



Monitoring the Microevolution of *Salmonella enterica* in Healthy Dairy Cattle Populations at the Individual Farm Level Using Whole-Genome Sequencing

Laura M. Carroll^{1*}, Ariel J. Buehler¹, Ahmed Gaballa¹, Julie D. Siler², Kevin J. Cummings², Rachel A. Cheng¹ and Martin Wiedmann¹

¹ Department of Food Science, Cornell University, Ithaca, NY, United States, ² Department of Population Medicine and Diagnostic Sciences, Cornell University, Ithaca, NY, United States

OPEN ACCESS

Edited by:

Bao-Tao Liu,
Qingdao Agricultural University, China

Reviewed by:

Xiang Chen,
Yangzhou University, China
Alexandra Lianou,
University of Patras, Greece

*Correspondence:

Laura M. Carroll
lmc297@cornell.edu

† Present address:

Laura M. Carroll,
Structural and Computational Biology
Unit, EMBL, Heidelberg, Germany

Specialty section:

This article was submitted to
Antimicrobials, Resistance
and Chemotherapy,
a section of the journal
Frontiers in Microbiology

Received: 24 August 2021

Accepted: 28 September 2021

Published: 18 October 2021

Citation:

Carroll LM, Buehler AJ, Gaballa A,
Siler JD, Cummings KJ, Cheng RA
and Wiedmann M (2021) Monitoring
the Microevolution of *Salmonella*
enterica in Healthy Dairy Cattle
Populations at the Individual Farm
Level Using Whole-Genome
Sequencing.
Front. Microbiol. 12:763669.
doi: 10.3389/fmicb.2021.763669

Livestock represent a possible reservoir for facilitating the transmission of the zoonotic foodborne pathogen *Salmonella enterica* to humans; there is also concern that strains can acquire resistance to antimicrobials in the farm environment. Here, whole-genome sequencing (WGS) was used to characterize *Salmonella* strains ($n = 128$) isolated from healthy dairy cattle and their associated environments on 13 New York State farms to assess the diversity and microevolution of this important pathogen at the level of the individual herd. Additionally, the accuracy and concordance of multiple *in silico* tools are assessed, including: (i) two *in silico* serotyping tools, (ii) combinations of five antimicrobial resistance (AMR) determinant detection tools and one to five AMR determinant databases, and (iii) one antimicrobial minimum inhibitory concentration (MIC) prediction tool. For the isolates sequenced here, *in silico* serotyping methods outperformed traditional serotyping and resolved all un-typable and/or ambiguous serotype assignments. Serotypes assigned *in silico* showed greater congruency with the *Salmonella* whole-genome phylogeny than traditional serotype assignments, and *in silico* methods showed high concordance (99% agreement). *In silico* AMR determinant detection methods additionally showed a high degree of concordance, regardless of the pipeline or database used ($\geq 98\%$ agreement among susceptible/resistant assignments for all pipeline/database combinations). For AMR detection methods that relied exclusively on nucleotide BLAST, accuracy could be maximized by using a range of minimum nucleotide identity and coverage thresholds, with thresholds of 75% nucleotide identity and 50–60% coverage adequate for most pipeline/database combinations. *In silico* characterization of the microevolution and AMR dynamics of each of six serotype groups (*S. Anatum*, *Cerro*, *Kentucky*, *Meleagridis*, *Newport*, *Typhimurium*/*Typhimurium* variant *Copenhagen*) revealed that some lineages were strongly associated with individual farms, while others were distributed across multiple farms. Numerous AMR determinant acquisition and loss events were identified, including the recent acquisition of cephalosporin resistance-conferring *bla*_{CMY}- and

*bla*_{CTX-M}-type beta-lactamases. The results presented here provide high-resolution insight into the temporal dynamics of AMR *Salmonella* at the scale of the individual farm and highlight both the strengths and limitations of WGS in tracking zoonotic pathogens and their associated AMR determinants at the livestock-human interface.

Keywords: *Salmonella*, antimicrobial resistance, serotyping, dairy cattle, whole-genome sequencing, evolution, livestock

INTRODUCTION

The foodborne pathogen *Salmonella enterica* is estimated to be responsible for 1.35 million infections, 26,500 hospitalizations, and 420 deaths each year in the United States alone (Centers for Disease Control and Prevention, 2021). Despite the fact that over 2,600 *Salmonella* serotypes have been described (Issenhuth-Jeanjean et al., 2014), fewer than 100 of these serotypes are responsible for the majority of human infections (Centers for Disease Control and Prevention, 2020). In line with this, some *Salmonella* serotypes may share strong associations with a specific host, an extreme example of which can be seen in the human-restricted nature of *Salmonella* Typhi (Uzzau et al., 2000; Boore et al., 2015). Other serotypes, while not confined exclusively to infection of a single host, may be adapted to a given reservoir; for example, *Salmonella* Choleraesuis, while largely adapted to swine, occasionally infects humans (Uzzau et al., 2000; Chiu et al., 2004).

Cattle are a potential reservoir from which humans can acquire salmonellosis, and infected animals can shed *Salmonella* at irregular intervals for varying periods of time, regardless of whether they express clinical signs of bovine salmonellosis or not (Cummings et al., 2010b; Davidson et al., 2018; Holschbach and Peek, 2018). The bovine reservoir boasts its own repertoire of serotypes that can infect humans, with bovine-associated *Salmonella* serotype Dublin, known for its rare but frequently invasive infections in humans, being arguably the most noteworthy (Taylor et al., 1982; Uzzau et al., 2000; Rodriguez-Rivera et al., 2014; Harvey et al., 2017; Mohammed et al., 2017). However, a range of *Salmonella* serotypes can persist and thrive in cattle, potentially infecting humans via either direct contact with infected animals or through food (Gutema et al., 2019). In a previous survey of 46 dairy cattle herds in New York State, *Salmonella* strains isolated from subclinically infected dairy cattle and associated farm environments spanned 26 serotypes, the most common being Cerro, Kentucky, Typhimurium, Newport, and Anatum (Rodriguez-Rivera et al., 2014). Additionally, antimicrobial resistant (AMR) isolates were observed on several farms, on numerous occasions, suggesting subclinically infected dairy cattle as a potential source of AMR *Salmonella* (Rodriguez-Rivera et al., 2014).

Numerous studies have employed whole-genome sequencing (WGS) to characterize *Salmonella* from bovine sources (Mather et al., 2013; Agren et al., 2016; Carroll et al., 2017b; Delgado-Suarez et al., 2018; Liao et al., 2019); however, little is known regarding the evolution and AMR acquisition and loss dynamics of *Salmonella* at the single herd/farm level. Furthermore, the bulk of bovine-associated *Salmonella* WGS efforts have focused on clinical veterinary samples and/or

epidemic lineages (e.g., *S.* Typhimurium DT104). In this study, 128 non-typhoidal *S. enterica* strains isolated from repeated sampling on 13 New York State dairy cattle farms between 2007 and 2009 were characterized using WGS. All strains were isolated from apparently healthy, subclinically infected bovine hosts, as well as the associated farm environment (Rodriguez-Rivera et al., 2014). Using WGS, the microevolution of these persistent lineages within each herd is characterized, as well as the temporal acquisition and loss of AMR determinants among them. In addition to offering insight into the genomics of *Salmonella* isolated from healthy bovine populations at the individual herd/farm level, the accuracy and concordance of multiple *in silico* serotyping and AMR prediction tools are evaluated. Finally, an in-depth, critical analysis of the strengths and limitations of the methods used here is provided, which includes guidance to researchers who wish to employ WGS for herd-level pathogen monitoring.

MATERIALS AND METHODS

Isolate Selection

Salmonella enterica isolates ($n = 128$) obtained from one of 13 dairy farms in New York State were selected to undergo WGS for this study (Supplementary Table 1). All strains were isolated from farms that had undergone surveillance for *Salmonella* for a period of at least 12 months as described previously (Cummings et al., 2010a; Rodriguez-Rivera et al., 2014). Strains were isolated from repeated sampling on each farm between October 2007 and August 2009, from either (i) fecal samples from healthy, subclinically infected dairy cows (referred to hereafter as “bovine” isolates), or (ii) farm environmental swabs (referred to hereafter as “farm environmental” isolates) (Cummings et al., 2010a). All isolates underwent serotyping, phenotypic antimicrobial susceptibility testing, and pulsed-field gel electrophoresis (PFGE) as described previously (Rodriguez-Rivera et al., 2014).

Whole-Genome Sequencing and Data Pre-processing

Genomic DNA extraction and sequencing library preparation were performed as described previously (Carroll et al., 2017b), and the genomes of all 128 *Salmonella* isolates were sequenced using an Illumina HiSeq platform and 2×250 bp paired-end reads. Illumina sequencing adapters and low-quality bases were trimmed using Trimmomatic version 0.33 (using default parameters for Nextera paired-end reads) (Bolger et al., 2014), and FastQC version 0.11.9 (Andrews, 2019) was used to confirm adapter removal and assess read quality. SPAdes version 3.8.0

(Bankevich et al., 2012) was used to assemble genomes *de novo* (using the “careful” option and *k*-mer sizes of 21, 33, 55, 77, 99, and 127), and QCAST version 4.5 (Gurevich et al., 2013) and the “lineage_wf” workflow implemented in CheckM version 1.1.3 (Parks et al., 2015) were used to assess the quality of the resulting assemblies. MultiQC version 1.8 (Ewels et al., 2016) was used to aggregate genome quality metrics. Genome quality statistics are available for all isolates (**Supplementary Table 1**).

In silico Serotyping

In addition to undergoing traditional serotyping in a laboratory setting (i.e., serological detection of expressed O and H antigens using the White-Kauffmann-Le Minor scheme) as described previously (Rodriguez-Rivera et al., 2014), all 128 assembled *Salmonella* genomes (see section “Whole-Genome Sequencing and Data Pre-processing” above) underwent *in silico* serotyping using the command line implementations of (i) the *Salmonella in silico* Typing Resource (SISTR) version 1.0.2 (Yoshida et al., 2016) and (ii) SeqSero2 version 1.1.1 (Zhang et al., 2019) (using SeqSero2’s *k*-mer based workflow). Serotypes assigned using all three methods are available for all 128 isolates (**Supplementary Table 1**). In cases where a discrepancy existed among the traditional serotype designation and one or more of the *in silico* methods, the serotype assigned using two out of the three methods was selected as the final serotype to be reported (e.g., when assigning strain names to isolates in the manuscript, for phylogeny annotation). To confirm that all serotype assignments were reasonable, a phylogeny was constructed using core single nucleotide polymorphisms (SNPs) detected in all *Salmonella* genomes in this study (see section “Reference-Free Single Nucleotide Polymorphism Identification and Phylogeny Construction” below).

In silico Antimicrobial Resistance Determinant Detection

Antimicrobial resistance determinants were detected in each of the 128 *Salmonella* genomes using five separate approaches: (i) ABRicate¹ version 0.8 (Seemann, 2018), (ii) AMRFinderPlus version 3.2.3 (Feldgarden et al., 2019), (iii) ARIBA version 2.14.1 (Hunt et al., 2017), (iv) BTyper version 2.3.3 (Carroll et al., 2017a), and (v) SRST2 version 0.2.0 (Inouye et al., 2014). Assembled genomes were used as input for the ABRicate and BTyper approaches, while trimmed Illumina reads were used as input for the SRST2 and ARIBA approaches. Prokka version 1.12 (Seemann, 2014) was used to annotate each assembled genome, and the resulting GFF (.gff) and FASTA (.faa and .fn) files were used as input for the AMRFinderPlus approach. For the ABRicate approach, the following AMR gene databases were tested (each accessed June 11, 2018 via ABRicate’s abricate-get_db command): (i) the Antibiotic Resistance Gene-ANNOTation database (ARG-ANNOT) (Gupta et al., 2014), (ii) the Comprehensive Antibiotic Resistance Database (CARD) (Jia et al., 2017), (iii) the National Center for Biotechnology Information’s (NCBI’s) Bacterial AMR Reference Gene Database (NCBI) (Feldgarden et al., 2019), and (iv) the ResFinder database (ResFinder) (Zankari et al., 2012).

¹<https://github.com/tseemann/abricate>

For each genome and database combination, minimum AMR gene identity and coverage thresholds ranging from 50 to 100% (5% increments) and 0–100% (10% increments) were tested, respectively. For the BTyper approach, the (i) ARG-ANNOT v3 and (ii) MEGARes version 1.0.1 (Lakin et al., 2017) databases available with BTyper version 2.3.3 were used, with the minimum AMR gene identity and coverage thresholds varied in a manner identical to the ABRicate approach. For the SRST2 approach, the (i) ARG-ANNOT and (ii) ResFinder databases available with SRST2 version 0.2.0 were tested, using default thresholds. For the ARIBA approach, the following databases were tested (each accessed June 13, 2019 using ARIBA’s getref command): (i) the version of ARG-ANNOT available with SRST2, (ii) CARD, (iii) MEGARes, (iv) NCBI, and (v) ResFinder, with all default thresholds used. For the AMRFinder approach, the latest version of the AMRFinder database was used (accessed December 6, 2019), along with the organism-specific database for *Salmonella*.

In silico Prediction of Antimicrobial Minimum Inhibitory Concentration Values

The PATRIC3 antimicrobial minimum inhibitory concentration (MIC) prediction model for *Salmonella* (Nguyen et al., 2019) (accessed June 13, 2019) was used to predict MIC values for each of the 128 *Salmonella* isolates in this study, using the assembled genome of each as input (**Supplementary Text**).

Prediction of Phenotypic Susceptible-Intermediate-Resistant Classifications Using in silico Methods

All 128 *Salmonella* isolates underwent phenotypic antimicrobial susceptibility testing with a panel of 15 antimicrobials (i.e., amikacin, amoxicillin-clavulanic acid, ampicillin, cefoxitin, ceftiofur, ceftriaxone, chloramphenicol, ciprofloxacin, gentamicin, kanamycin, nalidixic acid, streptomycin, sulfamethoxazole-trimethoprim, sulfisoxazole, and tetracycline) using the Sensititre system (Trek Diagnostic Systems Ltd., Cleveland, OH, United States) available at Cornell University’s Animal Health Diagnostic Center as described previously (Rodriguez-Rivera et al., 2014). A “true” (i.e., phenotypic) susceptible-intermediate-resistant (SIR) classification for each of the 15 antimicrobials was obtained for 126 *Salmonella* isolates by comparing raw MIC values to NARMS breakpoints for *Salmonella* (accessed March 23, 2020; **Supplementary Table 1**). For streptomycin, the 1996–2013 NARMS breakpoints were used, as this was compatible with the concentrations used at the time of phenotypic testing (Rodriguez-Rivera et al., 2014). For sulfisoxazole, isolates with MIC > 256 were classified as resistant, as a concentration of 512 µg/mL was not tested. While raw MIC values were unavailable for two isolates (BOV_KENT_16_04-03-08_R8-0967 and ENV_MELA_01_01-10-08_R8-0165; **Supplementary Table 1**), both isolates had previously been categorized as pan-susceptible to all 15 antimicrobials (a classification that was maintained here, as all *in silico* methods correctly classified these isolates as pan-susceptible).

Known AMR determinant/phenotype associations for AMR determinants detected by each of the AMR determinant detection pipeline/database combinations described above (see section “*In silico* Antimicrobial Resistance Determinant Detection”) were obtained from (i) Supplementary Table 4 of the AMRFinder validation paper (Feldgarden et al., 2019) and (ii) CARD (Supplementary Table 2 and Supplementary Text). An isolate was predicted to be resistant to a particular antimicrobial if it possessed one or more AMR determinants known to confer resistance to that antimicrobial; if it did not possess any AMR determinants known to confer resistance to that antimicrobial, the isolate was predicted to be susceptible to that antimicrobial (Supplementary Table 2). For each AMR determinant detection pipeline/database combination, the caret package (Kuhn, 2008) in R version 3.6.1 (R Core Team, 2019) was used to construct a confusion matrix and calculate accuracy scores, Cohen’s kappa coefficients, and other statistics (Supplementary Table 3) by treating “true” susceptible/resistant classifications obtained using phenotypic susceptibility testing as a reference. Cases of intermediate phenotypic resistance were treated as susceptible, as it resulted in slightly higher accuracy scores for all pipeline/database combinations for this particular data set. Because *in silico* prediction of susceptibility/resistance was highly dependent on prior knowledge of AMR determinants and the antimicrobials to which they conferred resistance, the concordance of all pipeline/database combinations was assessed by comparing each pipeline/database combination to results obtained using the SRST2 pipeline/ARG-ANNOT database combination.

To assess the ability of the MIC prediction method implemented in PATRIC3 to predict *Salmonella* SIR classification (see section “*In silico* Prediction of Antimicrobial Minimum Inhibitory Concentration Values” above), predicted MIC values for 14 antimicrobials produced using PATRIC3 were used to predict the SIR status of each of the 128 *Salmonella* isolates using the same NARMS breakpoints used for phenotypic testing. Azithromycin MICs produced by PATRIC3 were excluded, as azithromycin was not among the 15 antimicrobials used here for phenotypic testing. The ability of PATRIC3 to predict amikacin resistance was also not evaluated, as amikacin is not among the antimicrobials queried by PATRIC3. A confusion matrix was constructed as described above, using predicted SIR classifications derived from predicted MIC values produced by PATRIC3 and NARMS breakpoints. Additionally, the deviation of raw MIC predictions produced by PATRIC3 ($MIC_{PATRIC3}$) from “true” raw MIC predictions produced using phenotypic testing ($MIC_{Phenotypic}$) in number of dilution factors ($N_{dilution\ factors}$) was assessed using the following equation:

$$N_{dilution\ factors} = \frac{\ln\left(\frac{MIC_{PATRIC3}}{MIC_{Phenotypic}}\right)}{\ln(2)}$$

where \ln corresponds to the natural logarithm. For example: if PATRIC3 predicted an MIC value of 8 and the “true” MIC value obtained with phenotypic testing was 2, then $\ln(8/2)/\ln(2) = 2$; this means that the PATRIC3 prediction of 8 is +2 dilution factors

away from the “true” MIC of 2 (as dilution used for MIC are 2 fold serial dilutions, e.g., 2, 4, and 8 $\mu\text{g}/\text{mL}$).

Re-testing of Isolates With Highly Incongruent Antimicrobial Resistance Phenotypes

Several ($n = 21$) isolates possessed a phenotypic AMR SIR profile which was deemed to be highly incongruent with its predicted *in silico* AMR profile, regardless of the *in silico* pipeline/database used (Supplementary Table 4). For example, S. Cerro isolate BOV_CERO_35_10-02-08_R8-2685 was resistant to nine antimicrobials but did not harbor any known acquired AMR genes (Supplementary Table 4). Similarly, S. Newport isolate ENV_NEWP_62_03-05-09_R8-3442 itself was pan-susceptible, but harbored multiple acquired AMR genes (e.g., *bla*_{CMY-2}, *floR*, *sul2*, and *tetA*), which conferred multidrug resistance in closely related S. Newport isolates (Supplementary Table 4). To address these incongruencies, 21 selected *Salmonella* isolates underwent phenotypic antimicrobial susceptibility re-testing (conducted September 16, 2020) as described above (see section “Prediction of Phenotypic Susceptible-Intermediate-Resistant Classifications Using *in silico* Methods”), with the exception of amikacin and kanamycin, as the contemporary panel did not include these antimicrobials (Supplementary Table 4).

Kanamycin testing was conducted separately using a gradient diffusion assay (Jorgensen and Ferraro, 2009) according to the manufacturer’s instructions (BioMérieux Kanamycin Strip KM 256, product number 412381). Briefly, bacterial isolates were streaked for single colonies onto Brain Heart Infusion [BHI, Becton Dickinson (BD), Franklin Lakes, NJ, United States] agar plates from frozen glycerol stocks. Pre-cultures were prepared by inoculating a single colony in 3 mL Mueller-Hinton (MH) broth (BD Difco), followed by incubating at 37°C with shaking at 200 rpm for 12–14 h. The pre-cultures were used to inoculate tubes with 5 mL MH broth at 1:200 dilution, and the tubes were incubated at 37°C with shaking at 200 rpm for 5 h. Four mL of melted MH soft agar medium (0.7% agar) were mixed with 100 μL of culture and poured onto Petri plates containing 15 mL of MH agar medium (0.7% agar), and the plates were dried for 5 min. Kanamycin gradient strips were laid on top of the soft agar, and the plates were incubated at 35°C for 18 h. MIC values were determined by evaluating the inhibition zone using a magnifying lens according to the manufacturer’s instructions.

Minimum inhibitory concentration values obtained from re-testing these isolates were interpreted within NARMS breakpoints as described above (see section “Prediction of Phenotypic Susceptible-Intermediate-Resistant Classifications Using *in silico* Methods”) and are reported in the main manuscript (with the exception of amikacin; due to its exclusion from the contemporary panel, original MIC values are reported). Original and updated MIC and SIR values for all 21 isolates are available in Supplementary Table 4.

In silico Plasmid Replicon Detection

Plasmid replicons were detected in all *Salmonella* genome assemblies using ABRicate and the PlasmidFinder database

(accessed June 11, 2018 via ABRicate's abricate-get_db command). For a plasmid replicon to be considered present in a genome, minimum nucleotide BLAST (BLASTN) (Camacho et al., 2009) identity and coverage values of 80 and 60%, respectively, were used (Carattoli et al., 2014).

Reference-Free Single Nucleotide Polymorphism Identification and Phylogeny Construction

A reference-free approach was used to compare the 128 *Salmonella* genomes sequenced in this study to 442 of the 445 *Salmonella* genomes described by Worley et al. (2018); three genomes were omitted because their Sequence Read Archive (SRA) data was not publicly available at the time of access (February 20, 2019). Raw reads for each of the 442 publicly available genomes were downloaded from SRA (Leinonen et al., 2011; Kodama et al., 2012) and processed and assembled as described above (see section "Whole-Genome Sequencing and Data Pre-processing" described above). kSNP3 version 3.1 (Gardner and Hall, 2013; Gardner et al., 2015) was used to identify core SNPs among all 570 assembled *Salmonella* genomes, using the optimal *k*-mer size determined by Kchooser ($k = 19$). IQ-TREE version 1.6.10 (Nguyen et al., 2015) was used to construct a maximum likelihood (ML) phylogeny using the resulting core SNPs and the optimal nucleotide substitution model identified using ModelFinder [determined using model Bayesian Information Criteria (BIC) values; **Supplementary Text**]. Bootstrapping was performed using 1,000 replicates of the Ultrafast Bootstrap method (Minh et al., 2013; Hoang et al., 2018). The resulting ML phylogeny was annotated in R using the bactaxR package (Carroll et al., 2020b; **Supplementary Text**).

Pan-Genome Characterization

GFF files produced by Prokka (see section "In silico Antimicrobial Resistance Determinant Detection" above) were used as input for Roary version 3.12.0 (Page et al., 2015), which was used to identify orthologous gene clusters at a 70% protein BLAST (BLASTP) identity threshold. The resulting gene presence/absence matrix produced by Roary was used as input for besPLOT² (Carroll et al., 2020a), which was used to perform non-metric multidimensional scaling (NMDS) (Kruskal, 1964) and construct plots in two dimensions using a Jaccard distance metric (**Supplementary Text**).

Clustering based on gene presence/absence was assessed for each of the following grouping factors: (i) serotype, (ii) farm, and (iii) isolation source (i.e., bovine or farm environmental). For each of the three grouping factors, the following three statistical tests were performed, using the gene presence/absence matrix produced by Roary, a Jaccard distance metric, and 10,000 permutations: (i) the permutest and betadisper functions in R's vegan package (Oksanen et al., 2019) were used to conduct an ANOVA-like permutation test (Anderson, 2006) to test if group dispersions were homogenous (referred to hereafter as the PERMDISP2 test); (ii) analysis of similarity (ANOSIM)

(Clarke, 1993) using the ANOSIM function in the vegan package in R was used to determine if the average of the ranks of within-group distances was greater than or equal to the average of the ranks of between-group distances (Anderson and Walsh, 2013); (iii) permutational analysis of variance (PERMANOVA) (Anderson, 2001) using the adonis2 function in the vegan package in R was used to determine if group centroids were equivalent. For all tests, a Bonferroni correction was applied to correct for multiple comparisons.

Potential clustering based on AMR gene presence/absence was additionally assessed for the same three grouping factors (serotype, farm, and isolation source), using the presence and absence of AMR determinants detected by AMRFinderPlus as input (i.e., AMR and stress response determinants identified using the "plus" option in AMRFinderPlus). All steps were performed as described above, and a Bonferroni correction was used to correct for multiple comparisons.

Reference-Based Core Single Nucleotide Polymorphism Identification Within Serotypes

For each individual serotype, core SNPs were identified among genomes assigned to that serotype using a reference-based approach. For each serotype, Snippy version 4.3.6³ (Seemann, 2019) was used to identify core SNPs among all representatives assigned to the serotype, using the trimmed Illumina paired-end reads of each genome as input (see section "Whole-Genome Sequencing and Data Pre-processing" above) and one of six high-quality assembled genomes from isolates in this study as a reference genome (**Supplementary Table 1** and **Supplementary Text**). Gubbins version 2.3.4 (Croucher et al., 2015) was used to identify and remove recombination within the full alignment that resulted, and the filtered alignment produced by Gubbins was queried using snp-sites version 2.4.0 (Page et al., 2016) to produce an alignment of core SNPs for each serotype.

Construction of Within-Serotype Phylogenies

For each serotype, IQ-TREE version 1.6.10 was used to construct a ML phylogeny, using core SNPs detected among all isolates assigned to the serotype as input (see "Reference-Based Core Single Nucleotide Polymorphism Identification Within Serotypes" section above), the optimal ascertainment bias-aware nucleotide substitution model selected using ModelFinder, and 1,000 replicates of the UltraFast bootstrap approximation. The temporal structure of each resulting ML phylogeny was assessed using the R^2 value produced by the best-fitting root in TempEst version 1.5.1 (**Supplementary Table 5**; Rambaut et al., 2016).

A tip-dated phylogeny was then constructed for each serotype using BEAST version 2.5.0 (Bouckaert et al., 2014, 2019), using the serotype's corresponding core SNP alignment as input (**Supplementary Text**, **Supplementary Table 5**, and **Supplementary Figure 1**). For a detailed

²<https://github.com/lmc297/besPLOT>

³<https://github.com/tseemann/snippy>

description of all temporal phylogeny construction steps, see the **Supplementary Text**.

Data Availability

Illumina reads are available for all isolates sequenced in this study under NCBI Bioproject Accession PRJNA756552. NCBI BioSample accession numbers for each individual isolate, as well as all associated metadata and genome quality statistics, are available in **Supplementary Table 1**. All BEAST 2 XML files used for temporal phylogeny construction are available at https://github.com/lmc297/zru_farms.

RESULTS

In silico Serotyping of Bovine-Associated *Salmonella* Resolves Incongruencies Between Traditional Serotyping and Whole-Genome Phylogeny

A total of 128 *Salmonella* strains isolated from healthy (i.e., subclinically infected) dairy cattle ($n = 39$) and their associated farm environments ($n = 89$) on 13 different New York State farms underwent WGS (**Supplementary Table 1**). In addition to undergoing traditional serotyping in a laboratory setting, all isolates were assigned serotypes *in silico* using both (i) SISTR and (ii) SeqSero2 (**Supplementary Table 1**). Importantly, serotypes assigned *in silico* using SISTR and/or SeqSero2 were able to resolve all un-typable and/or ambiguous serotypes assigned using traditional serotyping (**Supplementary Table 1**). Furthermore, *in silico* serotypes assigned using (i) SISTR's core-genome multi-locus sequence typing (cgMLST) approach and (ii) SeqSero2 were both highly congruent with the *Salmonella* whole-genome phylogeny (**Figure 1**) and highly concordant with each other: 127 of 128 (99.2%) *Salmonella* isolates sequenced in this study were assigned to identical *in silico* serotypes using both SISTR cgMLST and SeqSero2 (**Supplementary Table 1**), with 100% concordance observed for six of seven observed *in silico* serotype groups (i.e., *S. Anatum*, *S. Cerro*, *S. Meleagridis*, *S. Minnesota*, *S. Newport*, and *S. Typhimurium* and its variants, assigned to $n = 15, 13, 20, 1, 16$, and 27 isolates, respectively). Among *S. Kentucky* ($n = 36$), a single incongruent isolate was observed (ENV_KENT_16_12-04-07_R8-0061), as SeqSero2 could not detect an O-antigen within the genome and was thus unable to assign this isolate to any serotype. This isolate was assigned a serotype of 8,20::z6 using traditional serotyping (*S. Kentucky* has antigenic formula 8,20:i:z6); SISTR classified the isolate as *S. Kentucky*, and the isolate clustered among the *S. Kentucky* isolates sequenced in this study within the *Salmonella* whole-genome phylogeny (**Figure 1** and **Supplementary Table 1**).

When variants of the *S. Typhimurium* serotype ($n = 27$) were considered, discrepancies were observed among traditional serotype assignments and both *in silico* methods (**Supplementary Table 1**). While SeqSero2 could differentiate between *S. Typhimurium* and the O5- variant of *S. Typhimurium* (also known as *S. Typhimurium* variant Copenhagen; “*S. Typhimurium* Copenhagen” is used hereafter), SISTR

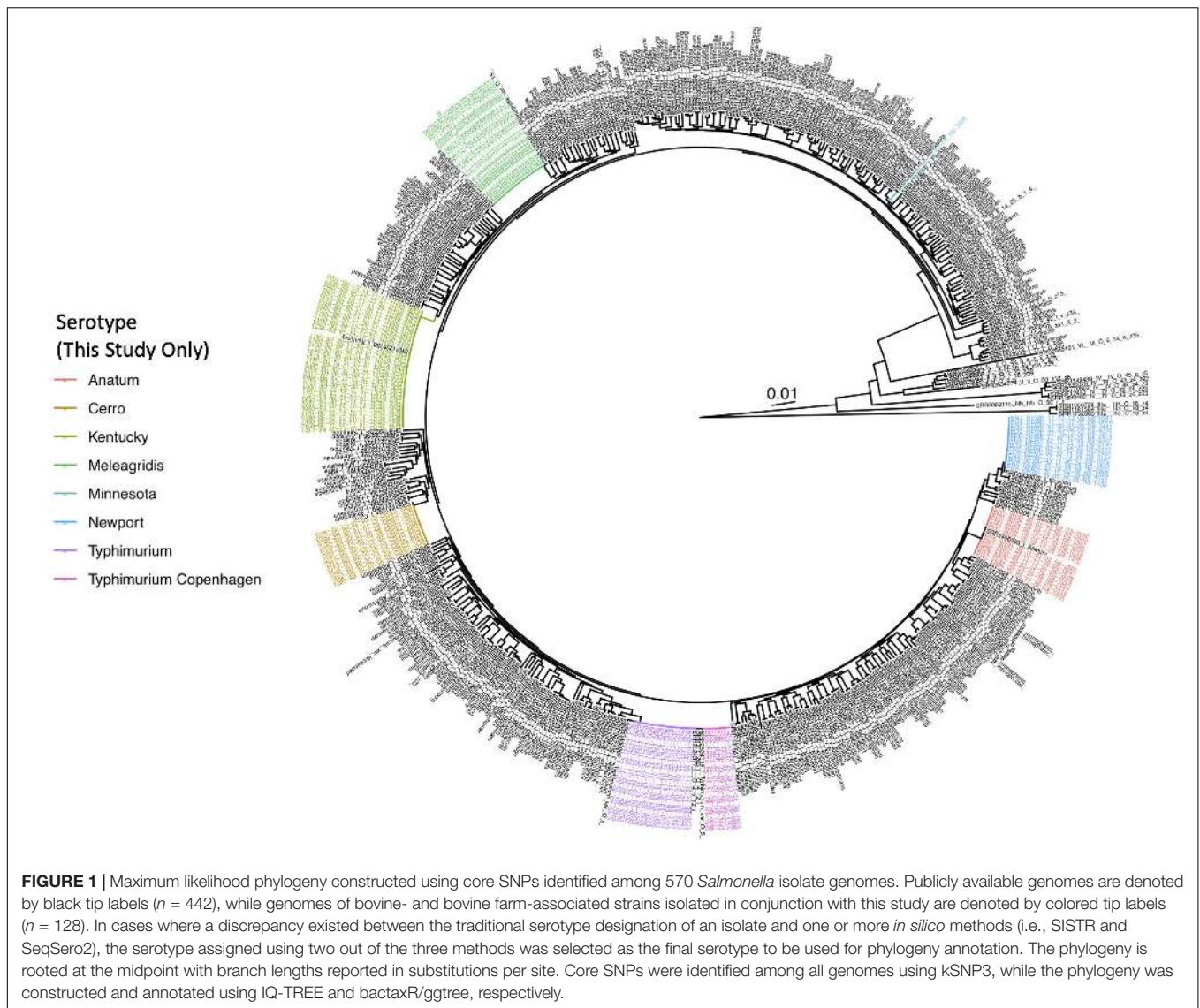
was unable to differentiate the two (**Supplementary Table 1**), as noted previously (Ibrahim and Morin, 2018; Zhang et al., 2019). However, *S. Typhimurium* and *S. Typhimurium* Copenhagen serotype assignments obtained using SeqSero2 and traditional serotyping did not always agree, as five of 27 *S. Typhimurium*/*S. Typhimurium* Copenhagen assignments (18.5%) differed between the two methods (**Supplementary Table 1**). For four of the five incongruent isolates, SeqSero2 assigned an isolate to *S. Typhimurium* Copenhagen, while traditional serotyping assigned a serotype of *S. Typhimurium*; for one isolate, the opposite scenario applied (**Supplementary Table 1**). Furthermore, the lineages formed by isolates classified here as *S. Typhimurium* Copenhagen using either traditional serotyping or SeqSero2, as well as two *S. Typhimurium* Copenhagen genomes from a previous study (Worley et al., 2018), were polyphyletic (**Figure 1**); consequently, the whole-genome phylogeny could not be used to reliably differentiate these two variants.

For the remainder of this study, a serotype assigned consistently with at least two out of the three methods (i.e., traditional serotyping, SeqSero2, and SISTR cgMLST) was selected as the final serotype to be reported for each isolate. Nine of the 13 farms surveyed here harbored *Salmonella* isolates that belonged to a single serotype, while two farms harbored two serotypes or serotype variants (Farms 25 and 35 harbored Typhimurium/Typhimurium Copenhagen and Cerro/Newport, respectively; **Supplementary Table 1**). The remaining two farms harbored three *Salmonella* serotypes (Farms 17 and 62 harbored Kentucky/Newport/Typhimurium and Cerro/Minnesota/Newport, respectively; **Supplementary Table 1**).

In silico Methods Predict Antimicrobial Susceptibility and Resistance Among Bovine-Associated *Salmonella* With High Accuracy and Concordance

Using a 15-antimicrobial panel and NARMS breakpoints for *Salmonella*, more than half of all isolates in this study (81 of 128; 63.3%) were classified as susceptible to all 15 antimicrobials tested, while 38 isolates (29.7%) were classified as resistant to two or more antimicrobials (obtained after the 15-antimicrobial panel was re-run for 22 isolates to resolve discrepancies between *in silico* predictions and phenotypic AMR data; **Supplementary Tables 1, 4**).

Regardless of choice of AMR determinant detection pipeline and AMR determinant database, all pipeline/database combinations performed nearly identically when given the task of predicting phenotypic AMR susceptibility/resistance to 15 antimicrobials using known AMR determinant-phenotype associations (**Figure 2**, **Table 1**, and **Supplementary Tables 2, 3**). Furthermore, all pipeline/database combinations showed an extremely high degree of concordance (98.0% or greater for all pipeline/database combinations; **Supplementary Figure 2**). The overall accuracy of all *in silico* AMR determinant detection pipeline/database combinations ranged from 95.8 to 97.4%, with the SRST2 AMR detection tool/ARG-ANNOT AMR

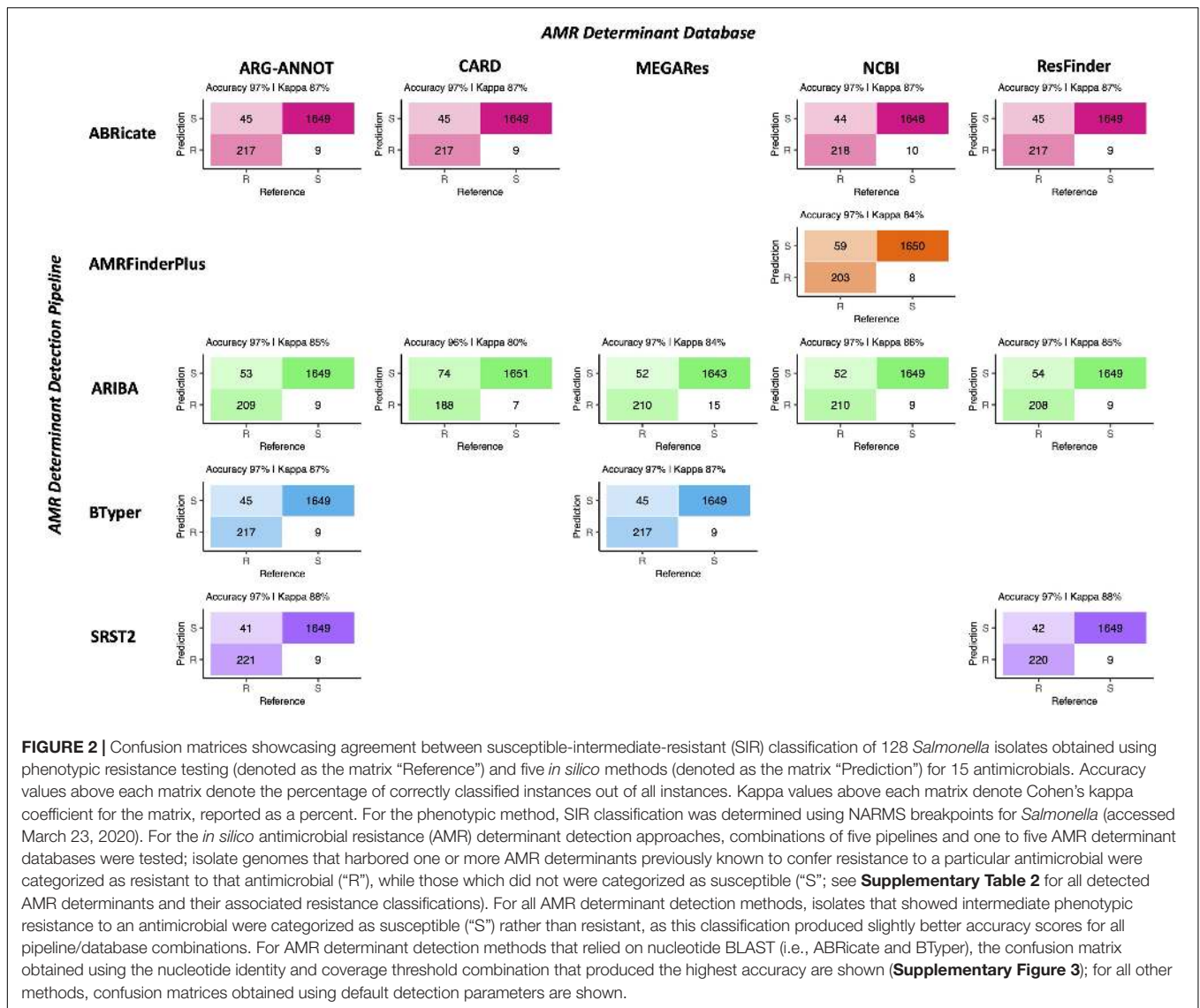


determinant database combination achieving the highest accuracy for this data set (Figure 2, Table 1, and Supplementary Table 3). The ARIBA/CARD pipeline/database combination achieved the highest specificity, although all pipeline/database combinations were able to predict phenotypic AMR with high specificity (>99.0%; Figure 2, Table 1, and Supplementary Table 3). Sensitivity ranged from 71.8 to 84.4%, with SRST2 achieving the highest sensitivities (84.4 and 84.0% for the ARG-ANNOT and ResFinder databases, respectively; Table 1 and Supplementary Table 3).

For the AMR determinant detection pipelines that relied on nucleotide BLAST (i.e., ABRicate and BTyper), a range of minimum percent nucleotide identity and coverage thresholds were additionally tested (i.e., all combinations of 50–100% nucleotide identity in increments of 5% and 0–100% coverage in increments of 10%; Supplementary Figure 3) so that the optimal combination(s) could be established for the isolate genomes sequenced here. For ABRicate/ARG-ANNOT, ABRicate/NCBI,

and ABRicate/ResFinder, maximum accuracy was achieved using minimum coverage thresholds of 60, 50, and 50–60%, respectively, and 75–95% nucleotide identity thresholds (Supplementary Figure 3). For ABRicate/CARD, minimum thresholds of 60% coverage and 75% nucleotide identity were optimal (Supplementary Figure 3). For BTyper/ARG-ANNOT, maximum accuracy was achieved using 60% coverage and 50–95% nucleotide identity; for BTyper/MEGARes, 50–60% coverage and 95% nucleotide identity were the optimal thresholds (Supplementary Figure 3).

The performance of the PATRIC3 *in silico* MIC prediction method was additionally evaluated (Figure 3 and Supplementary Figure 4). PATRIC3 was able to correctly classify *Salmonella* isolates as SIR based on NARMS breakpoints with an overall accuracy of 92.9% [95% confidence interval 91.6–94.1%, accuracy P -value (accuracy > no information rate) < 1.25E-26; Figure 3]. At the individual antimicrobial level, PATRIC3 achieved >90% SIR prediction accuracy for 12 of 14 antimicrobials; only



sulfisoxazole and tetracycline resistance prediction accuracies were <90% (83.6 and 68.0%, respectively; **Figure 3** and **Supplementary Figure 4**).

Genomic Antimicrobial Resistance Determinants of Bovine-Associated *Salmonella* Are Serotype-Associated

Based on the presence and absence of pan-genome elements among all 128 *Salmonella* isolates sequenced here, the *Salmonella* pan-genome was more similar within serotype and within farm than between serotype and between farm, respectively (PERMANOVA and ANOSIM $P < 0.05$ after a Bonferroni correction; **Figure 4** and **Table 2**), with serotypes showing a higher degree of pan-genome dissimilarity (ANOSIM $R = 0.99$) and accounting for a larger proportion of the variance (PERMANOVA $R^2 = 0.93$) than farms (**Figure 4** and **Table 2**); however, dispersion among both serotypes and farms differed

(PERMDISP2 $P < 0.05$ after a Bonferroni correction; **Table 2**), indicating that the ANOSIM and/or PERMANOVA tests could potentially be confounding dispersion with serotype/farm. Additionally, subclinical bovine *Salmonella* isolates did not significantly differ from strains isolated from the associated farm environment based on pan-genome element presence/absence (PERMANOVA, ANOSIM, and PERMDISP2 $P > 0.05$ after a Bonferroni correction; **Figure 4** and **Table 2**).

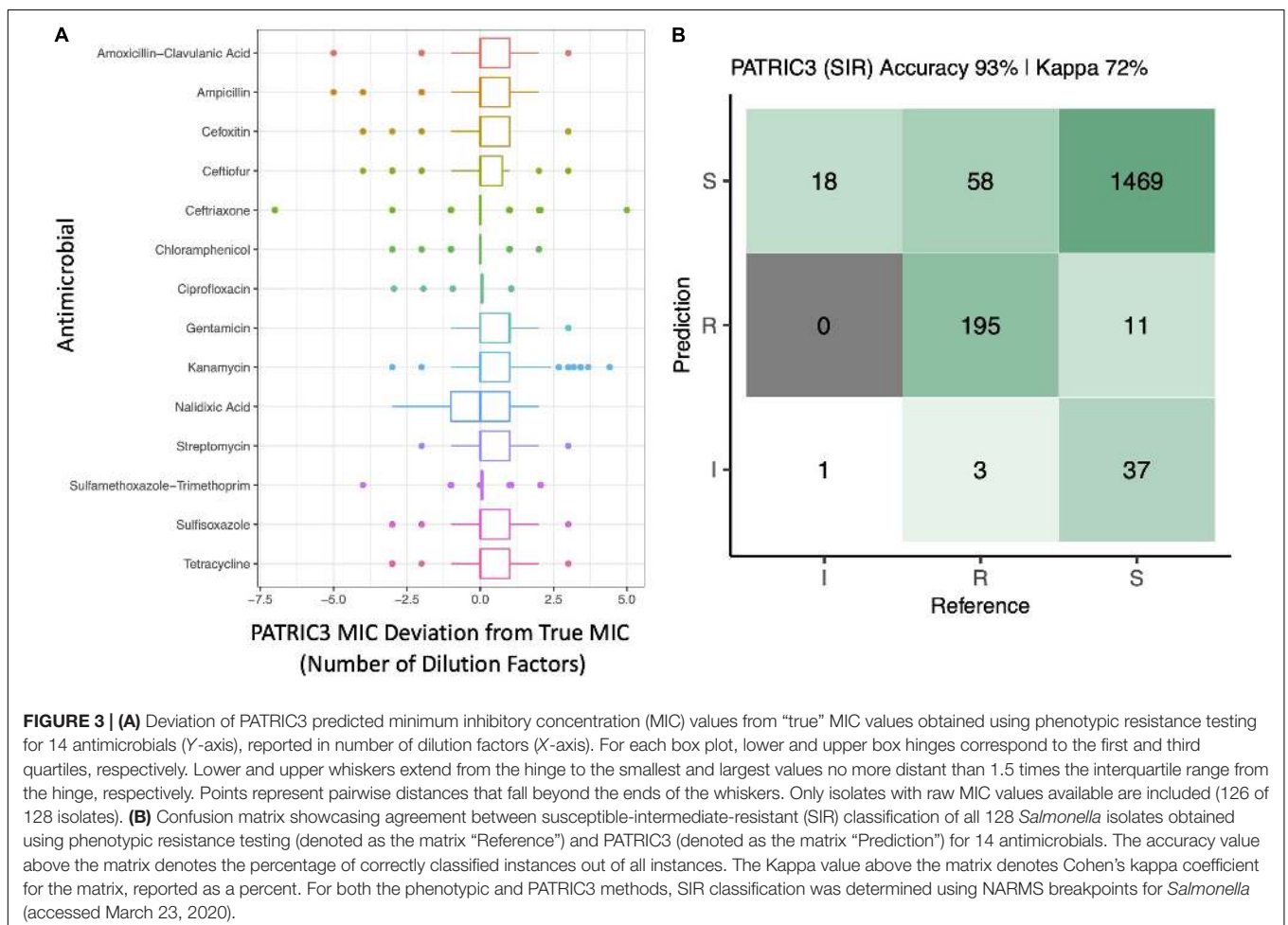
Based on the presence and absence of AMR and stress response determinants detected among all 128 *Salmonella* genomes, isolates were more similar within serotype than between serotypes (PERMANOVA and ANOSIM $P < 0.05$ and PERMDISP2 $P > 0.05$ after a Bonferroni correction; **Figure 4** and **Table 2**). Additionally, isolates were more similar within farm than between farm based on their AMR and stress response gene presence/absence profiles (PERMANOVA and ANOSIM $P < 0.05$; **Figure 4** and **Table 2**), although significant, potentially confounding dispersion differences among farms were present

TABLE 1 | Statistics for the 15-antimicrobial phenotypic susceptibility/resistance prediction task for all antimicrobial resistance (AMR) determinant pipeline/database combinations^a.

AMR Pipeline	AMR Database	% Accuracy (95% Confidence Interval)	Cohen's Kappa (%)	Corrected Accuracy P-Value ^b	Corrected McNemar's Test P-Value ^b	Sensitivity (%)	Specificity (%)
ABRicate	ARG-ANNOT	97.2 (96.3–97.9)	87.3	1.72E-59	2.67E-05	82.8	99.5
ABRicate	CARD	97.2 (96.3–97.9)	87.3	1.72E-59	2.67E-05	82.8	99.5
ABRicate	NCBI	97.2 (96.3–97.9)	87.4	1.72E-59	9.94E-05	83.2	99.4
ABRicate	ResFinder	97.2 (96.3–97.9)	87.3	1.72E-59	2.67E-05	82.8	99.5
AMRFinderPlus	NCBI	96.5 (95.6–97.3)	83.9	1.41E-50	1.41E-08	77.5	99.5
ARIBA	ARG-ANNOT	96.8 (95.9–97.5)	85.3	7.37E-54	6.63E-07	79.8	99.5
ARIBA	CARD	95.8 (94.8–96.6)	79.9	3.08E-42	3.14E-12	71.8	99.6
ARIBA	MEGARes	96.5 (95.6–97.3)	84.3	1.41E-50	1.53E-04	80.2	99.1
ARIBA	NCBI	96.8 (95.9–97.6)	85.5	1.55E-54	1.06E-06	80.2	99.5
ARIBA	ResFinder	96.7 (95.8–97.5)	85.0	3.45E-53	4.15E-07	79.4	99.5
BTypeR	ARG-ANNOT	97.2 (96.3–97.9)	87.3	1.72E-59	2.67E-05	82.8	99.5
BTypeR	MEGARes	97.2 (96.3–97.9)	87.3	1.72E-59	2.67E-05	82.8	99.5
SRST2	ARG-ANNOT	97.4 (96.6–98.1)	88.4	1.69E-62	1.63E-04	84.4	99.5
SRST2	ResFinder	97.3 (96.5–98.0)	88.1	9.85E-62	1.04E-04	84.0	99.5

^aStatistics were calculated using the `confusionMatrix` function in the `caret` package in R, with resistant (“R”) phenotypes/genotypes treated as the “positive” result and susceptible (“S”) phenotypes/genotypes treated as the “negative” result; See **Supplementary Table 3** for an extended version of this table.

^bAdjusted using a Bonferroni correction.



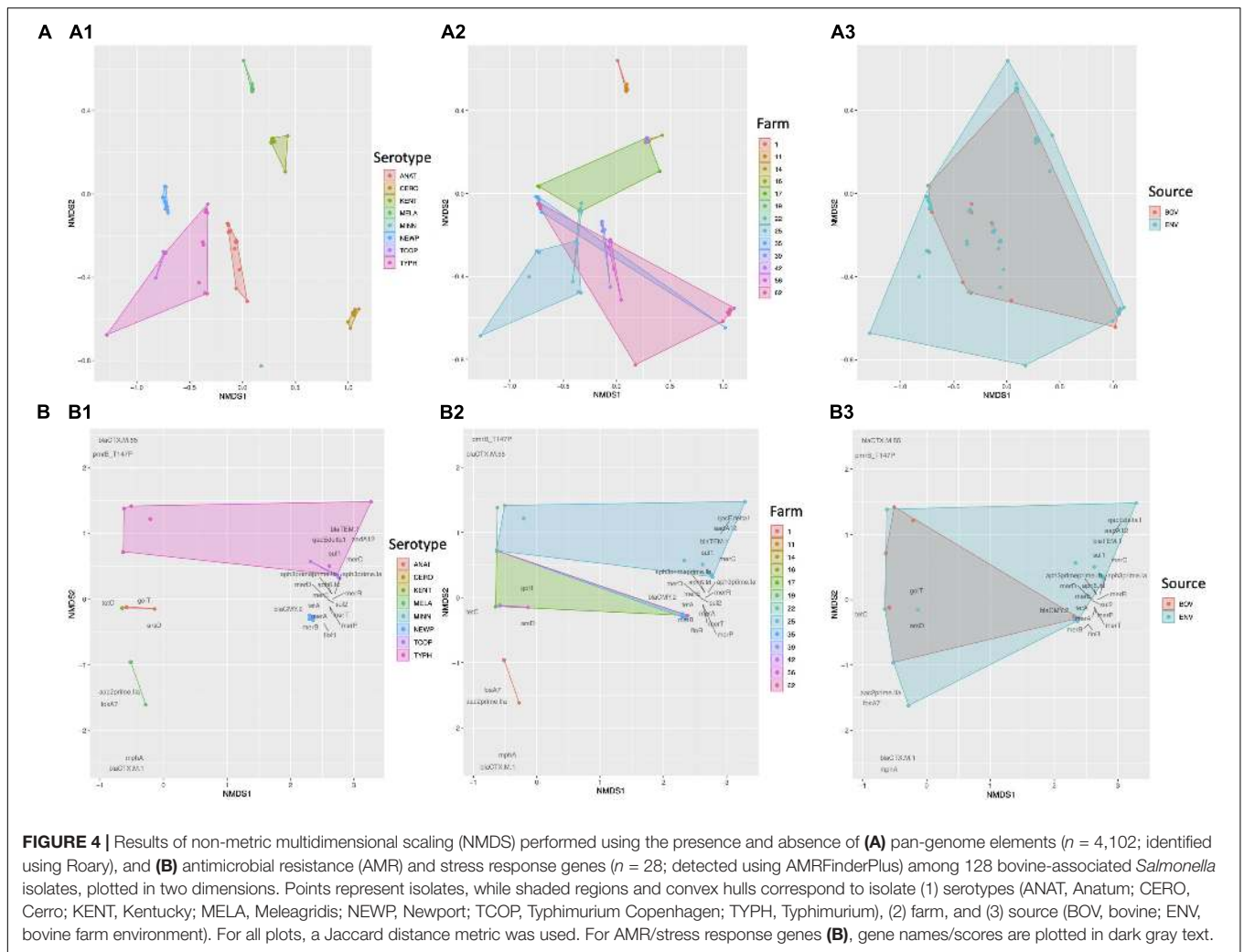


FIGURE 4 | Results of non-metric multidimensional scaling (NMDS) performed using the presence and absence of **(A)** pan-genome elements ($n = 4,102$; identified using Roary), and **(B)** antimicrobial resistance (AMR) and stress response genes ($n = 28$; detected using AMRFinderPlus) among 128 bovine-associated *Salmonella* isolates, plotted in two dimensions. Points represent isolates, while shaded regions and convex hulls correspond to isolate (1) serotypes (ANAT, Anatum; CERO, Cerro; KENT, Kentucky; MELA, Meleagridis; NEWP, Newport; TCOP, Typhimurium Copenhagen; TYPH, Typhimurium), (2) farm, and (3) source (BOV, bovine; ENV, bovine farm environment). For all plots, a Jaccard distance metric was used. For AMR/stress response genes **(B)**, gene names/scores are plotted in dark gray text.

TABLE 2 | Results of PERMDISP2, ANOSIM, and PERMANOVA tests^a.

Group	PERMDISP2 Raw P-Value (F) ^b	ANOSIM Raw P-Value (R) ^c	PERMANOVA Raw P-Value (R^2) ^d
Pan-genome element presence/absence ($n = 4,102$)^e			
Serotype	2.0E-4 (14.3)*	<1.0E-4 (0.99)*	<1.0E-4 (0.93)*
Farm	<1.0E-4 (4.43)*	<1.0E-4 (0.54)*	<1.0E-4 (0.73)*
Source	0.071 (3.40)	0.99 (-0.07)	0.32 (0.01)
Antimicrobial resistance and stress response gene presence/absence ($n = 28$)^f			
Serotype	0.013 (4.46)	<1.0E-4 (0.79)*	<1.0E-4 (0.85)*
Farm	<1.0E-4 (5.52)*	<1.0E-4 (0.30)*	<1.0E-4 (0.54)*
Source	0.74 (0.01)	0.79 (-0.03)	0.31 (0.01)

^aAll tests were performed using a Jaccard dissimilarity metric and 10,000 permutations; raw P-values are reported for all tests, with significant P-values ($P < 0.05$ after a Bonferroni correction was applied to all values) denoted with an asterisk (*).

^bANOVA-like permutation test applied to results obtained using the PERMDISP2 procedure for the analysis of multivariate homogeneity of group dispersions (i.e., variances), obtained using the betadisper and permtest functions in the vegan package in R; betadisper is a multivariate analog of Levene's test for homogeneity of variances.

^cAnalysis of similarities (ANOSIM) test results obtained using the ANOSIM function in the vegan package in R.

^dPermutational analysis of variance (PERMANOVA) test results obtained using the adonis2 function in the vegan package in R.

^eIdentified using Roary and a 70% protein BLAST (BLASTP) identity threshold.

^fDetected using AMRFinderPlus.

(PERMDISP2 $P < 0.05$; **Table 2**). As was the case with the pan-genome in its entirety, subclinical bovine *Salmonella* isolates did not significantly differ from farm environmental isolates based on AMR and stress response gene presence/absence (PERMANOVA, ANOSIM, and PERMDISP2 $P > 0.05$ after a Bonferroni correction; **Figure 4** and **Table 2**).

Each of Two New York State Dairy Farms Harbors a Unique, Bovine-Associated *Salmonella* Anatum Lineage

Fifteen *S. Anatum* strains encompassing four PFGE types (**Supplementary Table 1**) were isolated from subclinical bovine

sources and their associated farm environments on two different New York State dairy farms (i.e., Farms 39 and 56; **Figure 5** and **Table 3**). Notably, the *S. Anatum* lineages circulating on each farm were distinct at a genomic level, with isolates from each farm forming a separate clade [posterior probability (PP) = 1 for each; **Figure 5**]. The two farm-associated lineages were predicted to share a common ancestor circa 1836 (node age 1836.28 using median node heights; **Figure 5**), although the age of the common ancestor could not be dated reliably [node height 95% highest posterior density (HPD) interval 540.85–1978.42; **Supplementary Figure 5**].

Salmonella Anatum isolates from Farm 39 shared a common ancestor circa 2005 (node age 2004.69, node height 95% HPD

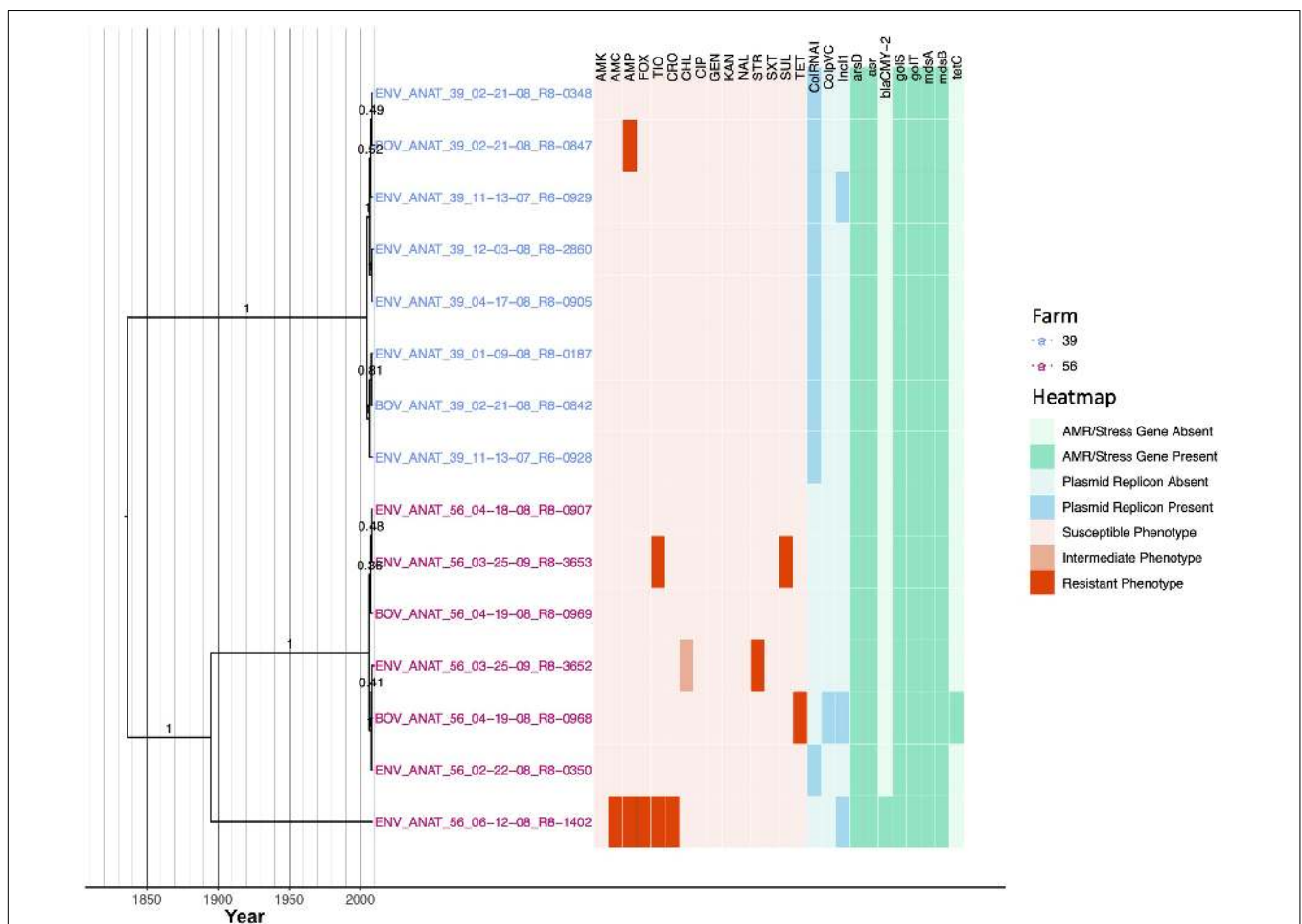


FIGURE 5 | Rooted, time-scaled maximum clade credibility (MCC) phylogeny constructed using core SNPs identified among 15 *Salmonella* Anatum genomes isolated from subclinical bovine sources and the surrounding bovine farm environment. Tip label colors denote the ID of the farm from which each strain was isolated. Branch labels denote posterior probabilities of branch support. Time in years is plotted along the X-axis, and branch lengths are reported in years. The heatmap to the right of the phylogeny denotes (i) the susceptible-intermediate-resistant (SIR) classification of each isolate for each of 15 antimicrobials (obtained using phenotypic testing and NARMS breakpoints; orange); (ii) presence and absence of plasmid replicons (detected using ABRicate/PlasmidFinder and minimum nucleotide identity and coverage thresholds of 80 and 60%, respectively; blue); (iii) presence and absence of antimicrobial resistance (AMR) and stress response genes (identified using AMRFinderPlus and default parameters; green). Core SNPs were identified using Snippy. The phylogeny was constructed using the results of ten independent runs using a strict clock model, the Standard_TVMef nucleotide substitution model, and the Coalescent Bayesian Skyline population model implemented in BEAST version 2.5.1, with 10% burn-in applied to each run. LogCombiner-2 was used to combine BEAST 2 log files, and TreeAnnotator-2 was used to construct the phylogeny using median node heights. Abbreviations for the 15 antimicrobials are: AMK, amikacin; AMC, amoxicillin-clavulanic acid; AMP, ampicillin; FOX, cefoxitin; TIO, ceftiofur; CRO, ceftriaxone; CHL, chloramphenicol; CIP, ciprofloxacin; GEN, gentamicin; KAN, kanamycin; NAL, nalidixic acid; STR, streptomycin; SXT, sulfamethoxazole-trimethoprim; SUL, sulfisoxazole; TET, tetracycline.

1978.46–2007.69; **Figure 5** and **Supplementary Figure 5**). All Farm 39 *S. Anatum* isolates possessed identical AMR/stress response gene profiles, and all isolates were pan-susceptible except for a single isolate that was resistant to ampicillin (**Figure 5**). All Farm 39 *S. Anatum* isolates additionally harbored ColRNAI plasmids; a single isolate additionally harbored an IncII plasmid that appeared to harbor no AMR genes (**Figure 5**).

Salmonella Anatum isolates from Farm 56, however, were considerably more diverse than their Farm 39 counterparts; while a clade containing six of seven strains shared a very recent common ancestor (node age 2006.16, node height 95% HPD 1985.37–2008.11; **Figure 5** and **Supplementary Figure 5**), a unique lineage represented by a single environmental isolate (ENV_ANAT_56_06-12-08_R8-1402) was present among *S. Anatum* from Farm 56 (**Figure 5**). All *S. Anatum* isolates from Farm 56 were predicted to have evolved from a common ancestor that existed circa 1895 (node age 1895.03), although this node could not be reliably dated (node height 95% HPD 1036.89–1989.944; **Supplementary Figure 5**). Additionally, *S. Anatum* isolated from Farm 56 showcased a greater degree of AMR heterogeneity than those from Farm 39 (**Figure 5**). Notably, the isolate comprising the unique Farm 56 *S. Anatum* lineage possessed an IncII plasmid and *bla*_{CMY-2} and was multidrug resistant (MDR) (resistant to amoxicillin-clavulanic acid, ampicillin, cefoxitin, ceftiofur, and ceftriaxone; **Figure 5**). Three of six *S. Anatum* strains comprising the major Farm 56 *S. Anatum* lineage were pan-susceptible. The remaining three isolates were resistant to one of (i) tetracycline, (ii) streptomycin, or (iii) ceftiofur and sulfisoxazole; the streptomycin-resistant isolate additionally exhibited reduced susceptibility to chloramphenicol (**Figure 5**). The tetracycline-resistant isolate additionally possessed both ColpVC and IncII plasmids and harbored tetracycline resistance gene *tetC* (**Figure 5**). While these data suggest some *S. Anatum* lineages queried here have recently acquired AMR, the limited number of isolates and the large degree of uncertainty for some phylogeny node ages preclude reliable estimation of AMR acquisition timeframes.

A Closely Related *Salmonella Cerro* Lineage Spans Two New York State Dairy Farms

Thirteen *S. Cerro* strains encompassing two PFGE types (**Supplementary Table 1**) isolated from two dairy farms (12 from Farm 62 and one from Farm 35) were found to share a high degree of genomic similarity; isolates differed by, at most, 12 core SNPs and evolved from a common ancestor that existed circa March 2008 [node age 2008.21, common ancestor (CA) node height 95% HPD interval 2007.6–2008.6; **Figure 6**, **Table 3**, **Supplementary Figure 6**, and **Supplementary Table 5**]. While IncII and ColRNAI plasmid replicons were detected in all 13 *S. Cerro* isolates, only one isolate was not pan-susceptible (**Figure 6**). Notably, the isolate from Farm 35 (BOV_CERO_35_10-02-08_R8-2685) was classified as resistant to nine antimicrobials using phenotypic methods (i.e., amoxicillin-clavulanic acid, ampicillin, cefoxitin, ceftiofur, ceftriaxone, chloramphenicol, streptomycin, sulfisoxazole, and

tetracycline); based on the most parsimonious explanation for AMR acquisition, this lineage acquired AMR after July 2008 (node age 2008.51, CA node height 95% HPD interval 2008.14–2008.75; **Figure 6** and **Supplementary Figure 6**). However, no genomic determinants known to confer resistance to these antimicrobials were detected in the genome of the MDR isolate (**Figure 6**), and the MDR phenotype was confirmed in a second, independent phenotypic AMR test (**Supplementary Table 4**).

Salmonella Kentucky Strains Isolated Across Five Different New York State Dairy Farms Evolved From a Common Ancestor That Existed Circa 2004

Thirty-six *S. Kentucky* isolates encompassing two PFGE types (**Supplementary Table 1**) isolated across five New York State dairy farms (i.e., five, seven, nine, seven, and eight isolates from each of Farm 14, 16, 17, 19, and 42, respectively) were similar at a genomic level; isolates differed by between 0 and 30 core SNPs and shared a common ancestor that was predicted to have existed circa January/February 2004 (node age 2004.07, CA node height 95% HPD interval 2000.73–2006.8; **Figure 7**, **Table 3**, **Supplementary Figure 7**, and **Supplementary Table 5**). Two farms harbored a total of three *S. Kentucky* isolates, which were not pan-susceptible (two isolates from Farm 16 and one from Farm 17; **Figure 7**). Farm 17 harbored a tetracycline-resistant isolate (ENV_KENT_17_03-11-08_R8-0815), which possessed an IncII plasmid and *tetC* (**Figure 7**). The lineage represented by this isolate was predicted to have acquired tetracycline resistance after March 2007 (node height 2007.19, CA node height 95% HPD interval 2006.43–2007.84; **Figure 7** and **Supplementary Figure 7**). The two *S. Kentucky* isolates from Farm 16 additionally showed reduced susceptibility to chloramphenicol, a trait predicted to have been acquired by these lineages after December 2006/January 2007 (for the lineage represented by isolate ENV_KENT_16_12-04-07_R8-0061; node height 2006.98, CA node height 95% HPD interval 2005.95–2007.85) and May 2007 (for the lineage represented by isolate BOV_KENT_16_02-13-08_R8-0838; node height 2007.38, CA node height 95% HPD interval 2006.59–2008.10; **Figure 7** and **Supplementary Figure 7**). No corresponding genes that may encode for reduced chloramphenicol susceptibility were identified in these two isolates.

A Clonal *Salmonella Meleagridis* Lineage Is Distributed Across Two New York State Dairy Farms and Encompasses Isolates Carrying *bla*_{CTX-M-1}

Nineteen *S. Meleagridis* isolates encompassing two PFGE types (**Supplementary Table 1**) were isolated from two dairy farms (13 and six isolates from Farms 01 and 11, respectively) and were highly clonal: isolates differed by fewer than ten core SNPs and evolved from a common ancestor that existed circa May/June 2007 (node age 2007.42, CA node height 95% HPD interval 2006.91–2007.75; **Figure 8**, **Table 3**, **Supplementary Figure 8**, and **Supplementary Table 5**). All but three branches within the *S. Meleagridis* phylogeny had low

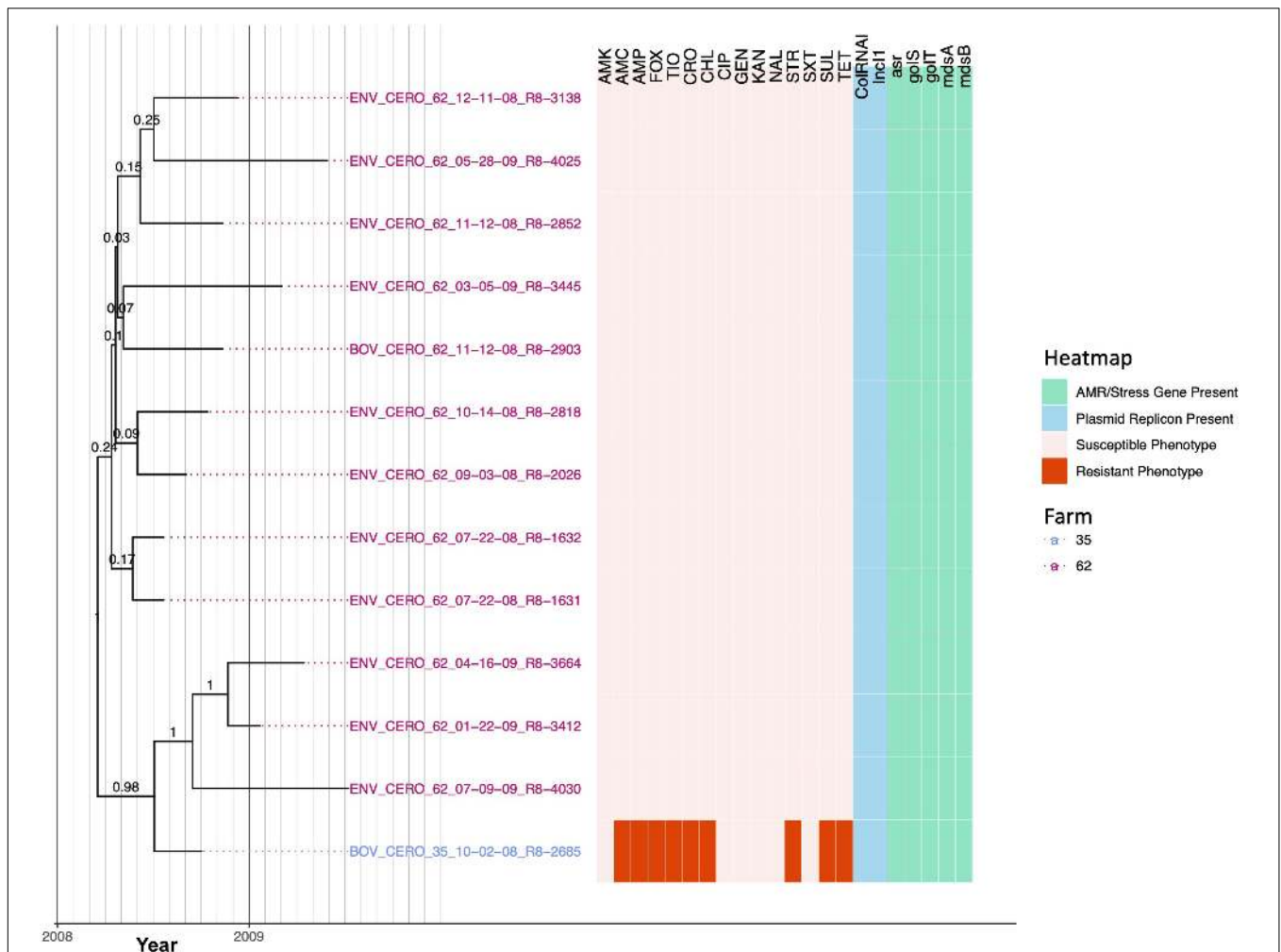


FIGURE 6 | Rooted, time-scaled maximum clade credibility (MCC) phylogeny constructed using core SNPs identified among 13 *Salmonella* Cerro genomes isolated from subclinical bovine sources and the surrounding bovine farm environment. Tip label colors denote the ID of the farm from which each strain was isolated. Branch labels denote posterior probabilities of branch support. Time in years is plotted along the X-axis, and branch lengths are reported in years. The heatmap to the right of the phylogeny denotes (i) the susceptible-intermediate-resistant (SIR) classification of each isolate for each of 15 antimicrobials (obtained using phenotypic testing and NARMS breakpoints; orange); (ii) presence of plasmid replicons (detected using ABRicate/PlasmidFinder and minimum nucleotide identity and coverage thresholds of 80 and 60%, respectively; blue); (iii) presence of antimicrobial resistance (AMR) and stress response genes (identified using AMRFinderPlus and default parameters; green). Core SNPs were identified using Snippy. The phylogeny was constructed using the results of ten independent runs using a strict clock model, the Standard_TPM1 nucleotide substitution model, and the Coalescent Bayesian Skyline population model implemented in BEAST version 2.5.1, with 10% burn-in applied to each run. LogCombiner-2 was used to combine BEAST 2 log files, and TreeAnnotator-2 was used to construct the phylogeny using common ancestor node heights. Abbreviations for the 15 antimicrobials are: AMK, amikacin; AMC, amoxicillin-clavulanic acid; AMP, ampicillin; FOX, cefoxitin; TIO, ceftiofur; CRO, ceftriaxone; CHL, chloramphenicol; CIP, ciprofloxacin; GEN, gentamicin; KAN, kanamycin; NAL, nalidixic acid; STR, streptomycin; SXT, sulfamethoxazole-trimethoprim; SUL, sulfisoxazole; TET, tetracycline.

support ($PP \leq 0.41$; **Figure 8** and **Supplementary Figure 8**), indicating that most nodes were unreliable, likely due to the isolates being highly clonal. All *S. Meleagridis* isolates from Farm 11 were pan-susceptible, possessed no plasmid replicons, and did not possess any acquired AMR genes (**Figure 8**). Among the *S. Meleagridis* isolates from Farm 01, one isolate (ENV_MELA_01_10-02-07_R6-0938) was resistant to ampicillin, ceftiofur, and ceftriaxone, and possessed an IncN plasmid, macrolide resistance gene *mph(A)*, and beta-lactamase *bla_{CTX-M-1}* (**Figure 8**). Two additional *S. Meleagridis* isolates from Farm 01 each exhibited reduced susceptibility

to either (i) ceftiofur, sulfisoxazole, and tetracycline, or (ii) ceftiofur (**Figure 8**).

Kanamycin Resistance Among Each of Three New York State Dairy Farms Harboring a Distinct, Multidrug-Resistant *Salmonella* Newport Lineage Is Farm-Associated

Sixteen *S. Newport* isolates encompassing three PFGE types (**Supplementary Table 1**) were isolated from one

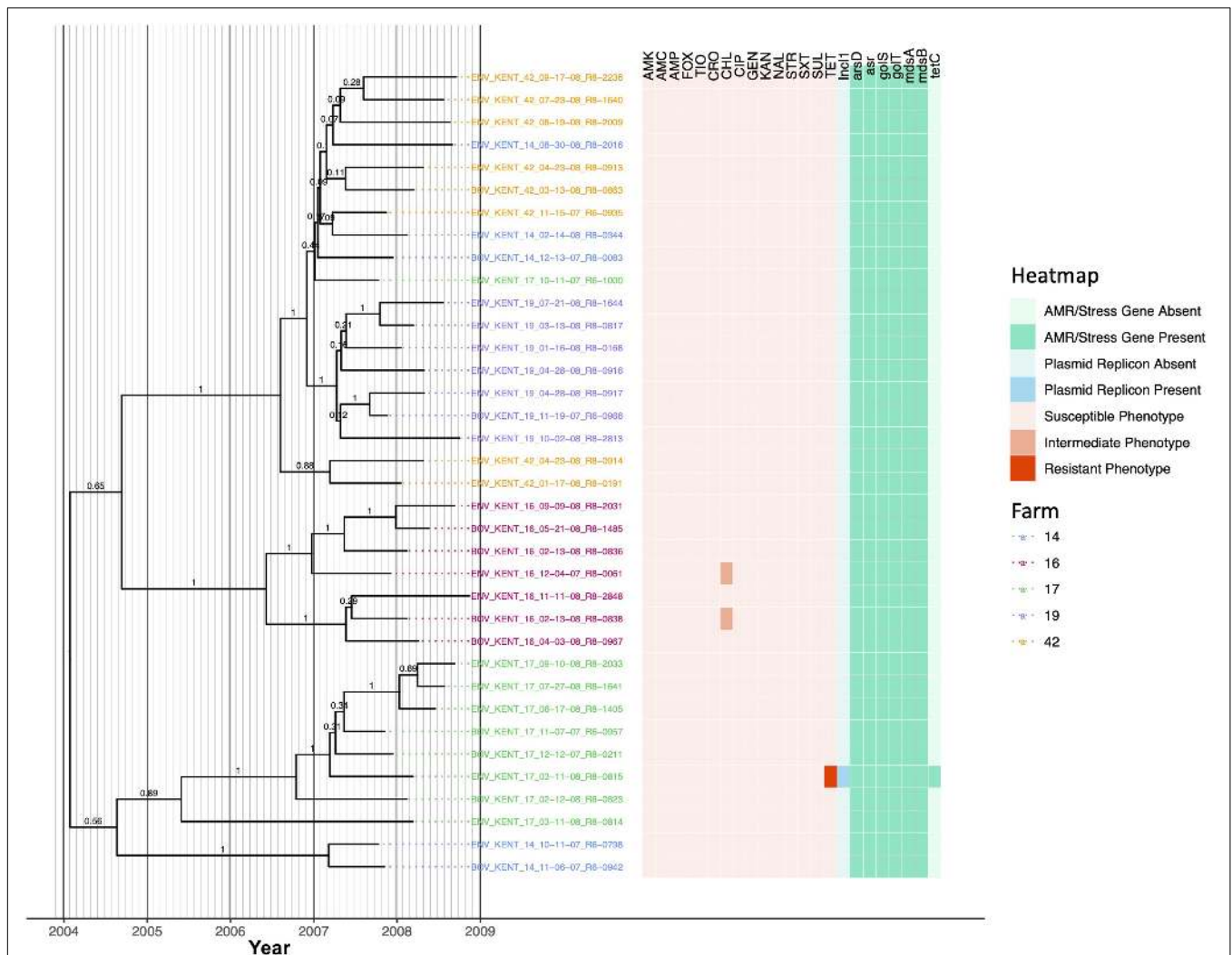


FIGURE 7 | Rooted, time-scaled maximum clade credibility (MCC) phylogeny constructed using core SNPs identified among 36 *Salmonella* Kentucky genomes isolated from subclinical bovine sources and the surrounding bovine farm environment. Tip label colors denote the ID of the farm from which each strain was isolated. Branch labels denote posterior probabilities of branch support. Time in years is plotted along the X-axis, and branch lengths are reported in years. The heatmap to the right of the phylogeny denotes (i) the susceptible-intermediate-resistant (SIR) classification of each isolate for each of 15 antimicrobials (obtained using phenotypic testing and NARMS breakpoints; orange); (ii) presence of plasmid replicons (detected using ABRicate/PlasmidFinder and minimum nucleotide identity and coverage thresholds of 80 and 60%, respectively; blue); (iii) presence of antimicrobial resistance (AMR) and stress response genes (identified using AMRFinderPlus and default parameters; green). Core SNPs were identified using Snippy. The phylogeny was constructed using the results of ten independent runs using a relaxed lognormal clock model, the Standard_TVMEf nucleotide substitution model, and the Coalescent Bayesian Skyline population model implemented in BEAST version 2.5.1, with 10% burn-in applied to each run. LogCombiner-2 was used to combine BEAST 2 log files, and TreeAnnotator-2 was used to construct the phylogeny using common ancestor node heights. Abbreviations for the 15 antimicrobials are: AMK, amikacin; AMC, amoxicillin-clavulanic acid; AMP, ampicillin; FOX, cefoxitin; TIO, ceftiofur; CRO, ceftriaxone; CHL, chloramphenicol; CIP, ciprofloxacin; GEN, gentamicin; KAN, kanamycin; NAL, nalidixic acid; STR, streptomycin; SXT, sulfamethoxazole-trimethoprim; SUL, sulfisoxazole; TET, tetracycline.

of three farms (four, five, and seven isolates from Farms 17, 35, and 62, respectively); all isolates were resistant to amoxicillin-clavulanic acid, ampicillin, cefoxitin, ceftiofur, ceftriaxone, streptomycin, sulfisoxazole, and tetracycline (Figure 9). All *S. Newport* genomes harbored IncA/C2 and ColRNAI plasmids, as well as streptomycin resistance genes *APH(3'')-Ib* and *APH(6)-Id* (i.e., *strAB*), beta-lactamase *bla_{CMY-2}*, sulfonamide resistance gene *sul2*, and tetracycline resistance gene *tetA* (Figure 9). Notably, the

S. Newport lineage circulating on each farm formed one of three separate clades (PP = 0.99–1.0) that evolved from a common ancestor that existed circa March/April 2004 (node age 2004.23, CA node height 95% HPD interval 2000.42–2007.85; Figure 9, Table 3, Supplementary Figure 9, and Supplementary Table 5).

The *S. Newport* lineages present on Farm 17 and Farm 62 were additionally resistant to chloramphenicol and kanamycin and possessed chloramphenicol and kanamycin resistance genes

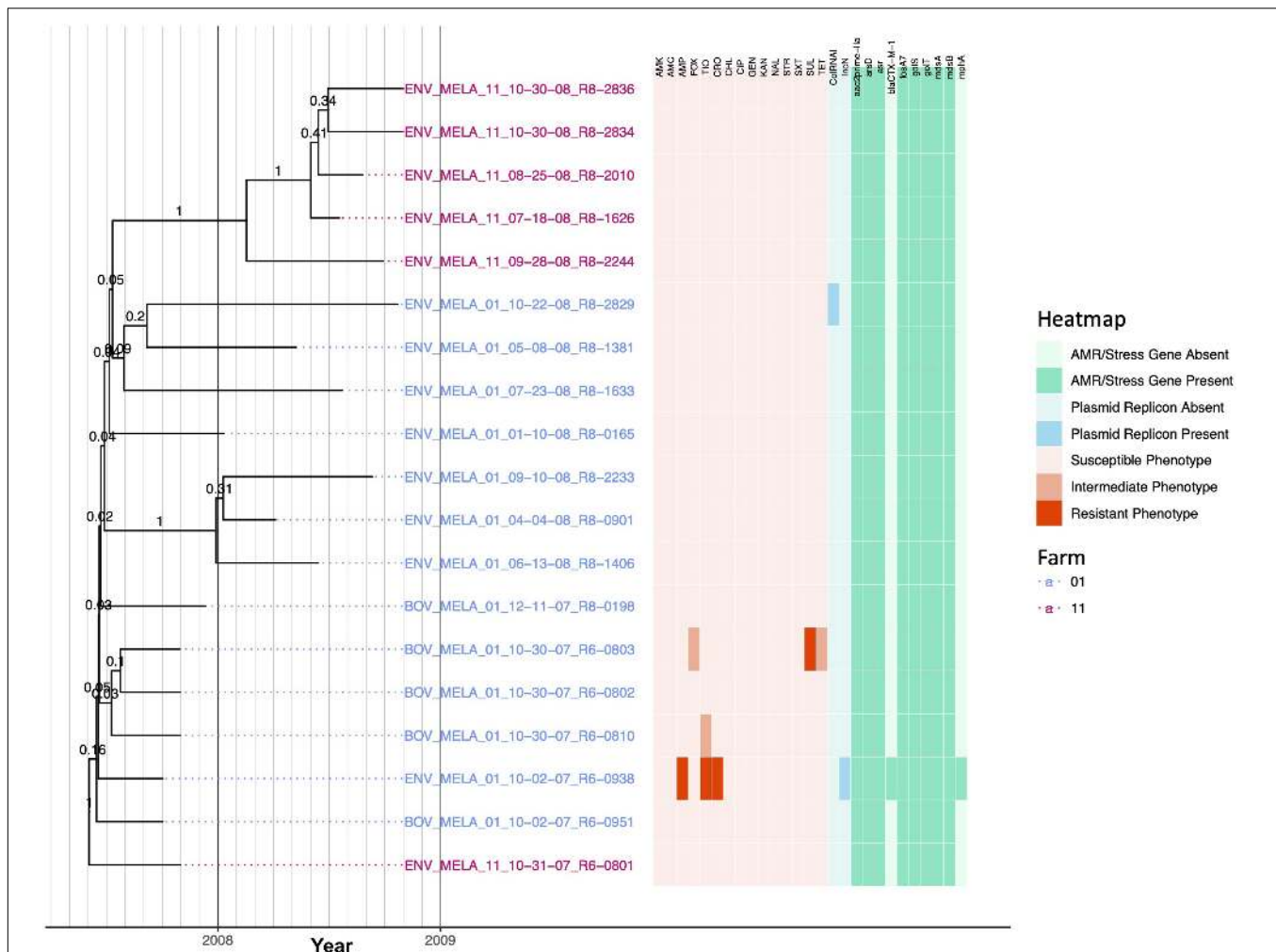


FIGURE 8 | Rooted, time-scaled maximum clade credibility (MCC) phylogeny constructed using core SNPs identified among 19 *Salmonella* Meleagridis genomes isolated from subclinical bovine sources and the surrounding bovine farm environment. Tip label colors denote the ID of the farm from which each strain was isolated. Branch labels denote posterior probabilities of branch support. Time in years is plotted along the X-axis, and branch lengths are reported in years. The heatmap to the right of the phylogeny denotes (i) the susceptible-intermediate-resistant (SIR) classification of each isolate for each of 15 antimicrobials (obtained using phenotypic testing and NARMS breakpoints; orange); (ii) presence and absence of plasmid replicons (detected using ABRicate/PlasmidFinder and minimum nucleotide identity and coverage thresholds of 80 and 60%, respectively; blue); (iii) presence and absence of antimicrobial resistance (AMR) and stress response genes (identified using AMRFinderPlus and default parameters; green). Core SNPs were identified using Snippy. The phylogeny was constructed using the results of ten independent runs using a strict clock model, the Standard_TPM2 nucleotide substitution model, and the Coalescent Bayesian Skyline population model implemented in BEAST version 2.5.1, with 10% burn-in applied to each run. LogCombiner-2 was used to combine BEAST 2 log files, and TreeAnnotator-2 was used to construct the phylogeny using common ancestor node heights. Abbreviations for the 15 antimicrobials are: AMK, amikacin; AMC, amoxicillin-clavulanic acid; AMP, ampicillin; FOX, cefoxitin; TIO, ceftiofur; CRO, ceftriaxone; CHL, chloramphenicol; CIP, ciprofloxacin; GEN, gentamicin; KAN, kanamycin; NAL, nalidixic acid; STR, streptomycin; SXT, sulfamethoxazole-trimethoprim; SUL, sulfisoxazole; TET, tetracycline.

floR and *APH(3')-Ia*, respectively (Figure 9). The Farm 17 and Farm 62 lineages evolved from a common ancestor predicted to have existed circa November/December 2005 (node age 2005.91, CA node height 95% HPD interval 2003.77–2007.85; Figure 9 and Supplementary Figure 9). All members of the Farm 17 lineage additionally harbored a ColpVC plasmid and shared a common ancestor dated to circa August/September 2007 (node age 2007.65, CA node height 95% HPD interval 2007.29–2007.85; Figure 9 and Supplementary Figure 9). The Farm 62 lineage, which did not possess the ColpVC plasmid, evolved from a common

ancestor circa August/September 2008 (node age 2008.68, CA node height 95% HPD interval 2008.42–2008.78; Figure 9 and Supplementary Figure 9).

Unlike the *S. Newport* lineages present on Farm 17 and Farm 62, the Farm 35 *S. Newport* lineage did not possess kanamycin resistance gene *APH(3')-Ia* and was kanamycin-susceptible (Figure 9). The common ancestor of the Farm 35 *S. Newport* lineage was dated circa May 2008 (node age 2008.38, CA node height 95% HPD interval 2008.06–2008.55). All but one Farm 35 *S. Newport* isolates were additionally resistant to chloramphenicol and possessed *floR*;

TABLE 3 | Summary of within-serotype evolutionary analyses^a.

Serotype	Isolates	Core SNPs (Pairwise Range) ^b	Clock/Population Model ^c	Mean/Median Tree Height in Years (95% HPD Interval) ^d	Mean/Median Evolutionary Rate in Substitutions/Site/Year (95% HPD Interval) ^e
Anatum	15	337 (0–257)	Strict/Skyline	1484.9/1837.0 (549.6–1980.1)	$1.67 \times 10^{-7}/1.48 \times 10^{-7}$ (6.92×10^{-11} – 3.86×10^{-7})
Cerro	13	21 (0–12)	Strict/Skyline	2008.2/2008.4 (2007.6–2008.6)	$9.11 \times 10^{-7}/8.94 \times 10^{-7}$ (3.06×10^{-7} – 1.57×10^{-6})
Kentucky	36	102 (0–30)	Relaxed/Skyline	2004.1/2005.0 (2000.8–2006.8)	$6.39 \times 10^{-7}/6.34 \times 10^{-7}$ (2.05×10^{-7} – 1.07×10^{-6})
Meleagridis	19	27 (0–9)	Strict/Skyline	2007.4/2007.5 (2006.9–2007.7)	$6.88 \times 10^{-7}/6.66 \times 10^{-7}$ (2.90×10^{-7} – 1.12×10^{-6})
Newport	16	52 (0–38)	Relaxed/Skyline	2004.2/2004.6 (2000.4–2007.8)	$9.02 \times 10^{-7}/8.22 \times 10^{-7}$ (2.64×10^{-7} – 1.65×10^{-6})
Typhimurium (Copenhagen)	27	732 (0–634)	Relaxed/Skyline	1936.0/1943.0 (1864.7–1991.4)	$1.07 \times 10^{-6}/9.66 \times 10^{-7}$ (2.84×10^{-7} – 2.05×10^{-6})

^aSee **Supplementary Table 5** for an extended version of this table; note that evolutionary rates may be higher than previously reported estimates for *Salmonella* populations isolated over a longer time frame, due to the small sample sizes and short temporal period characterized here (Moller et al., 2018).

^bNumber of core SNPs identified among all genomes within the serotype after removing recombination with Gubbins; the range of pairwise SNP differences between isolates was calculated using the *dist.gene* function in the *ape* package in R.

^cThe optimal model selected for the data set; can be a combination of a strict or lognormal relaxed molecular clock (“Strict” or “Relaxed,” respectively) and a Constant Coalescent or Coalescent Bayesian Skyline population model (“CC” or “Skyline,” respectively); see **Supplementary Table 5** for more details.

^dThe tree height parameter and its respective 95% highest posterior density (HPD) interval reported by Tracer.

^eCorresponds to the clock Rate and rate.mean parameters estimated by BEAST2 for models using strict and lognormal relaxed molecular clock models, respectively, as reported by Tracer.

BOV_NEWP_35_10-02-08_R8-2688 did not possess *floR* and was chloramphenicol-susceptible (**Figure 9**).

Each of Four Major Lineages Composed of *Salmonella* Typhimurium and Its O5-Copenhagen Variant Is Associated With One of Three New York State Dairy Farms

Twenty-seven bovine and farm environmental *S. Typhimurium* and *S. Typhimurium* Copenhagen isolates that encompassed five PFGE types (**Supplementary Table 1**) were isolated from three dairy farms (1, 10, and 16 strains isolated from Farm 17, 22, and 25, respectively). All isolates queried here shared a common ancestor that existed circa 1936 (node age 1935.62, CA node height 95% HPD interval 1864.84–1991.86; **Figure 10**, **Table 3**, **Supplementary Figure 10**, and **Supplementary Table 5**). Notably, the *S. Typhimurium* Copenhagen variant was polyphyletic (**Figure 10**), regardless of whether traditional or *in silico* (i.e., SeqSero2) methods had been used for serotype variant assignment. Additionally, the *S. Typhimurium*/*S. Typhimurium* Copenhagen isolates sequenced here showcased the most diverse AMR phenotypic profiles and AMR gene presence/absence profiles (**Figures 4**, **10**).

Isolates from Farm 25 were partitioned into two clades: one containing *S. Typhimurium* isolates, and one containing *S. Typhimurium* Copenhagen isolates (based on SeqSero2’s *in silico* serotype assignments; **Figure 10**). Farm 25 isolates assigned to the *S. Typhimurium* Copenhagen variant (i) shared a common ancestor that existed circa December 2007/January 2008 (node age 2007.99, CA node height 95% HPD interval 2007.68–2008.21); (ii) were all resistant to ampicillin, kanamycin,

streptomycin, sulfisoxazole, and tetracycline, with reduced susceptibility to additional antimicrobials observed sporadically; (iii) all possessed replicons for IncA/C2, IncFIB(AP001918), and IncFII(s) plasmids; and (iv) all possessed streptomycin resistance genes *aadA12*, *APH(3’)-Ib* and *APH(6)-Id* (i.e., *strAB*), beta-lactamase *bla*_{TEM-1}, and antiseptic resistance gene *qacE* delta 1, with other AMR/stress response genes present sporadically (**Figure 10** and **Supplementary Figure 10**). Farm 25 isolates assigned to the *S. Typhimurium* clade shared a common ancestor that existed circa July 2007 (node age 2007.55, CA node height 95% HPD interval 2007.18–2007.78; **Figure 10** and **Supplementary Figure 10**). All Farm 25 *S. Typhimurium* isolates were resistant to ceftiofur; resistance to additional antimicrobials, along with presence of IncI1 plasmids and *bla*_{CMY-2}, was observed sporadically (**Figure 10**).

The isolate from Farm 17 was predicted to belong to the *S. Typhimurium* Copenhagen serotype variant using SeqSero2 and shared a common ancestor with the *S. Typhimurium* isolates from Farm 22, which existed circa 2000 (node age 1999.92, CA node height 95% HPD interval 1991.72–2006.21; **Figure 10** and **Supplementary Figure 10**). Of the ten *S. Typhimurium* strains from Farm 22, seven were pan-susceptible (**Figure 10**). A bovine strain (BOV_TYPH_22_03–14–08_R8–0865) was resistant to ampicillin, ceftiofur, and ceftriaxone and was found to harbor IncI1 and IncI2 plasmids, as well as beta-lactamase *bla*_{CTX-M-55} (**Figure 10**). The remaining two bovine isolates were intermediately resistant to chloramphenicol and additionally resistant to either (i) amoxicillin-clavulanic acid, ampicillin, and sulfisoxazole, or (ii) tetracycline (**Figure 10**). Overall, isolates from Farm 17 and Farm 22 shared a common ancestor with the Farm 25 *S. Typhimurium* clade that existed circa 1988 (node age 1988.02, CA node height 95%

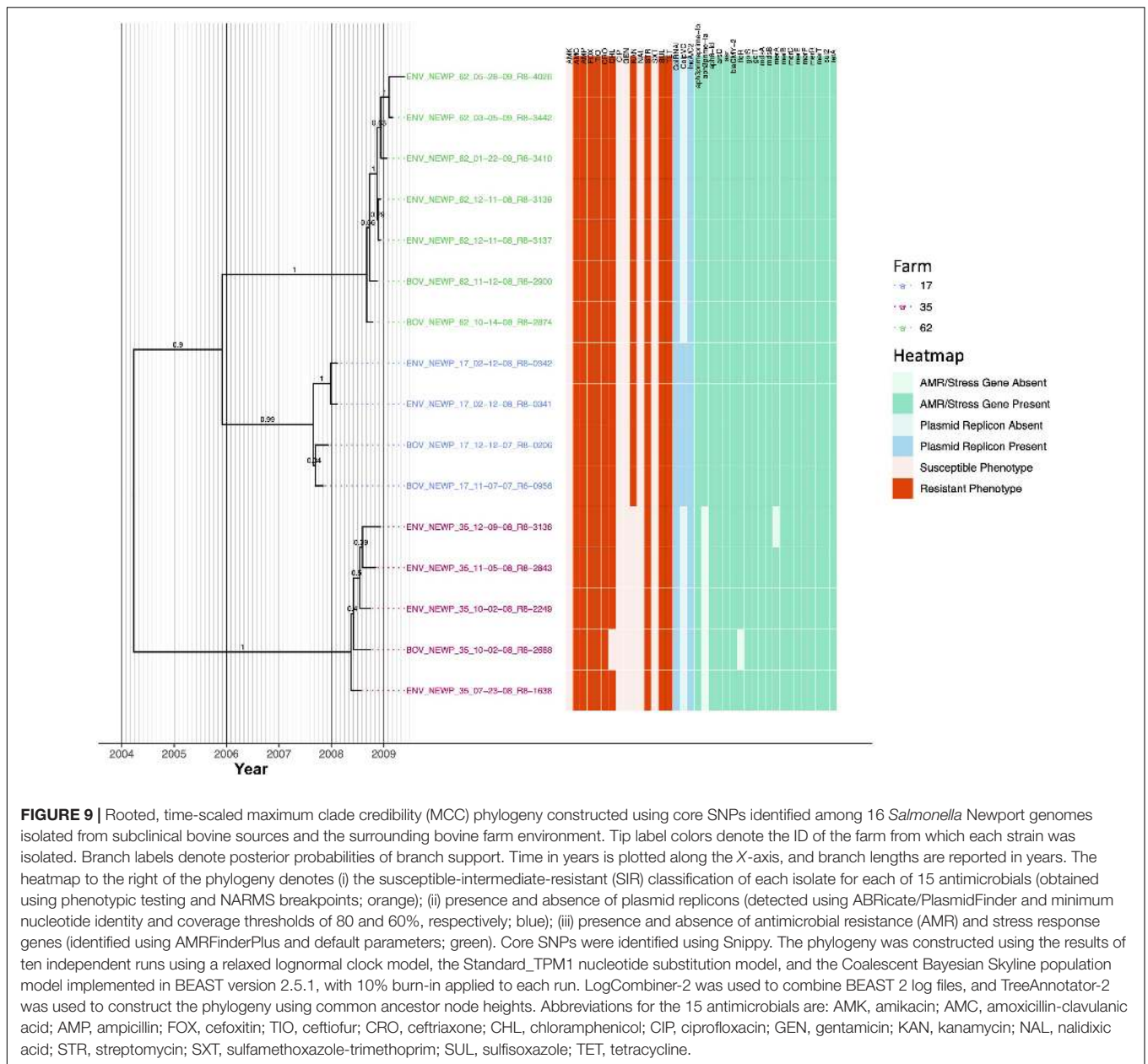


FIGURE 9 | Rooted, time-scaled maximum clade credibility (MCC) phylogeny constructed using core SNPs identified among 16 *Salmonella* Newport genomes isolated from subclinical bovine sources and the surrounding bovine farm environment. Tip label colors denote the ID of the farm from which each strain was isolated. Branch labels denote posterior probabilities of branch support. Time in years is plotted along the X-axis, and branch lengths are reported in years. The heatmap to the right of the phylogeny denotes (i) the susceptible-intermediate-resistant (SIR) classification of each isolate for each of 15 antimicrobials (obtained using phenotypic testing and NARMS breakpoints; orange); (ii) presence and absence of plasmid replicons (detected using ABRicate/PlasmidFinder and minimum nucleotide identity and coverage thresholds of 80 and 60%, respectively; blue); (iii) presence and absence of antimicrobial resistance (AMR) and stress response genes (identified using AMRFinderPlus and default parameters; green). Core SNPs were identified using Snippy. The phylogeny was constructed using the results of ten independent runs using a relaxed lognormal clock model, the Standard_TPM1 nucleotide substitution model, and the Coalescent Bayesian Skyline population model implemented in BEAST version 2.5.1, with 10% burn-in applied to each run. LogCombiner-2 was used to combine BEAST 2 log files, and TreeAnnotator-2 was used to construct the phylogeny using common ancestor node heights. Abbreviations for the 15 antimicrobials are: AMK, amikacin; AMC, amoxicillin-clavulanic acid; AMP, ampicillin; FOX, cefoxitin; TIO, ceftiofur; CRO, ceftriaxone; CHL, chloramphenicol; CIP, ciprofloxacin; GEN, gentamicin; KAN, kanamycin; NAL, nalidixic acid; STR, streptomycin; SXT, sulfamethoxazole-trimethoprim; SUL, sulfisoxazole; TET, tetracycline.

HPD interval 1969.24–2002.47; **Figure 10** and **Supplementary Figure 10**).

DISCUSSION

Whole-Genome Sequencing Can Be Used to Monitor Pathogen Microevolution and Temporal Antimicrobial Resistance Dynamics in Animal Reservoirs

Cattle may act as a reservoir for *Salmonella* and may facilitate its transmission to other animals (Mentaberre et al., 2013;

Wiethoelter et al., 2015) or humans, either through direct contact or via the food supply chain (Hoelzer et al., 2011; Cummings et al., 2012; Mughini-Gras et al., 2014; An et al., 2017; Gutema et al., 2019). Even outside of a bovine host, *Salmonella* can survive in the farm environment for a prolonged amount of time, making persistent strains a particularly relevant threat to animal and human health (Rodriguez et al., 2006; Cummings et al., 2010b; Gorski et al., 2011; Toth et al., 2011; Tassinari et al., 2019). This threat can be compounded when persistent strains are exposed to antimicrobials, as a number of studies have linked antimicrobial exposure to the emergence of AMR in different foodborne pathogens, including *Salmonella*, *Escherichia coli*, and *Campylobacter* (Boerlin et al., 2001; McDermott et al., 2002; Delsol et al., 2003; Dutil et al., 2010; Hoelzer et al., 2017).

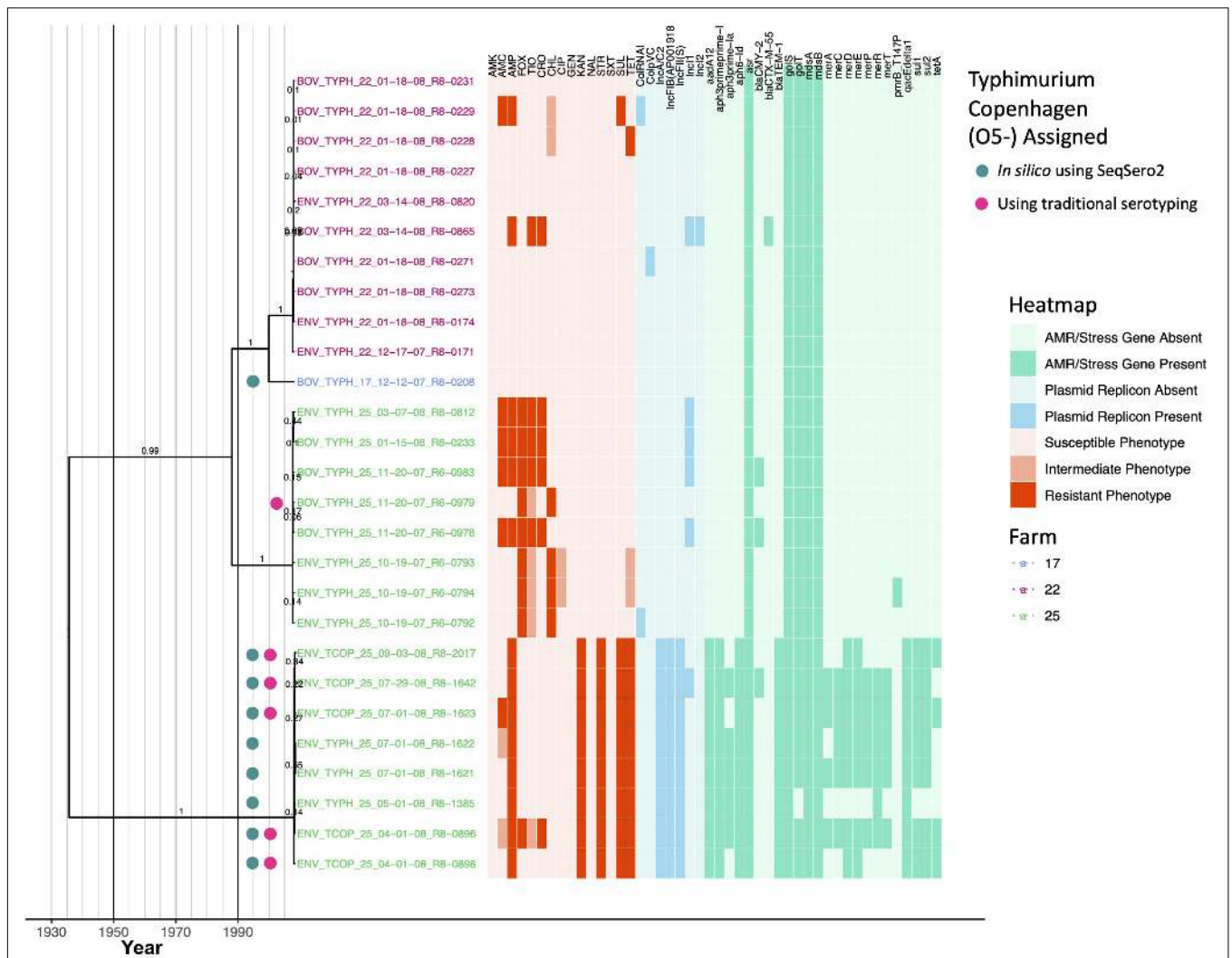


FIGURE 10 | Rooted, time-scaled maximum clade credibility (MCC) phylogeny constructed using core SNPs identified among 27 *Salmonella* Typhimurium and Typhimurium Copenhagen genomes isolated from subclinical bovine sources and the surrounding bovine farm environment. Tip label colors denote the ID of the farm from which each strain was isolated. Circle colors to the left of tip labels denote isolates that were assigned to the Typhimurium Copenhagen variant of *S. Typhimurium* using SeqSero2 (teal) and/or traditional serotyping (pink). Branch labels denote posterior probabilities of branch support. Time in years is plotted along the X-axis, and branch lengths are reported in years. The heatmap to the right of the phylogeny denotes (i) the susceptible-intermediate-resistant (SIR) classification of each isolate for each of 15 antimicrobials (obtained using phenotypic testing and NARMS breakpoints; orange); (ii) presence and absence of plasmid replicons (detected using ABRicate/PlasmidFinder and minimum nucleotide identity and coverage thresholds of 80 and 60%, respectively; blue); (iii) presence and absence of antimicrobial resistance (AMR) and stress response genes (identified using AMRFinderPlus and default parameters; green). Core SNPs were identified using Snippy. The phylogeny was constructed using the results of ten independent runs using a relaxed lognormal clock model, the Standard_TPM1 nucleotide substitution model, and the Coalescent Bayesian Skyline population model implemented in BEAST version 2.5.1, with 10% burn-in applied to each run. LogCombiner-2 was used to combine BEAST 2 log files, and TreeAnnotator-2 was used to construct the phylogeny using common ancestor node heights. Abbreviations for the 15 antimicrobials are: AMK, amikacin; AMC, amoxicillin-clavulanic acid; AMP, ampicillin; FOX, cefoxitin; TIO, ceftiofur; CRO, ceftriaxone; CHL, chloramphenicol; CIP, ciprofloxacin; GEN, gentamicin; KAN, kanamycin; NAL, nalidixic acid; STR, streptomycin; SXT, sulfamethoxazole-trimethoprim; SUL, sulfisoxazole; TET, tetracycline.

However, AMR acquisition among pathogens in livestock environments is far from absolute; in the absence of selective pressures (e.g., antimicrobial exposure), some AMR traits may be associated with a fitness cost for a given organism (Melnik et al., 2015; Hoelzer et al., 2017; San Millan and MacLean, 2017). Consequently, interventions or changes in farm management practices (e.g., limiting antimicrobial use for all or selected antimicrobials, targeted use of

some antimicrobials) may lead to reduced selection of AMR bacteria (Aarestrup, 2015; Tang et al., 2017; Scott et al., 2018). As such, the dynamics of AMR acquisition and loss among livestock-associated bacterial pathogens are complex and influenced by a wide range of factors, including the antimicrobials and treatment regimens used, farm management practices, environmental conditions, and the biology of the pathogens themselves (Aarestrup, 2015;

Hoelzer et al., 2017; Davidson et al., 2018; Pereira et al., 2019; Clarke et al., 2020).

Using a WGS-based approach applied to serially sampled *Salmonella* strains isolated over a short time frame (i.e., less than 2 years), the study detailed here reveals that sporadic acquisition and loss of acquired AMR genes can occur within closely related populations over a short timescale. One particularly notable observation is represented by multiple, independent acquisitions of the beta-lactamase *bla*_{CMY} among *S. Typhimurium* and *S. Typhimurium* Copenhagen, as all *bla*_{CMY} acquisition events within this serotype group were confined to the 2000s. *bla*_{CMY} can confer resistance to cephalosporins, including (i) ceftriaxone, which has been used in human medicine since the early 1980s, and is used to treat invasive salmonellosis cases when fluoroquinolones cannot be used (e.g., for pediatric salmonellosis cases), and (ii) ceftiofur, which has been used in veterinary settings since the late 1980s to treat disease cases among dairy cattle and other animals (Hornish and Kotarski, 2002; Alcaine et al., 2005; Liebana et al., 2013; Yang et al., 2016; Carroll et al., 2017b, 2020a). Because *bla*_{CMY} often confers resistance to both ceftriaxone and ceftiofur, there has been concern that the use of ceftiofur in livestock can contribute to the dissemination of *bla*_{CMY} and thus yield bacterial populations that are co-resistant to ceftriaxone (Alcaine et al., 2005; Tragesser et al., 2006; Carroll et al., 2017b, 2020a).

Two independent *bla*_{CTX-M} acquisition events among *S. Meleagridis* and *S. Typhimurium* were additionally observed. *bla*_{CTX-M}, which also confers resistance to cephalosporins, was rarely detected in the United States in the 1990s (Lewis et al., 2007; Canton et al., 2012). However, *bla*_{CTX-M} rapidly increased in prevalence in the United States between 2000 and 2005 (Lewis et al., 2007; Canton et al., 2012), and there is evidence that bacterial populations associated with dairy cattle may have been affected as well. In a study of *E. coli* isolated from dairy cattle in the western United States, the prevalence of *bla*_{CTX-M} was found to have increased between 2008 and 2012 (Afema et al., 2018). The results of our study are congruent with these findings, as all observed *bla*_{CTX-M} acquisition events were estimated to have occurred in the 2000s.

Antimicrobial resistance loss events were additionally observed among the bovine-associated, MDR *S. Newport* isolates sequenced here. Prevalence of MDR *S. Newport* among humans increased rapidly in the United States within the late 1990s and early 2000s and was linked to cattle exposure, farm/petting zoo exposure, unpasteurized milk consumption, and ground beef consumption (Spika et al., 1987; Gupta et al., 2003; Karon et al., 2007). While chloramphenicol resistance is often a hallmark characteristic of MDR *S. Newport*, the MDR *S. Newport* lineage represented by an isolate in this study was chloramphenicol-susceptible and was predicted to have lost chloramphenicol resistance gene *floR* after 2008. These results indicate that even well-established MDR pathogens can still be subjected to temporal changes in AMR profile.

Due to the global burden that AMR pathogens impose on the health of humans and animals, numerous agencies have called for improved monitoring of pathogens and their associated AMR determinants along the food supply chain

(World Health Organization, 2014, 2017; Centers for Disease Control and Prevention, 2019). The study detailed here showcases how WGS can be used to identify temporal changes in the resistomes of livestock-associated pathogens at the farm level. However, further sequencing efforts querying (i) a larger selection of *Salmonella* strains isolated from livestock on individual farms (ii) over a longer timeframe are needed to determine whether the AMR dynamics observed here are merely sporadic, or rather are indications of larger trends.

Bovine-Associated *Salmonella* Lineages With Heterogeneous Antimicrobial Resistance Profiles May Be Present Across Multiple Farms or Strongly Farm-Associated

Geography has been shown to play an important role in shaping bacterial populations (Achtman, 2008; Strachan et al., 2015), including some *Salmonella* lineages (Carroll et al., 2017b; Palma et al., 2018; Fenske et al., 2019; Liao et al., 2020). However, for some foodborne pathogens, including some *Salmonella* populations, global spread of lineages due to human migration and movement of food and animals can often obfuscate local phylogeographic signals (Wong et al., 2015; Llerena et al., 2016; The et al., 2016; Palma et al., 2018).

In the study detailed here, *Salmonella* lineages isolated from cattle and their associated environments on 13 separate farms in a confined geographic location (i.e., New York State) were found to vary in terms of the farm-specific signal they possessed; some lineages (i.e., *S. Anatum*, *S. Newport*, *S. Typhimurium*, *S. Typhimurium* Copenhagen, some *S. Kentucky* populations) were found to be strongly associated with a particular farm, while other lineages (i.e., *S. Cerro*, *S. Meleagridis*, some *S. Kentucky* populations) were distributed across multiple farms. Multiple scenarios may explain the existence of *Salmonella* lineages distributed across multiple farms, including movement of livestock, humans, pets, and/or wildlife (Skov et al., 2008; Hoelzer et al., 2011; Palma et al., 2018) or introduction via feed; however, additional metadata (e.g., farm geography, proximity to other farms in the study, and management practices) are needed to draw further conclusions. Even with limited metadata available, WGS data can provide important insights into *Salmonella* transmission and introduction on farms, as shown in this study. For example, for one farm (i.e., Farm 25), two *Salmonella* Typhimurium clonal groups were present (i.e., one representing Typhimurium and one representing Typhimurium Copenhagen), each of which shared a common ancestor dated circa 2007. WGS data can be used to identify time frames in which *Salmonella* lineages may have emerged in a given farm or region, which could help pinpoint root causes (e.g., changes in management practices that occurred around the predicted time of emergence).

While the characterization of additional, larger strain sets from more geographically diverse farms is essential, our data suggest that specific *Salmonella* clones may persist on a given farm. This suggests that WGS databases covering isolates from a large number of farms could be used to develop initial hypotheses

about farm sources of *Salmonella* strains. While such applications are tempting, it is crucial that these types of data are only used for initial hypothesis generation; rigorous, critical epidemiological investigations are essential before any conclusions regarding strain source are drawn.

***In silico* Serotyping of Bovine-Associated *Salmonella* Can Outperform Traditional Serotyping**

Well into the genomic era, serotyping remains a vital microbiological assay that allows *Salmonella* isolates to be classified into meaningful, evolutionary units. Serotype assignments are used to facilitate outbreak investigations and surveillance efforts, construct salmonellosis risk assessment frameworks, and inform food safety and public health policy and decision-making efforts (Yoshida et al., 2016; Gutema et al., 2019). Importantly, serotyping is used worldwide to monitor salmonellosis cases among humans and animals, including cattle (Gutema et al., 2019; Centers for Disease Control and Prevention, 2020).

In this study, serotypes assigned using traditional phenotypic methods were compared to serotypes assigned using two *in silico* methods (i.e., SISTR and SeqSero2). Notably, both *in silico* serotyping approaches outperformed traditional *Salmonella* serotyping for this data set. Serotypes assigned using SISTR's cgMLST approach and/or SeqSero2 were congruent with the *Salmonella* whole-genome phylogeny and were able to resolve all un-typable, ambiguous, and incorrectly assigned serotypes (Supplementary Table 1). It is essential to note that the data set queried here is far too small and, thus, inadequate to formally benchmark these tools. Furthermore, all serotypes studied here were among the ten most frequently reported serotypes of *Salmonella* isolated from subclinical cattle between 2000 and 2017 (Gutema et al., 2019), indicating that they are well-represented in public databases and thus likely do not pose a significant challenge to *in silico* tools. However, the results observed here reflect observations made in several recent studies, which queried greater numbers of isolate genomes and/or a wider array of diverse *Salmonella* serotypes (Yachison et al., 2017; Ibrahim and Morin, 2018; Diep et al., 2019; Banerji et al., 2020; Uelze et al., 2020). In their analysis of 1,624 animal- and food-associated (i.e., non-human) *Salmonella* isolate genomes assigned to 72 serotypes, Uelze et al. (2020) reported that SISTR and SeqSero2 achieved the highest and second-highest accuracy of all tested *in silico* *Salmonella* serotype assignment tools, correctly serotyping 94 and 87% of isolates, respectively. However, unlike the results observed here, the authors note that neither tool outperformed traditional serotyping conducted by *Salmonella* reference laboratories. Similarly, in a study of 813 *Salmonella* isolates, SISTR outperformed the original version of SeqSero (i.e., SeqSero 1.0) with serotype prediction accuracies of 94.8 and 88.2%, respectively (Yachison et al., 2017).

With WGS data in hand, *in silico* serotyping is rapid, scalable, inexpensive, and reproducible (Uelze et al., 2020). Nevertheless, it is important to be mindful of the strengths and weaknesses of different *in silico* serotyping tools. In their benchmarking

study, Uelze et al. (2020) recommended SISTR as the optimal contemporary tool for routine *in silico* *Salmonella* serotyping based on overall accuracy; however, they additionally report that the raw read mapping approach implemented in SeqSero2 (i.e., "allele mode") outperforms SISTR for prediction of monophasic variants. Banerji et al. (2020) did not assess the performance of SISTR on their data set, as it requires assembled genomes and not raw reads (another potential drawback if a high-quality assembly is not available or obtainable for an isolate of interest); however, they found that both SeqSero and MLST approaches misidentified monophasic variants, particularly among the important monophasic *S. Typhimurium* lineage. Among the bovine-associated *Salmonella* strains sequenced here, a combination of *S. Typhimurium* strains that possessed the O5 epitope, and those that did not (i.e., *S. Typhimurium* Copenhagen) was observed. Importantly, SISTR was unable to differentiate *S. Typhimurium* from *S. Typhimurium* Copenhagen, while SeqSero2 could, as reported previously (Ibrahim and Morin, 2018; Zhang et al., 2019). While the differentiation of *S. Typhimurium* from its O5- counterpart may not be essential for all microbiological applications, it is important to be aware of this limitation; *S. Typhimurium* Copenhagen has been responsible for outbreaks and illnesses around the world (Luceron et al., 2017; Tack et al., 2020) and can be multidrug-resistant, as demonstrated here and elsewhere (Frech et al., 2003; Tack et al., 2020).

While serotypes assigned *in silico* using SISTR and SeqSero2 are highly accurate and congruent, each tool has strengths and limitations; as such, an approach that utilizes both methods, such as the one employed here, may increase accuracy and minimize potential misclassifications. Results from other studies support this (Yachison et al., 2017; Banerji et al., 2020). For example, in an analysis of 520 primarily human-associated *Salmonella* isolate genomes, Banerji et al. (2020) found that serotypes assigned *in silico* using SeqSero showed 98% concordance with traditional serotyping and outperformed serotype assignment using seven-gene MLST. However, when SeqSero and seven-gene MLST were used in combination, *in silico* serotyping accuracy surpassed 99%, consistent with our results that a combination of SeqSero2 and cgMLST-based serotyping (as implemented in SISTR) improved *in silico* serotyping accuracy. Overall, the results provided here lend further support to the idea that *in silico* serotyping may eventually replace traditional serotyping as WGS becomes more widely used and accessible (Yachison et al., 2017; Banerji et al., 2020; Uelze et al., 2020).

Limitations of the *in silico* Antimicrobial Resistance Method Evaluation Presented Here, and Considerations for Future Antimicrobial Resistance Monitoring Efforts Among Livestock and Beyond

In addition to studying the microevolution and AMR dynamics of bovine *Salmonella* on a genomic scale, the study presented here compared results obtained from numerous *in silico* AMR characterization pipelines that attempt to replicate traditional

microbiological assays used to characterize AMR *Salmonella*. More specifically, each of the following tools was applied to the set of 128 bovine-associated *Salmonella* genomes sequenced here: (i) combinations of five *in silico* AMR determinant detection pipelines (i.e., ABRicate, AMRFinderPlus, ARIBA, BTyper, and SRST2) and one to five AMR determinant databases (i.e., ARG-ANNOT, CARD, MEGARes, NCBI, and ResFinder); and (ii) an *in silico* MIC prediction tool (i.e., PATRIC3).

Here, all AMR determinant detection pipelines and AMR determinant databases showed an extremely high degree of concordance; regardless of pipeline or database selection, all tools performed nearly identically on an SIR-prediction task relative to (i) “true” SIR classifications based on NARMS breakpoints and “true” MIC values obtained for a panel of 15 antimicrobials, and (ii) each other. A previous, small-scale ($n = 111$) WGS-based study of AMR *Salmonella* observed similarly high rates of concordance among several *in silico* AMR determinant detection tools (Cooper et al., 2020). However, in addition to its small sample size, this study also relied on *Salmonella* strains isolated from a single source (broiler chickens) in a single country (Canada) over an extremely short temporal range (December 2012–2013). Similarly, the study detailed here is not a formal benchmarking study, and it is essential that its numerous limitations are pointed out.

First and foremost, the study conducted here relied on WGS data from an extremely small sample of *Salmonella* isolates ($n = 128$) from a single source (dairy cattle and their surrounding farm environments) in a confined geographic area (New York State, United States) isolated over a short temporal range (fewer than 2 years). While all isolates were “unique” (i.e., each strain was isolated from a separate sampling event of a unique source), many isolates were highly similar at both the genomic and pan-genomic level (e.g., *S. Cerro*, *S. Meleagridis*), indicating that, in some cases, the same lineage was being sampled repeatedly over time. Consequently, this relatively minuscule sample is unrepresentative of AMR pathogens and, more specifically, *Salmonella* as a whole; readers should not infer the general superiority or inferiority of any AMR detection tool or database tested, and the results obtained here should not be extrapolated to external data sets.

Secondly, the data set queried here was heavily biased toward susceptible isolates. More than half of all isolates were pan-susceptible to the 15 antimicrobials included on the panel, and only 21 unique phenotypic SIR profiles were observed. Congruent with this, relatively little diversity was observed in terms of AMR gene profile (e.g., the AMRFinderPlus pipeline produced 20 unique AMR/stress response determinant presence/absence profiles among the 128 isolates sequenced here). This is not particularly surprising; numerous studies have shown that the resistomes of bovine-associated *Salmonella* tend to be less diverse than *Salmonella* isolated from humans (Afema et al., 2015; Carroll et al., 2017b), as well as some other animals (Mellor et al., 2019). Furthermore, the resistomes of *Salmonella* isolated from subclinical cattle, such as the isolates queried in this study, have been shown to be less diverse than the resistomes of

Salmonella isolated from cattle showing clinical signs of disease (Afema et al., 2015).

The relative homogeneity of the subclinical *Salmonella* bovine resistome and bias toward antimicrobial-susceptible isolates have important implications for the AMR pipeline/database comparison conducted here. For this data set, stringent and conservative approaches are rewarded, as isolates that do not possess AMR determinants are more likely to be predicted to be susceptible. While it is possible that different AMR detection tools may perform better on WGS data from pathogens with more diverse resistomes, very few formal benchmarking studies of *in silico* AMR determinant detection tools currently exist (Hendriksen et al., 2019). The choice of AMR determinant database in combination with the choice of pipeline, on the other hand, can clearly affect AMR determinant identification in a critical way. For example, the ARG-ANNOT database (Gupta et al., 2014), a manually curated AMR determinant database first published in 2014, is not updated as frequently as other AMR databases (e.g., CARD, NCBI, and ResFinder). Since it was last updated in May 2018 (accessed May 25, 2020), at the time of our study, the database does not yet include three novel plasmid-mediated genes (*mcr-8*, *-9*, and *-10*) that can confer resistance to colistin, a last-resort antibiotic used to treat MDR and extensively drug resistant infections (Wang et al., 2018, 2020; Carroll et al., 2019). Similarly, versions of tools that rely on even older versions of this database would not be able to detect all members of the continuously growing repertoire of *mcr* genes. For the low-diversity subclinical bovine *Salmonella* resistomes queried here, the use of a smaller database was inconsequential, as reflected in the high congruency of all methods and databases observed here. For some studies, a manually curated database of AMR genes that is updated conservatively may possibly even be desirable, as such a database may yield less noise and improve interpretability and reproducibility. However, for pathogens with more diverse resistomes (e.g., human clinical isolates, isolates from geographic regions with different antibiotic use practices), the omission of critically important genes could be a disastrous flaw. Similar to our results, a large study ($n = 6,242$) querying NARMS isolates belonging primarily to the *S. enterica* species ($n = 5,425$) observed a high degree of concordance between NCBI’s AMRFinder tool and ResFinder (this study was used to validate AMRFinder and the NCBI AMR determinant database) (Feldgarden et al., 2019). However, when differences between tools were observed, the vast majority (81%) were attributed to differences in database composition (Feldgarden et al., 2019; Hendriksen et al., 2019).

Thirdly, the small sample size ($n = 128$) and sparsity of AMR isolates available in this study limited the methods that could be used to formulate the AMR tool/database comparison. The approach used here was similar to the one used to validate the AMRFinder tool (Feldgarden et al., 2019) in that it relied on known AMR-determinant/AMR phenotype associations available in the literature (see Supplementary Table 4 of Feldgarden et al.) (Feldgarden et al., 2019). As such, the approach used here does not account for previously unobserved genotype/phenotype associations. Furthermore, different variants of the same AMR gene may

yield different AMR phenotypes; for example, some variants of the OXA beta-lactamases are able to confer resistance to cephalosporins, while others are not (Evans and Amyes, 2014). All AMR determinant detection pipelines tested here produced nearly identical genes calls among the 128 isolates sequenced here, and all detected AMR determinants were manually annotated in a consistent fashion; while the overall accuracy of all pipeline/database combinations tested here could likely be optimized if more accurate, data set-specific AMR genotype/phenotype associations could be derived, congruency among tools/databases would likely remain high.

Classifying bacterial pathogens into discrete SIR groups using AMR determinant detection methods is challenging, as it requires users to have a great degree of prior knowledge regarding the AMR determinants that are detected, the antimicrobials of interest, and the pathogen being studied. PATRIC3's MIC prediction tool (Nguyen et al., 2019) offers a promising departure from this framework, as it allows for the prediction of MIC values directly from WGS data. Interpreting the resulting *in silico* MIC values does not require any prior knowledge on the user end, and results can be harmoniously integrated into the SIR framework using clinical breakpoints. Among the *Salmonella* isolates sequenced here, SIR classification using PATRIC3 resulted in an overall accuracy of 93%. However, all of the limitations of this study described above for AMR determinant detection (e.g., small sample size, AMR sparsity bias, single-source, single geographic region, small temporal range) apply to *in silico* MIC prediction as well; for example, when "intermediate" resistance predictions produced via PATRIC3 are re-classified as "susceptible" (as was done for the AMR determinant detection approaches used here), PATRIC3's accuracy for this data set increases to 96.0% and is on par with all other AMR prediction methods tested here. Readers should thus interpret comparisons among these methods with caution.

Benchmarking and validating AMR detection and prediction tools is notoriously challenging (Feldgarden et al., 2019), and very few researchers have undertaken this task (Hendriksen et al., 2019). While high congruency may be observed among tools (Clausen et al., 2016), identification of a clear "optimal" method for *in silico* AMR characterization has remained elusive, as the few available benchmarking studies differ in terms of the tools tested, the AMR database(s) used, and the data set(s) chosen for benchmarking. Furthermore, the underlying WGS data can affect pipeline performance (Clausen et al., 2016; Feldgarden et al., 2019). For example, assembly quality has been shown to influence AMR determinant detection for methods that rely on assembled genomes (Clausen et al., 2016; Hendriksen et al., 2018, 2019). Thus, whether a read- or assembly-based method performs optimally can depend on a given data set (e.g., sequencing quality, the organism being studied). Another criticism of BLAST-based AMR gene detection methods among assembled genomes has been the choice of thresholds used for considering AMR determinants present or absent (Hendriksen et al., 2019). Here, no significant differences were observed between the accuracy of read-based ARIBA and SRST2 and assembly-based ABRicate, AMRFinderPlus, and BType. Additionally, for BLAST-based methods, a relatively wide range of optimal nucleotide identity

and coverage values were found to maximize accuracy, with thresholds of 75% identity and 50–60% coverage adequate for most pipeline/database combinations. Overall, when selecting an *in silico* AMR characterization method, researchers should take into account not only practical considerations (e.g., whether reads or assembled genomes are available, the quality of reads and/or assembled genomes), but also the biology of the pathogen being studied (e.g., by querying organism-specific, AMR-conferring point mutations). To assess the robustness of *in silico* AMR predictions, researchers may additionally consider employing multiple *in silico* AMR characterization tools and/or databases in combination, as well as testing various AMR gene detection thresholds.

Finally, it is essential to note that accuracy estimates for *in silico* AMR characterization tools relative to gold-standard phenotypic methods are only as reliable as the phenotypic data they rely on. Previous studies of *Salmonella* that attempted to predict phenotypic AMR using *in silico* methods (McDermott et al., 2016; Cooper et al., 2020) have reported accuracy values between 98 and 100%. For this data set, the highest accuracy achieved was 97.4% (for SRST2/ARG-ANNOT). However, sensitivity (i.e., the ability of an *in silico* pipeline/database combination to correctly classify an isolate as phenotypically resistant to an antimicrobial) was lower for this study (71.8–84.4%) than sensitivity estimates calculated in a study of MDR *Salmonella* that included bovine isolates from New York State (97.2%) (Carroll et al., 2017b). As mentioned above, this could be due to the sparsity of AMR among isolates in this data set (i.e., predicting susceptible, rather than resistant phenotypes, is incentivized here). However, it is important to note that the AMR phenotypes of several isolates were highly incongruent with their respective AMR genotypes, regardless of the tool/database used for *in silico* AMR prediction. For example, one *S. Cerro* isolate (BOV_CERO_35_10-02-08_R8-2685) was reported to be phenotypically resistant to nine antimicrobials but did not possess any acquired AMR determinants known to produce this phenotypic AMR profile. A recent case study in which WGS and phenotypic methods were used to characterize *Salmonella* isolates from raw chicken identified numerous AMR genotype/phenotype discrepancies resulting in both false negative and false positive predictions for *in silico* methods (Zwe et al., 2020). In this case study, the authors attributed these discrepancies to heteroresistant *Salmonella* subpopulations (i.e., a subpopulation of bacteria that exhibits a range of susceptibility to a particular antimicrobial). The possibility that several heteroresistant *Salmonella* populations were characterized here cannot be discounted, as isolates underwent phenotypic AMR characterization and WGS separately (i.e., years apart). Other biological phenomena, such as plasmid loss during storage or culturing, or unknown/undetected resistance genes or mutations, could also contribute to discrepancies (Hendriksen et al., 2018). However, it is also possible that one or more incongruent isolates was mislabeled and/or mishandled during AMR phenotyping, genomic DNA extraction, and/or WGS. While removal of these isolates from the data set would increase overall prediction accuracy, the high congruency among AMR genotyping methods would be unaffected.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

LC designed and carried out all computational analyses. AB, AG, JS, KC, and RC collected, analyzed, and/or interpreted all microbiological data. LC and MW conceived the study and co-wrote the manuscript with input from all authors. All authors contributed to the article and approved the submitted version.

REFERENCES

- Aarestrup, F. M. (2015). 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. doi: 10.1098/rstb.2014.0085
- Achtman, M. (2008). Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu. Rev. Microbiol.* 62, 53–70. doi: 10.1146/annurev.micro.62.081307.162832
- Afema, J. A., Ahmed, S., Besser, T. E., Jones, L. P., Sischo, W. M., and Davis, M. A. (2018). Molecular epidemiology of dairy cattle-associated *Escherichia coli* carrying blaCTX-M genes in Washington State. *Appl. Environ. Microbiol.* 84:e02430-17. doi: 10.1128/AEM.02430-17
- Afema, J. A., Mather, A. E., and Sischo, W. M. (2015). Antimicrobial resistance profiles and diversity in *Salmonella* from humans and Cattle, 2004–2011. *Zoonoses Public Health* 62, 506–517. doi: 10.1111/zph.12172
- Agren, E. C., Wahlstrom, H., Vesterlund-Carlson, C., Lahti, E., Melin, L., and Soderlund, R. (2016). Comparison of whole genome sequencing typing results and epidemiological contact information from outbreaks of *Salmonella* Dublin in Swedish cattle herds. *Infect. Ecol. Epidemiol.* 6:31782. doi: 10.3402/iee.v6.31782
- Alcaine, S. D., Sukhnanand, S. S., Warnick, L. D., Su, W. L., McGann, P., McDonough, P., et al. (2005). Ceftiofur-resistant *Salmonella* strains isolated from dairy farms represent multiple widely distributed subtypes that evolved by independent horizontal gene transfer. *Antimicrob. Agents Chemother.* 49, 4061–4067. doi: 10.1128/AAC.49.10.4061-4067.2005
- An, R., Alshalchi, S., Breimhurst, P., Munoz-Aguayo, J., Flores-Figueroa, C., and Vidovic, S. (2017). Strong influence of livestock environments on the emergence and dissemination of distinct multidrug-resistant phenotypes among the population of non-typhoidal *Salmonella*. *PLoS One* 12:e0179005. doi: 10.1371/journal.pone.0179005
- Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* 26, 32–46. doi: 10.1046/j.1442-9993.2001.01070.x
- Anderson, M. J. (2006). Distance-Based tests for homogeneity of multivariate dispersions. *Biometrics* 62, 245–253. doi: 10.1111/j.1541-0420.2005.00440.x
- Anderson, M. J., and Walsh, D. C. I. (2013). PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: what null hypothesis are you testing? *Ecol. Monogr.* 83, 557–574. doi: 10.1890/12-2010.1
- Andrews, S. (2019). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. 0.11.8 ed.
- Banerji, S., Simon, S., Tille, A., Fruth, A., and Fliieger, A. (2020). Genome-based *Salmonella* serotyping as the new gold standard. *Sci. Rep.* 10:4333. doi: 10.1038/s41598-020-67917-3
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

FUNDING

This material is based on work supported by the National Science Foundation Graduate Research Fellowship Program under grant No. DGE-1650441. This project was supported in part by the Cornell University Zoonoses Research Units of the Food and Waterborne Diseases Integrated Research Network, funded by the National Institute of Allergy and Infectious Diseases, National Institutes of Health, under contract No. N01-AI-30054.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2021.763669/full#supplementary-material>

- Boerlin, P., Wissing, A., Aarestrup, F. M., Frey, J., and Nicolet, J. (2001). Antimicrobial growth promoter ban and resistance to macrolides and vancomycin in enterococci from pigs. *J. Clin. Microbiol.* 39, 4193–4195. doi: 10.1128/JCM.39.11.4193-4195.2001
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Boore, A. L., Hoekstra, R. M., Iwamoto, M., Fields, P. I., Bishop, R. D., and Swerdlow, D. L. (2015). *Salmonella enterica* infections in the united states and assessment of coefficients of variation: a novel approach to identify epidemiologic characteristics of individual serotypes, 1996–2011. *PLoS One* 10:e0145416. doi: 10.1371/journal.pone.0145416
- Bouckaert, R., Heled, J., Kuhnert, D., Vaughan, T., Wu, C. H., Xie, D., et al. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 10:e1003537. doi: 10.1371/journal.pcbi.1003537
- Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchene, S., Fourment, M., Gavryushkina, A., et al. (2019). BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 15:e1006650. doi: 10.1371/journal.pcbi.1006650
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421
- Canton, R., Gonzalez-Alba, J. M., and Galán, J. C. (2012). CTX-M enzymes: origin and diffusion. *Front. Microbiol.* 3:110. doi: 10.3389/fmicb.2012.00110
- Carattoli, A., Zankari, E., Garcia-Fernandez, A., Voldby Larsen, M., Lund, O., Villa, L., et al. (2014). *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* 58, 3895–3903. doi: 10.1128/AAC.02412-14
- Carroll, L. M., Gaballa, A., Guldemann, C., Sullivan, G., Henderson, L. O., and Wiedmann, M. (2019). Identification of novel mobilized colistin resistance gene mcr-9 in a multidrug-resistant, colistin-susceptible *Salmonella enterica* serotype Typhimurium isolate. *mBio* 10:e00853-19. doi: 10.1128/mBio.00853-19
- Carroll, L. M., Huisman, J. S., and Wiedmann, M. (2020a). Twentieth-century emergence of antimicrobial resistant human- and bovine-associated *Salmonella enterica* serotype Typhimurium lineages in New York State. *Sci. Rep.* 10:14428. doi: 10.1038/s41598-020-71344-9
- Carroll, L. M., Wiedmann, M., and Kovac, J. (2020b). Proposal of a taxonomic nomenclature for the *Bacillus cereus* group which reconciles genomic definitions of bacterial species with clinical and industrial phenotypes. *mBio* 11:e00034-20. doi: 10.1101/779199
- Carroll, L. M., Kovac, J., Miller, R. A., and Wiedmann, M. (2017a). Rapid, high-throughput identification of anthrax-causing and emetic *Bacillus cereus* group genome assemblies via BType, a computational tool for virulence-based classification of *Bacillus cereus* group isolates by using nucleotide sequencing data. *Appl. Environ. Microbiol.* 83:e01096-17. doi: 10.1128/AEM.01096-17
- Carroll, L. M., Wiedmann, M., Den Bakker, H., Siler, J., Warchocki, S., Kent, D., et al. (2017b). Whole-Genome sequencing of drug-resistant *Salmonella*

- enterica* isolates from dairy cattle and humans in New York and Washington States reveals source and geographic associations. *Appl. Environ. Microbiol.* 83:e00140-17. doi: 10.1128/AEM.00140-17
- Centers for Disease Control and Prevention (2019). *Antibiotic Resistance Threats in the United States, 2019*. Atlanta, GA: U.S. Department of Health and Human Services.
- Centers for Disease Control and Prevention (2020). *Serotypes and the Importance of Serotyping Salmonella*. CDC. Available online at: <https://www.cdc.gov/salmonella/reportspubs/salmonella-atlas/serotyping-importance.html> (accessed May 25, 2021).
- Centers for Disease Control and Prevention (2021). *Salmonella*. Available online at: <https://www.cdc.gov/salmonella/index.html> (accessed September 24, 2021).
- Chiu, C. H., Su, L. H., and Chu, C. (2004). *Salmonella enterica* serotype Choleraesuis: epidemiology, pathogenesis, clinical disease, and treatment. *Clin. Microbiol. Rev.* 17, 311–322. doi: 10.1128/CMR.17.2.311-322.2004
- Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* 18, 117–143. doi: 10.1111/j.1442-9993.1993.tb00438.x
- Clarke, L., Pelin, A., Phan, M., and Wong, A. (2020). The effect of environmental heterogeneity on the fitness of antibiotic resistance mutations in *Escherichia coli*. *Evol. Ecol.* 34, 379–390. doi: 10.1007/s10682-019-10027-y
- Clausen, P. T., Zankari, E., Aarestrup, F. M., and Lund, O. (2016). Benchmarking of methods for identification of antimicrobial resistance genes in bacterial whole genome data. *J. Antimicrob. Chemother.* 71, 2484–2488. doi: 10.1093/jac/dkw184
- Cooper, A. L., Low, A. J., Koziol, A. G., Thomas, M. C., Leclair, D., Tamber, S., et al. (2020). Systematic evaluation of whole genome sequence-based predictions of *Salmonella* serotype and antimicrobial resistance. *Front. Microbiol.* 11:549. doi: 10.3389/fmicb.2020.00549
- Croucher, N. J., Page, A. J., Connor, T. R., Delaney, A. J., Keane, J. A., Bentley, S. D., et al. (2015). Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* 43:e15. doi: 10.1093/nar/gku1196
- Cummings, K. J., Warnick, L. D., Davis, M. A., Eckmann, K., Grohn, Y. T., Hoelzer, K., et al. (2012). Farm animal contact as risk factor for transmission of bovine-associated *Salmonella* subtypes. *Emerg. Infect. Dis.* 18, 1929–1936. doi: 10.3201/eid1812.110831
- Cummings, K. J., Warnick, L. D., Elton, M., Grohn, Y. T., McDonough, P. L., and Siler, J. D. (2010a). The effect of clinical outbreaks of salmonellosis on the prevalence of fecal *Salmonella* shedding among dairy cattle in New York. *Foodborne Pathog. Dis.* 7, 815–823. doi: 10.1089/fpd.2009.0481
- Cummings, K. J., Warnick, L. D., Elton, M., Rodriguez-Rivera, L. D., Siler, J. D., Wright, E. M., et al. (2010b). *Salmonella enterica* serotype Cerro among dairy cattle in New York: an emerging pathogen? *Foodborne Pathog. Dis.* 7, 659–665. doi: 10.1089/fpd.2009.0462
- Davidson, K. E., Byrne, B. A., Pires, A. F. A., Magdesian, K. G., and Pereira, R. V. (2018). Antimicrobial resistance trends in fecal *Salmonella* isolates from northern California dairy cattle admitted to a veterinary teaching hospital, 2002–2016. *PLoS One* 13:e0199928. doi: 10.1371/journal.pone.0199928
- Delgado-Suarez, E. J., Selem-Mojica, N., Ortiz-Lopez, R., Gebreyes, W. A., Allard, M. W., Barona-Gomez, F., et al. (2018). Whole genome sequencing reveals widespread distribution of typhoidal toxin genes and VirB/D4 plasmids in bovine-associated nontyphoidal *Salmonella*. *Sci. Rep.* 8:9864. doi: 10.1038/s41598-018-28169-4
- Delsol, A. A., Anjum, M., Woodward, M. J., Sunderland, J., and Roe, J. M. (2003). The effect of chlortetracycline treatment and its subsequent withdrawal on multi-resistant *Salmonella enterica* serovar Typhimurium DT104 and commensal *Escherichia coli* in the pig. *J. Appl. Microbiol.* 95, 1226–1234. doi: 10.1046/j.1365-2672.2003.02088.x
- Diep, B., Barretto, C., Portmann, A. C., Fournier, C., Karczmarek, A., Voets, G., et al. (2019). *Salmonella* serotyping: comparison of the traditional method to a microarray-based method and an *in silico* platform using whole genome sequencing data. *Front. Microbiol.* 10:2554. doi: 10.3389/fmicb.2019.02554
- Dutil, L., Irwin, R., Finley, R., Ng, L. K., Avery, B., Boerlin, P., et al. (2010). Ceftiofur resistance in *Salmonella enterica* serovar Heidelberg from chicken meat and humans, Canada. *Emerg. Infect. Dis.* 16, 48–54. doi: 10.3201/eid1601.090729
- Evans, B. A., and Amyes, S. G. B. (2014). OXA β -Lactamases. *Clin. Microbiol. Rev.* 27, 241–263. doi: 10.1128/CMR.00117-13
- Ewels, P., Magnusson, M., Lundin, S., and Kaller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048. doi: 10.1093/bioinformatics/btw354
- Feldgarden, M., Brover, V., Haft, D. H., Prasad, A. B., Slotta, D. J., Tolstoy, I., et al. (2019). Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrob. Agents Chemother.* 63:e00483-19. doi: 10.1128/AAC.00483-19
- Fenske, G. J., Thachil, A., McDonough, P. L., Glaser, A., and Scaria, J. (2019). Geography shapes the population genomics of *Salmonella enterica* dublin. *Genome Biol. Evol.* 11, 2220–2231. doi: 10.1093/gbe/evz158
- Frech, G., Kehrenberg, C., and Schwarz, S. (2003). Resistance phenotypes and genotypes of multiresistant *Salmonella enterica* subsp. *enterica* serovar Typhimurium var. Copenhagen isolates from animal sources. *J. Antimicrob. Chemother.* 51, 180–182. doi: 10.1093/jac/dkg058
- Gardner, S. N., and Hall, B. G. (2013). When whole-genome alignments just won't work: kSNP v2 software for alignment-free SNP discovery and phylogenetics of hundreds of microbial genomes. *PLoS One* 8:e81760. doi: 10.1371/journal.pone.0081760
- Gardner, S. N., Slezak, T., and Hall, B. G. (2015). kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics* 31, 2877–2878. doi: 10.1093/bioinformatics/btv271
- Gorski, L., Parker, C. T., Liang, A., Cooley, M. B., Jay-Russell, M. T., Gordus, A. G., et al. (2011). Prevalence, distribution, and diversity of *Salmonella enterica* in a major produce region of California. *Appl. Environ. Microbiol.* 77, 2734–2748. doi: 10.1128/AEM.02321-10
- Gupta, A., Fontana, J., Crowe, C., Bolstorff, B., Stout, A., Duyn, S. V., et al. (2003). Emergence of multidrug-resistant *Salmonella enterica* serotype newport infections resistant to expanded-spectrum Cephalosporins in the United States. *J. Infect. Dis.* 188, 1707–1716. doi: 10.1086/379668
- Gupta, S. K., Padmanabhan, B. R., Diene, S. M., Lopez-Rojas, R., Kempf, M., Landraud, L., et al. (2014). ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob. Agents Chemother.* 58, 212–220. doi: 10.1128/AAC.01310-13
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUASt: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086
- Gutema, F. D., Agga, G. E., Abdi, R. D., De Zutter, L., Duchateau, L., and Gabriel, S. (2019). Prevalence and serotype diversity of *Salmonella* in apparently healthy cattle: systematic review and meta-analysis of published studies, 2000–2017. *Front. Vet. Sci.* 6:102. doi: 10.3389/fvets.2019.00184
- Harvey, R. R., Friedman, C. R., Crim, S. M., Judd, M., Barrett, K. A., Tolar, B., et al. (2017). Epidemiology of *Salmonella enterica* serotype Dublin infections among humans, United States, 1968–2013. *Emerg. Infect. Dis.* 23, 1493–1501. doi: 10.3201/eid2309.170136
- Hendriksen, R. S., Bortolaia, V., Tate, H., Tyson, G. H., Aarestrup, F. M., and Mcdermott, P. F. (2019). Using genomics to track global antimicrobial resistance. *Front. Public Health* 7:242. doi: 10.3389/fpubh.2019.00242
- Hendriksen, R. S., Pedersen, S. K., Leekitcharoenphon, P., Malorny, B., Borowiak, M., Battisti, A., et al. (2018). Final report of ENGAGE – establishing next generation sequencing ability for genomic analysis in Europe. *EFSA Support. Publ.* 15:1431E. doi: 10.2903/sp.efsa.2018.EN-1431
- Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281
- Hoelzer, K., Moreno Switt, A. I., and Wiedmann, M. (2011). Animal contact as a source of human non-typhoidal salmonellosis. *Vet. Res.* 42:34. doi: 10.1186/1297-9716-42-34
- Hoelzer, K., Wong, N., Thomas, J., Talkington, K., Jungman, E., and Coukell, A. (2017). Antimicrobial drug use in food-producing animals and associated human health risks: what, and how strong, is the evidence? *BMC Vet. Res.* 13:211. doi: 10.1186/s12917-017-1131-3
- Holschbach, C. L., and Peek, S. F. (2018). *Salmonella* in dairy cattle. *Vet. Clin. North Am. Food Anim. Pract.* 34, 133–154. doi: 10.1016/j.cvfa.2017.10.005

- Hornish, R. E., and Kotarski, S. F. (2002). Cephalosporins in veterinary medicine - ceftiofur use in food animals. *Curr. Top. Med. Chem.* 2, 717–731. doi: 10.2174/1568026023393679
- Hunt, M., Mather, A. E., Sanchez-Buso, L., Page, A. J., Parkhill, J., Keane, J. A., et al. (2017). ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb. Genom.* 3:e000131. doi: 10.1099/mgen.0.000131
- Ibrahim, G. M., and Morin, P. M. (2018). *Salmonella* serotyping using whole genome sequencing. *Front. Microbiol.* 9:2993. doi: 10.3389/fmicb.2018.02993
- Inouye, M., Dashnow, H., Raven, L. A., Schultz, M. B., Pope, B. J., Tomita, T., et al. (2014). SRST2: rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med.* 6:90. doi: 10.1186/s13073-014-0090-6
- Issenhuth-Jeanjean, S., Roggentin, P., Mikoleit, M., Guibourdenche, M., De Pinna, E., Nair, S., et al. (2014). Supplement 2008-2010 (no. 48) to the White-Kauffmann-Le Minor scheme. *Res. Microbiol.* 165, 526–530. doi: 10.1016/j.resmic.2014.07.004
- Jia, B., Raphenya, A. R., Alcock, B., Wagglechner, N., Guo, P., Tsang, K. K., et al. (2017). CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 45, D566–D573. doi: 10.1093/nar/gkw1004
- Jorgensen, J. H., and Ferraro, M. J. (2009). Antimicrobial susceptibility testing: a review of general principles and contemporary practices. *Clin. Infect. Dis.* 49, 1749–1755. doi: 10.1086/647952
- Karon, A. E., Archer, J. R., Sotir, M. J., Monson, T. A., and Kazmierczak, J. J. (2007). Human multidrug-resistant *Salmonella* Newport infections, Wisconsin, 2003–2005. *Emerg. Infect. Dis.* 13, 1777–1780. doi: 10.3201/eid1311.061138
- Kodama, Y., Shumway, M., Leinonen, R., and International Nucleotide Sequence Database Collaboration (2012). The sequence read archive: explosive growth of sequencing data. *Nucleic Acids Res.* 40, D54–D56. doi: 10.1093/nar/gkr854
- Kruskal, J. B. (1964). Nonmetric multidimensional scaling: a numerical method. *Psychometrika* 29, 115–129. doi: 10.1007/BF02289694
- Kuhn, M. (2008). Building predictive models in R using the caret package. *J. Stat. Softw.* 28, 1–26. doi: 10.18637/jss.v028.i05
- Lakin, S. M., Dean, C., Noyes, N. R., Dettenwanger, A., Ross, A. S., Doster, E., et al. (2017). MEGARes: an antimicrobial resistance database for high throughput sequencing. *Nucleic Acids Res.* 45, D574–D580. doi: 10.1093/nar/gkw1009
- Leinonen, R., Sugawara, H., Shumway, M., and International Nucleotide Sequence Database Collaboration (2011). The sequence read archive. *Nucleic Acids Res.* 39, D19–D21. doi: 10.1093/nar/gkq1019
- Lewis, J. S. II, Herrera, M., Wickes, B., Patterson, J. E., and Jorgensen, J. H. (2007). First report of the emergence of CTX-M-type extended-spectrum beta-lactamases (ESBLs) as the predominant ESBL isolated in a U.S. health care system. *Antimicrob. Agents Chemother.* 51, 4015–4021. doi: 10.1128/AAC.00576-07
- Liao, J., Orsi, R. H., Carroll, L. M., and Wiedmann, M. (2020). Comparative genomics reveals different population structures associated with host and geographic origin in antimicrobial-resistant *Salmonella enterica*. *Environ. Microbiol.* 22, 2811–2828. doi: 10.1111/1462-2920.15014
- Liao, J., Orsi, R. H., Carroll, L. M., Kovac, J., Ou, H., Zhang, H., et al. (2019). Serotype-specific evolutionary patterns of antimicrobial-resistant *Salmonella enterica*. *BMC Evol. Biol.* 19:132. doi: 10.1186/s12862-019-1457-5
- Liebana, E., Carattoli, A., Coque, T. M., Hasman, H., Magiorakos, A. P., Mevius, D., et al. (2013). Public health risks of enterobacterial isolates producing extended-spectrum beta-lactamases or AmpC beta-lactamases in food and food-producing animals: an EU perspective of epidemiology, analytical methods, risk factors, and control options. *Clin. Infect. Dis.* 56, 1030–1037. doi: 10.1093/cid/cis1043
- Llarena, A. K., Zhang, J., Vehkala, M., Valimaki, N., Hakkinen, M., Hanninen, M. L., et al. (2016). Monomorphic genotypes within a generalist lineage of *Campylobacter jejuni* show signs of global dispersion. *Microb. Genom.* 2:e000088. doi: 10.1099/mgen.0.000088
- Luceron, C. O., Meixeira, A. P., Sanz, I. A., Deleyto, V. C., Leon, S. H., and Ruiz, L. G. (2017). Notes from the field: an outbreak of *Salmonella* Typhimurium associated with playground sand in a preschool setting - Madrid, Spain, September–October 2016. *MMWR Morb. Mortal Wkly. Rep.* 66, 256–257. doi: 10.15585/mmwr.mm6609a3
- Mather, A. E., Reid, S. W., Maskell, D. J., Parkhill, J., Fookes, M. C., Harris, S. R., et al. (2013). Distinguishable epidemics of multidrug-resistant *Salmonella* Typhimurium DT104 in different hosts. *Science* 341, 1514–1517. doi: 10.1126/science.1240578
- McDermott, P. F., Bodeis, S. M., English, L. L., White, D. G., Walker, R. D., Zhao, S., et al. (2002). Ciprofloxacin resistance in *Campylobacter jejuni* evolves rapidly in chickens treated with fluoroquinolones. *J. Infect. Dis.* 185, 837–840. doi: 10.1086/339195
- McDermott, P. F., Tyson, G. H., Kabera, C., Chen, Y., Li, C., Folster, J. P., et al. (2016). Whole-genome sequencing for detecting antimicrobial resistance in nontyphoidal *Salmonella*. *Antimicrob. Agents Chemother.* 60, 5515–5520. doi: 10.1128/AAC.01030-16
- Mellor, K. C., Petrovska, L., Thomson, N. R., Harris, K., Reid, S. W. J., and Mather, A. E. (2019). Antimicrobial resistance diversity suggestive of distinct *Salmonella* Typhimurium sources or selective pressures in food-production animals. *Front. Microbiol.* 10:708. doi: 10.3389/fmicb.2019.00708
- Melnyk, A. H., Wong, A., and Kassen, R. (2015). The fitness costs of antibiotic resistance mutations. *Evol. Appl.* 8, 273–283. doi: 10.1111/eva.12196
- Mentaberre, G., Porrero, M. C., Navarro-Gonzalez, N., Serrano, E., Dominguez, L., and Lavin, S. (2013). Cattle drive *Salmonella* infection in the wildlife-livestock interface. *Zoonoses Public Health* 60, 510–518. doi: 10.1111/zph.12028
- Minh, B. Q., Nguyen, M. A., and Von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 30, 1188–1195. doi: 10.1093/molbev/mst024
- Mohammed, M., Le Hello, S., Leekitcharoenphon, P., and Hendriksen, R. (2017). The invasome of *Salmonella* Dublin as revealed by whole genome sequencing. *BMC Infect. Dis.* 17:544. doi: 10.1186/s12879-017-2628-x
- Moller, S., Du Plessis, L., and Stadler, T. (2018). Impact of the tree prior on estimating clock rates during epidemic outbreaks. *Proc. Natl. Acad. Sci. U.S.A.* 115, 4200–4205. doi: 10.1073/pnas.1713314115
- Mughini-Gras, L., Enserink, R., Friesema, I., Heck, M., Van Duynhoven, Y., and Van Pelt, W. (2014). Risk factors for human salmonellosis originating from pigs, cattle, broiler chickens and egg laying hens: a combined case-control and source attribution analysis. *PLoS One* 9:e87933. doi: 10.1371/journal.pone.0087933
- Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Nguyen, M., Long, S. W., McDermott, P. F., Olsen, R. J., Olson, R., Stevens, R. L., et al. (2019). Using machine learning to predict antimicrobial MICs and associated genomic features for nontyphoidal *Salmonella*. *J. Clin. Microbiol.* 57:e01260-18. doi: 10.1128/JCM.01260-18
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., et al. (2019). *vegan: Community Ecology Package*. R package version 2.5-6. Available online at: <https://CRAN.R-project.org/package=vegan> (accessed May 22, 2020).
- Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693. doi: 10.1093/bioinformatics/btv421
- Page, A. J., Taylor, B., Delaney, A. J., Soares, J., Seemann, T., Keane, J. A., et al. (2016). SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb. Genom.* 2:e000056. doi: 10.1099/mgen.0.000056
- Palma, F., Manfreda, G., Silva, M., Parisi, A., Barker, D. O. R., Taboada, E. N., et al. (2018). Genome-wide identification of geographical segregated genetic markers in *Salmonella enterica* serovar Typhimurium variant 4,[5],12:i. *Sci. Rep.* 8:15251. doi: 10.1038/s41598-018-33266-5
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114
- Pereira, R., Williams, D. R., Rossitto, P., Adaska, J., Okello, E., Champagne, J., et al. (2019). Association between herd management practices and antimicrobial resistance in *Salmonella* spp. from cull dairy cattle in Central California. *PeerJ* 7:e6546. doi: 10.7717/peerj.6546
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. 3.6.1 ed. Vienna: R Foundation for Statistical Computing.
- Rambaut, A., Lam, T. T., Max Carvalho, L., and Pybus, O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* 2:vev007. doi: 10.1093/ve/vev007
- Rodriguez, A., Pangloli, P., Richards, H. A., Mount, J. R., and Draughon, F. A. (2006). Prevalence of *Salmonella* in diverse environmental farm samples. *J. Food Prot.* 69, 2576–2580. doi: 10.4315/0362-028X-69.11.2576

- Rodriguez-Rivera, L. D., Wright, E. M., Siler, J. D., Elton, M., Cummings, K. J., Warnick, L. D., et al. (2014). Subtype analysis of *Salmonella* isolated from subclinically infected dairy cattle and dairy farm environments reveals the presence of both human- and bovine-associated subtypes. *Vet. Microbiol.* 170, 307–316. doi: 10.1016/j.vetmic.2014.02.013
- San Millan, A., and MacLean, R. C. (2017). Fitness costs of plasmids: a limit to plasmid transmission. *Microbiol. Spectr.* 5, 1–12. doi: 10.1128/microbiolspec.MTBP-0016-2017
- Scott, A. M., Beller, E., Glasziou, P., Clark, J., Ranakusuma, R. W., Byambasuren, O., et al. (2018). Is antimicrobial administration to food animals a direct threat to human health? A rapid systematic review. *Int. J. Antimicrob. Agents* 52, 316–323. doi: 10.1016/j.ijantimicag.2018.04.005
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Seemann, T. (2018). *ABRicate: Mass Screening of Contigs for Antimicrobial Resistance or Virulence Genes*.
- Seemann, T. (2019). *Snippy: Rapid Haploid Variant Calling and Core Genome Alignment*. 4.3.6 ed.
- Skov, M. N., Madsen, J. J., Rahbek, C., Lodal, J., Jespersen, J. B., Jorgensen, J. C., et al. (2008). Transmission of *Salmonella* between wildlife and meat-production animals in Denmark. *J. Appl. Microbiol.* 105, 1558–1568. doi: 10.1111/j.1365-2672.2008.03914.x
- Spika, J. S., Waterman, S. H., Hoo, G. W., St Louis, M. E., Pacer, R. E., James, S. M., et al. (1987). Chloramphenicol-resistant *Salmonella* Newport traced through hamburger to dairy farms. A major persisting source of human salmonellosis in California. *N. Engl. J. Med.* 316, 565–570. doi: 10.1056/NEJM198703053161001
- Strachan, N. J. C., Rotariu, O., Lopes, B., Macrae, M., Fairley, S., Laing, C., et al. (2015). Whole genome sequencing demonstrates that geographic variation of *Escherichia coli* O157 genotypes dominates host association. *Sci. Rep.* 5:14145. doi: 10.1038/srep14145
- Tack, B., Phoba, M. F., Barbe, B., Kalonji, L. M., Hardy, L., Van Puyvelde, S., et al. (2020). Non-typhoidal *Salmonella* bloodstream infections in Kisantu, DR Congo: emergence of O5-negative *Salmonella* Typhimurium and extensive drug resistance. *PLoS Negl. Trop. Dis.* 14:e0008121. doi: 10.1371/journal.pntd.0008121
- Tang, K. L., Caffrey, N. P., Nobrega, D. B., Cork, S. C., Ronksley, P. E., Barkema, H. W., et al. (2017). Restricting the use of antibiotics in food-producing animals and its associations with antibiotic resistance in food-producing animals and human beings: a systematic review and meta-analysis. *Lancet Planet Health* 1, e316–e327. doi: 10.1016/S2542-5196(17)30141-9
- Tassinari, E., Duffy, G., Bawn, M., Burgess, C. M., McCabe, E. M., Lawlor, P. G., et al. (2019). Microevolution of antimicrobial resistance and biofilm formation of *Salmonella* Typhimurium during persistence on pig farms. *Sci. Rep.* 9:8832. doi: 10.1038/s41598-019-45216-w
- Taylor, D. N., Bied, J. M., Munro, J. S., and Feldman, R. A. (1982). *Salmonella* Dublin infections in the United States, 1979–1980. *J. Infect. Dis.* 146, 322–327. doi: 10.1093/infdis/146.3.322
- The, H. C., Thanh, D. P., Holt, K. E., Thomson, N. R., and Baker, S. (2016). The genomic signatures of *Shigella* evolution, adaptation and geographical spread. *Nat. Rev. Microbiol.* 14, 235–250. doi: 10.1038/nrmicro.2016.10
- Toth, J. D., Aceto, H. W., Rankin, S. C., and Dou, Z. (2011). Survival characteristics of *Salmonella enterica* serovar Newport in the dairy farm environment. *J. Dairy Sci.* 94, 5238–5246. doi: 10.3168/jds.2011-4493
- Tragesser, L. A., Wittum, T. E., Funk, J. A., Winokur, P. L., and Rajala-Schultz, P. J. (2006). Association between ceftiofur use and isolation of *Escherichia coli* with reduced susceptibility to ceftriaxone from fecal samples of dairy cows. *Am. J. Vet. Res.* 67, 1696–1700. doi: 10.2460/ajvr.67.10.1696
- Uelze, L., Borowiak, M., Deneke, C., Szabo, I., Fischer, J., Tausch, S. H., et al. (2020). Performance and accuracy of four open-source tools for *in silico* serotyping of *salmonella* spp. based on whole-genome short-read sequencing data. *Appl. Environ. Microbiol.* 86:e02265-19. doi: 10.1128/AEM.02265-19
- Uzzau, S., Brown, D. J., Wallis, T., Rubino, S., Leori, G., Bernard, S., et al. (2000). Host adapted serotypes of *Salmonella enterica*. *Epidemiol. Infect.* 125, 229–255. doi: 10.1017/S0950268899004379
- Wang, C., Feng, Y., Liu, L., Wei, L., Kang, M., and Zong, Z. (2020). Identification of novel mobile colistin resistance gene mcr-10. *Emerg. Microbes Infect.* 9, 508–516. doi: 10.1080/22221751.2020.1732231
- Wang, X., Wang, Y., Zhou, Y., Li, J., Yin, W., Wang, S., et al. (2018). Emergence of a novel mobile colistin resistance gene, mcr-8, in NDM-producing *Klebsiella pneumoniae*. *Emerg. Microbes Infect.* 7:122. doi: 10.1038/s41426-018-0124-z
- Wiethoelter, A. K., Beltran-Alcrudo, D., Kock, R., and Mor, S. M. (2015). Global trends in infectious diseases at the wildlife-livestock interface. *Proc. Natl. Acad. Sci. U.S.A.* 112, 9662–9667. doi: 10.1073/pnas.1422741112
- Wong, V. K., Baker, S., Pickard, D. J., Parkhill, J., Page, A. J., Feasey, N. A., et al. (2015). Phylogeographical analysis of the dominant multidrug-resistant H58 clade of *Salmonella* Typhi identifies inter- and intracontinental transmission events. *Nat. Genet.* 47, 632–639. doi: 10.1038/ng.3281
- World Health Organization (2014). *Antimicrobial Resistance: Global Report on Surveillance*. Geneva: World Health Organization (WHO).
- World Health Organization (2017). *Critically Important Antimicrobials for Human Medicine, 5th Revision*. Geneva: World Health Organization.
- Worley, J., Meng, J., Allard, M. W., Brown, E. W., and Timme, R. E. (2018). *Salmonella enterica* phylogeny based on whole-genome sequencing reveals two new clades and novel patterns of horizontally acquired genetic elements. *MBio* 9:e02303-18. doi: 10.1128/mBio.02303-18
- Yachison, C. A., Yoshida, C., Robertson, J., Nash, J. H. E., Kruczkiewicz, P., Taboada, E. N., et al. (2017). The validation and implications of using whole genome sequencing as a replacement for traditional serotyping for a national *Salmonella* reference laboratory. *Front. Microbiol.* 8:1044. doi: 10.3389/fmicb.2017.01044
- Yang, W.-C., Chan, O.-W., Wu, T.-L., Chen, C.-L., Su, L.-H., and Chiu, C.-H. (2016). Development of ceftriaxone resistance in *Salmonella enterica* serotype Oranienburg during therapy for bacteremia. *J. Microbiol. Immunol. Infect.* 49, 41–45. doi: 10.1016/j.jmii.2014.01.011
- Yoshida, C. E., Kruczkiewicz, P., Laing, C. R., Lingohr, E. J., Gannon, V. P., Nash, J. H., et al. (2016). The *Salmonella in silico* typing resource (SISTR): an open web-accessible tool for rapidly typing and subtyping draft *Salmonella* genome assemblies. *PLoS One* 11:e0147101. doi: 10.1371/journal.pone.0147101
- Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., et al. (2012). Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* 67, 2640–2644. doi: 10.1093/jac/dks261
- Zhang, S., Den Bakker, H. C., Li, S., Chen, J., Dinsmore, B. A., Lane, C., et al. (2019). SeqSero2: rapid and improved *Salmonella* serotype determination using whole-genome sequencing data. *Appl. Environ. Microbiol.* 85:e01746-19. doi: 10.1128/AEM.01746-19
- Zwe, Y. H., Chin, S. F., Kohli, G. S., Aung, K. T., Yang, L., and Yuk, H.-G. (2020). Whole genome sequencing (WGS) fails to detect antimicrobial resistance (AMR) from heteroresistant subpopulation of *Salmonella enterica*. *Food Microbiol.* 91:103530. doi: 10.1016/j.fm.2020.103530

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Carroll, Buehler, Gaballa, Siler, Cummings, Cheng and Wiedmann. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.