Monitoring the Antimicrobial Resistance Dynamics 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, Martin Wiedmann¹

¹Department of Food Science, Cornell University, Ithaca, New York, USA

²Department of Population Medicine and Diagnostic Sciences, Cornell University, Ithaca, New York, USA

* Correspondence:

Laura M. Carroll
lmc297@cornell.edu

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

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

Abstract

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, we use whole-genome sequencing (WGS) 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 between 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.

1 Introduction

The foodborne pathogen *Salmonella enterica* is estimated to be responsible for 1.2 million illnesses and 450 deaths each year in the U.S. alone (Scallan et al., 2011). 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 *Salmonella 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, we characterized the microevolution of these persistent lineages within each herd, 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, we evaluate the accuracy and concordance of multiple *in silico* serotyping and AMR prediction tools. Finally, we provide an in-depth, critical analysis of the strengths and

limitations of the methods used here, and we offer guidance to researchers who wish to employ WGS for herd-level pathogen monitoring.

2 Materials and Methods

2.1 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 S1). 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).

2.2 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 x 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 QUAST 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 S1).

2.3 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 S1). In cases where a discrepancy existed between 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 SNP Identification and Phylogeny Construction” below).

2.4 *In Silico* Antimicrobial Resistance Determinant Detection

Antimicrobial resistance (AMR) determinants were detected in each of the 128 *Salmonella* genomes using five separate approaches: (i) ABRicate (<https://github.com/tseemann/abricate>) 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 .ffn) 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 Antimicrobial Resistance Reference Gene Database (NCBI) (Feldgarden et al., 2019), and (iv) the ResFinder database (ResFinder) (Zankari et al., 2012). For each genome and database combination, minimum AMR gene identity and coverage thresholds ranging from 50-100% (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*.

2.5 *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. The following were used as dependencies: `kmc` version 3.0 (Kokot et al., 2017) and `XGBoost` version 0.82 (Chen and Guestrin, 2016).

2.6 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) 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 S1). 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 S1), 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).

Supplemental Table S4 of the AMRFinder validation paper (Feldgarden et al., 2019) was used to identify 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”) and all of the 15 antimicrobials tested in this study (Supplementary Table S2). If a detected AMR gene was not identified in the AMRFinder literature search, the gene name was queried in CARD, and the resulting literature linked to the CARD entry was searched for known AMR genotype/phenotype associations (Supplementary Table S2). 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 S2). 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 S3) 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 µg/mL, 4 µg/mL, 8 µg/mL).

2.7 Re-testing of Isolates with Highly Incongruent AMR 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 S4). 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 S4). 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*, *tetA*), which conferred multidrug resistance in closely related *S. Newport* isolates (Supplementary Table S4). To address these incongruities, 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 S4).

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, USA) agar plates from frozen glycerol stocks. Precultures 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 12-14h. The precultures 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 hours. 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 hours. MIC values were determined by evaluating the inhibition zone using a magnifying lens according to the manufacturer’s instructions.

MIC 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 S4.

2.8 *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).

2.9 Reference-Free SNP 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. (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); the optimal model was the transversion model with unequal, empirical base frequencies, an ascertainment bias correction, and the FreeRate model with six categories, i.e., the TVM +F+ASC+R6 model (Yang, 1995; Lewis, 2001; Soubrier et al., 2012; Kalyaanamoorthy et al., 2017). 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) and the following dependencies: ape (Paradis et al., 2004; Paradis and Schliep, 2019), dplyr (Wickham et al., 2020), ggtree (Yu et al., 2017; Yu et al., 2018), phylobase (R Hackathon, 2019), phytools (Revell, 2012), and reshape2 (Wickham, 2007).

2.10 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 besPLOT (<https://github.com/lmc297/besPLOT>) (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 and the following dependencies in R: vegan version 2.5-6 (Oksanen et al., 2019), shiny version 1.4.0.2 (Chang et al., 2020), ggplot2 version 3.3.0 (Wickham, 2016), plyr version 1.8.6 (Wickham, 2011), dplyr version 0.8.5 (Wickham et al., 2020), cluster version 2.1.0 (Maechler et al., 2019), and ggrepel version 0.8.2 (Slowikowski, 2020).

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.

2.11 Reference-Based Core SNP 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 4.3.6 (<https://github.com/tseemann/snippy>) (Seemann, 2019b) 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), one of six high-quality assembled genomes from isolates in this study as a reference genome (Supplementary Table S1), and the following dependencies: BWA MEM version 0.7.13-r1126 (Li and Durbin, 2009; Li, 2013), Minimap2 version 2.15 (Li, 2018), SAMtools version 1.8 (Li et al., 2009), BEDtools version 2.27.1 (Quinlan and Hall, 2010; Quinlan, 2014), BCFtools version 1.8 (Li, 2011), FreeBayes version v1.1.0-60-gc15b070 (Garrison and Marth, 2012), vcflib version v1.0.0-rc2 (Cleary et al., 2015), vt version 0.57721 (Tan et al., 2015), SnpEff version 4.3T (Cingolani et al., 2012), samclip version 0.2 (Seemann, 2019a), seqtk version 1.2-r102-dirty (Li, 2019), and snp-sites version 2.4.0 (Page et al., 2016). 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 to produce an alignment of core SNPs for each serotype.

2.12 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 SNP 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 S5) (Rambaut et al., 2016).

A tip-dated phylogeny was then constructed for each serotype using BEAST version 2.5.0 (Bouckaert et al., 2014; Bouckaert et al., 2019), the appropriate core SNP alignment, an initial clock rate of 2.79×10^{-7} substitutions/site/year (Leekitcharoenphon et al., 2016), and an ascertainment bias correction to account for the use of solely variant sites (Bouckaert, 2014). The modelTest function in R’s phangorn package (Schliep, 2011) was used to select an appropriate nucleotide substitution model for each serotype, and the implementation of the model with the lowest BIC value in BEAST 2’s SSN package (Bouckaert and Xie, 2017) was selected. For each serotype, combinations of either a strict or lognormal relaxed molecular clock (Drummond et al., 2006) and either a coalescent constant or coalescent Bayesian Skyline (Drummond et al., 2005) population model were tested. For all models, a lognormal prior was used for the clockRate/uclMean parameters (in real space, $M = 2.058 \times 10^{-6}$ and $S = 2.0$, yielding a median value of 2.79×10^{-7} substitutions/site/year, and 2.5% and 97.5% quantiles of 5.53×10^{-9} and 1.40×10^{-5} substitutions/site/year, respectively). For each molecular clock and population model combination for each serotype, three independent stepping stone sampling analyses (Xie et al., 2011) were performed with BEAST 2, using 10 steps of at least 100 million generations. Bayes factors (Kass and Raftery, 1995) were calculated to determine the optimal combination of molecular clock and population model for each serotype (Supplementary Table S5). Ten independent BEAST 2 runs were then performed using the optimal model for each serotype, using chain lengths of at least 100 million generations, sampling every 10,000 generations. Tracer

version 1.7.1 (Rambaut et al., 2018) was used to ensure adequate mixing of each independent run. The resulting log and tree files were aggregated using LogCombiner-2, and TreeAnnotator-2 (Heled and Bouckaert, 2013) was used to produce a maximum clade credibility tree using 10% burn-in. The phylogenies were annotated using R and the following packages: ggplot2, ggtree, and phylobase. Bayesian skyline plots for all serotype groups analyzed using BEAST2 are also available (Supplementary Figure S1); however, due to the limited number of available isolates surveyed among each serotype and the short temporal range queried here, potential changes in effective population sizes may not be robust and are thus not discussed in this study.

2.13 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 S1. All BEAST 2 XML files used for temporal phylogeny construction are available at https://github.com/lmc297/zru_farms.

3 Results

3.1 *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 S1). 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 S1). 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 S1). 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 S1), 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 S1).

When variants of the *S. Typhimurium* serotype ($n = 27$) were considered, discrepancies were observed between traditional serotype assignments and both *in silico* methods (Supplementary Table S1). 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 S1), 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 S1). 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 S1). 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, serotypes assigned consistently with at least two out of the three methods (i.e., traditional serotyping, SeqSero2, and SISTR cgMLST) were 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 S1). The remaining two farms harbored three *Salmonella* serotypes (Farms 17 and 62 harbored Kentucky/Newport/Typhimurium and Cerro/Minnesota/Newport, respectively; Supplementary Table S1).

3.2 *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 S1 and S4).

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 S2 and S3). Furthermore, all pipeline/database combinations showed an extremely high degree of concordance (98.0% or greater for all pipeline/database combinations; Supplementary Figure S2). The overall accuracy of all *in silico* AMR determinant detection pipeline/database combinations ranged from 95.8-97.4%, 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 S3). 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 S3). Sensitivity ranged from 71.8-84.4%, with SRST2 achieving the highest sensitivities (84.4 and 84.0% for the ARG-ANNOT and ResFinder databases, respectively; Table 1 and Supplementary Table S3).

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 S3) 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 S3). For ABRicate/CARD, minimum thresholds of 60% coverage and 75% nucleotide identity were optimal (Supplementary Figure S3). 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 S3).

The performance of the PATRIC3 *in silico* MIC prediction method was additionally evaluated (Figure 3 and Supplementary Figure S4). PATRIC3 was able to correctly classify *Salmonella* isolates as susceptible-intermediate-resistant (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 S4).

3.3 Genomic AMR 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 (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).

3.4 Each of two New York State dairy farms harbors a unique, bovine-associated *S. Anatum* lineage

Fifteen *S. Anatum* strains encompassing four PFGE types (Supplementary Table S1) 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 S5).

S. Anatum isolates from Farm 39 shared a common ancestor circa 2005 (node age 2004.69, node height 95% HPD 1978.46-2007.69; Figure 5 and Supplementary Figure S5). 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).

S. 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 S5), 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 S5). 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.

3.5 A closely related *S. Cerro* lineage spans two New York State dairy farms

Thirteen *S. Cerro* strains encompassing two PFGE types (Supplementary Table S1) 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 S6, and Supplementary Table S5). 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 S6). 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 S4).

3.6 *S. 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 S1) 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 S7, and Supplementary Table S5). 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 IncI1 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 S7). 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 S7). No corresponding genes that may encode for reduced chloramphenicol susceptibility were identified in these two isolates.

3.7 A clonal *S. 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 S1) 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 S8, and Supplementary Table S5). All but three branches within the *S. Meleagridis* phylogeny had low support ($PP \leq 0.41$; Figure 8 and Supplementary Figure S8), 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) cefoxitin, sulfisoxazole, and tetracycline, or (ii) ceftiofur (Figure 8).

3.8 Kanamycin resistance among each of three New York State dairy farms harboring a distinct, multidrug-resistant *S. Newport* lineage is farm-associated

Sixteen *S. Newport* isolates encompassing three PFGE types (Supplementary Table S1) were isolated from one 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 S9, and Supplementary Table S5).

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 *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 S9). 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 S9). 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 S9).

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

3.9 Each of four major lineages composed of *S. 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 S1) 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 S10, and Supplementary Table S5). 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 and 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 S10). 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 S10). All Farm 25 *S. Typhimurium* isolates were resistant to cefoxitin; 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 S10). 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% HPD interval 1969.24-2002.47; Figure 10 and Supplementary Figure S10).

4 Discussion

4.1 WGS can be used to monitor pathogen microevolution and temporal AMR 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).

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 two 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; Carroll et al., 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; Carroll et al., 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.

AMR 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.

4.2 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

694 movement of food and animals can often obfuscate local phylogeographic signals (Wong et al., 2015;
695 Llarena et al., 2016; The et al., 2016; Palma et al., 2018).

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

713 While the characterization of additional, larger strain sets from more geographically diverse
714 farms is essential, our data suggest that specific *Salmonella* clones may persist on a given farm. This
715 suggests that WGS databases covering isolates from a large number of farms could be used to
716 develop initial hypotheses about farm sources of *Salmonella* strains. While such applications are
717 tempting, it is crucial that these types of data are only used for initial hypothesis generation; rigorous,
718 critical epidemiological investigations are essential before any conclusions regarding strain source
719 are drawn.

720 **4.3 *In silico* serotyping of bovine-associated *Salmonella* can outperform traditional serotyping**

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

728 In this study, serotypes assigned using traditional phenotypic methods were compared to
729 serotypes assigned using two *in silico* methods (i.e., SISTR and SeqSero2). Notably, both *in silico*
730 serotyping approaches outperformed traditional *Salmonella* serotyping for this data set. Serotypes
731 assigned using SISTR's cgMLST approach and/or SeqSero2 were congruent with the *Salmonella*
732 whole-genome phylogeny and were able to resolve all un-typable, ambiguous, and incorrectly
733 assigned serotypes (Supplementary Table S1). It is essential to note that the data set queried here is
734 far too small and, thus, inadequate to formally benchmark these tools. Furthermore, all serotypes
735 studied here were among the ten most frequently reported serotypes of *Salmonella* isolated from
736 subclinical cattle between 2000 and 2017 (Gutema et al., 2019), indicating that they are well-
737 represented in public databases and thus likely do not pose a significant challenge to *in silico* tools.

However, the results observed here reflect results observed in several recent studies querying larger 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. (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. (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. (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, we observed a combination of *S. Typhimurium* strains that possessed the O5 epitope, and those that did not (i.e., *S. Typhimurium* Copenhagen). 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. (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).

4.4 Limitations of the *in silico* AMR method evaluation presented here, and considerations for future AMR 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, BType, 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-December 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 two 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 miniscule 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 towards 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 towards 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, 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; Carroll et al., 2019; Wang et al., 2020). 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 *Salmonella 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 Supplemental Table S4 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 between 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 between 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 between 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; Hendriksen et al., 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 BTypers. 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 between AMR genotyping methods would be unaffected.

5 Author Contributions

LMC designed and carried all computational analyses. AJB, AG, JDS, KJC, and RAC collected, analyzed, and/or interpreted all microbiological data. LMC and MW conceived the study and co-wrote the manuscript with input from all authors.

6 Funding

This material is based on work supported by the National Science Foundation Graduate Research Fellowship Program under grant no. DGE-1650441.

7 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.

8 Acknowledgments

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.

9 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.
- Achtman, M. (2008). Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu Rev Microbiol* 62, 53-70.
- 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 *bla*CTX-M Genes in Washington State. *Appl Environ Microbiol* 84.
- 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.
- 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

960 information from outbreaks of *Salmonella* Dublin in Swedish cattle herds. *Infect Ecol*
961 *Epidemiol* 6, 31782.

962 Alcaine, S.D., Sukhnanand, S.S., Warnick, L.D., Su, W.L., McGann, P., McDonough, P., and
963 Wiedmann, M. (2005). Ceftiofur-resistant *Salmonella* strains isolated from dairy farms
964 represent multiple widely distributed subtypes that evolved by independent horizontal gene
965 transfer. *Antimicrob Agents Chemother* 49, 4061-4067.

966 An, R., Alshalchi, S., Breimhurst, P., Munoz-Aguayo, J., Flores-Figueroa, C., and Vidovic, S.
967 (2017). Strong influence of livestock environments on the emergence and dissemination of
968 distinct multidrug-resistant phenotypes among the population of non-typhoidal *Salmonella*.
969 *PLoS One* 12, e0179005.

970 Anderson, M.J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral*
971 *Ecology* 26, 32-46.

972 Anderson, M.J. (2006). Distance-Based Tests for Homogeneity of Multivariate Dispersions.
973 *Biometrics* 62, 245-253.

974 Anderson, M.J., and Walsh, D.C.I. (2013). PERMANOVA, ANOSIM, and the Mantel test in the face
975 of heterogeneous dispersions: What null hypothesis are you testing? *Ecological Monographs*
976 83, 557-574.

977 Andrews, S. (2019). "FastQC: a quality control tool for high throughput sequence data". 0.11.8 ed.

978 Banerji, S., Simon, S., Tille, A., Fruth, A., and Flieger, A. (2020). Genome-based *Salmonella*
979 serotyping as the new gold standard. *Sci Rep* 10, 4333.

980 Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M.,
981 Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler,
982 G., Alekseyev, M.A., and Pevzner, P.A. (2012). SPAdes: a new genome assembly algorithm
983 and its applications to single-cell sequencing. *J Comput Biol* 19, 455-477.

984 Boerlin, P., Wissing, A., Aarestrup, F.M., Frey, J., and Nicolet, J. (2001). Antimicrobial growth
985 promoter ban and resistance to macrolides and vancomycin in enterococci from pigs. *J Clin*
986 *Microbiol* 39, 4193-4195.

987 Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina
988 sequence data. *Bioinformatics* 30, 2114-2120.

989 Boore, A.L., Hoekstra, R.M., Iwamoto, M., Fields, P.I., Bishop, R.D., and Swerdlow, D.L. (2015).
990 *Salmonella enterica* Infections in the United States and Assessment of Coefficients of
991 Variation: A Novel Approach to Identify Epidemiologic Characteristics of Individual
992 Serotypes, 1996-2011. *PLoS One* 10, e0145416.

993 Bouckaert, R. (2014). *Correcting for constant sites in BEAST2* [Online]. Available:
994 <https://groups.google.com/forum/#!topic/beast-users/QfBHMOqImFE> [Accessed September
995 9, 2020].

996 Bouckaert, R., Heled, J., Kuhnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A.,
997 and Drummond, A.J. (2014). BEAST 2: a software platform for Bayesian evolutionary
998 analysis. *PLoS Comput Biol* 10, e1003537.

999 Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchene, S., Fourment, M., Gavryushkina, A.,
1000 Heled, J., Jones, G., Kuhnert, D., De Maio, N., Matschiner, M., Mendes, F.K., Muller, N.F.,
1001 Ogilvie, H.A., Du Plessis, L., Poppinga, A., Rambaut, A., Rasmussen, D., Siveroni, I.,
1002 Suchard, M.A., Wu, C.H., Xie, D., Zhang, C., Stadler, T., and Drummond, A.J. (2019).

- 1003 BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS*
1004 *Comput Biol* 15, e1006650.
- 1005 Bouckaert, R., and Xie, D. (2017). "SSN: Standard Nucleotide Substitution Models".
- 1006 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L.
1007 (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421.
- 1008 Canton, R., Gonzalez-Alba, J.M., and Galán, J.C. (2012). CTX-M Enzymes: Origin and Diffusion.
1009 *Frontiers in Microbiology* 3.
- 1010 Carattoli, A., Zankari, E., Garcia-Fernandez, A., Voldby Larsen, M., Lund, O., Villa, L., Moller
1011 Aarestrup, F., and Hasman, H. (2014). *In silico* detection and typing of plasmids using
1012 PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 58,
1013 3895-3903.
- 1014 Carroll, L.M., Gaballa, A., Guldemann, C., Sullivan, G., Henderson, L.O., and Wiedmann, M. (2019).
1015 Identification of Novel Mobilized Colistin Resistance Gene *mcr-9* in a Multidrug-Resistant,
1016 Colistin-Susceptible *Salmonella enterica* Serotype Typhimurium Isolate. *mBio* 10.
- 1017 Carroll, L.M., Huisman, J.S., and Wiedmann, M. (2020a). Twentieth-century emergence of
1018 antimicrobial resistant human- and bovine-associated *Salmonella enterica* serotype
1019 Typhimurium lineages in New York State. *Sci Rep* 10, 14428.
- 1020 Carroll, L.M., Kovac, J., Miller, R.A., and Wiedmann, M. (2017a). Rapid, High-Throughput
1021 Identification of Anthrax-Causing and Emetic *Bacillus cereus* Group Genome Assemblies via
1022 BTyper, a Computational Tool for Virulence-Based Classification of *Bacillus cereus* Group
1023 Isolates by Using Nucleotide Sequencing Data. *Appl Environ Microbiol* 83.
- 1024 Carroll, L.M., Wiedmann, M., Den Bakker, H., Siler, J., Warchock, S., Kent, D., Lyalina, S., Davis,
1025 M., Sicho, W., Besser, T., Warnick, L.D., and Pereira, R.V. (2017b). Whole-Genome
1026 Sequencing of Drug-Resistant *Salmonella enterica* Isolates from Dairy Cattle and Humans in
1027 New York and Washington States Reveals Source and Geographic Associations. *Appl*
1028 *Environ Microbiol* 83.
- 1029 Carroll, L.M., Wiedmann, M., and Kovac, J. (2020b). Proposal of a Taxonomic Nomenclature for the
1030 *Bacillus cereus* Group Which Reconciles Genomic Definitions of Bacterial Species with
1031 Clinical and Industrial Phenotypes. *mBio* 11, e00034-00020.
- 1032 Centers for Disease Control and Prevention (2019). *Antibiotic Resistance Threats in the United*
1033 *States, 2019*. Atlanta, GA: U.S. Department of Health and Human Services, CDC.
- 1034 Centers for Disease Control and Prevention (2020). *Serotypes and the Importance of Serotyping*
1035 *Salmonella* [Online]. CDC. Available:
1036 <https://www.cdc.gov/salmonella/reportspubs/salmonella-atlas/serotyping-importance.html>
1037 [Accessed 25 May 2021 2021].
- 1038 Chang, W., Cheng, J., Allaire, J., Xie, Y., and McPherson, J. (2020). "shiny: Web Application
1039 Framework for R. R package version 1.4.0.2. <https://CRAN.R-project.org/package=shiny>".
- 1040 Chen, T., and Guestrin, C. (2016). "XGBoost: A Scalable Tree Boosting System", in: *Proceedings of*
1041 *the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data*
1042 *Mining*. (San Francisco, California, USA: Association for Computing Machinery).
- 1043 Chiu, C.H., Su, L.H., and Chu, C. (2004). *Salmonella enterica* serotype Choleraesuis: epidemiology,
1044 pathogenesis, clinical disease, and treatment. *Clin Microbiol Rev* 17, 311-322.

- 1045 Cingolani, P., Platts, A., Wang Le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and
1046 Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide
1047 polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-
1048 2; iso-3. *Fly (Austin)* 6, 80-92.
- 1049 Clarke, K.R. (1993). Non-parametric multivariate analyses of changes in community structure.
1050 *Australian Journal of Ecology* 18, 117-143.
- 1051 Clarke, L., Pelin, A., Phan, M., and Wong, A. (2020). The effect of environmental heterogeneity on
1052 the fitness of antibiotic resistance mutations in *Escherichia coli*. *Evolutionary Ecology* 34,
1053 379-390.
- 1054 Clausen, P.T., Zankari, E., Aarestrup, F.M., and Lund, O. (2016). Benchmarking of methods for
1055 identification of antimicrobial resistance genes in bacterial whole genome data. *J Antimicrob*
1056 *Chemother* 71, 2484-2488.
- 1057 Cleary, J.G., Braithwaite, R., Gaastra, K., Hilbush, B.S., Inglis, S., Irvine, S.A., Jackson, A., Littin,
1058 R., Rathod, M., Ware, D., Zook, J.M., Trigg, L., and De La Vega, F.M. (2015). Comparing
1059 Variant Call Files for Performance Benchmarking of Next-Generation Sequencing Variant
1060 Calling Pipelines. *bioRxiv*, 023754.
- 1061 Cooper, A.L., Low, A.J., Koziol, A.G., Thomas, M.C., Leclair, D., Tamber, S., Wong, A., Blais,
1062 B.W., and Carrillo, C.D. (2020). Systematic Evaluation of Whole Genome Sequence-Based
1063 Predictions of *Salmonella* Serotype and Antimicrobial Resistance. *Front Microbiol* 11, 549.
- 1064 Croucher, N.J., Page, A.J., Connor, T.R., Delaney, A.J., Keane, J.A., Bentley, S.D., Parkhill, J., and
1065 Harris, S.R. (2015). Rapid phylogenetic analysis of large samples of recombinant bacterial
1066 whole genome sequences using Gubbins. *Nucleic Acids Res* 43, e15.
- 1067 Cummings, K.J., Warnick, L.D., Davis, M.A., Eckmann, K., Grohn, Y.T., Hoelzer, K., Macdonald,
1068 K., Root, T.P., Siler, J.D., McGuire, S.M., Wiedmann, M., Wright, E.M., Zansky, S.M., and
1069 Besser, T.E. (2012). Farm animal contact as risk factor for transmission of bovine-associated
1070 *Salmonella* subtypes. *Emerg Infect Dis* 18, 1929-1936.
- 1071 Cummings, K.J., Warnick, L.D., Elton, M., Grohn, Y.T., McDonough, P.L., and Siler, J.D. (2010a).
1072 The effect of clinical outbreaks of salmonellosis on the prevalence of fecal *Salmonella*
1073 shedding among dairy cattle in New York. *Foodborne Pathog Dis* 7, 815-823.
- 1074 Cummings, K.J., Warnick, L.D., Elton, M., Rodriguez-Rivera, L.D., Siler, J.D., Wright, E.M., Grohn,
1075 Y.T., and Wiedmann, M. (2010b). *Salmonella enterica* serotype Cerro among dairy cattle in
1076 New York: an emerging pathogen? *Foodborne Pathog Dis* 7, 659-665.
- 1077 Davidson, K.E., Byrne, B.A., Pires, A.F.A., Magdesian, K.G., and Pereira, R.V. (2018).
1078 Antimicrobial resistance trends in fecal *Salmonella* isolates from northern California dairy
1079 cattle admitted to a veterinary teaching hospital, 2002-2016. *PLoS One* 13, e0199928.
- 1080 Delgado-Suarez, E.J., Selem-Mojica, N., Ortiz-Lopez, R., Gebreyes, W.A., Allard, M.W., Barona-
1081 Gomez, F., and Rubio-Lozano, M.S. (2018). Whole genome sequencing reveals widespread
1082 distribution of typhoidal toxin genes and VirB/D4 plasmids in bovine-associated
1083 nontyphoidal *Salmonella*. *Sci Rep* 8, 9864.
- 1084 Delsol, A.A., Anjum, M., Woodward, M.J., Sunderland, J., and Roe, J.M. (2003). The effect of
1085 chlortetracycline treatment and its subsequent withdrawal on multi-resistant *Salmonella*
1086 *enterica* serovar Typhimurium DT104 and commensal *Escherichia coli* in the pig. *J Appl*
1087 *Microbiol* 95, 1226-1234.

- 1088 Diep, B., Barretto, C., Portmann, A.C., Fournier, C., Karczmarek, A., Voets, G., Li, S., Deng, X., and
1089 Klijn, A. (2019). *Salmonella* Serotyping; Comparison of the Traditional Method to a
1090 Microarray-Based Method and an *in silico* Platform Using Whole Genome Sequencing Data.
1091 *Front Microbiol* 10, 2554.
- 1092 Drummond, A.J., Ho, S.Y., Phillips, M.J., and Rambaut, A. (2006). Relaxed phylogenetics and
1093 dating with confidence. *PLoS Biol* 4, e88.
- 1094 Drummond, A.J., Rambaut, A., Shapiro, B., and Pybus, O.G. (2005). Bayesian coalescent inference
1095 of past population dynamics from molecular sequences. *Mol Biol Evol* 22, 1185-1192.
- 1096 Dutil, L., Irwin, R., Finley, R., Ng, L.K., Avery, B., Boerlin, P., Bourgault, A.M., Cole, L.,
1097 Daignault, D., Desruisseau, A., Demczuk, W., Hoang, L., Horsman, G.B., Ismail, J.,
1098 Jamieson, F., Maki, A., Pacagnella, A., and Pillai, D.R. (2010). Ceftiofur resistance in
1099 *Salmonella enterica* serovar Heidelberg from chicken meat and humans, Canada. *Emerg*
1100 *Infect Dis* 16, 48-54.
- 1101 Evans, B.A., and Amyes, S.G.B. (2014). OXA β -Lactamases. *Clinical Microbiology Reviews* 27,
1102 241-263.
- 1103 Ewels, P., Magnusson, M., Lundin, S., and Kaller, M. (2016). MultiQC: summarize analysis results
1104 for multiple tools and samples in a single report. *Bioinformatics* 32, 3047-3048.
- 1105 Feldgarden, M., Brover, V., Haft, D.H., Prasad, A.B., Slotta, D.J., Tolstoy, I., Tyson, G.H., Zhao, S.,
1106 Hsu, C.H., Mcdermott, P.F., Tadesse, D.A., Morales, C., Simmons, M., Tillman, G.,
1107 Wasilenko, J., Folster, J.P., and Klimke, W. (2019). Validating the AMRFinder Tool and
1108 Resistance Gene Database by Using Antimicrobial Resistance Genotype-Phenotype
1109 Correlations in a Collection of Isolates. *Antimicrob Agents Chemother* 63.
- 1110 Fenske, G.J., Thachil, A., Mcdonough, P.L., Glaser, A., and Scaria, J. (2019). Geography Shapes the
1111 Population Genomics of *Salmonella enterica* Dublin. *Genome Biol Evol* 11, 2220-2231.
- 1112 Frech, G., Kehrenberg, C., and Schwarz, S. (2003). Resistance phenotypes and genotypes of
1113 multiresistant *Salmonella enterica* subsp. *enterica* serovar Typhimurium var. Copenhagen
1114 isolates from animal sources. *J Antimicrob Chemother* 51, 180-182.
- 1115 Gardner, S.N., and Hall, B.G. (2013). When whole-genome alignments just won't work: kSNP v2
1116 software for alignment-free SNP discovery and phylogenetics of hundreds of microbial
1117 genomes. *PLoS One* 8, e81760.
- 1118 Gardner, S.N., Slezak, T., and Hall, B.G. (2015). kSNP3.0: SNP detection and phylogenetic analysis
1119 of genomes without genome alignment or reference genome. *Bioinformatics* 31, 2877-2878.
- 1120 Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing.
1121 *arXiv*, 1207.3907.
- 1122 Gorski, L., Parker, C.T., Liang, A., Cooley, M.B., Jay-Russell, M.T., Gordus, A.G., Atwill, E.R., and
1123 Mandrell, R.E. (2011). Prevalence, distribution, and diversity of *Salmonella enterica* in a
1124 major produce region of California. *Appl Environ Microbiol* 77, 2734-2748.
- 1125 Gupta, A., Fontana, J., Crowe, C., Bolstorff, B., Stout, A., Duyne, S.V., Hoekstra, M.P., Whichard,
1126 J.M., Barrett, T.J., and Group, T.N.a.R.M.S.P.W. (2003). Emergence of Multidrug-Resistant
1127 *Salmonella enterica* Serotype Newport Infections Resistant to Expanded-Spectrum
1128 Cephalosporins in the United States. *The Journal of Infectious Diseases* 188, 1707-1716.

- 1129 Gupta, S.K., Padmanabhan, B.R., Diene, S.M., Lopez-Rojas, R., Kempf, M., Landraud, L., and
1130 Rolain, J.M. (2014). ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance
1131 genes in bacterial genomes. *Antimicrob Agents Chemother* 58, 212-220.
- 1132 Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for
1133 genome assemblies. *Bioinformatics* 29, 1072-1075.
- 1134 Gutema, F.D., Agga, G.E., Abdi, R.D., De Zutter, L., Duchateau, L., and Gabriel, S. (2019).
1135 Prevalence and Serotype Diversity of *Salmonella* in Apparently Healthy Cattle: Systematic
1136 Review and Meta-Analysis of Published Studies, 2000-2017. *Front Vet Sci* 6, 102.
- 1137 Harvey, R.R., Friedman, C.R., Crim, S.M., Judd, M., Barrett, K.A., Tolar, B., Folster, J.P., Griffin,
1138 P.M., and Brown, A.C. (2017). Epidemiology of *Salmonella enterica* Serotype Dublin
1139 Infections among Humans, United States, 1968-2013. *Emerg Infect Dis* 23.
- 1140 Heled, J., and Bouckaert, R.R. (2013). Looking for trees in the forest: summary tree from posterior
1141 samples. *BMC Evol Biol* 13, 221.
- 1142 Hendriksen, R.S., Bortolaia, V., Tate, H., Tyson, G.H., Aarestrup, F.M., and Mcdermott, P.F. (2019).
1143 Using Genomics to Track Global Antimicrobial Resistance. *Front Public Health* 7, 242.
- 1144 Hendriksen, R.S., Pedersen, S.K., Leekitcharoenphon, P., Malorny, B., Borowiak, M., Battisti, A.,
1145 Franco, A., Alba, P., Carfora, V., Ricci, A., Mastroilli, E., Losasso, C., Longo, A., Petrin, S.,
1146 Barco, L., Wołkowicz, T., Gierczyński, R., Zacharczuk, K., Wolaniuk, N., Wasyl, D., Zajac,
1147 M., Wiczorek, K., Półtorak, K., Petrovska-Holmes, L., Davies, R., Tang, Y., Grant, K.,
1148 Underwood, A., Dallman, T., Painset, A., Hartman, H., Al-Shabib, A., and Cowley, L.
1149 (2018). Final report of ENGAGE - Establishing Next Generation sequencing Ability for
1150 Genomic analysis in Europe. *EFSA Supporting Publications* 15, 1431E.
- 1151 Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2:
1152 Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* 35, 518-522.
- 1153 Hoelzer, K., Moreno Switt, A.I., and Wiedmann, M. (2011). Animal contact as a source of human
1154 non-typhoidal salmonellosis. *Vet Res* 42, 34.
- 1155 Hoelzer, K., Wong, N., Thomas, J., Talkington, K., Jungman, E., and Coukell, A. (2017).
1156 Antimicrobial drug use in food-producing animals and associated human health risks: what,
1157 and how strong, is the evidence? *BMC Veterinary Research* 13, 211.
- 1158 Holschbach, C.L., and Peek, S.F. (2018). *Salmonella* in Dairy Cattle. *Vet Clin North Am Food Anim*
1159 *Pract* 34, 133-154.
- 1160 Hornish, R.E., and Kotarski, S.F. (2002). Cephalosporins in veterinary medicine - ceftiofur use in
1161 food animals. *Curr Top Med Chem* 2, 717-731.
- 1162 Hunt, M., Mather, A.E., Sanchez-Buso, L., Page, A.J., Parkhill, J., Keane, J.A., and Harris, S.R.
1163 (2017). ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads.
1164 *Microb Genom* 3, e000131.
- 1165 Ibrahim, G.M., and Morin, P.M. (2018). *Salmonella* Serotyping Using Whole Genome Sequencing.
1166 *Front Microbiol* 9, 2993.
- 1167 Inouye, M., Dashnow, H., Raven, L.A., Schultz, M.B., Pope, B.J., Tomita, T., Zobel, J., and Holt,
1168 K.E. (2014). SRST2: Rapid genomic surveillance for public health and hospital microbiology
1169 labs. *Genome Med* 6, 90.

- 1170 Issenhuth-Jeanjean, S., Roggentin, P., Mikoleit, M., Guibourdenche, M., De Pinna, E., Nair, S.,
1171 Fields, P.I., and Weill, F.X. (2014). Supplement 2008-2010 (no. 48) to the White-Kauffmann-
1172 Le Minor scheme. *Res Microbiol* 165, 526-530.
- 1173 Jia, B., Raphenya, A.R., Alcock, B., Wagglechner, N., Guo, P., Tsang, K.K., Lago, B.A., Dave, B.M.,
1174 Pereira, S., Sharma, A.N., Doshi, S., Courtot, M., Lo, R., Williams, L.E., Frye, J.G.,
1175 Elsayegh, T., Sardar, D., Westman, E.L., Pawlowski, A.C., Johnson, T.A., Brinkman, F.S.,
1176 Wright, G.D., and McArthur, A.G. (2017). CARD 2017: expansion and model-centric curation
1177 of the comprehensive antibiotic resistance database. *Nucleic Acids Res* 45, D566-D573.
- 1178 Jorgensen, J.H., and Ferraro, M.J. (2009). Antimicrobial susceptibility testing: a review of general
1179 principles and contemporary practices. *Clin Infect Dis* 49, 1749-1755.
- 1180 Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Von Haeseler, A., and Jermini, L.S. (2017).
1181 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14, 587-
1182 589.
- 1183 Karon, A.E., Archer, J.R., Sotir, M.J., Monson, T.A., and Kazmierczak, J.J. (2007). Human
1184 multidrug-resistant *Salmonella* Newport infections, Wisconsin, 2003-2005. *Emerg Infect Dis*
1185 13, 1777-1780.
- 1186 Kass, R.E., and Raftery, A.E. (1995). Bayes Factors. *Journal of the American Statistical Association*
1187 90, 773-795.
- 1188 Kodama, Y., Shumway, M., Leinonen, R., and International Nucleotide Sequence Database, C.
1189 (2012). The Sequence Read Archive: explosive growth of sequencing data. *Nucleic Acids Res*
1190 40, D54-56.
- 1191 Kokot, M., Dlugosz, M., and Deorowicz, S. (2017). KMC 3: counting and manipulating k-mer
1192 statistics. *Bioinformatics* 33, 2759-2761.
- 1193 Kruskal, J.B. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika* 29,
1194 115-129.
- 1195 Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *2008* 28, 26.
- 1196 Lakin, S.M., Dean, C., Noyes, N.R., Dettenwanger, A., Ross, A.S., Doster, E., Rovira, P., Abdo, Z.,
1197 Jones, K.L., Ruiz, J., Belk, K.E., Morley, P.S., and Boucher, C. (2017). MEGARes: an
1198 antimicrobial resistance database for high throughput sequencing. *Nucleic Acids Res* 45,
1199 D574-D580.
- 1200 Leekitcharoenphon, P., Hendriksen, R.S., Le Hello, S., Weill, F.X., Baggesen, D.L., Jun, S.R.,
1201 Ussery, D.W., Lund, O., Crook, D.W., Wilson, D.J., and Aarestrup, F.M. (2016). Global
1202 Genomic Epidemiology of *Salmonella enterica* Serovar Typhimurium DT104. *Appl Environ*
1203 *Microbiol* 82, 2516-2526.
- 1204 Leinonen, R., Sugawara, H., Shumway, M., and International Nucleotide Sequence Database, C.
1205 (2011). The sequence read archive. *Nucleic Acids Res* 39, D19-21.
- 1206 Lewis, J.S., 2nd, Herrera, M., Wickes, B., Patterson, J.E., and Jorgensen, J.H. (2007). First report of
1207 the emergence of CTX-M-type extended-spectrum beta-lactamases (ESBLs) as the
1208 predominant ESBL isolated in a U.S. health care system. *Antimicrob Agents Chemother* 51,
1209 4015-4021.
- 1210 Lewis, P.O. (2001). A likelihood approach to estimating phylogeny from discrete morphological
1211 character data. *Syst Biol* 50, 913-925.

1212 Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and
1213 population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987-
1214 2993.

1215 Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
1216 *arXiv*, 1303.3997.

1217 Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094-
1218 3100.

1219 Li, H. (2019). "Seqtk: a fast and lightweight tool for processing sequences in the FASTA or FASTQ
1220 format". 1.2-r102-dirty ed.).

1221 Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler
1222 transform. *Bioinformatics* 25, 1754-1760.

1223 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G.,
1224 Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map
1225 format and SAMtools. *Bioinformatics* 25, 2078-2079.

1226 Liao, J., Orsi, R.H., Carroll, L.M., Kovac, J., Ou, H., Zhang, H., and Wiedmann, M. (2019).
1227 Serotype-specific evolutionary patterns of antimicrobial-resistant *Salmonella enterica*. *BMC*
1228 *Evol Biol* 19, 132.

1229 Liao, J., Orsi, R.H., Carroll, L.M., and Wiedmann, M. (2020). Comparative genomics reveals
1230 different population structures associated with host and geographic origin in antimicrobial-
1231 resistant *Salmonella enterica*. *Environ Microbiol*.

1232 Liebana, E., Carattoli, A., Coque, T.M., Hasman, H., Magiorakos, A.P., Mevius, D., Peixe, L., Poirer,
1233 L., Schuepbach-Regula, G., Torneke, K., Torren-Edo, J., Torres, C., and Threlfall, J. (2013).
1234 Public health risks of enterobacterial isolates producing extended-spectrum beta-lactamases or
1235 AmpC beta-lactamases in food and food-producing animals: an EU perspective of
1236 epidemiology, analytical methods, risk factors, and control options. *Clin Infect Dis* 56, 1030-
1237 1037.

1238 Llarena, A.K., Zhang, J., Vehkala, M., Valimaki, N., Hakkinen, M., Hanninen, M.L., Roasto, M.,
1239 Maesaar, M., Taboada, E., Barker, D., Garofolo, G., Camma, C., Di Giannatale, E., Corander,
1240 J., and Rossi, M. (2016). Monomorphic genotypes within a generalist lineage of
1241 *Campylobacter jejuni* show signs of global dispersion. *Microb Genom* 2, e000088.

1242 Luceron, C.O., Meixeira, A.P., Sanz, I.A., Deleyto, V.C., Leon, S.H., and Ruiz, L.G. (2017). Notes
1243 from the Field: An Outbreak of *Salmonella* Typhimurium Associated with Playground Sand
1244 in a Preschool Setting - Madrid, Spain, September-October 2016. *MMWR Morb Mortal Wkly*
1245 *Rep* 66, 256-257.

1246 Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2019). "cluster: Cluster
1247 Analysis Basics and Extensions". 2.1.0 ed.).

1248 Mather, A.E., Reid, S.W., Maskell, D.J., Parkhill, J., Fookes, M.C., Harris, S.R., Brown, D.J., Coia,
1249 J.E., Mulvey, M.R., Gilmour, M.W., Petrovska, L., De Pinna, E., Kuroda, M., Akiba, M.,
1250 Izumiya, H., Connor, T.R., Suchard, M.A., Lemey, P., Mellor, D.J., Haydon, D.T., and
1251 Thomson, N.R. (2013). Distinguishable epidemics of multidrug-resistant *Salmonella*
1252 Typhimurium DT104 in different hosts. *Science* 341, 1514-1517.

- 1253 Mcdermott, P.F., Bodeis, S.M., English, L.L., White, D.G., Walker, R.D., Zhao, S., Simjee, S., and
1254 Wagner, D.D. (2002). Ciprofloxacin resistance in *Campylobacter jejuni* evolves rapidly in
1255 chickens treated with fluoroquinolones. *J Infect Dis* 185, 837-840.
- 1256 Mcdermott, P.F., Tyson, G.H., Kabera, C., Chen, Y., Li, C., Folster, J.P., Ayers, S.L., Lam, C., Tate,
1257 H.P., and Zhao, S. (2016). Whole-Genome Sequencing for Detecting Antimicrobial
1258 Resistance in Nontyphoidal *Salmonella*. *Antimicrobial Agents and Chemotherapy* 60, 5515-
1259 5520.
- 1260 Mellor, K.C., Petrovska, L., Thomson, N.R., Harris, K., Reid, S.W.J., and Mather, A.E. (2019).
1261 Antimicrobial Resistance Diversity Suggestive of Distinct *Salmonella* Typhimurium Sources
1262 or Selective Pressures in Food-Production Animals. *Front Microbiol* 10, 708.
- 1263 Melnyk, A.H., Wong, A., and Kassen, R. (2015). The fitness costs of antibiotic resistance mutations.
1264 *Evol Appl* 8, 273-283.
- 1265 Mentaberre, G., Porrero, M.C., Navarro-Gonzalez, N., Serrano, E., Dominguez, L., and Lavin, S.
1266 (2013). Cattle drive *Salmonella* infection in the wildlife-livestock interface. *Zoonoses Public*
1267 *Health* 60, 510-518.
- 1268 Minh, B.Q., Nguyen, M.A., and Von Haeseler, A. (2013). Ultrafast approximation for phylogenetic
1269 bootstrap. *Mol Biol Evol* 30, 1188-1195.
- 1270 Mohammed, M., Le Hello, S., Leekitcharoenphon, P., and Hendriksen, R. (2017). The invasome of
1271 *Salmonella* Dublin as revealed by whole genome sequencing. *BMC Infect Dis* 17, 544.
- 1272 Moller, S., Du Plessis, L., and Stadler, T. (2018). Impact of the tree prior on estimating clock rates
1273 during epidemic outbreaks. *Proc Natl Acad Sci U S A* 115, 4200-4205.
- 1274 Mughini-Gras, L., Enserink, R., Friesema, I., Heck, M., Van Duynhoven, Y., and Van Pelt, W.
1275 (2014). Risk factors for human salmonellosis originating from pigs, cattle, broiler chickens
1276 and egg laying hens: a combined case-control and source attribution analysis. *PLoS One* 9,
1277 e87933.
- 1278 Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and
1279 effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*
1280 32, 268-274.
- 1281 Nguyen, M., Long, S.W., Mcdermott, P.F., Olsen, R.J., Olson, R., Stevens, R.L., Tyson, G.H., Zhao,
1282 S., and Davis, J.J. (2019). Using Machine Learning To Predict Antimicrobial MICs and
1283 Associated Genomic Features for Nontyphoidal *Salmonella*. *Journal of Clinical Microbiology*
1284 57, e01260-01218.
- 1285 Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., Minchin, P.R.,
1286 O'hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., and Wagner, H.
1287 (2019). "vegan: Community Ecology Package. R package version 2.5-6. [https://CRAN.R-](https://CRAN.R-project.org/package=vegan)
1288 [project.org/package=vegan](https://CRAN.R-project.org/package=vegan)".).
- 1289 Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T., Fookes, M., Falush, D.,
1290 Keane, J.A., and Parkhill, J. (2015). Roary: rapid large-scale prokaryote pan genome analysis.
1291 *Bioinformatics* 31, 3691-3693.
- 1292 Page, A.J., Taylor, B., Delaney, A.J., Soares, J., Seemann, T., Keane, J.A., and Harris, S.R. (2016).
1293 SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb Genom*
1294 2, e000056.

1295 Palma, F., Manfreda, G., Silva, M., Parisi, A., Barker, D.O.R., Taboada, E.N., Pasquali, F., and
1296 Rossi, M. (2018). Genome-wide identification of geographical segregated genetic markers in
1297 *Salmonella enterica* serovar Typhimurium variant 4,[5],12:i. *Scientific Reports* 8, 15251.

1298 Paradis, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R
1299 language. *Bioinformatics* 20, 289-290.

1300 Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and
1301 evolutionary analyses in R. *Bioinformatics* 35, 526-528.

1302 Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015). CheckM:
1303 assessing the quality of microbial genomes recovered from isolates, single cells, and
1304 metagenomes. *Genome Res* 25, 1043-1055.

1305 Pereira, R., Williams, D.R., Rossitto, P., Adaska, J., Okello, E., Champagne, J., Lehenbauer, T.W.,
1306 Li, X., Chase, J., Nguyen, T., Pires, A.F.A., Atwill, E.R., and Aly, S.S. (2019). Association
1307 between herd management practices and antimicrobial resistance in *Salmonella* spp. from cull
1308 dairy cattle in Central California. *PeerJ* 7, e6546.

1309 Quinlan, A.R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc*
1310 *Bioinformatics* 47, 11 12 11-34.

1311 Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic
1312 features. *Bioinformatics* 26, 841-842.

1313 R Core Team (2019). "R: A Language and Environment for Statistical Computing". 3.6.1 ed.
1314 (Vienna, Austria: R Foundation for Statistical Computing).

1315 R Hackathon (2019). "phylobase: Base Package for Phylogenetic Structures and Comparative Data".
1316 0.8.6 ed.).

1317 Rambaut, A., Drummond, A.J., Xie, D., Baele, G., and Suchard, M.A. (2018). Posterior
1318 Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst Biol* 67, 901-904.

1319 Rambaut, A., Lam, T.T., Max Carvalho, L., and Pybus, O.G. (2016). Exploring the temporal
1320 structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* 2,
1321 vew007.

1322 Revell, L.J. (2012). phytools: an R package for phylogenetic comparative biology (and other things).
1323 *Methods in Ecology and Evolution* 3, 217-223.

1324 Rodriguez, A., Pangloli, P., Richards, H.A., Mount, J.R., and Draughon, F.A. (2006). Prevalence of
1325 *Salmonella* in diverse environmental farm samples. *J Food Prot* 69, 2576-2580.

1326 Rodriguez-Rivera, L.D., Wright, E.M., Siler, J.D., Elton, M., Cummings, K.J., Warnick, L.D., and
1327 Wiedmann, M. (2014). Subtype analysis of *Salmonella* isolated from subclinically infected
1328 dairy cattle and dairy farm environments reveals the presence of both human- and bovine-
1329 associated subtypes. *Vet Microbiol* 170, 307-316.

1330 San Millan, A., and Maclean, R.C. (2017). Fitness Costs of Plasmids: a Limit to Plasmid
1331 Transmission. *Microbiology Spectrum* 5.

1332 Scallan, E., Hoekstra, R.M., Angulo, F.J., Tauxe, R.V., Widdowson, M.A., Roy, S.L., Jones, J.L.,
1333 and Griffin, P.M. (2011). Foodborne illness acquired in the United States--major pathogens.
1334 *Emerg Infect Dis* 17, 7-15.

1335 Schliep, K.P. (2011). phangorn: phylogenetic analysis in R. *Bioinformatics* 27, 592-593.

- 1336 Scott, A.M., Beller, E., Glasziou, P., Clark, J., Ranakusuma, R.W., Byambasuren, O., Bakhit, M.,
1337 Page, S.W., Trott, D., and Mar, C.D. (2018). Is antimicrobial administration to food animals a
1338 direct threat to human health? A rapid systematic review. *International Journal of*
1339 *Antimicrobial Agents* 52, 316-323.
- 1340 Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068-2069.
- 1341 Seemann, T. (2018). "ABRicate: Mass screening of contigs for antimicrobial resistance or virulence
1342 genes".
- 1343 Seemann, T. (2019a). "samclip: Filter SAM file for soft and hard clipped alignments". 0.2 ed.
- 1344 Seemann, T. (2019b). "Snippy: Rapid haploid variant calling and core genome alignment". 4.3.6 ed.
- 1345 Skov, M.N., Madsen, J.J., Rahbek, C., Lodal, J., Jespersen, J.B., Jorgensen, J.C., Dietz, H.H., Chriel,
1346 M., and Baggesen, D.L. (2008). Transmission of *Salmonella* between wildlife and meat-
1347 production animals in Denmark. *J Appl Microbiol* 105, 1558-1568.
- 1348 Slowikowski, K. (2020). "ggrepel: Automatically Position Non-Overlapping Text Labels with
1349 'ggplot2'. R package version 0.8.2. <https://CRAN.R-project.org/package=ggrepel>".
- 1350 Soubrier, J., Steel, M., Lee, M.S.Y., Der Sarkissian, C., Guindon, S., Ho, S.Y.W., and Cooper, A.
1351 (2012). The Influence of Rate Heterogeneity among Sites on the Time Dependence of
1352 Molecular Rates. *Molecular Biology and Evolution* 29, 3345-3358.
- 1353 Spika, J.S., Waterman, S.H., Hoo, G.W., St Louis, M.E., Pacer, R.E., James, S.M., Bissett, M.L.,
1354 Mayer, L.W., Chiu, J.Y., Hall, B., and Et Al. (1987). Chloramphenicol-resistant *Salmonella*
1355 newport traced through hamburger to dairy farms. A major persisting source of human
1356 salmonellosis in California. *N Engl J Med* 316, 565-570.
- 1357 Strachan, N.J.C., Rotariu, O., Lopes, B., Macrae, M., Fairley, S., Laing, C., Gannon, V., Allison, L.J.,
1358 Hanson, M.F., Dallman, T., Ashton, P., Franz, E., Van Hoek, A.H.a.M., French, N.P.,
1359 George, T., Biggs, P.J., and Forbes, K.J. (2015). Whole Genome Sequencing demonstrates
1360 that Geographic Variation of *Escherichia coli* O157 Genotypes Dominates Host Association.
1361 *Scientific Reports* 5, 14145.
- 1362 Tack, B., Phoba, M.F., Barbe, B., Kalonji, L.M., Hardy, L., Van Puyvelde, S., Ingelbeen, B., Falay,
1363 D., Ngonda, D., Van Der Sande, M.a.B., Deborggraeve, S., Jacobs, J., and Lunguya, O.
1364 (2020). Non-typhoidal *Salmonella* bloodstream infections in Kisantu, DR Congo: Emergence
1365 of O5-negative *Salmonella* Typhimurium and extensive drug resistance. *PLoS Negl Trop Dis*
1366 14, e0008121.
- 1367 Tan, A., Abecasis, G.R., and Kang, H.M. (2015). Unified representation of genetic variants.
1368 *Bioinformatics* 31, 2202-2204.
- 1369 Tang, K.L., Caffrey, N.P., Nobrega, D.B., Cork, S.C., Ronksley, P.E., Barkema, H.W., Polachek,
1370 A.J., Ganshorn, H., Sharma, N., Kellner, J.D., and Ghali, W.A. (2017). Restricting the use of
1371 antibiotics in food-producing animals and its associations with antibiotic resistance in food-
1372 producing animals and human beings: a systematic review and meta-analysis. *Lancet Planet*
1373 *Health* 1, e316-e327.
- 1374 Tassinari, E., Duffy, G., Bawn, M., Burgess, C.M., McCabe, E.M., Lawlor, P.G., Gardiner, G., and
1375 Kingsley, R.A. (2019). Microevolution of antimicrobial resistance and biofilm formation of
1376 *Salmonella* Typhimurium during persistence on pig farms. *Sci Rep* 9, 8832.
- 1377 Taylor, D.N., Bied, J.M., Munro, J.S., and Feldman, R.A. (1982). *Salmonella* dublin infections in the
1378 United States, 1979-1980. *J Infect Dis* 146, 322-327.

1379 The, H.C., Thanh, D.P., Holt, K.E., Thomson, N.R., and Baker, S. (2016). The genomic signatures of
1380 *Shigella* evolution, adaptation and geographical spread. *Nat Rev Microbiol* 14, 235-250.

1381 Toth, J.D., Aceto, H.W., Rankin, S.C., and Dou, Z. (2011). Survival characteristics of *Salmonella*
1382 *enterica* serovar Newport in the dairy farm environment. *J Dairy Sci* 94, 5238-5246.

1383 Tragesser, L.A., Wittum, T.E., Funk, J.A., Winokur, P.L., and Rajala-Schultz, P.J. (2006).
1384 Association between ceftiofur use and isolation of *Escherichia coli* with reduced
1385 susceptibility to ceftriaxone from fecal samples of dairy cows. *Am J Vet Res* 67, 1696-1700.

1386 Uelze, L., Borowiak, M., Deneke, C., Szabo, I., Fischer, J., Tausch, S.H., and Malorny, B. (2020).
1387 Performance and Accuracy of Four Open-Source Tools for *In Silico* Serotyping of *Salmonella*
1388 spp. Based on Whole-Genome Short-Read Sequencing Data. *Appl Environ Microbiol* 86.

1389 Uzzau, S., Brown, D.J., Wallis, T., Rubino, S., Leori, G., Bernard, S., Casadesus, J., Platt, D.J., and
1390 Olsen, J.E. (2000). Host adapted serotypes of *Salmonella enterica*. *Epidemiol Infect* 125, 229-
1391 255.

1392 Wang, C., Feng, Y., Liu, L., Wei, L., Kang, M., and Zong, Z. (2020). Identification of novel mobile
1393 colistin resistance gene *mcr-10*. *Emerg Microbes Infect* 9, 508-516.

1394 Wang, X., Wang, Y., Zhou, Y., Li, J., Yin, W., Wang, S., Zhang, S., Shen, J., Shen, Z., and Wang, Y.
1395 (2018). Emergence of a novel mobile colistin resistance gene, *mcr-8*, in NDM-producing
1396 *Klebsiella pneumoniae*. *Emerg Microbes Infect* 7, 122.

1397 Wickham, H. (2007). Reshaping Data with the reshape Package. 2007 21, 20.

1398 Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. 2011 40, 29.

1399 Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

1400 Wickham, H., François, R., Henry, L., and Müller, K. (2020). "dplyr: A Grammar of Data
1401 Manipulation". 0.8.5 ed.).

1402 Wiethoelter, A.K., Beltran-Alcrudo, D., Kock, R., and Mor, S.M. (2015). Global trends in infectious
1403 diseases at the wildlife-livestock interface. *Proc Natl Acad Sci U S A* 112, 9662-9667.

1404 Wong, V.K., Baker, S., Pickard, D.J., Parkhill, J., Page, A.J., Feasey, N.A., Kingsley, R.A.,
1405 Thomson, N.R., Keane, J.A., Weill, F.X., Edwards, D.J., Hawkey, J., Harris, S.R., Mather,
1406 A.E., Cain, A.K., Hadfield, J., Hart, P.J., Thieu, N.T., Klemm, E.J., Glinos, D.A., Breiman,
1407 R.F., Watson, C.H., Kariuki, S., Gordon, M.A., Heyderman, R.S., Okoro, C., Jacobs, J.,
1408 Lunguya, O., Edmunds, W.J., Msefula, C., Chabalgoity, J.A., Kama, M., Jenkins, K., Dutta,
1409 S., Marks, F., Campos, J., Thompson, C., Obaro, S., MacLennan, C.A., Dolecek, C., Keddy,
1410 K.H., Smith, A.M., Parry, C.M., Karkey, A., Mulholland, E.K., Campbell, J.I., Dongol, S.,
1411 Basnyat, B., Dufour, M., Bandaranayake, D., Naseri, T.T., Singh, S.P., Hatta, M., Newton, P.,
1412 Onsare, R.S., Isaia, L., Dance, D., Davong, V., Thwaites, G., Wijedoru, L., Crump, J.A., De
1413 Pinna, E., Nair, S., Nilles, E.J., Thanh, D.P., Turner, P., Soeng, S., Valcanis, M., Powling, J.,
1414 Dimovski, K., Hogg, G., Farrar, J., Holt, K.E., and Dougan, G. (2015). Phylogeographical
1415 analysis of the dominant multidrug-resistant H58 clade of *Salmonella* Typhi identifies inter-
1416 and intracontinental transmission events. *Nat Genet* 47, 632-639.

1417 World Health Organization (2014). *Antimicrobial resistance: global report on surveillance*. Geneva:
1418 World Health Organization (WHO).

1419 World Health Organization (2017). *Critically Important Antimicrobials for Human Medicine, 5th*
1420 *Revision*. Geneva: World Health Organization.

- 1421 Worley, J., Meng, J., Allard, M.W., Brown, E.W., and Timme, R.E. (2018). *Salmonella enterica*
1422 Phylogeny Based on Whole-Genome Sequencing Reveals Two New Clades and Novel
1423 Patterns of Horizontally Acquired Genetic Elements. *MBio* 9.
- 1424 Xie, W., Lewis, P.O., Fan, Y., Kuo, L., and Chen, M.H. (2011). Improving marginal likelihood
1425 estimation for Bayesian phylogenetic model selection. *Syst Biol* 60, 150-160.
- 1426 Yachison, C.A., Yoshida, C., Robertson, J., Nash, J.H.E., Kruczkiewicz, P., Taboada, E.N., Walker,
1427 M., Reimer, A., Christianson, S., Nichani, A., Pulsenet Canada Steering, C., and Nadon, C.
1428 (2017). The Validation and Implications of Using Whole Genome Sequencing as a
1429 Replacement for Traditional Serotyping for a National *Salmonella* Reference Laboratory.
1430 *Front Microbiol* 8, 1044.
- 1431 Yang, W.-C., Chan, O.-W., Wu, T.-L., Chen, C.-L., Su, L.-H., and Chiu, C.-H. (2016). Development
1432 of ceftriaxone resistance in *Salmonella enterica* serotype Oranienburg during therapy for
1433 bacteremia. *Journal of Microbiology, Immunology and Infection* 49, 41-45.
- 1434 Yang, Z. (1995). A space-time process model for the evolution of DNA sequences. *Genetics* 139,
1435 993-1005.
- 1436 Yoshida, C.E., Kruczkiewicz, P., Laing, C.R., Lingohr, E.J., Gannon, V.P., Nash, J.H., and Taboada,
1437 E.N. (2016). The *Salmonella In Silico* Typing Resource (SISTR): An Open Web-Accessible
1438 Tool for Rapidly Typing and Subtyping Draft *Salmonella* Genome Assemblies. *PLoS One* 11,
1439 e0147101.
- 1440 Yu, G., Lam, T.T., Zhu, H., and Guan, Y. (2018). Two Methods for Mapping and Visualizing
1441 Associated Data on Phylogeny Using Ggtree. *Mol Biol Evol* 35, 3041-3043.
- 1442 Yu, G., Smith, D.K., Zhu, H., Guan, Y., and Lam, T.T.-Y. (2017). ggtree: an r package for
1443 visualization and annotation of phylogenetic trees with their covariates and other associated
1444 data. *Methods in Ecology and Evolution* 8, 28-36.
- 1445 Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F.M.,
1446 and Larsen, M.V. (2012). Identification of acquired antimicrobial resistance genes. *J*
1447 *Antimicrob Chemother* 67, 2640-2644.
- 1448 Zhang, S., Den Bakker, H.C., Li, S., Chen, J., Dinsmore, B.A., Lane, C., Lauer, A.C., Fields, P.I.,
1449 and Deng, X. (2019). SeqSero2: Rapid and Improved *Salmonella* Serotype Determination
1450 Using Whole-Genome Sequencing Data. *Appl Environ Microbiol* 85.
- 1451 Zwe, Y.H., Chin, S.F., Kohli, G.S., Aung, K.T., Yang, L., and Yuk, H.-G. (2020). Whole genome
1452 sequencing (WGS) fails to detect antimicrobial resistance (AMR) from heteroresistant
1453 subpopulation of *Salmonella enterica*. *Food Microbiology* 91, 103530.

1455

1456 **10 Tables**

1457 **Table 1.** Statistics for the 15-antimicrobial phenotypic susceptibility/resistance prediction task for all antimicrobial resistance (AMR)
1458 determinant pipeline/database combinations.^a

AMR Pipeline	AMR Database	% Accuracy (95% Confidence Interval)	Cohen's Kappa (%)	Corrected Accuracy <i>P</i> -Value ^b	Corrected McNemar's Test <i>P</i> -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
AMRFinder Plus	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

1459 ^aStatistics were calculated using the confusionMatrix function in the caret package in R, with resistant (“R”) phenotypes/genotypes treated as
1460 the “positive” result and susceptible (“S”) phenotypes/genotypes treated as the “negative” result; See Supplementary Table S3 for an
1461 extended version of this table.

1462 ^bAdjusted using a Bonferroni correction.

1463

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

Group	PERMDISP2 Raw <i>P</i> -Value (<i>F</i>) ^b	ANOSIM Raw <i>P</i> - Value (<i>R</i>) ^c	PERMANOVA Raw <i>P</i> - Value (<i>R</i> ²) ^d
Pan-genome element presence/absence (<i>n</i> = 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 (<i>n</i> = 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 permutest functions in the vegan package in R; betadisper is a multivariate analogue 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

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)	1.67×10 ⁻⁷ /1.48×10 ⁻⁷ (6.92×10 ⁻¹¹ -3.86×10 ⁻⁷)
Cerro	13	21 (0-12)	Strict/Skyline	2008.2/2008.4 (2007.6-2008.6)	9.11×10 ⁻⁷ /8.94×10 ⁻⁷ (3.06×10 ⁻⁷ -1.57×10 ⁻⁶)
Kentucky	36	102 (0-30)	Relaxed/Skyline	2004.1/2005.0 (2000.8-2006.8)	6.39×10 ⁻⁷ /6.34×10 ⁻⁷ (2.05×10 ⁻⁷ -1.07×10 ⁻⁶)
Meleagridis	19	27 (0-9)	Strict/Skyline	2007.4/2007.5 (2006.9-2007.7)	6.88×10 ⁻⁷ /6.66×10 ⁻⁷ (2.90×10 ⁻⁷ -1.12×10 ⁻⁶)
Newport	16	52 (0-38)	Relaxed/Skyline	2004.2/2004.6 (2000.4-2007.8)	9.02×10 ⁻⁷ /8.22×10 ⁻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)	1.07×10 ⁻⁶ /9.66×10 ⁻⁷ (2.84×10 ⁻⁷ -2.05×10 ⁻⁶)

^aSee Supplementary Table S5 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 S5 for more details.

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

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

11 Figure Legends

Figure 1. Maximum likelihood phylogeny constructed using core SNPs identified among 570 *Salmonella* isolate genomes. Publicly available genomes are denoted by black tip labels ($n = 442$), while genomes of bovine- and bovine farm-associated strains isolated in conjunction with this study are denoted by colored tip labels ($n = 128$). In cases where a discrepancy existed between the traditional serotype designation of an isolate and one or more *in silico* methods (i.e., SISTR and SeqSero2), the serotype assigned using two out of the three methods was selected as the final serotype to be used for phylogeny annotation. The phylogeny is rooted at the midpoint with branch lengths reported in substitutions per site. Core SNPs were identified among all genomes using kSNP3, while the phylogeny was constructed and annotated using IQ-TREE and bactaxR/ggtree, respectively.

Figure 2. Confusion matrices showcasing agreement between susceptible-intermediate-resistant (SIR) classification of 128 *Salmonella* isolates obtained using phenotypic resistance testing (denoted as the matrix “Reference”) and five *in silico* methods (denoted as the matrix “Prediction”) for 15 antimicrobials. Accuracy values above each matrix denote the percentage of correctly classified instances out of all instances. Kappa values above each matrix denote Cohen’s kappa coefficient for the matrix, reported as a percent. For the phenotypic method, SIR classification was determined using NARMS breakpoints for *Salmonella* (accessed March 23, 2020). For the *in silico* antimicrobial resistance (AMR) determinant detection approaches, combinations of five pipelines and one to five AMR determinant databases were tested; isolate genomes that harbored one or more AMR determinants previously known to confer resistance to a particular antimicrobial were categorized as resistant to that antimicrobial (“R”), while those which did not were categorized as susceptible (“S”; see Supplementary Table S2 for all detected AMR determinants and their associated resistance classifications). For all AMR determinant detection methods, isolates that showed intermediate phenotypic resistance to an antimicrobial were categorized as susceptible (“S”) rather than resistant, as this classification produced slightly better accuracy scores for all pipeline/database combinations. For AMR determinant detection methods that relied on nucleotide BLAST (i.e., ABRicate and BTypser), the confusion matrix obtained using the nucleotide identity and coverage threshold combination that produced the highest accuracy are shown (Supplementary Figure S3); for all other methods, confusion matrices obtained using default detection parameters are shown.

Figure 3. (A) Deviation of PATRIC3 predicted minimum inhibitory concentration (MIC) values from “true” MIC values obtained using phenotypic resistance testing for 14 antimicrobials (Y-axis), reported in number of dilution factors (X-axis). For each box plot, lower and upper box hinges correspond to the first and third quartiles, respectively. Lower and upper whiskers extend from the hinge to the smallest and largest values no more distant than 1.5 times the interquartile range from the hinge, respectively. Points represent pairwise distances that fall beyond the ends of the whiskers. Only isolates with raw MIC values available are included (126 of 128 isolates). (B) Confusion matrix showcasing agreement between susceptible-intermediate-resistant (SIR) classification of all 128 *Salmonella* isolates obtained using phenotypic resistance testing (denoted as the matrix “Reference”) and PATRIC3 (denoted as the matrix “Prediction”) for 14 antimicrobials. The accuracy value above the matrix denotes the percentage of correctly classified instances out of all instances. The Kappa value above the matrix denotes Cohen’s kappa coefficient for the matrix, reported as a percent. For both the phenotypic and PATRIC3 methods, SIR classification was determined using NARMS breakpoints for *Salmonella* (accessed March 23, 2020).

Figure 4. Results of nonmetric 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.

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.

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;

1586 CIP, ciprofloxacin; GEN, gentamicin; KAN, kanamycin; NAL, nalidixic acid; STR, streptomycin;
1587 SXT, sulfamethoxazole-trimethoprim; SUL, sulfisoxazole; TET, tetracycline.

1588 **Figure 7.** Rooted, time-scaled maximum clade credibility (MCC) phylogeny constructed using core
1589 SNPs identified among 36 *Salmonella* Kentucky genomes isolated from subclinical bovine sources
1590 and the surrounding bovine farm environment. Tip label colors denote the ID of the farm from which
1591 each strain was isolated. Branch labels denote posterior probabilities of branch support. Time in years
1592 is plotted along the X-axis, and branch lengths are reported in years. The heatmap to the right of the
1593 phylogeny denotes (i) the susceptible-intermediate-resistant (SIR) classification of each isolate for
1594 each of 15 antimicrobials (obtained using phenotypic testing and NARMS breakpoints; orange); (ii)
1595 presence of plasmid replicons (detected using ABRicate/PlasmidFinder and minimum nucleotide
1596 identity and coverage thresholds of 80 and 60%, respectively; blue); (iii) presence of antimicrobial
1597 resistance (AMR) and stress response genes (identified using AMRFinderPlus and default
1598 parameters; green). Core SNPs were identified using Snippy. The phylogeny was constructed using
1599 the results of ten independent runs using a relaxed lognormal clock model, the Standard_TVMef
1600 nucleotide substitution model, and the Coalescent Bayesian Skyline population model implemented
1601 in BEAST version 2.5.1, with 10% burn-in applied to each run. LogCombiner-2 was used to combine
1602 BEAST 2 log files, and TreeAnnotator-2 was used to construct the phylogeny using common
1603 ancestor node heights. Abbreviations for the 15 antimicrobials are: AMK, amikacin; AMC,
1604 amoxicillin-clavulanic acid; AMP, ampicillin; FOX, cefoxitin; TIO, ceftiofur; CRO, ceftriaxone;
1605 CHL, chloramphenicol; CIP, ciprofloxacin; GEN, gentamicin; KAN, kanamycin; NAL, nalidixic
1606 acid; STR, streptomycin; SXT, sulfamethoxazole-trimethoprim; SUL, sulfisoxazole; TET,
1607 tetracycline.

1608 **Figure 8.** Rooted, time-scaled maximum clade credibility (MCC) phylogeny constructed using core
1609 SNPs identified among 19 *Salmonella* Meleagridis genomes isolated from subclinical bovine sources
1610 and the surrounding bovine farm environment. Tip label colors denote the ID of the farm from which
1611 each strain was isolated. Branch labels denote posterior probabilities of branch support. Time in years
1612 is plotted along the X-axis, and branch lengths are reported in years. The heatmap to the right of the
1613 phylogeny denotes (i) the susceptible-intermediate-resistant (SIR) classification of each isolate for
1614 each of 15 antimicrobials (obtained using phenotypic testing and NARMS breakpoints; orange); (ii)
1615 presence and absence of plasmid replicons (detected using ABRicate/PlasmidFinder and minimum
1616 nucleotide identity and coverage thresholds of 80 and 60%, respectively; blue); (iii) presence and
1617 absence of antimicrobial resistance (AMR) and stress response genes (identified using
1618 AMRFinderPlus and default parameters; green). Core SNPs were identified using Snippy. The
1619 phylogeny was constructed using the results of ten independent runs using a strict clock model, the
1620 Standard_TPM2 nucleotide substitution model, and the Coalescent Bayesian Skyline population
1621 model implemented in BEAST version 2.5.1, with 10% burn-in applied to each run. LogCombiner-2
1622 was used to combine BEAST 2 log files, and TreeAnnotator-2 was used to construct the phylogeny
1623 using common ancestor node heights. Abbreviations for the 15 antimicrobials are: AMK, amikacin;
1624 AMC, amoxicillin-clavulanic acid; AMP, ampicillin; FOX, cefoxitin; TIO, ceftiofur; CRO,
1625 ceftriaxone; CHL, chloramphenicol; CIP, ciprofloxacin; GEN, gentamicin; KAN, kanamycin; NAL,
1626 nalidixic acid; STR, streptomycin; SXT, sulfamethoxazole-trimethoprim; SUL, sulfisoxazole; TET,
1627 tetracycline.

1628 **Figure 9.** Rooted, time-scaled maximum clade credibility (MCC) phylogeny constructed using core
1629 SNPs identified among 16 *Salmonella* Newport genomes isolated from subclinical bovine sources
1630 and the surrounding bovine farm environment. Tip label colors denote the ID of the farm from which
1631 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.

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. Circles 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.

Serotype (This Study Only)

- Anatum
- Cerro
- Kentucky
- Meleagridis
- Minnesota
- Newport
- Typhimurium
- Typhimurium Copenhagen

bioRxiv preprint doi: <https://doi.org/10.1101/2021.08.21.457169>; this version posted August 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

A1

A2

A3

B1

B2

B3

Heatmap

Farm

Heatmap

- AMR/Stress Gene Absent
- AMR/Stress Gene Present
- Plasmid Replicon Absent
- Plasmid Replicon Present
- Susceptible Phenotype
- Intermediate Phenotype
- Resistant Phenotype

Farm

- 01
- 11

Farm

- a • 17
- a • 35
- a • 62

Heatmap

- AMR/Stress Gene Absent
- AMR/Stress Gene Present
- Plasmid Replicon Absent
- Plasmid Replicon Present
- Susceptible Phenotype
- Resistant Phenotype