A One Health Perspective on *Salmonella enterica* Serovar Infantis, an Emerging Human Multidrug-Resistant Pathogen

Jennifer Mattock,[1] Marie Anne Chattaway, Hassan Hartman, Timothy J. Dallman,[2] Anthony M. Smith, Karen Keddy, Liljana Petrovska,[3] Emma J. Manners,[4] Sanelisiwe T. Duze, Shannon Smouse, Nomsa Tau, Ruth Timme, Dave J. Baker, Alison E. Mather, John Wain, Gemma C. Langridge

Salmonella enterica serovar Infantis presents an ever-increasing threat to public health because of its spread throughout many countries and association with high levels of antimicrobial resistance (AMR). We analyzed whole-genome sequences of 5,284 *Salmonella* Infantis strains from 74 countries, isolated during 1989–2020 from a wide variety of human, animal, and food sources, to compare genetic phylogeny, AMR determinants, and plasmid presence. The global *Salmonella* Infantis population structure diverged into 3 clusters: a North American cluster, a European cluster, and a global cluster. The levels of AMR varied by *Salmonella* Infantis cluster and by isolation source; 73% of poultry isolates were multidrug resistant, compared with 35% of human isolates. This finding correlated with the presence of the pESI megaplasmid; 71% of poultry isolates contained pESI, compared with 32% of human isolates. This study provides key information for public health teams engaged in reducing the spread of this pathogen.

Nontyphoidal *Salmonella* infections place a large burden on public health; an estimated 79 million cases of foodborne nontyphoidal *Salmonella* infection occurred in 2010 (*1*). *Salmonella enterica* subspecies *enterica* serovar Infantis is becoming an increasingly prevalent serovar globally. A 167% increase in human infections was observed in the United States during 2001–2016 (*2*), and in European Union member states, Infantis is the predominant serovar isolated from broiler flocks and broiler meat, accounting for 56.7% of *Salmonella* isolates from broiler meat in 2018 (*3,4*). Higher levels have been observed in Japan, at 72.2% of isolates from ground chicken, and levels of 84% were seen in broilers in Ecuador (*5,6*).

Antimicrobial resistance (AMR) in *Salmonella* Infantis varies by location; in South Africa only 13.4% of 387 *Salmonella* Infantis isolates from humans had AMR (*7*). Conversely, in 2016 in European Union member states, 70% of *Salmonella* Infantis isolates from broiler meat were multidrug-resistant (MDR) (*8*). Of particular concern is the emergence of extended β-lactamases (ESBLs), such as the $bla_{CTX-M-65}$ gene, which has been reported in *Salmonella* Infantis from Ecuador, Peru, Switzerland, the United Kingdom, and the United States (*9–13*). The pESI megaplasmid has been found to be responsible for these high levels of AMR because it confers resistance to trimethoprim, streptomycin, sulfamethoxazole, and tetracycline; ESBLs have also been found to be carried by some pESI variants (*10,11,14*). Originally identified in Israel, pESI-like plasmids have since been reported in multiple countries (*14–19*).

Author affiliations: University of East Anglia, Norwich, UK (J. Mattock, E.J. Manners, A.E. Mather, J. Wain); UK Health Security Agency, London, UK (M.A. Chattaway, H. Hartman, T.J. Dallman); National Institute for Communicable Diseases, Johannesburg, South Africa (A.M. Smith, S. Smouse, N. Tau); University of Pretoria, Pretoria, South Africa (K. Keddy); Animal and Plant Health Agency, Addlestone, UK (L. Petrovska); University of the Witwatersrand, Johannesburg (S.T. Duze). US Food and Drug Administration, College Park, Maryland, USA (R. Timme); Quadram Institute Bioscience, Norwich (D.J. Baker, A.E. Mather, J. Wain, G.C. Langridge)

[1]Current affiliation: University of Edinburgh, Edinburgh, Scotland, UK.
[2]Current affiliation: Utrecht University, Utrecht, the Netherlands.
[3]Current affiliation: UK Health Security Agency, London, UK.
[4]Current affiliation: European Bioinformatics Institute, Cambridge, UK.

Salmonella Infantis has a polyphyletic population structure consisting of 2 eBurst Groups (eBG), eBG31 and eBG297, which differ by 5–7 multilocus sequence typing alleles (*16*). The dominant eBG (single locus variants around a central sequence type) globally is eBG31; eBG297 consisted of just 0.7% of *Salmonella* Infantis isolates in Enterobase on August 9, 2021 (*20*). However, higher levels (32%) of eBG297 have been reported in South Africa (*7*).

The population structure of *Salmonella* Infantis has been studied on a limited scale; whole-genome sequencing analysis of 100 *Salmonella* Infantis isolates from multiple continents and sources found no clustering by geographic location (*16*). *Salmonella* Infantis isolates were found to cluster by isolation source from human samples and chicken meat samples in Japan (*21*), by $bla_{CTX-M-65}$ presence in human and animal strains in the United States and Italy (*11*), and by pESI presence in human and poultry isolates from Switzerland (*10*).

Although MDR *Salmonella* Infantis is an emerging public health concern, no large-scale population structure study of this pathogen has been performed. Because eBG297 isolates have been analyzed in depth (*7*), our aim was to determine the global population structure of eBG31 from a One Health perspective, investigating whether population structure is associated with isolation source, location, MDR properties, or pESI presence.

Methods

Our eBG31 collection contained 5,284 isolates, sourced from the UK Health Security Agency (UKHSA), the National Institute for Communicable Diseases of South Africa, the Animal and Plant Health Agency, GenBank, and Enterobase (*20*) (Appendix 1, https://wwwnc.cdc.gov/EID/article/30/4/23-1031-App1.pdf). The collection contained strains isolated from 74 countries and spanned 4 decades and consisted of strains isolated during 1989–2020. The isolates were grouped into 8 sources: animal feed, human, environmental, food, other animals, poultry, poultry products, and unknown.

The whole-genome consensus FASTA sequences were grouped into clusters where all sequences in each cluster were <*n* single-nucleotide polymorphisms (SNPs) from another member. We generated a core SNP phylogeny of representatives of 25-SNP clusters; we identified clusters using fastbaps and used treedater to date the phylogeny (*22–25*). We used the ARIBA tool with the resfinder and plasmidfinder databases to screen for AMR determinants and plasmid presence; pESI was identified separately as described in Mattock et al. (*7,26–28*).

Ethical approval for the detection of gastrointestinal bacterial pathogens from fecal specimens, or the identification, characterization, and typing of cultures of gastrointestinal pathogens submitted to the Gastrointestinal Bacteria Reference Unit was not required because it is covered by UKHSA's surveillance mandate. Ethical approval for all laboratory-based surveillance and research activities was obtained from the Human Research Ethics Committee, University of the Witwatersrand, Johannesburg, South Africa (protocol reference nos. M060449 and M110499) by the Centre for Enteric Diseases, National Institute for Communicable Diseases. Ethical approval for characterization of the isolates was not required because of the surveillance mandate of Animal and Plant Health Agency.

Results

Demographics

We included isolates from a multitude of sources in the eBG31 collection. Most (60%, 3,150) of the 5,284 isolates were isolated from humans and associated with either noninvasive infections (samples from stool and urine), or invasive infection (samples from blood and cerebrospinal fluid). A further 6% (300) were from poultry and 13% (684) from poultry products, which included samples from poultry meat, eggs, and processed meals containing poultry meat. Isolates from other animals made up 6% (321) of the collection, 7% (390) were from food, 1% (74) from animal feed, and 5% (268) environmental, such as water, farm swab, and soil samples. A total of 97 isolates had no stated isolation source.

The number of isolates increased temporally until 2018; this increase was caused by isolates from public databases being included until February 2018 (Appendix 1 Figure 1). Only strains isolated by the UKHSA were included after that time. When categorized by continent, 54% (2,861) were from North America, 31% (1,642) from Europe, 6% (316) from Africa, 6% (312) from Asia, and 2% (128) from South America; origins were unknown for 0.47% (25). The United States contributed the largest number of isolates (n = 2,719) followed by the United Kingdom (n = 1,326) (Table; Figure 1).

Population Structure

At the sequence type (ST) level, most eBG31 isolates (99%, 5,205) belonged to ST32. The second most common was ST2283; 36 isolates belonged to this ST, all from Europe (17 from humans, 10 from other animals, 5 from food, and 4 from environmental samples). The

Table. Source group and country of isolates in One Health–focused analysis of *Salmonella enterica* serovar Infantis*

Country	Source group							
	Food	Environmental	Animal feed	Human	Poultry	Poultry products	Other animals	Unknown
Canada			4	21		1		
Cyprus								7
Denmark	18				1		36	11
Germany	16	7	8	1			15	1
Hungary				1				6
Japan		9		32	19	26		
Romania				2				6
South Africa		6		266				1
United Kingdom	75	52		1016	98	26	31	28
United States	234	185	57	1254	161	601	227	
Other	44	9	5	557	20	26	12	11
Unknown	3				1	4		26
Total	390	268	74	3150	300	684	321	97

*The number of isolates from each source group and the country of isolation were filtered to include countries that comprised >5% of ≥1 source group.

third most common was ST2146 with 26 isolates, all from North America (22 were from environmental samples, 3 from food, and 1 from a clinical sample). The 13 remaining STs were found in ≤3 isolates.

We identified 3 clusters of 250 SNPs; 1 contained just SRR8114924. For 50-SNP clusters, there were 408; for 25-SNP cluster, 1,288; for 10-SNP clusters, 2,876; and for 5-SNP clusters, 3,917. In a core SNP maximum-likelihood phylogeny of a member of each 25-SNP cluster, representing 5,283 eBG31 isolates (Figure 2; Appendix 1 Figure 2), Bayesian hierarchical clustering identified 3 clusters. Cluster A contained 348 sequences, representing 1,624 isolates (Figure 2, blue); cluster B had 831 sequences, representing 3,283 isolates (Figure 2, pink); and cluster C, which diverged from within cluster B, contained 109 sequences,

representing 376 isolates (Figure 2, purple). When annotated by ST, the phylogeny was dominated by ST32 (Appendix 1 Figure 3); 99% (1,269/1,288) of the 25-SNP clusters were exclusively ST32. The three 25-SNP clusters containing the ST2283 isolates clustered together in cluster C, and another 3 clusters comprising the ST3815 isolates clustered in cluster A.

In contrast to previous reports, a geographic signal was visible in the clustering of isolates in the phylogeny. Cluster A mainly consisted of North American isolates (Figure 3, panel A), of which 98% (1,180/1,203) were from the United States. Cluster B also contained a large percentage of North American isolates (50%, 1,657/3,238), but higher percentages of isolates from all other continents were observed. Conversely, cluster C consisted almost exclusively of
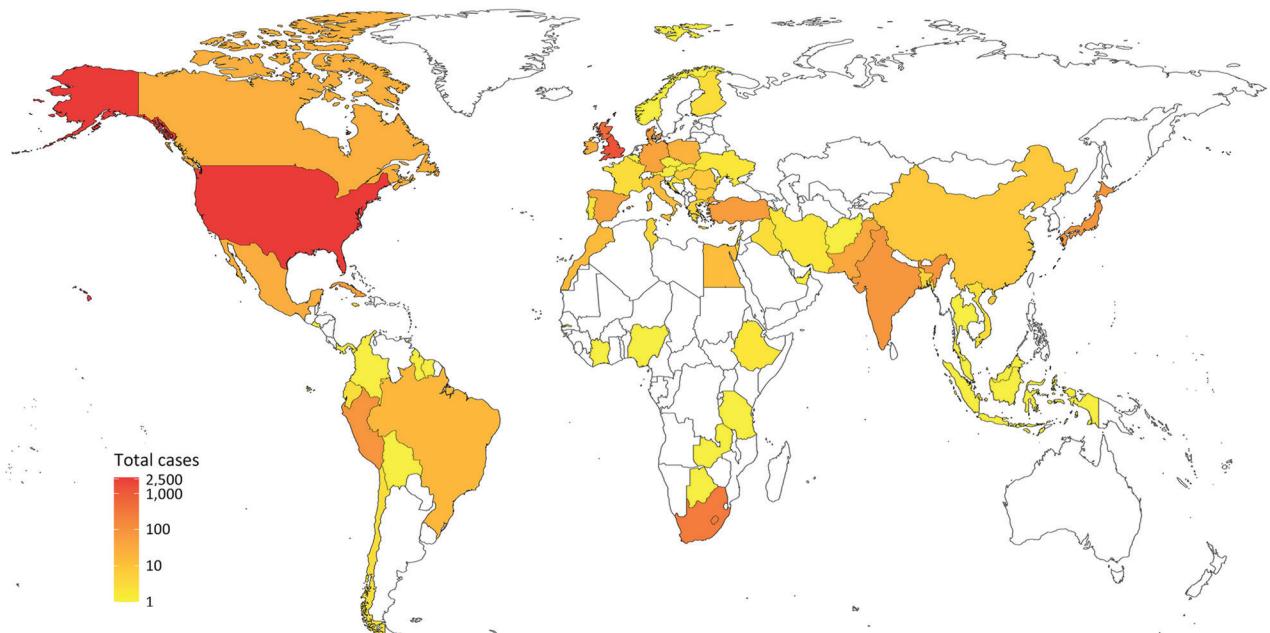

Figure 1. Heatmap indicating the number of isolates included in the dataset from each country in One Health–focused analysis.

European isolates, most of which were isolated in the United Kingdom (80%, 297/370).

The predominant isolation source in each of the clusters was humans (Figure 3, panel B). Isolates from poultry and poultry products were most often found in cluster B. However, isolates from other animals and animal feed made up a larger proportion of cluster A than the other clusters.

Minimal clustering by year was observed in the phylogeny (Appendix 1 Figure 4). The most common year range in each cluster was 2016–2020, representing 48% (785/1,624) of cluster A, 51% (1,664/3,238) of cluster B, and 67% (253/376) of cluster C. The earliest date of isolation varied; cluster B was the oldest with an isolate from 1989. Cluster A's oldest isolates were from 1996, and cluster C appeared more recently; its oldest isolate was from 2007. The time of the most recent common ancestor was calculated using relaxed clock dating of the phylogeny and estimated to be 1946; cluster A diverged in 1982 and cluster C in 1987.

We calculated the pairwise nucleotide distance between each whole-genome consensus FASTA to show the diversity within and between groups of isolates. Clusters B and C had the lowest median pairwise nucleotide distance of 191 (range 32–1,743);
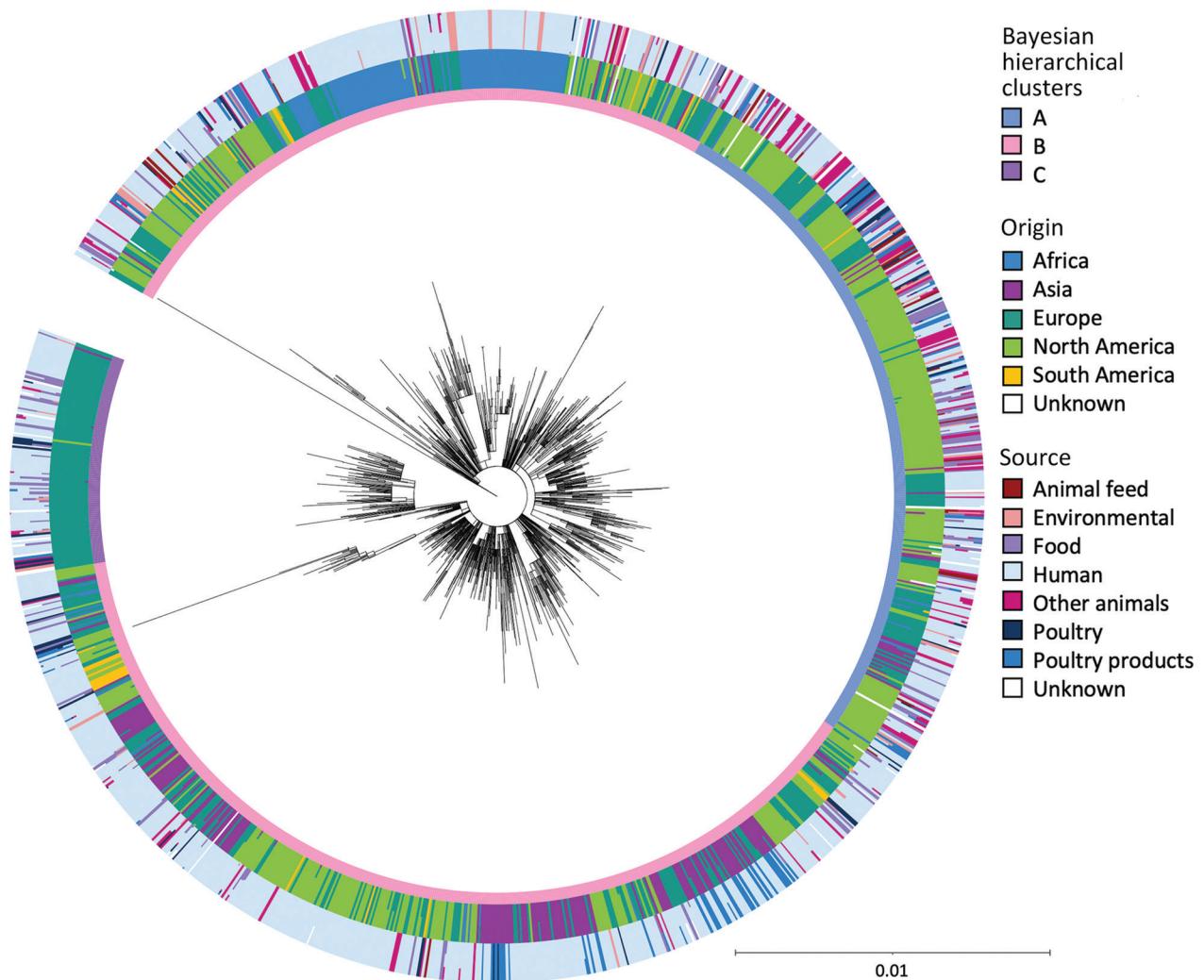

Figure 2. Core single-nucleotide polymorphism maximum-likelihood phylogeny of 1,288 representatives of 5,283 isolates in One Health–focused analysis of emerging multidrug-resistant pathogen *Salmonella enterica* serovar Infantis. The inner ring around the phylogeny is annotated with the Bayesian hierarchical clusters found by fastbaps. Cluster A consists of 348 representatives of 1,624 isolates; cluster B consists of 831 representatives of 3,283 isolates; and cluster C consists of 109 representatives of 376 isolates. The outer rings show the percentage of isolates in each 25-SNP cluster that were from each continent and source. Isolate origin was identified for Africa (n = 316), Asia (n = 312), Europe (n = 1,641), North America (n = 2,861), and South America (n = 128); the origin of the remaining isolates was unknown (n = 25). Sources were animal feed (n = 74), environmental (n = 268), food (n = 390), human (n = 3,149), other animals (n = 321), poultry (n = 300), poultry products (n = 684), and unknown (n = 97).

higher values were observed between clusters A and B (304, range 26–1,983) and clusters A and C (428, range 80–1,410). When comparing within isolation sources, strains from poultry products had the lowest median pairwise nucleotide distance at 145 (range 0–1,249) and human isolates the largest at 241 (range 0–2,060). The median pairwise nucleotide distance between human isolates and other sources ranged from 215 for environmental isolates (range 1–2,004) to 259 for poultry isolates (range 0–2,039). Poultry isolates had a similar median nucleotide distance to environmental isolates (251, range 4–1,822); a larger distance was observed between poultry isolates and those from other animals (310, range 4–1,817). The largest median pairwise nucleotide distance between source groups was poultry and animal feed at 334 (range 5–1,614). Low median pairwise nucleotide distances were observed within isolates from South America (44, range 0–920) and North America (159, range 0–1,883); the largest distance within isolates from a continent was Africa at 492 (range 0–1,885). Isolates from Africa also had the largest median pairwise nucleotide distances of all continents; distance was 395 (range 15–2,059) with North America and 422 (range 23–1,639) with South America.

AMR in *Salmonella* Infantis

In this collection, 44% (2,327/5,284) of the isolates contained ≥1 AMR gene; most of those (40%, 2,101/5,284) were MDR. Genes encoding AMR were identified in isolates throughout the phylogeny (Figure 4); 46.7% (602/1,288) of the 25-SNP clusters contained an isolate with AMR. Some 25-SNP clusters contained large numbers of isolates with AMR, such as 1 in cluster B that contained 734 isolates, of which 727 were MDR. A common resistance profile was visible in 25-SNP clusters across the phylogeny: 27.7% (357/1,288) contained an isolate with resistance to aminoglycosides, fluoroquinolones, sulphonamides, and tetracyclines (AFST). Of those, 56.9% (203/357) had an isolate with putative trimethoprim resistance, and 99.7% (356/357) contained a mutation in the quinolone-resistance determining region. The percentage of isolates with this resistance profile varied between the clusters: 1% (16/1,624) in cluster A, 43.5% (1,407/3,283) in cluster B, and 84% (316/376) in cluster C. The percentage of MDR isolates differed between the clusters; 7% (114/1,624) of isolates from cluster A were MDR, 50.5% (1,657/3,283) of isolates from cluster B were MDR, and 87.8% (330/376) from cluster C were MDR. More isolates were positive for ESBLs in cluster B (20.5%, 672/3,283) than in clusters A (0.2%, 4/1,624) and C (0%).
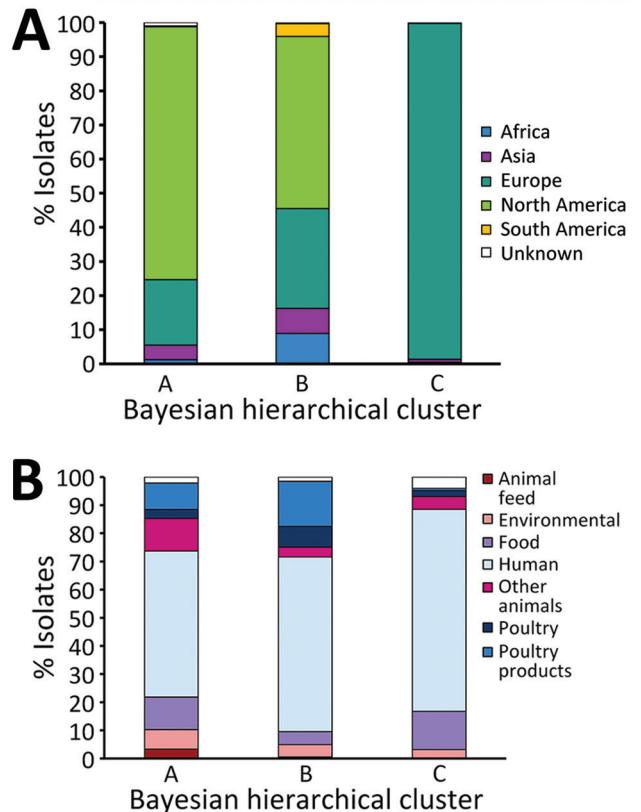

Figure 3. Source and continent composition of fastbaps clusters in study of One Health perspective of emerging multidrug-resistant pathogen *Salmonella enterica* serovar Infantis. A) Percentage of isolates from each continent in clusters A (n = 1,624), B (n = 3,283), and C (n = 376). B) Percentage of isolates from each source group in clusters A (n = 1,624), B (n = 3,283), and C (n = 376).

We observed variation in the distribution of AMR between isolation sources. Isolates from animal feed had the lowest amount of AMR; 4% (3/74) were predicted to have the AFST resistance profile. Higher levels of AMR were predicted in human isolates; 29% (913/3,150) had the AFST resistance profile, and 60% (552/913) of those also had trimethoprim resistance genes. Substantially more AMR was present in poultry and poultry product isolates: 61% (183/300) of poultry isolates and 64% (438/684) of poultry product isolates had the AFST resistance profile. β-lactam and chloramphenicol resistance was more common in poultry product isolates (43% [296/684] for β-lactam and 44% [302/684] for chloramphenicol) than in poultry isolates (21% [64/300] for β-lactam and 22% [65/300] for chloramphenicol). ESBLs were identified in 10% (312/3,150) of human isolates, 19% (58/300) of poultry isolates, and 40% (272/684) of isolates from poultry products. We observed multidrug resistance in 4% (3/74) of isolates from animal feed, 14% (38/268) of environmental isolates, 32%

(123/390) of food isolates, 35% (1,115/3,150) of human isolates, 21% (67/321) of isolates from other animals, 73% (218/300) of poultry isolates, and 73% (501/684) of poultry product isolates.

AMR profiles also varied by continent of isolation. The lowest levels of AMR were observed in isolates from Africa, of which 4% (14/316) had the AFST resistance profile, compared with 27% (763/2,861) of isolates from North America, 42% (692/1,642) of isolates from Europe, 55% (172/312) of isolates from Asia, and 76% (97/128) of isolates from South America. ESBLs were present in 0.3% (1/316) of isolates from Africa, 3% (9/312) of isolates from Asia, 4% (65/1,642) of isolates from Europe, 19% (531/2,861) of isolates from North America, and 55% (70/128) of isolates from South America. The percentage of MDR isolates was 20% (63/316) for Africa, 31% (891/2,861) for North America, 48% (793/1,642) for Europe, 80% (248/312) for Asia, and 81% (104/128) for South America.

The proportion of isolates with AMR fluctuated throughout the study period and trended upwards in the last 15 years of the collection period (Appendix 1 Figure 5). The earliest isolate in the collection (from 1989) was predicted to be resistant to 6 antimicrobial classes. AMR to aminoglycosides, sulphonamides, and tetracyclines were consistently the most common and appeared to follow a similar trend; after 2012, similar levels of AMR to fluoroquinolones were also present.

Plasmids in *Salmonella* Infantis

As observed with AMR, <50% (47%, 2502/5284) of *Salmonella* Infantis isolates contained a plasmid. Some of the most common types included IncA/C (n = 103), IncI1 (n = 251), and IncX1 (n = 65). As expected, pESI was the prevailing plasmid type, present in 36% (1,912/5,284) of *Salmonella* Infantis isolates. Low levels of IncA/C were observed in all isolation sources except animal feed; the highest level was just 4% (14/321) of other animal isolates, 2% of both human (64/3,150) and poultry product (15/684) isolates, and 0.6% (2/300) of poultry isolates. IncI1 was most
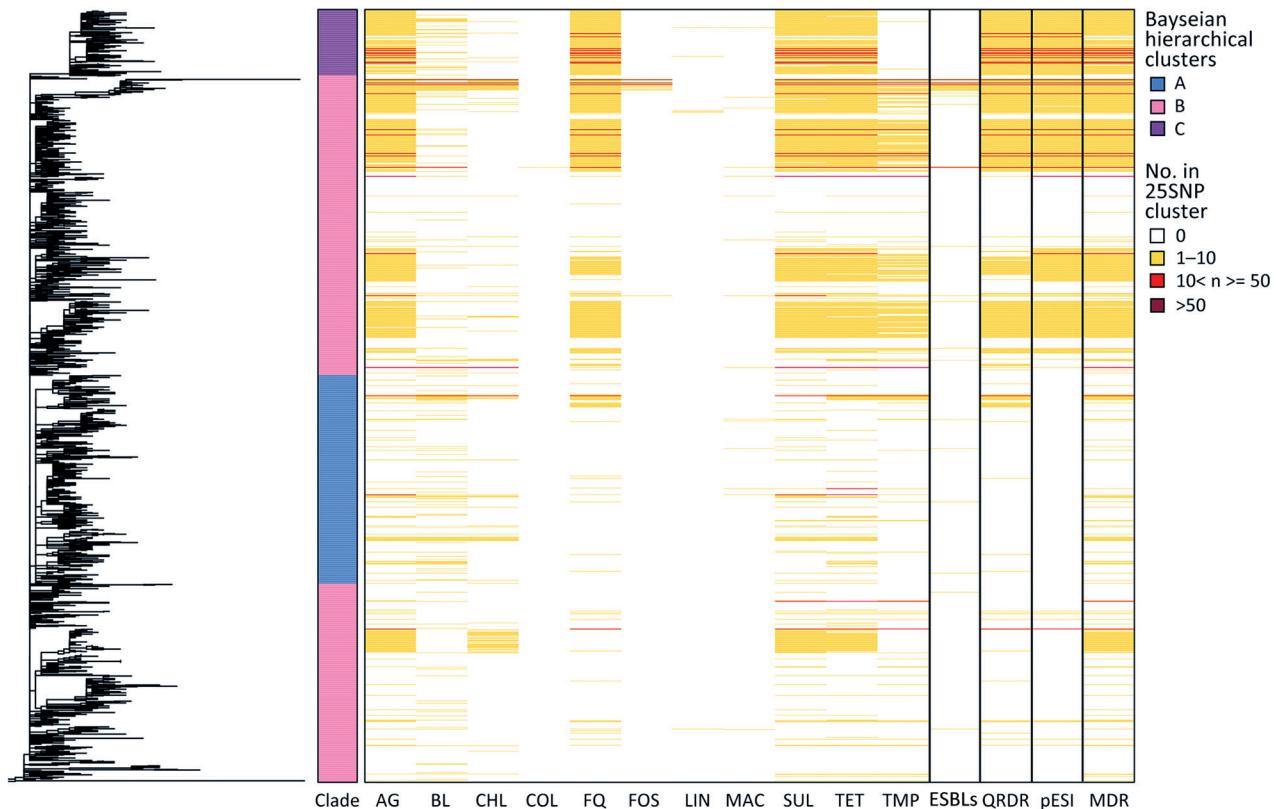

Figure 4. Phylogeny and heatmap of antimicrobial resistance and pESI in study of One Health perspective of emerging multidrug-resistant pathogen *Salmonella enterica* serovar Infantis. Heatmap shows the number of isolates in each 25 single-nucleotide polymorphism representative cluster (n = 1,288) of the eBG31 maximum-likelihood phylogeny with genes conferring resistance to antimicrobial drugs. Fastbaps clade and the number of isolates with MDR, ESBLs, mutations in the QRDR conferring resistance to fluoroquinolones and pESI presence are also shown. AG, aminoglycosides; BL, β-lactams; CHL, chloramphenicol; COL, colistin; ESBLs, extended β-lactamases; FQ, fluoroquinolones; FOS, fosfomycin; LIN, lincosamides; MAC, macrolides; MDR, multidrug-resistant; SUL, sulphonamides; TET, tetracyclines; TMP, trimethoprim.

common in other animal isolates at 9% (28/321) and IncX1 in human isolates (2%, 52/3,150). IncI1-positive isolates were found in all continents throughout the study period, mainly from humans (n = 184); IncX1 was observed in Asia, Europe, and North America.

The presence of pESI-like plasmids was observed in 71% (213/300) of poultry isolates, 71% (486/684) of isolates from poultry products, 32% (992/3,150) of human isolates, 4% (3/74) of animal feed isolates, 10% (31/321) of isolates from other animals, 31% (120/390) of food isolates, and 11% (30/268) of environmental isolates. Presence also varied by geographic location: the lowest percentage of pESI-positive isolates was from Africa at 4% (12/316), followed by North America at 28% (808/2,861), Europe at 47% (770/1,642), Asia at 71% (222/312), and South America at 77.3% (99/128). The earliest isolation of pESI in this collection was in 4 human isolates from Japan in 1999 (DRR022718, DRR022719, DRR022720, DRR022754). Although the common resistance profile AFST was distributed throughout the eBG31 phylogeny, cluster A lacked any pESI; most of cluster C contained pESI (99.7%, 375/376), and 46.8% (1,537/3,283) of cluster B contained pESI (Figure 4).

Discussion

In our large core SNP analysis of *Salmonella* Infantis, we determined that the global population structure of eBG31 consists of 3 clusters with varying isolation sources and levels of AMR. As observed previously, the dominant ST in eBG31 was ST32, comprising 99% of the isolates (*29–32*). The other STs were not observed in multiple continents, suggesting that those STs have emerged in specific areas but have not spread globally. A strong geographic signal was identified in the eBG31 phylogeny (Figure 2); we hypothesize that cluster B is an ancestor of the 2 other clusters, containing more genetic and geographic diversity in isolates from all continents, and we therefore designate this the global *Salmonella* Infantis cluster. Cluster A, estimated to have diverged from cluster B in 1982, mainly consisted of isolates from North America and hence is named the North American cluster. Cluster C, which diverged from cluster B in 1987, was dominated by isolates from Europe and is thus named the European cluster. This designation differs, perhaps because of our larger number of isolates, from Gymoese et al. (*16*), who found no geographic signal when examining isolates from 5 continents, and Alba et al. (*33*), who reported little clustering by location or source. Some clustering by country of isolation was described in Acar et al. (*34*); however, because that clustering was between isolates from the same region

in Turkey, the contribution to global clustering was not clear. Nucleotide distances relative to the reference showed that the African eBG31 isolates were both the most diverse and the most distant to isolates from other continents; our previous work identified that an increased proportion of isolates belonged to eBG297, and this study affirms that the African *Salmonella* Infantis population differs from that observed elsewhere (*7*).

Because most eBG31 isolates were from human sources, the phylogeny was dominated by this source. Cluster C in particular contained lower numbers of environmental, poultry, and poultry product isolates; that was possibly caused by bias in data sampling because the cluster contains strains from UKHSA that were isolated after the cutoff for inclusion from Enterobase. Most of those strains were from humans, as environmental sampling tends to be performed in association with an outbreak. Although cluster C contained strains isolated as early as 2007, it contained many of the newer strains isolated during 2018–2020; this group could represent an emerging clade of *Salmonella* Infantis. The nucleotide distance between source groups, relative to the reference, identified the least diversity in the poultry product isolates and the greatest diversity between poultry/poultry products and animal feed or other animal genomes. This finding could indicate a reduction of adaptation in poultry hosts and suggest that the different niches have encouraged adaptation; the reduced distance between poultry isolates could, however, be attributed to the large number of North American poultry isolates reducing the median (e.g., 519/734 of strains in the 25-SNP cluster with the representative SRR2537092 were from poultry, and 724 isolates in that cluster were from North America).

As described in many other studies, the *Salmonella* Infantis isolates in this project were associated with high levels of putative AMR (*8,18,35,36*). MDR genotypes were detected in 40% of the *Salmonella* Infantis isolates; their presence was notable in isolates from poultry and poultry products, in which 73% were MDR. In comparison, just 14% of environmental isolates, 21% of isolates from other animals, 32% of food isolates, and 35% of human isolates were MDR (Appendix 2 Table 1, https://wwwnc.cdc.gov/EID/article/30/4/23-1031-App2.xlsx); this finding could indicate that the source of human infection was the nonpoultry sources with similar levels of AMR. Although that hypothesis has been suggested in Slovenia, where most broiler isolates clustered separately from human isolates (*32*), many incidences of human outbreaks associated with poultry have been reported

(*21,37,38*). Both environmental and poultry sources could be contributing to human cases, but the pESI-positive strain circulating in the poultry industry could also have a selective disadvantage to causing infection in humans, leading to less frequent observation of those strains.

Levels of AMR varied by continent; the lowest levels were observed in Africa (20% of strains were MDR) and the highest were observed in South America (81%). The higher levels in South America concurs with other reports that found all but 1 isolate tested in Ecuador were MDR and observed multiple drug resistance profiles in *Salmonella* Infantis strains from Chile (*38,39*). AMR fluctuated temporally, increasing in the last 15 years of the collection period. The proportion of isolates with resistance to aminoglycosides, sulphonamides, and tetracyclines followed a similar trend, joined by fluoroquinolones after 2012. This trend could be attributed to pESI, which carries resistance genes for those antimicrobial drugs. Similarly, we noted an association between pESI presence and resistance to aminoglycosides, fluoroquinolones, sulphonamides, and tetracyclines; multidrug resistance; and mutations in the quinolone-resistance determining region (Figure 4). The pESI backbone has been confirmed to carry *aadA1*, *sul1*, and *tetA* (*14,19*), illustrating that pESI is a strong driver of AMR in the *Salmonella* Infantis population, which supports the suggestion of Alba et al. (*33*) that pESI acquisition could be the decisive factor in the spread of the serovar throughout Europe. Of note, cluster C, the clade we suspect is an emerging dominant strain in Europe, is dominated by AMR and pESI, but cluster A, the North American cluster with relatively low levels of MDR strains (7%), lacked any pESI-positive isolates. The pESI-like plasmid was present in 808 North American isolates in this dataset, but they belonged to either cluster B or C. This finding concurs with previous research that reported 2 clades within the US *Salmonella* Infantis population, 1 with and 1 without pESI, and suggests that 2 groups of *Salmonella* Infantis are circulating in North America: 1 associated with MDR strains and pESI, and the other endemic to North America and not carrying pESI (*40*). The presence of pESI-like plasmids has recently been identified in *Salmonella* serovars Agona, Muenchen, Schwarzengrund, and Senftenberg; the increased virulence of pESI-positive isolates and transmissibility of this plasmid within the *Salmonella* Infantis global population and to other *Salmonella* serovars is a grave public health concern (*41–43*).

In conclusion, most *Salmonella* Infantis isolates fall within eBG31, which consists of 3 clusters: a North American cluster (cluster A), a European cluster (cluster C), and an ancestral but still extant global cluster (cluster B). Isolates from Africa were genetically more diverse and distant from isolates from the other continents, further confirming previous work that identified a distinct population structure in *Salmonella* Infantis in South Africa. Using a One Health approach, we observed high levels of AMR in poultry and poultry products, highlighting the need to reduce the levels of this pathogen in poultry production premises and encouraging the development and use of a vaccine against *Salmonella* Infantis in poultry. Finally, pESI-like plasmids were shown to be a major driver for AMR in the global *Salmonella* Infantis population, posing a major threat to public health.

This article was preprinted at https://www.biorxiv.org/content/10.1101/2023.07.28.549231v1.

About the Author

Dr. Mattock is a bioinformatician at the Roslin Institute, University of Edinburgh, UK. Her research interests include the genetics, antimicrobial resistance, and epidemiology of gastrointestinal pathogens.

References

1. Havelaar AH, Kirk MD, Torgerson PR, Gibb HJ, Hald T, Lake RJ, et al.; World Health Organization Foodborne Disease Burden Epidemiology Reference Group. World Health Organization global estimates and regional comparisons of the burden of foodborne disease in 2010. PLoS Med. 2015;12:e1001923–1001923 https://doi.org/10.1371/journal.pmed.1001923
2. Centers for Disease Control and Prevention. National enteric disease surveillance: *Salmonella* annual report, 2016 [cited 2021 Aug 4]. https://www.cdc.gov/nationalsurveillance/pdfs/2016-Salmonella-report-508.pdf
3. European Food Safety Authority; European Centre for Disease Prevention and Control. The European Union One Health 2019 zoonoses report. EFSA J. 2021;19:e06406.
4. European Food Safety Authority; European Centre for Disease Prevention and Control. The European Union One Health 2018 Zoonoses Report. EFSA J. 2019;17:e05926.
5. Hara-Kudo Y, Konuma H, Kamata Y, Miyahara M, Takatori K, Onoue Y, et al. Prevalence of the main food-borne pathogens in retail food under the national food surveillance system in Japan. Food Addit Contam Part A Chem Anal Control Expo Risk Assess. 2013;30:1450–8. https://doi.org/10.1080/19440049.2012.745097
6. Vinueza-Burgos C, Cevallos M, Ron-Garrido L, Bertrand S, De Zutter L. Prevalence and diversity of *Salmonella* serotypes in Ecuadorian broilers at slaughter age. PLoS One. 2016;11:e0159567. https://doi.org/10.1371/journal.pone.0159567
7. Mattock J, Smith AM, Keddy KH, Manners EJ, Duze ST, Smouse S, et al. Genetic characterization of *Salmonella* Infantis from South Africa, 2004–2016. Access Microbiol. 2022;4:acmi000371. https://doi.org/10.1099/acmi.0.000371
8. European Food Safety Authority; European Centre for Disease Prevention and Control. The European Union summary report on antimicrobial resistance in zoonotic and indicator bacteria from humans, animals and food in 2016. EFSA J. 2018;16:e05182.
9. Granda A, Riveros M, Karen SM, Corujo A, Reyes I, Ruiz J, et al. Presence of extended-spectrum β -lactamase, CTX-M-65 in *Salmonella enterica* serovar Infantis isolated from children with diarrhea in Lima, Peru. J Pediatr Infect Dis. 2019;14:194–200. https://doi.org/10.1055/s-0039-1685502
10. Hindermann D, Gopinath G, Chase H, Negrete F, Althaus D, Zurfluh K, et al. *Salmonella enterica* serovar Infantis from food and human infections, Switzerland, 2010–2015: poultry-related multidrug resistant clones and an emerging ESBL producing clonal lineage. Front Microbiol. 2017;8:1322. https://doi.org/10.3389/fmicb.2017.01322
11. Tate H, Folster JP, Hsu C-H, Chen J, Hoffmann M, Li C, et al. Comparative analysis of extended-spectrum-β-Lactamase CTX-M-65-producing *Salmonella enterica* serovar Infantis isolates from humans, food animals, and retail chickens in the United States. Antimicrob Agents Chemother. 2017;61:e00488-17. https://doi.org/10.1128/AAC.00488-17
12. Burke L, Hopkins KL, Meunier D, de Pinna E, Fitzgerald-Hughes D, Humphreys H, et al. Resistance to third-generation cephalosporins in human non-typhoidal *Salmonella enterica* isolates from England and Wales, 2010–12. J Antimicrob Chemother. 2014;69:977–81. https://doi.org/10.1093/jac/dkt469
13. Cartelle Gestal M, Zurita J, Paz Y Mino A, Ortega-Paredes D, Alcocer I. Characterization of a small outbreak of *Salmonella enterica* serovar Infantis that harbour CTX-M-65 in Ecuador. Braz J Infect Dis. 2016;20:406–7. https://doi.org/10.1016/j.bjid.2016.03.007
14. Aviv G, Tsyba K, Steck N, Salmon-Divon M, Cornelius A, Rahav G, et al. A unique megaplasmid contributes to stress tolerance and pathogenicity of an emergent *Salmonella enterica* serovar Infantis strain. Environ Microbiol. 2014;16:977–94. https://doi.org/10.1111/1462-2920.12351
15. Franco A, Leekitcharoenphon P, Feltrin F, Alba P, Cordaro G, Iurescia M, *et al*. Emergence of a clonal lineage of multidrug-resistant ESBL-producing *Salmonella* Infantis transmitted from broilers and broiler meat to humans in Italy between 2011 and 2014. PLoS One. 2015;10:e0144802.
16. Gymoese P, Kiil K, Torpdahl M, Østerlund MT, Sørensen G, Olsen JE, et al. WGS based study of the population structure of *Salmonella enterica* serovar Infantis. BMC Genomics. 2019;20:870. https://doi.org/10.1186/s12864-019-6260-6
17. Iriarte A, Giner-Lamia J, Silva C, Betancor L, Astocondor L, Cestero JJ, et al.; SalmoIber CYTED Network. Draft genome sequence of *Salmonella enterica* subsp. *enterica* serovar Infantis strain SPE101, isolated from a chronic human infection. Genome Announc. 2017;5:e00679–17. https://doi.org/10.1128/genomeA.00679-17
18. Szmolka A, Szabó M, Kiss J, Pászti J, Adrián E, Olasz F, et al. Molecular epidemiology of the endemic multiresistance plasmid pSI54/04 of *Salmonella* Infantis in broiler and human population in Hungary. Food Microbiol. 2018;71:25–31. https://doi.org/10.1016/j.fm.2017.03.011
19. Lee WWY, Mattock J, Greig DR, Langridge GC, Baker D, Bloomfield S, *et al*. Characterization of a pESI-like plasmid and analysis of multidrug-resistant *Salmonella enterica* Infantis isolates in England and Wales. Microb Genom. 2021;7:000658.
20. Alikhan N-F, Zhou Z, Sergeant MJ, Achtman M. A genomic overview of the population structure of *Salmonella*. PLoS Genet. 2018;14:e1007261. https://doi.org/10.1371/journal.pgen.1007261
21. Yokoyama E, Murakami K, Shiwa Y, Ishige T, Ando N, Kikuchi T, et al. Phylogenetic and population genetic analysis of *Salmonella enterica* subsp. enterica serovar Infantis strains isolated in Japan using whole genome sequence data. Infect Genet Evol. 2014;27:62–8. https://doi.org/10.1016/j.meegid.2014.06.012
22. Seemann T. Snippy: rapid haploid variant calling [cited 2021 Feb 8]. https://github.com/tseemann/snippy
23. Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. Fast hierarchical Bayesian analysis of population structure. Nucleic Acids Res. 2019;47:5539–49. https://doi.org/10.1093/nar/gkz361
24. Volz EM, Frost SDW. Scalable relaxed clock phylogenetic dating. Virus Evol. 2017;3:vex025. https://doi.org/10.1093/ve/vex025
25. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3. https://doi.org/10.1093/bioinformatics/btu033
26. Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J, Keane JA, et al. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. Microb Genom. 2017;3:e000131.

27. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother. 2014;58:3895–903. https://doi.org/10.1128/AAC.02412-14

28. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. 2012;67:2640–4. https://doi.org/10.1093/jac/dks261

29. Kürekci C, Sahin S, Iwan E, Kwit R, Bomba A, Wasyl D. Whole-genome sequence analysis of *Salmonella* Infantis isolated from raw chicken meat samples and insights into pESI-like megaplasmid. Int J Food Microbiol. 2021;337:108956. https://doi.org/10.1016/j.ijfoodmicro.2020.108956

30. Mughini-Gras L, van Hoek AHAM, Cuperus T, Dam-Deisz C, van Overbeek W, van den Beld M, et al. Prevalence, risk factors and genetic traits of *Salmonella* Infantis in Dutch broiler flocks. Vet Microbiol. 2021; 258:109120. https://doi.org/10.1016/j.vetmic.2021.109120

31. Vilela FP, Rodrigues DDP, Allard MW, Falcão JP. Genomic characterization and antimicrobial resistance profiles of *Salmonella enterica* serovar Infantis isolated from food, humans and veterinary-related sources in Brazil. J Appl Microbiol. 2022;132:3327–42. https://doi.org/10.1111/jam.15430

32. Papić B, Kušar D, Mićunović J, Pirš M, Ocepek M, Avberšek J. Clonal spread of pESI-positive multidrug-resistant ST32 *Salmonella enterica* serovar Infantis isolates among broilers and humans in Slovenia. Microbiol Spectr. 2022;10:e0248122. https://doi.org/10.1128/spectrum.02481-22

33. Alba P, Leekitcharoenphon P, Carfora V, Amoruso R, Cordaro G, Di Matteo P, et al. Molecular epidemiology of *Salmonella* Infantis in Europe: insights into the success of the bacterial host and its parasitic pESI-like megaplasmid. Microb Genom. 2020;6:e000365.

34. Acar S, Bulut E, Stasiewicz MJ, Soyer Y. Genome analysis of antimicrobial resistance, virulence, and plasmid presence in Turkish *Salmonella* serovar Infantis isolates. Int J Food Microbiol. 2019;307:108275. https://doi.org/10.1016/j.ijfoodmicro.2019.108275

35. Askari Badouei M, Vaezi H, Nemati A, Ghorbanyoon E, Firoozeh F, Jajarmi M, et al. High prevalence of clonally related multiple resistant *Salmonella* Infantis carrying class 1 integrons in broiler farms. Vet Ital. 2021;57.

36. Di Marcantonio L, Romantini R, Marotta F, Chiaverini A, Zilli K, Abass A, et al. The current landscape of antibiotic resistance of *Salmonella* Infantis in Italy: the expansion of extended-spectrum beta-lactamase producers on a local scale. Front Microbiol. 2022;13:812481. https://doi.org/10.3389/fmicb.2022.812481

37. Martínez-Puchol S, Riveros M, Ruidias K, Granda A, Ruiz-Roldán L, Zapata-Cachay C, et al. Dissemination of a multidrug resistant CTX-M-65 producer *Salmonella enterica* serovar Infantis clone between marketed chicken meat and children. Int J Food Microbiol. 2021;344:109109. https://doi.org/10.1016/j.ijfoodmicro.2021.109109

38. Mejía L, Medina JL, Bayas R, Salazar CS, Villavicencio F, Zapata S, et al. Genomic epidemiology of *Salmonella* Infantis in Ecuador: from poultry farms to human infections. Front Vet Sci. 2020;7:547891. https://doi.org/10.3389/fvets.2020.547891

39. Pardo-Esté C, Lorca D, Castro-Severyn J, Krüger G, Alvarez-Thon L, Zepeda P, et al. Genetic characterization of *Salmonella* Infantis with multiple drug resistance profiles isolated from a poultry-farm in Chile. Microorganisms. 2021;9:2370. https://doi.org/10.3390/microorganisms9112370

40. McMillan EA, Weinroth MD, Frye JG. Increased prevalence of *Salmonella* Infantis isolated from raw chicken and turkey products in the United States is due to a single clonal lineage carrying the pESI plasmid. Microorganisms. 2022;10:1478. https://doi.org/10.3390/microorganisms10071478

41. Drauch V, Kornschober C, Palmieri N, Hess M, Hess C. Infection dynamics of *Salmonella* Infantis strains displaying different genetic backgrounds—with or without pESI-like plasmid—vary considerably. Emerg Microbes Infect. 2021;10:1471–80. https://doi.org/10.1080/22221751.2021.1951124

42. Dos Santos AMP, Panzenhagen P, Ferrari RG, Conte-Junior CA. Large-scale genomic analysis reveals the pESI-like megaplasmid presence in *Salmonella* Agona, Muenchen, Schwarzengrund, and Senftenberg. Food Microbiol. 2022;108:104112. https://doi.org/10.1016/j.fm.2022.104112

43. Cohen E, Kriger O, Amit S, Davidovich M, Rahav G, Gal-Mor O. The emergence of a multidrug resistant *Salmonella* Muenchen in Israel is associated with horizontal acquisition of the epidemic pESI plasmid. Clin Microbiol Infect. 2022;28:1499.e7–14. https://doi.org/10.1016/j.cmi.2022.05.029

Address for correspondence: Jennifer Mattock, Roslin Institute, The University of Edinburgh, Easter Bush Campus, Midlothian, EH25 9RG, UK; email: jennifer.mattock@roslin.ed.ac.uk

A One Health Perspective on *Salmonella enterica* Serovar Infantis, the Emerging Human Multidrug-Resistant Pathogen

Appendix 1

Methods

Isolate selection

Enterobase was searched on the 19th of February 2018, and all isolates belonging to eBG31 with sequence data available in the short read archive (SRA) were included; due to the computational cost of including more sequences we chose to not include isolates after this cutoff (*1*). All UK Health Security Agency (UKHSA) eBG31 sequence data reported by 31st of December 2020 was downloaded from the *UKHSA* Pathogens *BioProject* at NCBI (PRJNA248792). Sequence data for 62 Animal and Plant Health Agency (APHA) *S*. Infantis isolates were downloaded from the SRA on 28th of May 2019.

DNA of *S*. Infantis isolates was shared for sequencing by the APHA, UKHSA and National Institute of Communicable Diseases (NICD), South Africa. The APHA isolates included had been collected for surveillance up until March 2018. Strains, isolated between 2000 and 2014, were selected from the UKHSA culture store; ensuring that there was an even distribution by time and, if applicable, continent of travel. Strains isolated between 2004 and 2016 were selected from the NICD culture store for sequencing.

The metadata available was used to stratify the isolates into the following source groups: human, food, poultry, poultry products, animal feed, other animals, environmental and unknown (Appendix 2 Table 1). All clinical isolates and those with metadata on patient age or gender were

classified as human. Isolates from chickens, ducks, turkeys and quail were categorised as poultry. All strains from poultry products, such as eggs and chicken meat, were grouped together. Isolates from all other animals were grouped into 'other animals' and non-poultry animal products and other food samples were grouped into 'food'. All samples from animal feed were grouped and farm swabs, soil, water, air and sewage were designated environmental. Isolates missing source information to assign to these groups were classed as unknown. For the UKHSA isolates with travel history information, the continent of travel was used to designate the continent of isolate origin. The world map plot was generated with R (v.4.1.0) and the package ggplot2 (v.3.3.6) (*2,3*).

Whole Genome Sequencing

DNA from the NICD isolates was extracted using QIAamp DNA Mini Kit (Qiagen) and sequenced as described in Mattock *et al*., 2021 (*4*); libraries were prepared using either Illumina Nextera XT library preparation kits or using a custom library preparation method described in Rasheed *et al*., 2020 (*5*). Twelve NICD isolates not used in Mattock *et al*., 2021 are included here and a library was prepared using one of the aforementioned methods and sequenced on an Illumina NextSeq 500.

The DNA of the UKHSA samples was extracted using the Qiasymphony DSP DNA Midi Kit (Qiagen), following a protocol described in Nair *et al*., 2016 (*6*). The historical UKHSA samples, described in Lee *et al*., 2021, were also sequenced using either Illumina Nextera XT or the custom library preparation methods (*7*). Two additional historical UKHSA samples were included and sequenced using the custom library preparation method. The contemporary UKHSA samples were prepared using the Illumina Nextera XT protocol and sequenced on an Illumina HiSeq 2500 instrument.

DNA from the APHA isolates was extracted using the MagMAX CORE Nucleic Acid Purification Kit (ThermoFisher) with a KingFisher Flex Purification System (ThermoFisher). The DNA was then sequenced using the custom library preparation method and an Illumina NextSeq 500.

Phylogenetic analysis

All the sequence data generated by UKHSA were trimmed using Trimmomatic (v0.27) with leading and trailing at <Q30 (*8*). All other sequence data were trimmed with Trimmomatic

(v0.36) and the same options. Sequence type and eBG were determined for each isolate using Metric-Oriented Sequence Typer (MOST) (v.1.0) with a UKHSA *Salmonella* database (*9,10*). The quality of borderline sequence typing was assessed using Tablet (*11*).

Using the Cloud Infrastructure for Big Data Microbial Bioinformatics (*12*), each sample was mapped and variant called against the eBG31 reference CP070301 (*7*), using Snippy (v.4.6.0) with the options minfrac 0.9 and 30 (*13*). Due to the large number of sequences included, it was not computationally possible to produce a phylogeny including all of the isolates. Therefore, the SNP distance between all of the consensus FASTAs produced by Snippy was calculated using snp-dists (v.0.7) (https://github.com/tseemann/snp-dists) with the -m option to output in the molten format. Using MCL (v14–137) and the abc option, Markov clustering was performed on the pairwise SNP distance matrix with the following SNP distance thresholds: 0-SNPs, 5-SNPs, 10-SNPs, 25-SNPs, 50-SNPs, 100-SNPs and 250-SNPs (*10,14*). This resulted in clusters where each cluster member was less than 'n' SNPs from another member. The smallest SNP threshold that it was possible to generate a phylogeny with was 25-SNPs; the first isolate in each 25-SNP cluster was chosen to be the representative for that cluster.

The eBG31 reference genome, CP070301, was screened for prophages using PHASTER (*7,15*). Four complete prophages were identified and masked by Snippy when producing an alignment of the 25-SNP cluster representatives. The whole genome alignment was used by Gubbins (v.2.4.1) to identify recombination which was then removed during core-SNP alignment generation (*16*). A core SNP maximum likelihood phylogeny of the 25-SNP cluster representatives, excluding the reference isolate, was constructed using RaxML (v.8.2.12), rooted to its most ancestral node and annotated using iToL (*17,18*). Hierarchical Bayesian clustering within the phylogeny was identified using fastbaps (v.1.0.5) with ape (v.5.3) and R (v.3.4.1), with the optimised symmetric prior and k.init at 5 (*2,19,20*). Pairwise SNP distances were calculated using the snp-dists output, with self against self comparisons excluded.

The treedater relaxed clock test was performed to determine support for using a strict clock; the strict clock was deemed the best approach (*21*). Treedater (v.0.5.0) was used with a strict clock and the dates ranges of the isolates in each 25-SNP cluster, excluding leaves with an unknown isolation date, to date the phylogeny. Phangorn (v.2.10.0) was used to identify the nodes where the clades diverged (*22*).

AMR and plasmid determination

ARIBA (v.2.10.1) was used with the ResFinder and Plasmidfinder databases (downloaded 17th January, 2022) to identify AMR determinants and plasmids (*23–25*). A database of *gyrA*, *gyrB*, *parC* and *parE*, from the *S.* Typhimurium LT2 reference, was used with ARIBA to identify mutations within the Quinolone Resistance Determining Regions (QRDR). MDR, resistance to three or more antimicrobial classes, was calculated using AMR genes identified by ARIBA and mutations in QRDRs. As aminoglycoside resistance in *Salmonella* is rarely conferred by *aac(6')-Iaa* it was excluded from all calculations (*26*). pESI presence was determined by mapping each sequence against a pseudomolecule of the eBG31 reference (CP070301) and pESI contigs (ASRF01000099-ASRF01000108) with SMALT (v.0.7.6) and calculating coverage with Samtools (v.1.5) (*27,28*). Heatmaps were generated using phytools (v.0.6) and data.table (v.1.11.8) (*29,30*).

The proportion of isolates containing each AMR determinant and plasmid from each metadata group was calculated in R (v.4.1.0) with the packages janitor (v.2.1.0), tidyr (v.1.1.3) and magrittr (v.2.0.1) (*31–33*). The heatmap was created in R with ape and gplots (v3.1.3) (*34*).

References

1. Alikhan N-F, Zhou Z, Sergeant MJ, Achtman M. A genomic overview of the population structure of *Salmonella.* PLoS Genet. 2018;14:e1007261. PubMed https://doi.org/10.1371/journal.pgen.1007261

2. The R Foundation. The R Project for Statistical Computing [cited 2021 May 12]. https://www.r-project.org/

3. ggplot2: elegant graphics for data analysis [cited 2022 May 10]. https://ggplot2.tidyverse.org

4. Mattock J, Smith AM, Keddy KH, Manners EJ, Duze ST, Smouse S, et al. Genetic characterization of *Salmonella* Infantis from South Africa, 2004–2016. Access Microbiol. 2022;4:acmi000371. PubMed https://doi.org/10.1099/acmi.0.000371

5. Rasheed F, Saeed M, Alikhan N-F, Baker D, Khurshid M, Ainsworth EV, et al. Emergence of resistance to fluoroquinolones and third-generation cephalosporins in *Salmonella* Typhi in Lahore, Pakistan. Microorganisms. 2020;8:1336. PubMed https://doi.org/10.3390/microorganisms8091336

6. Nair S, Ashton P, Doumith M, Connell S, Painset A, Mwaigwisya S, et al. WGS for surveillance of antimicrobial resistance: a pilot study to detect the prevalence and mechanism of resistance to azithromycin in a UK population of non-typhoidal *Salmonella*. J Antimicrob Chemother. 2016;71:3400–8. PubMed https://doi.org/10.1093/jac/dkw318

7. Lee WWY, Mattock J, Greig DR, Langridge GC, Baker D, Bloomfield S, *et al*. Characterization of a pESI-like plasmid and analysis of multidrug-resistant *Salmonella enterica* Infantis isolates in England and Wales. Microb Genom. 2021;10:000658. **PMID 34647862**

8. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20. PubMed https://doi.org/10.1093/bioinformatics/btu170

9. Tewolde R, Dallman T, Schaefer U, Sheppard CL, Ashton P, Pichon B, et al. MOST: a modified MLST typing tool based on short read sequencing. PeerJ. 2016;4:e2308. PubMed https://doi.org/10.7717/peerj.2308

10. Chattaway MA, Dallman TJ, Larkin L, Nair S, McCormick J, Mikhail A, et al. The transformation of reference microbiology methods and surveillance for *Salmonella* with the use of whole genome sequencing in England and Wales. Front Public Health. 2019;7:317. PubMed https://doi.org/10.3389/fpubh.2019.00317

11. Milne I, Stephen G, Bayer M, Cock PJA, Pritchard L, Cardle L, et al. Using Tablet for visual exploration of second-generation sequencing data. Brief Bioinform. 2013;14:193–202. PubMed https://doi.org/10.1093/bib/bbs012

12. Connor TR, Loman NJ, Thompson S, Smith A, Southgate J, Poplawski R, *et al*. CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community. Microb Genom. 2016;2:e000086. **PMID 28785418**

13. Seemann T. Snippy: rapid haploid variant calling. 2015 [cited 2021 Feb 8]. https://github.com/tseemann/snippy

14. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 2002;30:1575–84. PubMed https://doi.org/10.1093/nar/30.7.1575

15. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 2016;44(W1):W16-21. PubMed https://doi.org/10.1093/nar/gkw387

16. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. 2015;43:e15–15. PubMed https://doi.org/10.1093/nar/gku1196

17. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. 2014;30(9):1312–3. https://doi.org/10.1093/bioinformatics/btu033 **PMID 24451623**

18. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016;44(W1):W242-5. PubMed https://doi.org/10.1093/nar/gkw290

19. Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. Fast hierarchical Bayesian analysis of population structure. Nucleic Acids Res. 2019;47:5539–49. PubMed https://doi.org/10.1093/nar/gkz361

20. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in {R}. Bioinformatics. 2018. PubMed

21. Volz EM, Frost SDW. Scalable relaxed clock phylogenetic dating. Virus Evol. 2017;3:vex025. https://doi.org/10.1093/ve/vex025

22. Schliep KP. phangorn: phylogenetic analysis in R. Bioinformatics. 2011;27:592–3. PubMed https://doi.org/10.1093/bioinformatics/btq706

23. Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J, Keane JA, et al. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. Microb Genom. 2017;3:e000131. **PMID 29177089**

24. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother. 2014;58:3895–903. PubMed https://doi.org/10.1128/AAC.02412-14

25. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. 2012;67:2640–4. PubMed https://doi.org/10.1093/jac/dks261

26. Neuert S, Nair S, Day MR, Doumith M, Ashton PM, Mellor KC, et al. Prediction of phenotypic antimicrobial resistance profiles from whole genome sequences of non-typhoidal *Salmonella enterica.* Front Microbiol. 2018;9:592. PubMed https://doi.org/10.3389/fmicb.2018.00592

27. Wellcome Sanger Institute. SMALT [cited 2021 May 31].
    https://www.sanger.ac.uk/science/tools/smalt-0

28. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al.; 1000 Genome Project Data
    Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics.
    2009;25:2078–9. PubMed https://doi.org/10.1093/bioinformatics/btp352

29. Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things). Methods
    Ecol Evol. 2012;3:217–23. https://doi.org/10.1111/j.2041-210X.2011.00169.x

30. Dowle M, Srinivasan A. data.table: Extension of 'data.frame.' R package version 1.11.8. 2018 [cited
    2021 May 6]. https://cran.r-project.org/web/packages/data.table/data.table.pdf

31. Firke S. janitor: simple tools for examining and cleaning dirty data [cited 2021 Oct 06]. https://cran.r-
    project.org/web/packages/janitor/index.html

32. Wickham H, Vaughan D, Girlich M. tidyr. 2023 [cited 2021 Oct 9]. https://tidyr.tidyverse.org

33. Bache SM, Wickham H. magrittr. 2022 [cited 2021 May 5]. https://magrittr.tidyverse.org

34. Warnes G, Bolker B, Bonebakker L, Gentleman R, Huber W. gplots. 2022 [cited 2021 May 5].
   https://github.com/talgalili/gplots
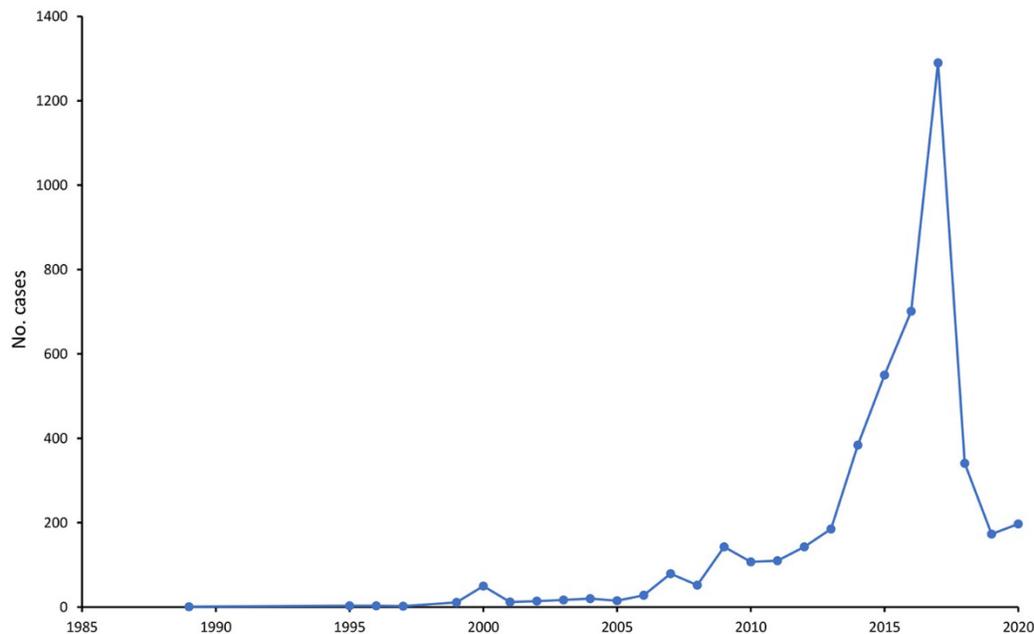
Appendix 1 Figure 1. Frequency of *S*. Infantis each year. Number of *S*. Infantis strains included in the dataset from 1989 to 2020.
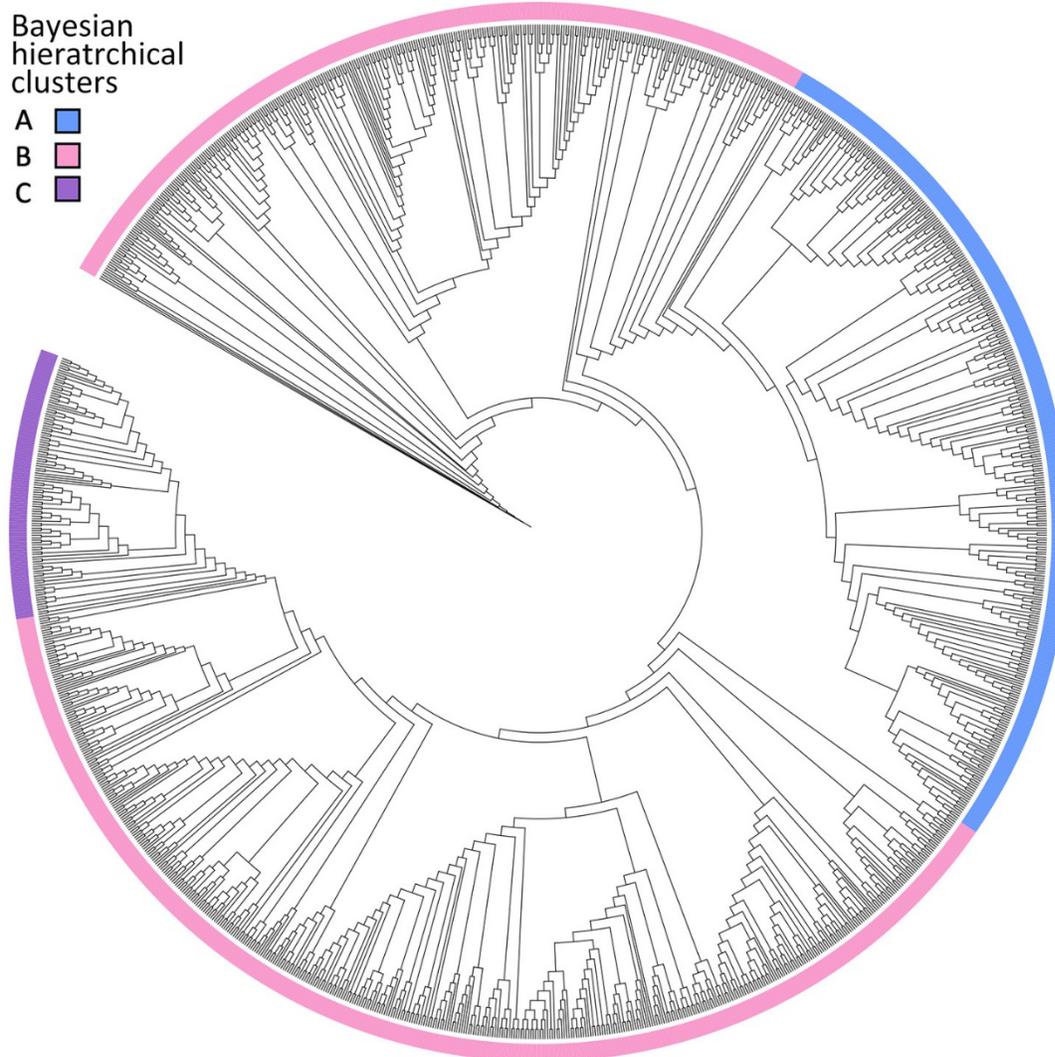
Appendix 1 Figure 2. Maximum likelihood cladogram of *S*. Infantis. Core SNP maximum likelihood cladogram of 1288 representatives of 5283 *S*. Infantis isolates. The ring around the cladogram is annotated with the Bayesian hierarchical clusters found by fastbaps. Bayesian hierarchical clusters: A, 348 representatives of 1624 isolates; B, 831 representatives of 3283 isolates and C 109 representatives of 376 isolates.
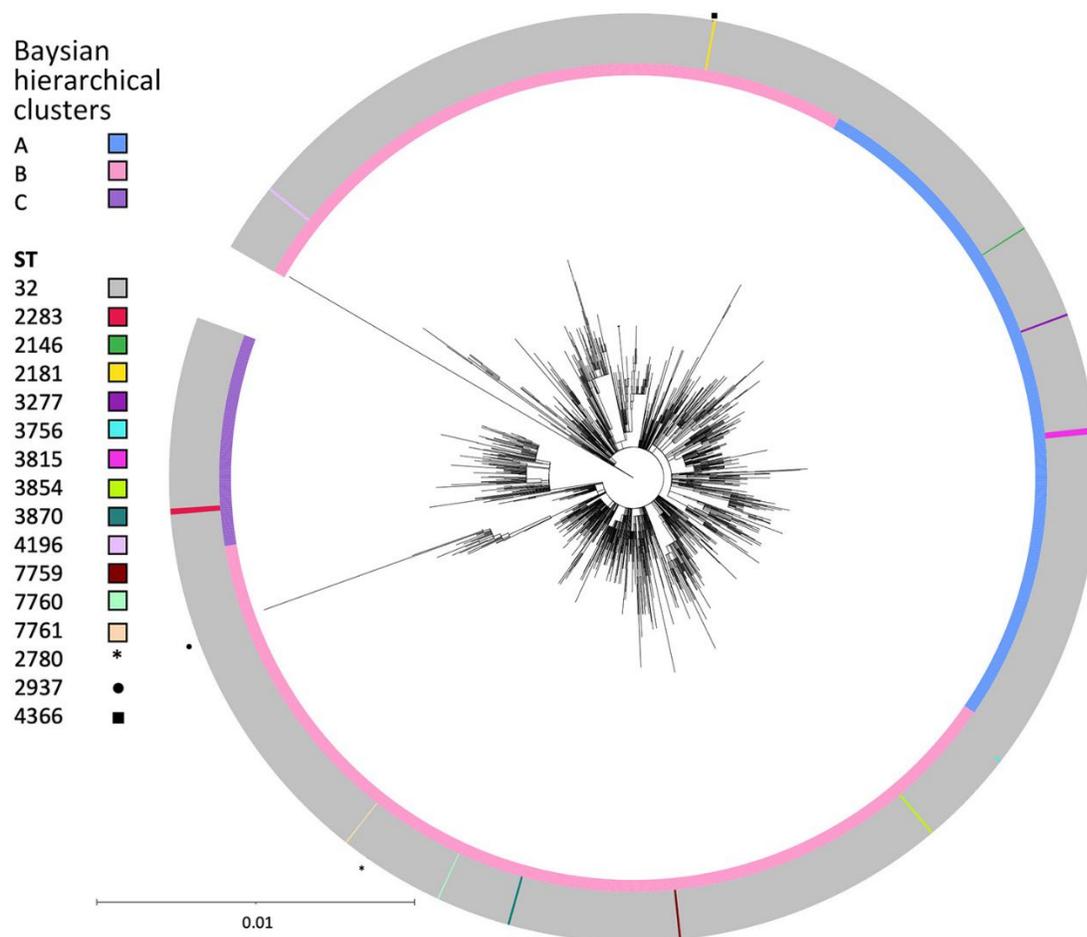
Appendix 1 Figure 3. Maximum likelihood phylogeny of *S*. Infantis with ST. Core SNP maximum likelihood phylogeny of 1288 representatives of 5283 *S*. Infantis isolates. The clusters in the phylogeny are annotated with the Bayesian hierarchical clusters found by fastbaps. The outer ring shows the percentage of isolates in each 25SNP cluster that belonged to each ST. Bayesian hierarchical clusters: A, 348 representatives of 1624 isolates; B, 831 representatives of 3283 isolates and C 109 representatives of 376 isolates. STs: 32 (n = 5204), 2283 (n = 36), 2780 (n = 1), 2181 (n = 1), 3756 (n = 2), 3277 (n = 1), 2146 (n = 26), 2937 (n = 2), 3815 (n = 3), 3870 (n = 1), 3854 (n = 1), 4196 (n = 1), 4366 (n = 1), 7759 (n = 1), 7760 (n = 1), 7761 (n = 1).
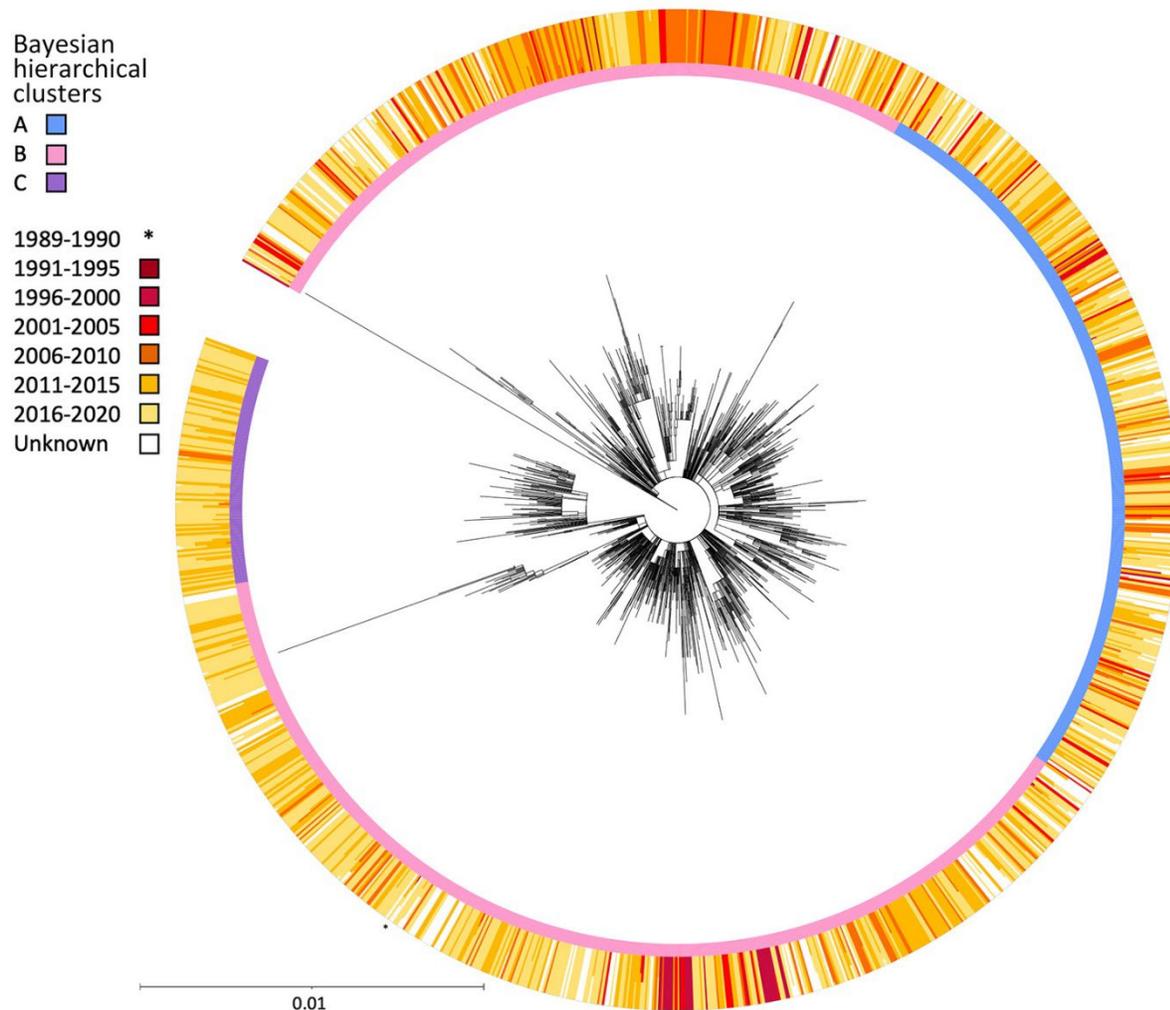
Appendix 1 Figure 4. Maximum likelihood phylogeny of *S*. Infantis with isolation year. Core SNP maximum likelihood phylogeny of 1288 representatives of 5283 *S*. Infantis isolates. The clusters in the phylogeny are annotated with the Bayesian hierarchical clusters found by fastbaps. The outer ring shows the percentage of isolates in each 25SNP cluster that were from each year group. Bayesian hierarchical clusters: A, 348 representatives of 1624 isolates; B, 831 representatives of 3283 isolates and C 109 representatives of 376 isolates. Years: 1989–1990 (n = 1), 1991–1995 (n = 3), 1996–2000 (n = 66), 2001–2005 (n = 78), 2006–2010 (n = 409), 2011–2015 (n = 1371), 2016–2020 (n = 2702), unknown (n = 653).
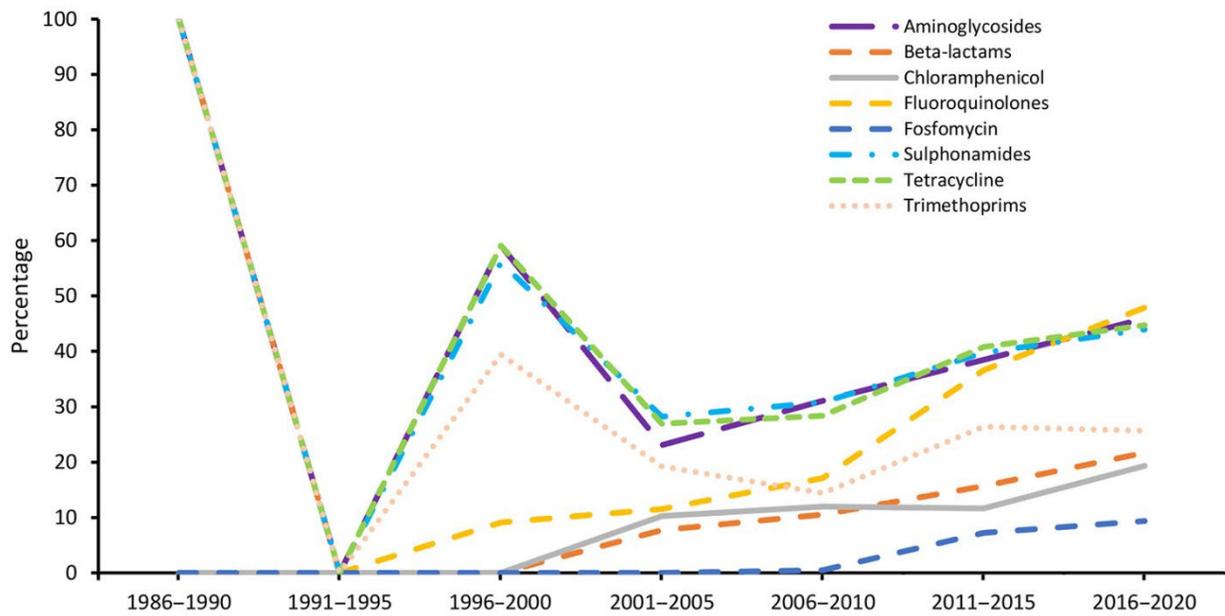
Appendix 1 Figure 5. AMR in *S*. Infantis each year. Proportion of *S*. Infantis isolates from each year with genes conferring resistance to the antimicrobial classes that had resistance in >5% of isolates: aminoglycosides, β-lactams, chloramphenicol, fluoroquinolones, sulphonamides, tetracyclines and trimethoprim. Years: 1989–1990 (n = 1), 1991–1995 (n = 3), 1996–2000 (n = 66), 2001–2005 (n = 78), 2006–2010 (n = 409), 2011–2015 (n = 1372), 2016–2020 (n = 2702), unknown (n = 653).