on the predicted MGE profile using Jaccard distance and complete linkage for all strains (Fig. S15) and isolates per country (Figs. S16–S21). The trees showed the distribution of MGEs within the *E. coli* population and any relationship with country, animal source, farm ID or ST type. *Escherichia coli* isolates carried a highly diverse set of MGEs that were seemingly independent of host or ST type. These *E. coli* isolates carried on average ~ 4 plasmids but a minority of the predicted MGEs were located on them (Fig. S22).

AMR gene was classified based on their location in respect to MGEs. A gene considered to be associated if it was located closer than 31 kbp to an MGE, mobilized if it was an accessory gene within the MGE, and having an unknown association if neither criterion were met. A gene associated with an MGE could be mobilized through formation of ComTn. Beta-lactamase type AMR genes were more frequently mobilized by MGEs more than genes from other AMR classes. In addition, more beta-lactamases were carried on transposons in poultry and swine than in veal calves (Fig. 5C). These carried beta-lactamases were mainly *bla*TEM-1 carried on transposon Tn2. In addition, three *tet*(A) genes were mobilized on Tn1721 (Table S4). Some putative ComTn were identified where the most abundant mobilized a variety of AMR genes (Table S5).



Figure 4. Distribution of resistance and susceptible *E. coli* isolates in different AMR class and host. Y-axis is number of isolates. Three columns are number of isolates from poultry, swine and veal. In each column, first bar is number of susceptible isolates and second bar is number of resistant isolates.

The distance between the gene and the MGE its associated with were shown in Fig. 6. MGEs such as IS26, ISVsa3 and Tn2 were associated with several AMR genes where IS26 was associated with the most MGEs (Fig. 6). MGEs tended to be located at a fixed distance from a gene. This could be an indication that they are transferred as a unit. One indication that the AMR genes was mobilized by the MGEs are their tendency to be located arrays of repeated genes with variable length and content. These arrays have MGE located up to 4151 bp downstream of the first AMR (Table S6). However due to the low number of observations these results are rather speculative.

Abundance of AMR from metagenomic and whole genome sequencing. Fecal samples used for *E. coli* isolation were sequenced using shotgun metagenomics⁸ and AMR genes were characterized in the metagenomics data using the ResFinder database. Abundance of AMR genes from metagenomic fecal samples and *E. coli* whole genome sequencing from the same farm was illustrated as scatter plot in Fig. 7. The AMR abundance in the metagenomic samples were calculated as fragments per kilobase reference per million bacterial fragments (FPKM), while the AMR abundance in *E. coli* WGS was determined as percentage of *E. coli* isolates carrying AMR genes. Correlation and *p* value from the regression analysis in Fig. 7 was shown in Table 1. There was significant correlation between AMR abundance of metagenomics and AMR prevalence in *E. coli* WGS from poultry in trimethoprim, aminoglycoside, beta-lactam, phenicol and quinolone (5 out of 8 AMR classes present in *E. coli* WGS from poultry). Correlation between AMR abundance in colistin was found only in swine samples. Significant correlation was not observed in veal samples. This suggested that AMR abundance from poultry feces could potentially represent AMR abundance in single isolation of *E. coli* from poultry.

Discussion

This study is a large-scale multinational study of *E. coli* isolates from several animal sources within the same year. The 627 *E. coli* exhibited a considerable degree of genetic diversity, differing by a total of 17,213 SNPs and representing 182 different ST types. A recent study on genetic diversification from a large collection of non-clinical *E.*

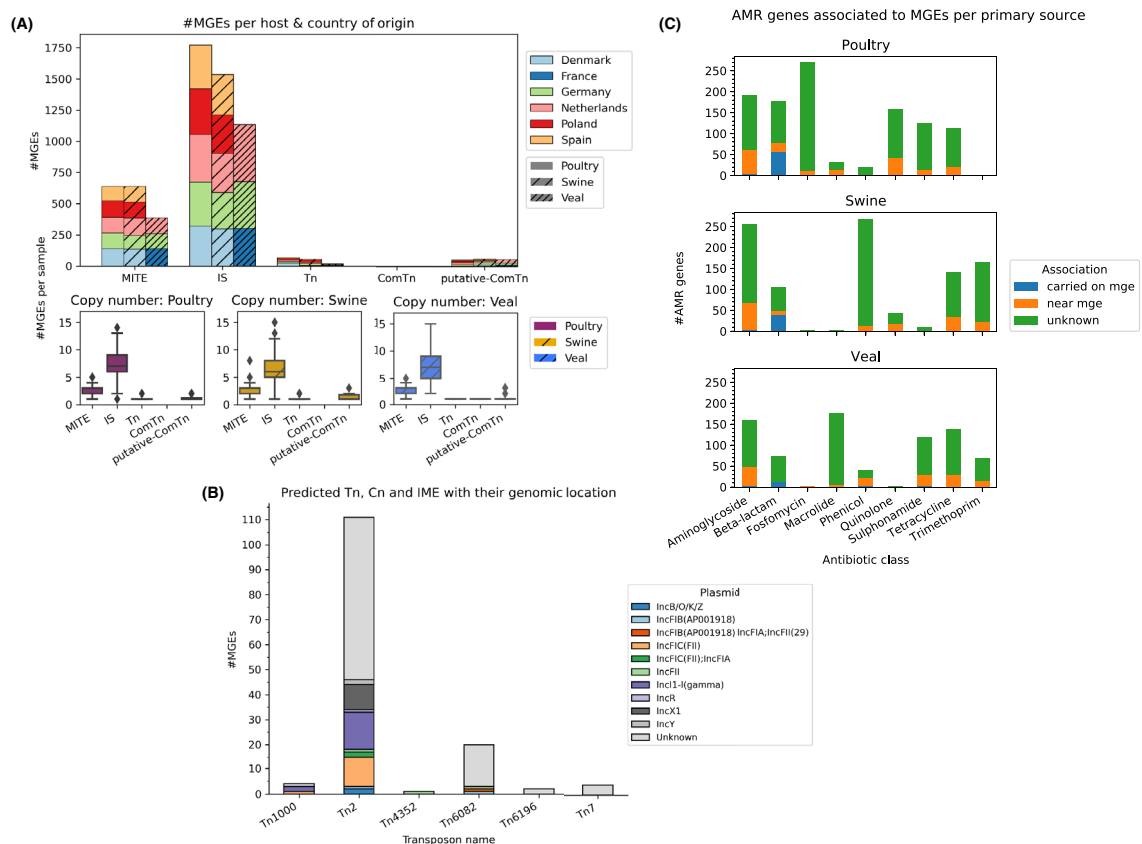


Figure 5. (A) Distribution of mobile genetic elements (MGEs) in *E. coli* isolates per animal host and country of origin (top). Distribution of MGEs per *E. coli* isolate in different animal hosts (bottom). *Cn* XX, *IS* insertion sequences, *MITE* miniature inverted-repeat transposable elements, and *Tn* transposons. (B) Distribution of transposons and their plasmids. (C) Number of MGEs associated with mobile elements per antimicrobial class and animal source. There are three levels of associations, carried on MGEs, if MGE is located within 31 kbp from an AMR and unknown association.

coli isolates also revealed 442 distinct ST types of which 61% were from a single strain¹². *Escherichia coli* strains isolated from humans, domesticated and wild animals represent a primary habitat of *E. coli*⁵. Most of the *E. coli* from these domesticated animals were not harmful pathogenic strains and 6% were identified as EAEC, ETEC or STEC. The high genetic diversity in the *E. coli* isolates resulted in phylogeny which showed no clustering based on country of origin, animal reservoir and antimicrobial resistance. Moreover, the strains within the same country could be more related to strains from different farms. The results suggested that the *E. coli* strains could be randomly spread between the farms in each country. A recent study on *E. coli* isolates collected from pig surveillance between 2013 and 2017 in the United Kingdom also found no correlation between phylogeny and year of isolation¹³. This suggests that *E. coli* isolates could be as well randomly spread in countries and through different years. An alternate explanation would be that the *E. coli* we have sequenced do not fully represent the extent of the diversity within each animal, farm, and country.

Average nucleotide identity (ANI) is a measure of genomic divergence between genomes. The ANI of *E. coli* isolates from different countries did not exhibit any clear differences among countries. Nonetheless, *E. coli* isolates from poultry showed lowest and broadest range of ANI suggesting that the poultry *E. coli* had higher genetic diversity and were more heterogeneous than isolates from swine and veal calves. Previous study concluded that *E. coli* genotypes and serotypes in pullets and layers are heterogeneous and do not maintain a single clonality within the same bird¹⁴. In addition, *E. coli* do not show definite clonal population structure based on geographical region and age of the host¹⁴. In animals, the main environmental force shaping the genetic structure of *E. coli* gut population is domestication status of the host¹⁵ and socioeconomic factors, such as diet¹⁶ and hygiene, are presumably the main factors accounting for phylogenetic group distribution rather than geographical or host genetic conditions⁵.

According to phylogenetic result and distribution of AMR genes in *E. coli* from different countries, we suggested that AMR seems to spread randomly in all the surveyed countries. Exceptions were higher fluoroquinolone resistance in Polish strains and higher sulphonamide resistant *E. coli* in Spain and France. Veal calves had a higher proportion of resistant *E. coli* strains. A previous study on commensal *E. coli* from a systematic resistance monitoring of primary food production, slaughterhouses and retail, they found that dairy cows had the lowest number of resistant *E. coli*, whereas veal and pork had higher frequency of resistance strains ranging between 43

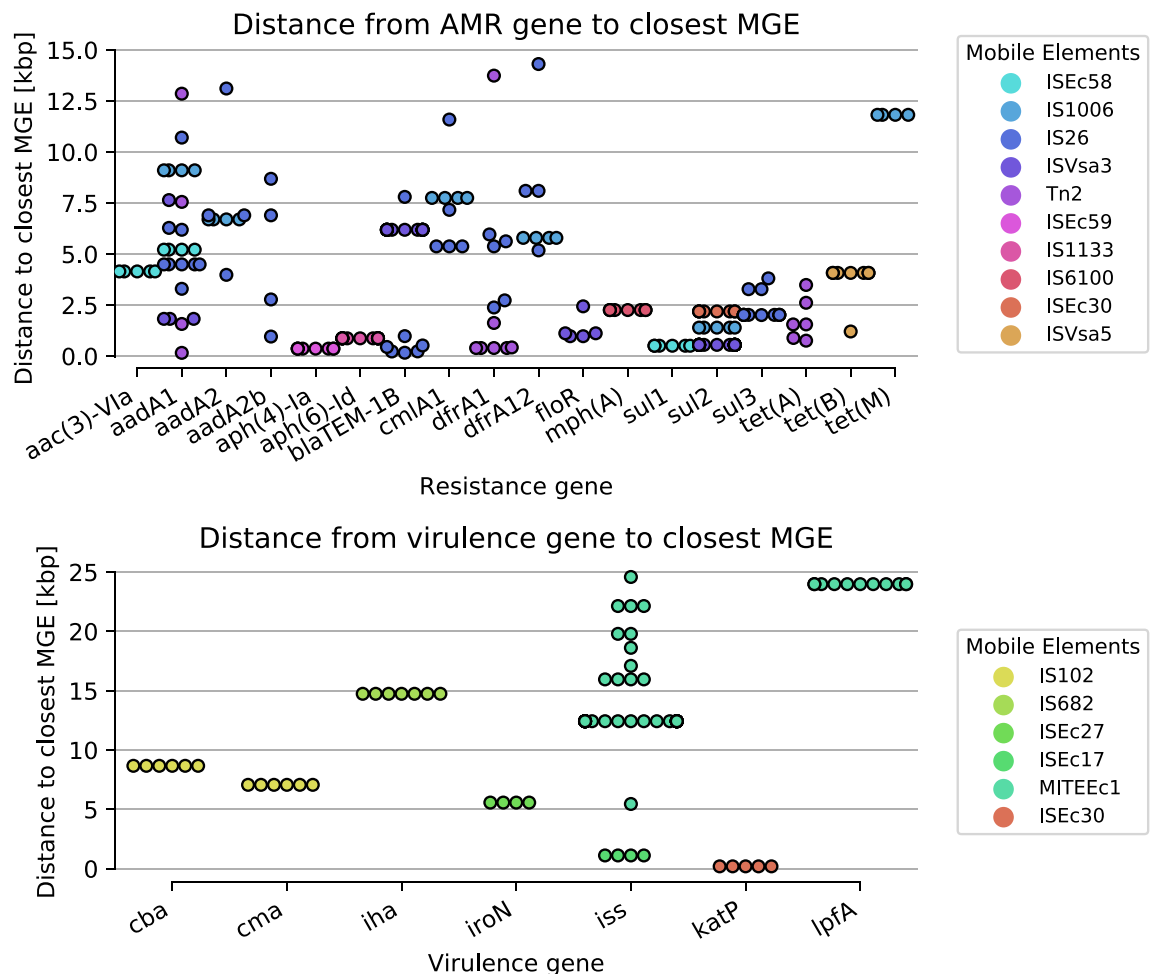


Figure 6. Distance from AMR gene (top) and virulence gene (bottom) to closest MGE.

and 73%¹⁷. In addition, our results showed that fluoroquinolone and beta-lactam resistant *E. coli* was higher in poultry. Similar results were found in a study in Czech Republic on *E. coli* from poultry meat, poultry farms and market-weight turkeys. They found that resistance of the poultry isolates to quinolones ranged from 53 to 73%¹⁸. Another study on *E. coli* from broilers, workers who worked in poultry slaughterhouses and clinical specimens in Romania, indicated a very high rate of ESBL and AmpC-producing *E. coli* in chickens¹⁹.

Genes are mobilized through the interplay of several MGEs of different types. Elements with intra-cellular transposition, such as IS and Tn, can mobilize new genes and transport them to conjugative/ mobilizable transposons and plasmids¹⁰. AMR genes associated with integrative MGEs (iMGEs) tended to be located at fixed distances from the iMGE, and was often part of an array of AMR genes. Some arrays exhibited greater variability of the genes contained the arrays than other. This indicates that these genes could be part of an integron, which in turn could be mobilized by the associated IS through the formation of a ComTn. Of these arrays the *aph(6)-Id* associated with IS1133 and *mph(A)* associated with IS6100 seemed to have less variation in the genes compared to the other arrays. The lack of variation might be caused by the associated MGEs are less prone to form ComTn.

The distribution of MGEs within the *E. coli* isolates revealed the lack of well-defined clusters and comparably dissimilar profiles suggesting that *E. coli* has a diverse and flexible mobilome. This could mean that mobile genetic elements are more mobile in *E. coli*, either due being frequently mobilized on plasmids or higher conjugation rate. Plasmids are important vehicles for the carriage of MGEs and AMR genes¹⁰, the result showed that the *E. coli* carried approximately four plasmids on average but a minority of the MGEs were located on them. However, the number of mobilized MGEs are likely underestimated due to a combination of the limitations in the detection method of PlasmidFinder²⁰ and the difficulty assembling repetitive DNA, and with that plasmids, from short read sequence data. Moreover, a significant number of plasmid sequences in INSDC databases (DDBJ, EMBL-EBI, NCBI) may be misassembled, as evidenced by the presence of fragments of MGEs that are not explained by truncation by other MGEs²¹. To further study the mobility and distribution of AMRs one would need to construct near complete the assemblies, preferably via hybrid assembly using long read sequencing as a template, and combine it with additional plasmid prediction tools that do bases its predictions on circular features in the data.

When looking at combinations of MGE, distances and AMR genes that were observed more than four times they were primarily found on a single type of plasmid which could explain the observed pattern.

MGEs can influence the expression and/or mobility of nearby genes through a variety of mechanisms. Through the formation of ComTn an IS, which lacks accessory genes, transpose several genes as unit within

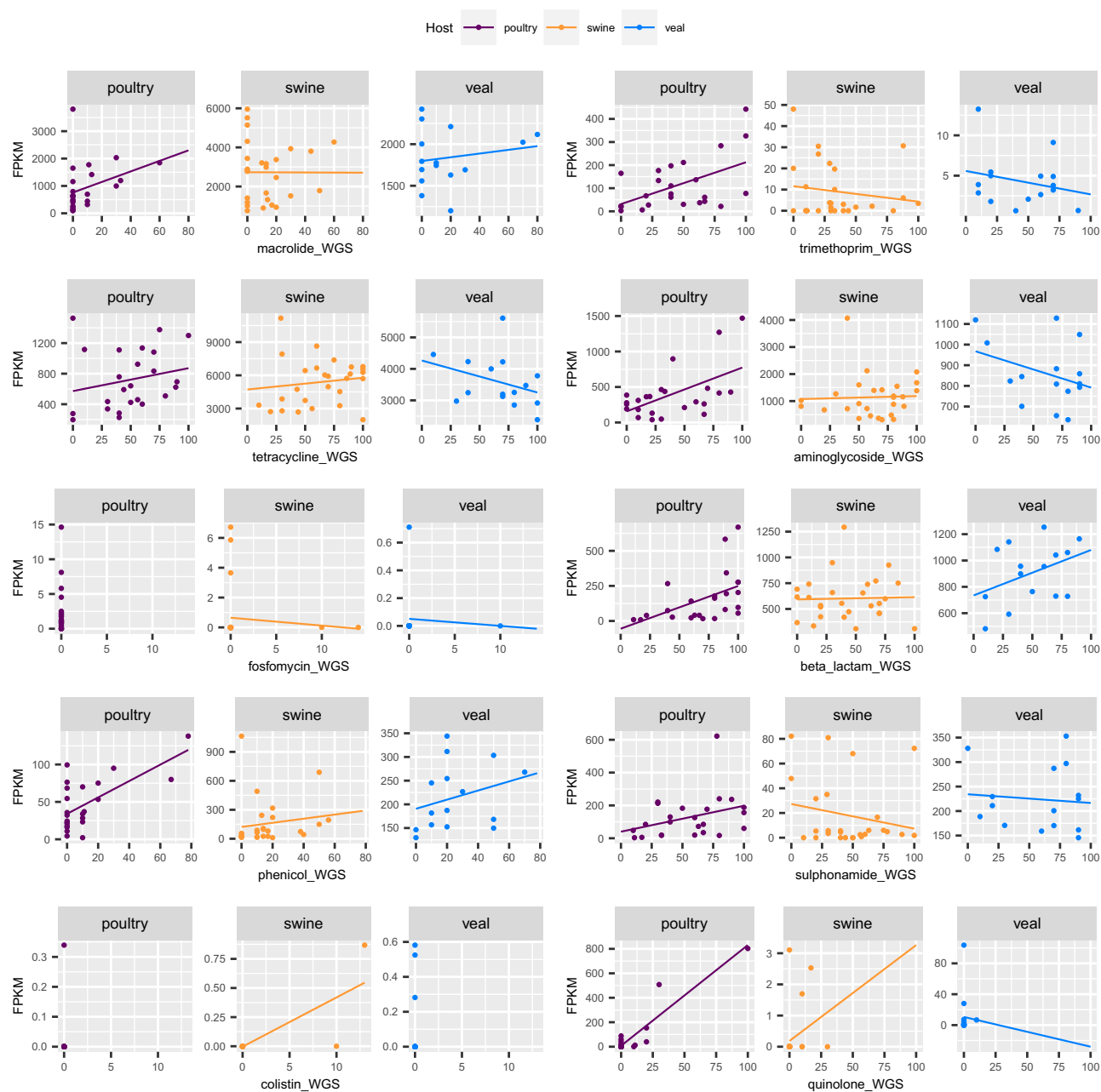


Figure 7. Scatter plot between abundance of AMR genes from metagenomic fecal samples and percentage of *E. coli* isolates carrying AMR genes from the same farm. X-axis is percentage of *E. coli* isolates carrying resistance genes. Y-axis represents abundance of resistance genes in FPKM (Fragments Per Kilo base per Million fragments) from metagenomic fecal samples.

the cell. We identified several putative composite transposons carrying AMR genes based on the location and completeness of IS elements. Of these were IS26 based elements both the smallest, only carrying one or two genes, and the one most frequently predicted to carry different genes. This is likely caused by the tendency of IS26 to form arrays of repeated MGEs²². However, further research is needed to verify if the putative ComTn can mobilize these genes.

AMR genes that were found within 31 kbp of a MGE was considered to be associated, meaning that it has the potential to be mobilized. The threshold corresponds to the length of the largest ComTn from *Enterobacteriaceae* in MobileElementFinders database (Tn6108). Association between MGEs and AMR genes showed that most of the AMR genes carried on MGEs were from beta-lactam class especially *bla*_{TEM-1} and beta-lactamases were mainly carried on transposons in poultry and swine. In addition, MGEs such as IS26, ISVsa3 and Tn2 were associated with many AMR genes. IS26 has played a crucial role in the dissemination of resistance determinants in Gram-negative bacteria^{10,22}. AMR genes such as *bla*_{TEM} genes, including extended-spectrum beta-lactamases (ESBL)

AMR	Poultry		Swine		Veal	
	Correlation	<i>p</i> value	Correlation	<i>p</i> value	Correlation	<i>p</i> value
Macroline	0.35	0.099	−0.0034	0.99	0.17	0.55
Trimethoprim	0.52	0.0099	−0.16	0.43	−0.24	0.38
Tetracycline	0.22	0.31	0.14	0.5	−0.34	0.22
Aminoglycoside	0.55	0.0053	0.039	0.84	−0.34	0.21
Fosfomycin	–	–	−0.096	0.64	−0.071	0.8
Beta lactam	0.49	0.014	0.029	0.89	0.4	0.14
Phenicol	0.65	0.00056	0.15	0.46	0.29	0.3
Sulphonamide	0.34	0.1	−0.2	0.32	−0.089	0.75
Colistin	–	–	0.78	1.3e−06	–	–
Quinolone	0.93	4.7e−11	0.27	0.18	−0.037	0.9

Table 1. Correlation and *p* value from regression analysis between abundance of AMR genes in metagenomic fecal samples and percentage of *E. coli* isolates carrying AMR genes. The bold value represents significant correlation between the abundance of AMR genes from metagenomic samples and single isolation.

encoding genes have been known that they have been found within Tn1, Tn2 and Tn3 transposons¹⁰. Analysis of the distance from AMR and virulence genes to the nearest MGEs indicated that MGEs tended to be located at a fixed distance from the AMR and virulence genes suggesting that the AMR and virulence genes in the *E. coli* isolates have been in part mobilized by the associated MGEs.

The predicted MGEs were only of types that are comparably and performs intra cellular transposition. We did not identify any conjugative/mobilizable transposons such as SXT or ICEEc2¹⁰ suggests that we are missing these classes MGEs which play an important role for the mobilization of genes. These elements are likely not being predicted due to the difficulty in assembling these elements with short read data. Like plasmids, are these elements large and contains repetitive sequences and other MGEs which could confuse the assembler resulting in fragmented assemblies. To overcome these shortcomings, one would likely need to employ long read sequencing. While phages play an important role movement of AMR genes they have not been addressed in this work.

Comparing abundance of AMR from metagenomic analysis of fecal samples and whole genome sequencing of *E. coli* isolated from the same fecal samples, it showed that 5 main AMR classes out of 10 analyzed classes exhibited significant correlation of AMR abundance between metagenomics and WGS results in *E. coli* poultry isolates. No significant correlation of AMR abundance was found in any AMR classes between metagenomics and WGS analysis in swine (except colistin) and veal calves. This result suggested that metagenomics analysis could be used as a predictor for abundance of AMR from bacteria isolates in poultry.

Conclusion

Commensal *E. coli* isolates from different farms in different European countries in this study indicated neither clonality nor clustering based on geographical location of countries and genetic markers such as antimicrobial resistance and virulence genes. The *E. coli* tended to sporadically spread internationally. Nonetheless, mobile genetic elements play a role in the acquisition of antimicrobial resistance and virulence genes. Abundance of AMR was agreeable between metagenomic and whole genome sequencing analysis in several AMR classes in poultry isolates.

Materials and methods

Selection and collection of isolates. As part of the European Union-funded EFFORT project (www.effort-against-amr.eu), fecal samples from animal farms were collected in 2014–2015 from Denmark (10 farms), France (5 farms), Germany (17 farms), The Netherlands (15 farms), Poland (10 farms), and Spain (10 farms) and. From each farm, up to 10 *E. coli* isolates were isolated and whole genome sequenced from poultry (*n* = 239 from 25 farms), swine (*n* = 238 from 28 farms), and veal (*n* = 150 from 15 farms). Farms were selected based on a ranking of antimicrobial use (AMU). Distribution of number of *E. coli* isolates and a list of strains including metadata can be found in Fig. S1 and Table S1 respectively. In addition, previously sequenced metagenomic samples (poultry⁸, swine⁸ and veal calf [PRJEB39685]) from the same farm and fecal samples that were used for *E. coli* isolation were included in this study to compare abundance of AMR from metagenomic and whole genome sequence analysis.

DNA extraction and whole genome sequencing. Genomic DNA was extracted from *E. coli* isolates using PureLink Genomic DNA Mini Kit (Thermo Fisher Scientific), Qiagen, Gentra, or Puregene bacterial kits and DNA concentrations were determined using the Qubit dsDNA BR assay kit (Thermo Fisher Scientific). Sequencing libraries were prepared with the Nextera XT DNA library preparation kit or the Nextera DNA Flex library preparation kit according to the manufacturer's protocol (Illumina, Inc., San Diego, CA, United States). Paired-end sequencing was performed on the Illumina HiSeq, MiSeq or NextSeq sequencing platform.

Raw sequence data were deposited in the European Nucleotide Archive (<http://www.ebi.ac.uk/ena>) under study accession no.: PRJEB41365. The raw reads were adapter-trimmed, quality filtered using bbdutk (part of

the suite bbttools version 36.49)²³ and de novo assembled using SPAdes version 3.11²⁴. Genomic sequence data including ENA accession numbers is available in the Table S1.

Species identification, conventional seven multilocus sequence typing, screening for AMR genes, chromosomal point mutations, plasmid and mobile genetic elements. Species identification was used to confirm *E. coli* species using KmerFinder 3.2^{24–26}. Conventional seven multilocus sequence typing (MLST) was performed using MLST 2.0²⁷. Acquired antimicrobial resistance genes as well as chromosomal point mutations causing resistance were determined using ResFinder 4.0²⁸. Virulence genes in *E. coli* were predicted using VirulenceFinder 2.0²⁹. Plasmid replicon was identified to infer plasmid presence using PlasmidFinder 2.1. Integrated mobile genetic elements were predicted in the assembled genomes using MobileElementFinder version 1.0 with MGEdB v1.0.2³⁰. With these tools we could identify plasmids, miniature inverted repeats, insertion sequences, transposons and conjugative transposons. If two complete copies of the same MGE, or one complete copy and one truncated with preserved IR region was predicted within 31 kbp from one another they were flagged as putative composite transposons. In addition was MGEs located within 31 kbp from an AMR or virulence gene was considered to be associated and to have a theoretical ability to mobilize the associated gene³⁰. The threshold corresponds to the longest composite transposon in Enterobacteriaceae (Tn6108).

Average nucleotide identity (ANI). Average nucleotide identity (ANI) was determined from assembled genomes of *E. coli* strains by all-against-all alignment using FastANI³¹ (<https://github.com/kbaseapps/FastANI>) to estimate whole-genome similarity among the strains from different countries, animal hosts and farms.

Single nucleotide polymorphisms tree. Single nucleotide polymorphisms (SNPs) were determined using CSI phylogeny³². Basically, paired-end reads were mapped to the reference genome, *Escherichia coli* str. K-12 (accession number; NC_000913.3), using Burrows-Wheeler Aligner (BWA) version 0.7.2³³. SNPs were identified using 'mpileup' module in SAMTools version 0.1.18³⁴. Qualified SNPs were chosen using following criteria: (1) a minimum distance of 10 bps between each SNP, (2) a minimum of 10% of the average depth, (3) the mapping quality was above 25, (4) the SNP quality was more than 30, and (5) all INDELs were excluded. The qualified core SNPs were concatenated to a single alignment corresponding to the position of the reference genome. The concatenated SNP sequences were subjected to parsimony tree construction using PhyML³⁵ with the HKY85 substitution model.

cgMLST tree. The cgMLST loci sequences were retrieved from Enterobase³⁶ (<https://enterobase.warwick.ac.uk>). Raw reads were mapped with KMA against the cgMLST sequences to identify the best-matched allele numbers. Sequences were indexed with kmer size 16 and threshold for read mapping were match = 1, mismatch = - 2, gap-opening = - 3, gap-extension = - 1 and no threshold for read depth was used²⁶. ConClave sorting scheme was used to pick the best matching allele for each cgMLST gene²⁶. While all equally well matching alleles were saved in the first iteration, in a second iteration the most likely allele was chosen as the one with the highest number of matches. KMA is freely available at: <https://bitbucket.org/genomicpidemiology/kma> and <https://cge.cbs.dtu.dk/services/kma>. The cgMLSTFinder tool, which is based on KMA, can be found at: <https://cge.cbs.dtu.dk/services/cgMLSTFinder-1.0/>. The pairwise dissimilarities (distances) between genomes were identified based on the allele profile using 'gower' distance method from 'cluster' package version 2.1.0 in R. The cgMLST trees were constructed from the distance matrix using hierarchical clustering from 'ape' package version 5.4-1 in R.

Received: 11 January 2021; Accepted: 5 July 2021

Published online: 23 July 2021

References

- Berg, R. D. The indigenous gastrointestinal microflora. *Trends Microbiol.* **4**, 430–435 (1996).
- Gordon, D. M. & Cowling, A. The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: Host and geographic effects. *Microbiology* **149**, 3575–3586 (2003).
- Kosek, M., Bern, C. & Guerrant, R. L. The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000. *Bull. World Health Organ.* **81**, 197–204 (2003).
- Russo, T. A. & Johnson, J. R. Medical and economic impact of extraintestinal infections due to *Escherichia coli*: Focus on an increasingly important endemic problem. *Microbes Infect.* **5**, 449–456 (2003).
- Tenaillon, O., Skurnik, D., Picard, B. & Denamur, E. The population genetics of commensal *Escherichia coli*. *Nat. Rev. Microbiol.* **8**, 207–217 (2010).
- O'Neill, J. *Tackling Drug-Resistant Infections Globally: Final Report and Recommendations. The Review on Antimicrobial Resistance.* (2016).
- Aarestrup, F. M. The livestock reservoir for antimicrobial resistance: A personal view on changing patterns of risks, effects of interventions and the way forward. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **370**, 20140085 (2015).
- Munk, P. *et al.* Abundance and diversity of the faecal resistome in slaughter pigs and broilers in nine European countries. *Nat. Microbiol.* **3**, 898–908 (2018).
- Andremont, A. Commensal flora may play key role in spreading antibiotic resistance. *ASM News* **63**, 601–607 (2003).
- Partridge, S. R., Kwong, S. M., Firth, N. & Jensen, S. O. Mobile genetic elements associated with antimicrobial resistance. *Clin. Microbiol. Rev.* **31**, e00088–17 (2018).
- Chen, S. T. & Clowes, R. C. Variations between the nucleotide sequences of Tn1, Tn2, and Tn3 and expression of beta-lactamase in *Pseudomonas aeruginosa* and *Escherichia coli*. *J. Bacteriol.* **169**, 913–916 (1987).
- Touchon, M. *et al.* Phylogenetic background and habitat drive the genetic diversification of *Escherichia coli*. *PLoS Genet.* **16**, e1008866 (2020).

13. Duggett, N. *et al.* The importance of using whole genome sequencing and extended spectrum beta-lactamase selective media when monitoring antimicrobial resistance. *Sci. Rep.* **10**, 19880 (2020).
14. Paudel, S., Stessl, B., Hess, C., Zloch, A. & Hess, M. High genetic diversity among extraintestinal *Escherichia coli* isolates in pullets and layers revealed by a longitudinal study. *BMC Vet. Res.* **12**, 221 (2016).
15. Escobar-Páramo, P. *et al.* Identification of forces shaping the commensal *Escherichia coli* genetic structure by comparing animal and human isolates. *Environ. Microbiol.* **8**, 1975–1984 (2006).
16. Skarżyńska, M., Leekitcharoenphon, P., Hendriksen, R. S., Aarestrup, F. M. & Wasyl, D. A metagenomic glimpse into the gut of wild and domestic animals: Quantification of antimicrobial resistance and more. *PLoS One* **15**, e0242987 (2020).
17. Kaesbohrer, A. *et al.* Emerging antimicrobial resistance in commensal *Escherichia coli* with public health relevance. *Zoonoses Public Health* **59**(Suppl 2), 158–165 (2012).
18. Hricová, K. *et al.* Quinolone-resistant *Escherichia coli* in poultry farming. *Cent. Eur. J. Public Health* **25**, 163–167 (2017).
19. Maciucă, I. E. *et al.* High prevalence of *Escherichia coli*-producing CTX-M-15 extended-spectrum beta-lactamases in poultry and human clinical isolates in Romania. *Microb. Drug Resist.* **21**, 651–662 (2015).
20. Arredondo-Alonso, S., Willems, R. J., van Schaik, W. & Schürch, A. C. On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. *Microb. Genom.* **3**, e000128 (2017).
21. Partridge, S. R. & Tsafnat, G. Automated annotation of mobile antibiotic resistance in Gram-negative bacteria: The Multiple Antibiotic Resistance Annotator (MARA) and database. *J. Antimicrob. Chemother.* **73**, 883–890 (2018).
22. Harmer, C. J., Moran, R. A. & Hall, R. M. Movement of IS26-associated antibiotic resistance genes occurs via a translocatable unit that includes a single IS26 and preferentially inserts adjacent to another IS26. *MBio* **5**, e01801–e01814 (2014).
23. Bushnell, B., Rood, J. & Singer, E. BBMerge—Accurate paired shotgun read merging via overlap. *PLoS One* **12**, e0185056 (2017).
24. Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
25. Larsen, M. V. *et al.* Benchmarking of methods for genomic taxonomy. *J. Clin. Microbiol.* **52**, 1529–1539 (2014).
26. Clausen, P. T. L. C., Aarestrup, F. M. & Lund, O. Rapid and precise alignment of raw reads against redundant databases with KMA. *BMC Bioinform.* **19**, 307 (2018).
27. Larsen, M. V. *et al.* Multilocus sequence typing of total-genome-sequenced bacteria. *J. Clin. Microbiol.* **50**, 1355–1361 (2012).
28. Bortolai, V. *et al.* ResFinder 4.0 for predictions of phenotypes from genotypes. *J. Antimicrob. Chemother.* <https://doi.org/10.1093/jac/dkaa345> (2020).
29. Joensen, K. G. *et al.* Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* **52**, 1501–1510 (2014).
30. Johansson, M. H. K. *et al.* Detection of mobile genetic elements associated with antibiotic resistance in *Salmonella enterica* using a newly developed web tool: MobileElementFinder. *J. Antimicrob. Chemother.* <https://doi.org/10.1093/jac/dkaa390> (2020).
31. Goris, J. *et al.* DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* **57**, 81–91 (2007).
32. Kaas, R. S., Leekitcharoenphon, P., Aarestrup, F. M. & Lund, O. Solving the problem of comparing whole bacterial genomes across different sequencing platforms. *PLoS One* **9**, e104984 (2014).
33. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
34. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
35. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
36. Zhou, Z. *et al.* The Enterobase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny, and *Escherichia coli* core genomic diversity. *Genome Res.* **30**, 138–152 (2020).

Acknowledgements

This work was supported by the European Union-funded EFFORT project (www.effort-against-amr.eu) (FP7-KBBE-2013-7, Grant Agreement 613754). By the EJP-one health project Full-force (Grant Agreement No. 773830) and by The Novo Nordisk Foundation (NNF16OC0021856: Global Surveillance of Antimicrobial Resistance). In addition, whole-genome sequencing done at National Veterinary Research Institute was possible due to support of Department of Omics Analyses from the Ministry of Science and Higher Education (Grant No. 4477/E-180/SPUB/2018).

Author contributions

P.L. and M.H.K.J. wrote the main manuscript text and prepared figures. P.M. analysed Metagenomics data. B.M., M.S., K.W., G.M., A.H., K.T.V., A.B., EFFORT Consortium, M.Z., D.W., P.S., B.G.Z., M.S.M.B., J.A.W., D.J.H., D.M., F.M.A. sampled, isolated, sequenced the isolates including quality assurance of the genomic data. F.M.A. supervised the study. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-93970-7>.

Correspondence and requests for materials should be addressed to P.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

EFFORT Consortium

Haitske Graveland⁷, Alieda van Essen⁶, Antonio Battisti¹⁰, Andrea Caprioli¹⁰, Thomas Blaha³, Tine Hald¹, Hristo Daskalov¹¹, Helmut W. Saatkamp⁶, Katharina D. C. Stärk¹², Roosmarijn E. C. Luiken⁷, Liese Van Gompel⁷, Rasmus Borup Hansen¹³, Jeroen Dewulf¹⁴ & Ana Sofia Ribeiro Duarte¹

¹⁰Department of General Diagnostics, Istituto Zooprofilattico Sperimentale del Lazio E Della Toscana, National Reference Laboratory for Antimicrobial Resistance, Rome, Italy. ¹¹National Diagnostic Research Veterinary Institute, Sofia, Bulgaria. ¹²SAFOSO AG, Liebefeld, Switzerland. ¹³Intomics A/S, Diplomvej 377, Kongens Lyngby, Denmark. ¹⁴Department of Reproduction, Obstetrics and Herd Health, Ghent University, Merelbeke, Belgium.