



# Genomic Characterization of Endemic and Ecdemic Non-typhoidal *Salmonella enterica* Lineages Circulating Among Animals and Animal Products in South Africa

Laura M. Carroll<sup>1</sup>, Rian Pierneef<sup>2</sup>, Masenyabu Mathole<sup>3</sup> and Itumeleng Matle<sup>3\*</sup>

## OPEN ACCESS

### Edited by:

Zhi Ruan,  
Zhejiang University, China

### Reviewed by:

Oliver Zishiri,  
University of KwaZulu-Natal,  
South Africa  
Stella Smith,  
Nigerian Institute of Medical Research  
(NIMR), Nigeria  
Manal Mohammed,  
University of Westminster,  
United Kingdom

### \*Correspondence:

Itumeleng Matle  
matlei@arc.agric.za

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

Received: 28 July 2021

Accepted: 09 September 2021

Published: 04 October 2021

### Citation:

Carroll LM, Pierneef R, Mathole M  
and Matle I (2021) Genomic  
Characterization of Endemic  
and Ecdemic Non-typhoidal  
*Salmonella enterica* Lineages  
Circulating Among Animals  
and Animal Products in South Africa.  
*Front. Microbiol.* 12:748611.  
doi: 10.3389/fmicb.2021.748611

<sup>1</sup> Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany, <sup>2</sup> Biotechnology Platform, Agricultural Research Council-Onderstepoort Veterinary Research, Onderstepoort, South Africa, <sup>3</sup> Bacteriology Division, Agricultural Research Council-Onderstepoort Veterinary Research, Onderstepoort, South Africa

In Africa, the burden of illness caused by non-typhoidal *Salmonella enterica* is disproportionately high; however, whole-genome sequencing (WGS) efforts are overwhelmingly concentrated in world regions with lower burdens. While WGS is being increasingly employed in South Africa to characterize *Salmonella enterica*, the bulk of these efforts have centered on characterizing human clinical strains. Thus, very little is known about lineages circulating among animals in the country on a genomic scale. Here, we used WGS to characterize 63 *Salmonella enterica* strains isolated from livestock, companion animals, wildlife, and animal products in South Africa over a 60-year period. Genomes were assigned to serotypes Dublin, Hadar, Enteritidis, and Typhimurium ( $n = 18, 8, 13,$  and  $24$  strains, respectively) and sequence types (STs) ST10 (all *S. Dublin*), ST33 (all *S. Hadar*), ST11/ST366 ( $n = 12$  and  $1$  *S. Enteritidis*, respectively), and ST19/ST34 ( $n = 23$  and  $1$  *S. Typhimurium*, respectively; via seven-gene multi-locus sequence typing). Within-ST phylogenies were constructed using genomes sequenced in this study, plus publicly available genomes representative of each ST's (i) global ( $n = 2,802$  and  $1,569$  *S. Dublin* and *Hadar* genomes, respectively) and (ii) African ( $n = 716$  and  $343$  *S. Enteritidis* and *Typhimurium* genomes, respectively) population. For *S. Dublin* ST10, a largely antimicrobial-susceptible, endemic lineage circulating among humans, animals, and food in South Africa was identified, as well as a lineage that was likely recently introduced from the United States. For *S. Hadar* ST33, multiple South African lineages harboring streptomycin and tetracycline resistance-conferring genes were identified. African *S. Enteritidis* ST11 could be primarily partitioned into one largely antimicrobial-susceptible and one largely multidrug-resistant (MDR) clade, with South African isolates confined to the largely antimicrobial-susceptible clade. *S. Typhimurium* ST19/ST34 strains sequenced here were distributed across the African *S. Typhimurium* ST19/ST34 phylogeny, representing a diverse range of lineages, including

numerous MDR lineages. Overall, this study provides critical insights into endemic and ecdemic non-typhoidal *Salmonella enterica* lineages circulating among animals, foods, and humans in South Africa and showcases the utility of WGS in characterizing animal-associated strains from a world region with a high salmonellosis burden.

**Keywords:** *Salmonella* Dublin, *Salmonella* Hadar, *Salmonella* Enteritidis, *Salmonella* Typhimurium, antimicrobial resistance, whole-genome sequencing, foodborne illness, Africa

## INTRODUCTION

Livestock, domestic animals, and wildlife can serve as potential reservoirs for non-typhoidal *Salmonella enterica* (Hoelzer et al., 2011; World Health Organization, 2018). As a zoonotic foodborne pathogen, *Salmonella enterica* can be transmitted from these animal reservoirs to humans, either via direct contact with infected animals or along the food supply chain (Hoelzer et al., 2011; Hald et al., 2016); however, evolutionary lineages within the *Salmonella enterica* species may vary in terms of their host specificity, geographic distribution, and the severity of illness that they cause in a given host (Hoelzer et al., 2011; Centers for Disease Control and Prevention, 2020). *Salmonella enterica* serotype Typhimurium (*S. Typhimurium*), for example, can infect a broad range of species, while serotype Dublin (*S. Dublin*) is largely adapted to cattle, but can cause rare but frequently invasive infections in humans (Taylor et al., 1982; Uzzau et al., 2000; Rabsch et al., 2002; Cummings et al., 2012; Rodriguez-Rivera et al., 2014; Harvey et al., 2017; Mohammed et al., 2017).

Due to its importance as a pathogen from both a human and animal health perspective, there is a strong incentive to monitor the evolution and spread of *Salmonella enterica* in animals and animal products (European Food Safety Authority, and European Centre for Disease Prevention and Control, 2018; Ferrari et al., 2019). Furthermore, there has been growing concern that *Salmonella enterica* can acquire antimicrobial resistance (AMR) determinants in livestock environments, which can make infections in humans and animals more difficult and costly to treat (World Health Organization, 2014; Centers for Disease Control and Prevention, 2019). To this end, whole-genome sequencing (WGS) is being increasingly employed to characterize *Salmonella enterica* from animals (e.g., livestock, companion animals, and wildlife) and animal products, as WGS can not only replicate many important microbiological assays *in silico* (e.g., prediction of serotype, AMR), but provide additional data that can be used to characterize isolates (e.g., identification of genome-wide single nucleotide polymorphisms [SNPs], core- and whole-genome multi-locus sequence typing [MLST], pan-genome characterization) (Yoshida et al., 2016; Ibrahim and Morin, 2018; Carroll, 2019; Zhang et al., 2019).

In South Africa, the bulk of *Salmonella enterica* WGS efforts have focused on characterizing human clinical strains associated with illnesses and/or outbreaks (Wong et al., 2015; Feasey et al., 2016; Smith et al., 2017, 2020; Park et al., 2018; Rule et al., 2019). WGS-based studies querying *Salmonella enterica* strains isolated from non-human sources in South Africa are limited (El Zowalaty et al., 2020), and little is known regarding which lineages are circulating among animals in the country (Mthembu et al., 2021). Here, we used WGS to characterize 63 South African

*Salmonella enterica* strains isolated from animals and animal products over the course of 60 years (i.e., between 1960 and 2019). Using phylogenomic approaches, we characterized the isolates sequenced here within the context of publicly available genomes representative of global (for *S. Dublin* and *S. Hadar*) and African (for *S. Enteritidis* and *S. Typhimurium*) *Salmonella enterica* populations. The results presented here will provide preliminary, critical insights into the evolution, population structure, and AMR dynamics of *S. Dublin*, *Hadar*, *Enteritidis*, and *Typhimurium* in Africa.

## MATERIALS AND METHODS

### Isolate Selection

The isolates used in this study were recovered from samples submitted between 1957 and 2019 at Bacteriology laboratory: Onderstepoort Veterinary Research, South Africa, as part of routine diagnostics services, which includes isolation and serotyping of *Salmonella* strains. Therefore, a total of 73 isolates representing (i) four major *Salmonella enterica* serotypes (i.e., Dublin, Enteritidis, Hadar, and Typhimurium) in the country (Kidanemariam et al., 2010; Gelaw et al., 2018) from (ii) various geographical locations in the country, (iii) different sources of isolation (animal and animal products), and (iv) animal species (livestock, companion animals, wildlife) were randomly selected for sequencing in this study. All strains had been isolated from animals or animal products; for four strains, the origin animal/animal product was unspecified (**Supplementary Table 1**). The isolates were preserved as lyophilized and revived by inoculation into brain heart infusion (BHI) broth and incubated at 37°C for 18–24 h.

### Whole-Genome Sequencing

Genomic DNA was extracted from BHI broth cultures using the High Pure PCR template preparation kit (Roche, Potsdam, Germany) according to the manufacturer's instructions. WGS of the isolates was performed at the Biotechnology Platform, Agricultural Research Council, South Africa. DNA libraries were prepared using TruSeq and Nextera DNA library preparation kits (Illumina, San Diego, CA, United States), followed by sequencing on Illumina HiSeq and MiSeq instruments (Illumina, San Diego, CA, United States).

### Initial Data Processing and Quality Control

Quality control, adapter removal, decontamination, and error correction of the raw sequencing data was performed using

BBDuk v. 37.90<sup>1</sup>, and SPAdes v. 3.12.0 (Bankevich et al., 2012) was used to create a *de novo* assembly for each isolate. FastQC v. 0.11.5<sup>2</sup> was used to assess the quality of the paired-end reads associated with each isolate ( $n = 73$  isolates total; 21, 15, 11, and 26 isolates assigned to serotypes Dublin, Enteritidis, Hadar, and Typhimurium, respectively) (Andrews, 2019), and QUAST v. 4.5 (Gurevich et al., 2013) was used to assess the quality of the associated assembled genome (**Supplementary Table 1**). The lineage workflow (i.e., “lineage\_wf”) implemented in CheckM v. 1.1.3 (Parks et al., 2015) was additionally used to identify potential contamination in each assembled genome, as well as to assess genome completeness (**Supplementary Table 1**). MultiQC v. 1.8 (Ewels et al., 2016) was used to assess the quality of all genomes in aggregate. Several low-quality isolate genomes with >5% contamination and/or <95% completeness were identified ( $n = 3, 2, 3,$  and  $2$  low-quality isolate genomes assigned to serotypes Dublin, Enteritidis, Hadar, and Typhimurium, respectively) and were thus omitted from further analysis, yielding a final set of 63 *Salmonella enterica* genomes used in subsequent steps (**Supplementary Table 1**).

### **In silico Serotyping and Multi-Locus Sequence Typing**

All 63 assembled *Salmonella enterica* genomes (see section “Initial Data Processing and Quality Control” above) underwent *in silico* serotyping using the command line implementations of (i) the *Salmonella In Silico* Typing Resource (SISTR) v. 1.1.1 (Yoshida et al., 2016) and (ii) the *k*-mer based workflow implemented in SeqSero2 v. 1.1.1 (Zhang et al., 2019; **Supplementary Table 1**). Each genome additionally underwent *in silico* seven-gene multi-locus sequence typing (MLST) using mlst v. 2.9<sup>3</sup> and the seven-gene scheme available for *Salmonella enterica* (–scheme ‘senterica’) in PubMLST (Jolley and Maiden, 2010; Jolley et al., 2018; **Supplementary Table 1**).

### **Reference-Free Single Nucleotide Polymorphism Identification and Phylogeny Construction**

The 63 *Salmonella enterica* genomes sequenced in this study were compared to 442 of the 445 *Salmonella* genomes described by Worley et al. (2018) (three genomes did not have publicly available sequence read archive [SRA] data at the time of access, i.e., 20 February 2019). The SRA toolkit v. 2.9.6 was used to download paired-end reads for each of the 442 publicly available genomes (Leinonen et al., 2011; Kodama et al., 2012), which were then assembled into contigs using SPAdes v. 3.8.0 (Bankevich et al., 2012), using *k*-mer sizes of 21, 33, 55, 77, 99, and 127, and the “careful” option. SNPs were identified among all 505 assembled *Salmonella* genomes with kSNP3 v. 3.92 (Gardner and Hall, 2013; Gardner et al., 2015), using the optimal *k*-mer size determined by Kchooser ( $k = 19$ ). The resulting core SNP alignment was supplied as input to

IQ-TREE v. 1.5.4 (Nguyen et al., 2015), which was used to construct a maximum likelihood (ML) phylogeny using the optimal ascertainment bias-aware nucleotide substitution model identified using ModelFinder (based on its Bayesian Information Criteria [BIC] value) (Kalyaanamoorthy et al., 2017) and 1,000 replicates of the Ultrafast Bootstrap method (Minh et al., 2013; Hoang et al., 2018). The resulting ML phylogeny (**Figure 1**) was annotated using FigTree v. 1.4.4<sup>4</sup>. All reference-free SNP identification and ML phylogeny construction steps described above were repeated to identify SNPs among the 63 *Salmonella enterica* genomes sequenced here, with publicly available genomes excluded; the resulting ML phylogeny was annotated in R v. 3.6.1 (R Core Team, 2019) using the bactaxR package (Carroll et al., 2020b) and its dependencies ggtree (Yu et al., 2017; Yu et al., 2018), ape (Paradis and Schliep, 2019), dplyr (Wickham et al., 2020), phylobase (R Hackathon, 2019), phytools (Revell, 2012), and reshape2 (Wickham, 2007; **Figure 2** and **Supplementary Table 2**).

### **In silico Antimicrobial Resistance Determinant, Plasmid Replicon, and Virulence Factor Detection**

Antimicrobial resistance determinants were identified within each of the 63 *Salmonella* genomes sequenced in this study, using each of the following pipelines: (i) AMRFinderPlus v. 3.9.3 (Feldgarden et al., 2019), (ii) ABRicate v. 1.0.1<sup>5</sup>, and (iii) ARIBA v. 2.14.6 (Hunt et al., 2017) (**Figure 2** and **Supplementary Figure 1**). For the AMRFinderPlus pipeline, Prokka v. 1.13 (Seemann, 2014) was used to annotate each of the 63 assembled genomes; the resulting GFF (.gff) and FASTA (.faa and .ffn) files were used as input for AMRFinderPlus, which was used to identify AMR and stress response determinants in each genome, using the *Salmonella* organism option and the most recent AMRFinderPlus database (database v. 2020-11-09.1, accessed 21 November 2020). For the ABRicate pipeline, AMR determinants were identified in each assembled genome using the NCBI AMR database (–db ncbi; accessed 19 April 2020) (Feldgarden et al., 2019) and minimum identity and coverage thresholds of 75 (–minid 75) and 50% (–mincov 50), respectively. For the ARIBA pipeline, ARIBA’s getref and prepareref commands were used to download and prepare the latest version of the ResFinder database (accessed 14 February 2021), respectively (Zankari et al., 2012). ARIBA’s run command was then used to identify AMR determinants in each genome, using the paired-end reads associated with each isolate as input.

ABRicate and ARIBA were additionally used to detect plasmid replicons within each of the 63 *Salmonella* genomes sequenced in this study using the PlasmidFinder database (Carattoli et al., 2014) (**Figure 2** and **Supplementary Figure 2**). For the ABRicate pipeline, assembled genomes were used as input, and plasmid replicons were detected in each genome (–db plasmidfinder; PlasmidFinder database accessed 19 April 2020) using minimum identity and coverage thresholds of 80 (–minid 80) and 60% (–mincov 60), respectively. For the ARIBA pipeline, ARIBA’s getref

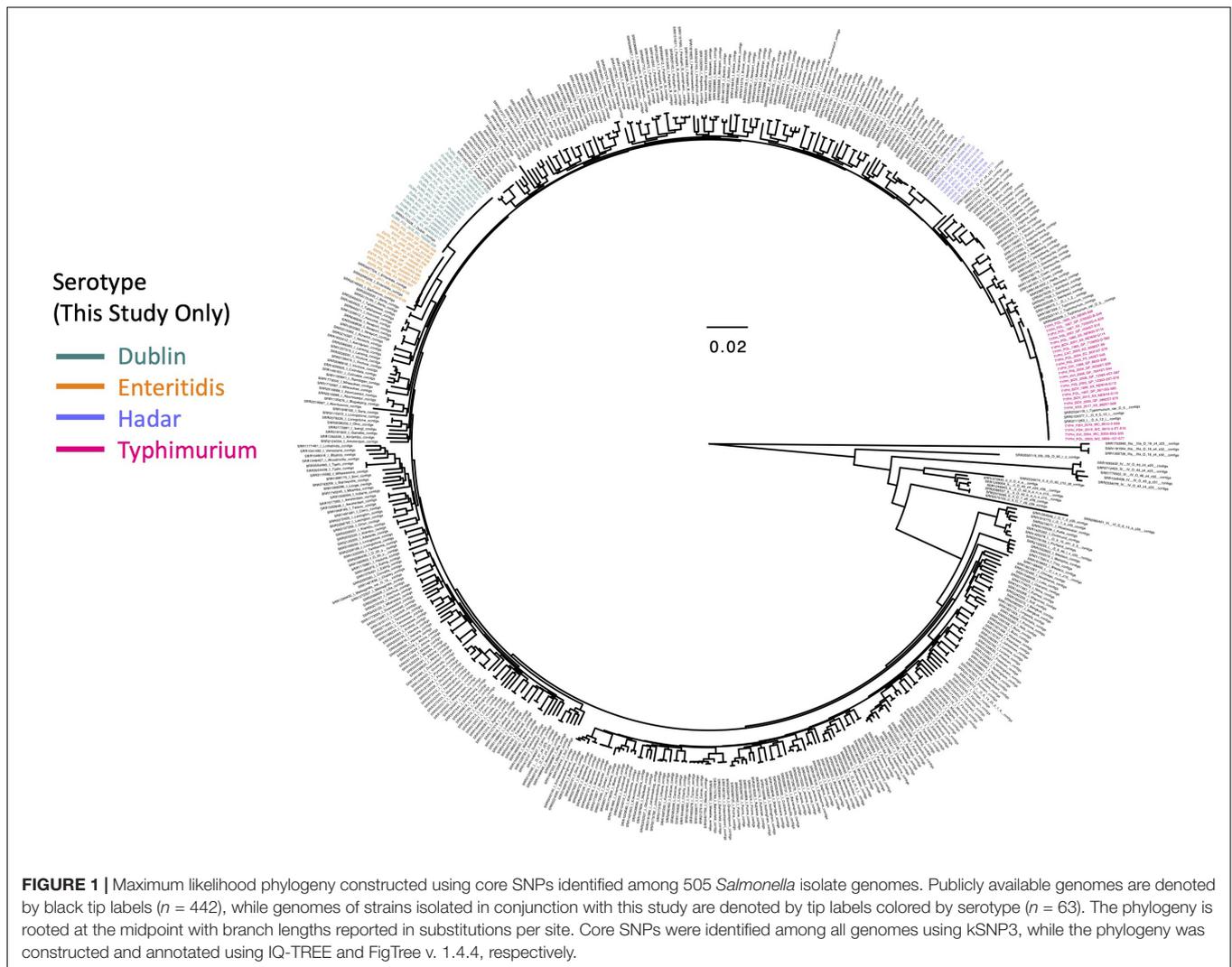
<sup>1</sup> <https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbdduk-guide/>

<sup>2</sup> <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

<sup>3</sup> <https://github.com/tseemann/mlst>

<sup>4</sup> <http://tree.bio.ed.ac.uk/software/figtree/>

<sup>5</sup> <https://github.com/tseemann/abricate>



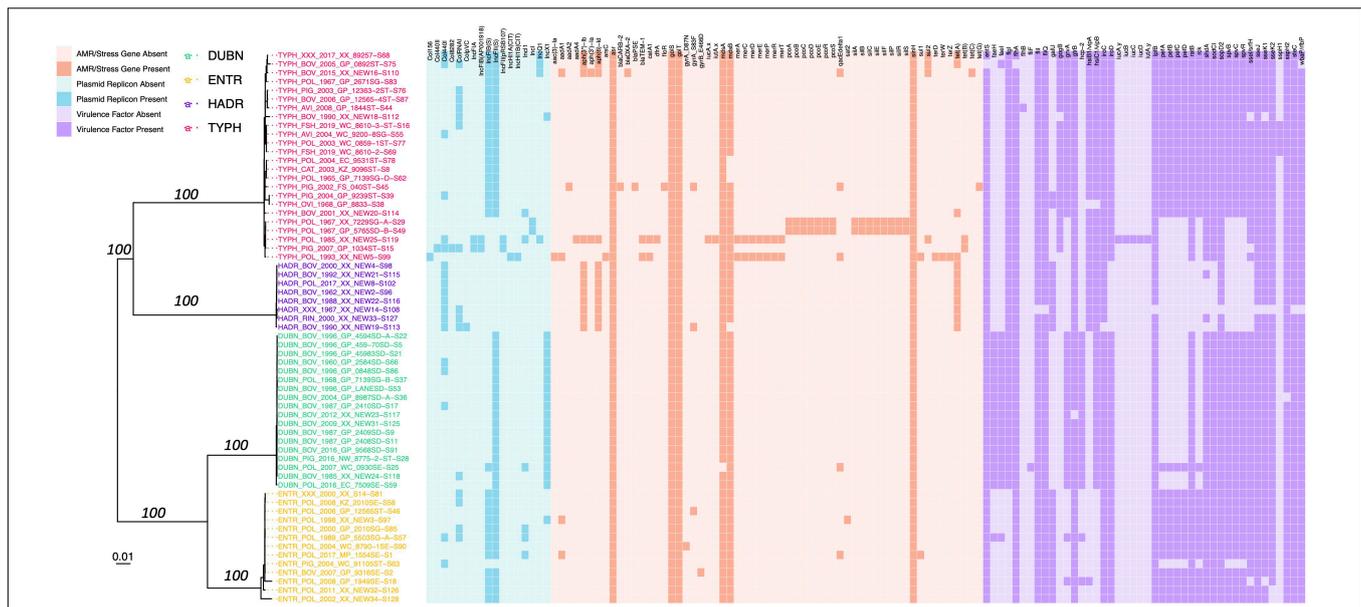
and preparer commands were used to download and prepare the latest version of the PlasmidFinder database (accessed 14 February 2021), respectively. ARIBA's run command was then used to identify plasmid replicons in each genome, using paired-end reads associated with each isolate as input. ABRicate was further used to detect virulence factors in each genome, using the Virulence Factor Database (VFDB; `-db vfdb`, accessed 19 April 2020) (Chen et al., 2005; Liu et al., 2019), using minimum identity and coverage thresholds of 70 (`-minid 70`) and 50% (`-mincov 50`), respectively (Figure 2 and Supplementary Table 2).

### Construction of Time-Scaled *S. Dublin* Phylogenies

To compare the 18 *S. Dublin* isolates sequenced in this study to publicly available *S. Dublin* genomes with known isolation sources and geographic origins, all genomes meeting each of the following conditions were downloaded via Enterobase (accessed 27 December 2020,  $n = 2,784$ ; Supplementary Table 3): (i) genomes were assigned to sequence type (ST) 10 (i.e., the ST

to which all *S. Dublin* isolates sequenced in this study were assigned/approximately assigned) using the Achtman seven-gene MLST scheme for *Salmonella*; (ii) genomes had an exact year of isolation reported in Enterobase's "Collection Year" field; (iii) genomes could be assigned to a known isolation source, with "Laboratory" strains excluded, per Enterobase's "Source Niche" field; (iv) genomes could be assigned to a known country of isolation, per Enterobase's "Country" field (Alikhan et al., 2018; Zhou et al., 2020). All 2,802 assembled *S. Dublin* genomes underwent *in silico* plasmid replicon and AMR determinant detection using ABRicate v. 1.0.1 and the PlasmidFinder and NCBI AMR databases, respectively, as described above (see section "In silico AMR Determinant, Plasmid Replicon, and Virulence Factor Detection" above).

Parsnp and HarvestTools v. 1.2 (Treangen et al., 2014) were used to identify core SNPs among all 2,802 *S. Dublin* genomes (2,784 publicly available genomes, plus the 18 sequenced here), using the closed chromosome of ST10 *S. Dublin* str. USMARC-69838 (NCBI Nucleotide Accession NZ\_CP032449.1) as a reference and Parsnp's implementation of PhiPack to remove



**FIGURE 2 |** Maximum likelihood phylogeny constructed using core SNPs identified among the genomes of 63 *Salmonella* strains isolated in conjunction with this study. Tip label colors denote isolate serotypes, and branch labels denote ultrafast bootstrap support percentages out of 1,000 replicates (selected for readability). The heatmap to the right of the phylogeny denotes the presence and absence of (i) plasmid replicons (blue), (ii) antimicrobial resistance (AMR) and stress response determinants (orange), and (iii) variably detected virulence factors (purple) in each genome. The phylogeny is rooted at the midpoint with branch lengths reported in substitutions per site. Core SNPs were identified among all genomes using kSNP3. Plasmid replicons were identified using ABRicate and the PlasmidFinder database, using minimum identity and coverage thresholds of 80 and 60%, respectively. AMR and stress response determinants were identified using AMRFinderPlus. Virulence factors were identified using ABRicate and VFDB, using minimum identity and coverage thresholds of 70 and 50%, respectively. Virulence factors detected in all genomes were excluded for readability (**Supporting Table 2**). The phylogeny was constructed and annotated using IQ-TREE and *bactaxR/ggtr*, respectively. DUBN, *S. Dublin*; ENTR, *S. Enteritidis*; HADR, *S. Hadar*; TYPH, *S. Typhimurium*.

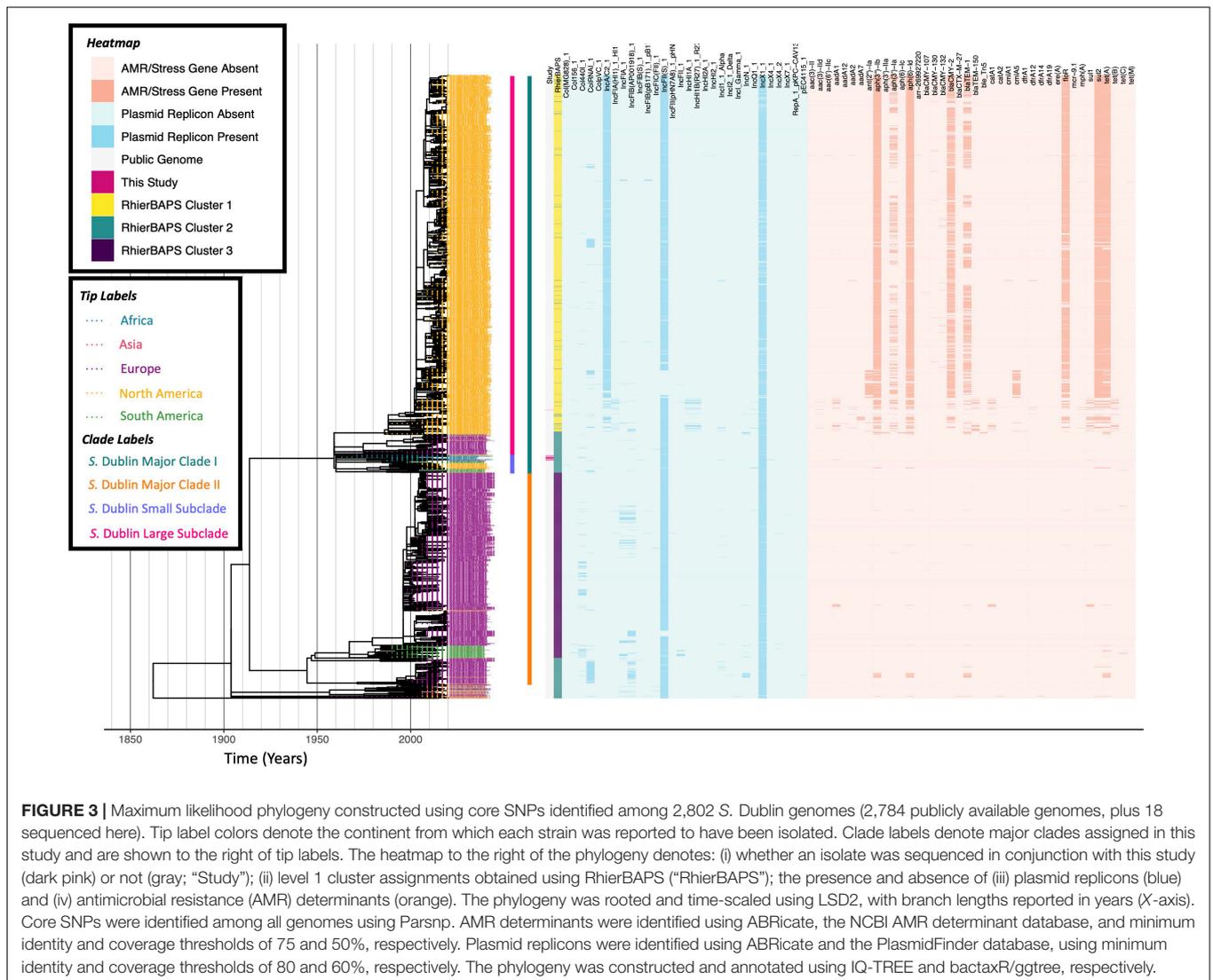
recombination (Bruen et al., 2006). Clusters were identified within the resulting core SNP alignment using RhierBAPs v. 1.1.3 (Cheng et al., 2013; Tonkin-Hill et al., 2018), R v. 4.0.0, and three clustering levels. IQ-TREE v. 1.5.4 (Nguyen et al., 2015) was used to construct a ML phylogeny using (i) the resulting core SNPs as input; (ii) an ascertainment bias correction (to account for the use of solely variant sites), corresponding to constant sites estimated using the GC content of the reference chromosome (-fconst 1171365,1282543,1281883,117722); (iii) the optimal nucleotide substitution model selected using ModelFinder (Kalyaanamoorthy et al., 2017), based on its corresponding BIC value (i.e., the TVM + I model); (iv) 1,000 replicates of the UltraFast bootstrap approximation (Minh et al., 2013).

The resulting ML phylogeny was rooted and time-scaled using LSD2 v. 1.4.2.2 (To et al., 2015) and the following parameters: (i) tip dates corresponding to the year of isolation associated with each genome; (ii) a fixed substitution rate of  $2.79 \times 10^{-7}$  substitutions/site/year (i.e., the substitution rate estimated in a previous study of *S. Typhimurium* phage type DT104) (Leekitcharoenphon et al., 2016); (iii) constrained mode (-c), with the root estimated using constraints on all branches (-r as); (iv) variances calculated using input branch lengths (-v 1); (v) 1,000 samples for calculating confidence intervals for estimated dates (-f 1000); (vi) a sequence length of 4,913,018 (i.e., the length of the reference chromosome; -s 4913018). The resulting phylogeny was annotated using the *bactaxR* package in R (Figure 3). All aforementioned *S. Dublin* SNP calling and

phylogeny construction steps were repeated to construct time-scaled ML phylogenies using the following subsets of *S. Dublin* genomes: (i) members of a large *S. Dublin* clade, which contained all 18 *S. Dublin* isolates sequenced in this study (i.e., “*S. Dublin* Major Clade I,”  $n = 1,787$  genomes; **Supporting Figure 3**); (ii) a smaller clade within *S. Dublin* Major Clade I, which contained 17 of the 18 *S. Dublin* isolates sequenced here (i.e., the “*S. Dublin* Small Subclade,”  $n = 78$ ; **Figure 4**); (iii) a larger clade within *S. Dublin* Major Clade I, which contained one *S. Dublin* isolate sequenced here (i.e., the “*S. Dublin* Large Subclade,”  $n = 1,709$ ; **Supporting Figure 4**).

## Construction of Time-Scaled *S. Hadar* Phylogeny

To compare the eight *S. Hadar* isolates sequenced in this study to publicly available *S. Hadar* genomes with known isolation sources and geographic origins, all genomes meeting each of the following conditions were downloaded via Enterobase (accessed 10 January 2021,  $n = 1,562$ ; **Supporting Table 4**): (i) genomes were assigned to ST33 (i.e., the ST to which all *S. Hadar* isolates sequenced in this study were assigned) using the Achtman seven-gene MLST scheme for *Salmonella*; (ii) genomes had an exact year of isolation reported in Enterobase’s “Collection Year” field; (iii) genomes could be assigned to a known isolation source, with “Laboratory” strains excluded, per Enterobase’s “Source Niche” field; (iv) genomes could be assigned to a known country



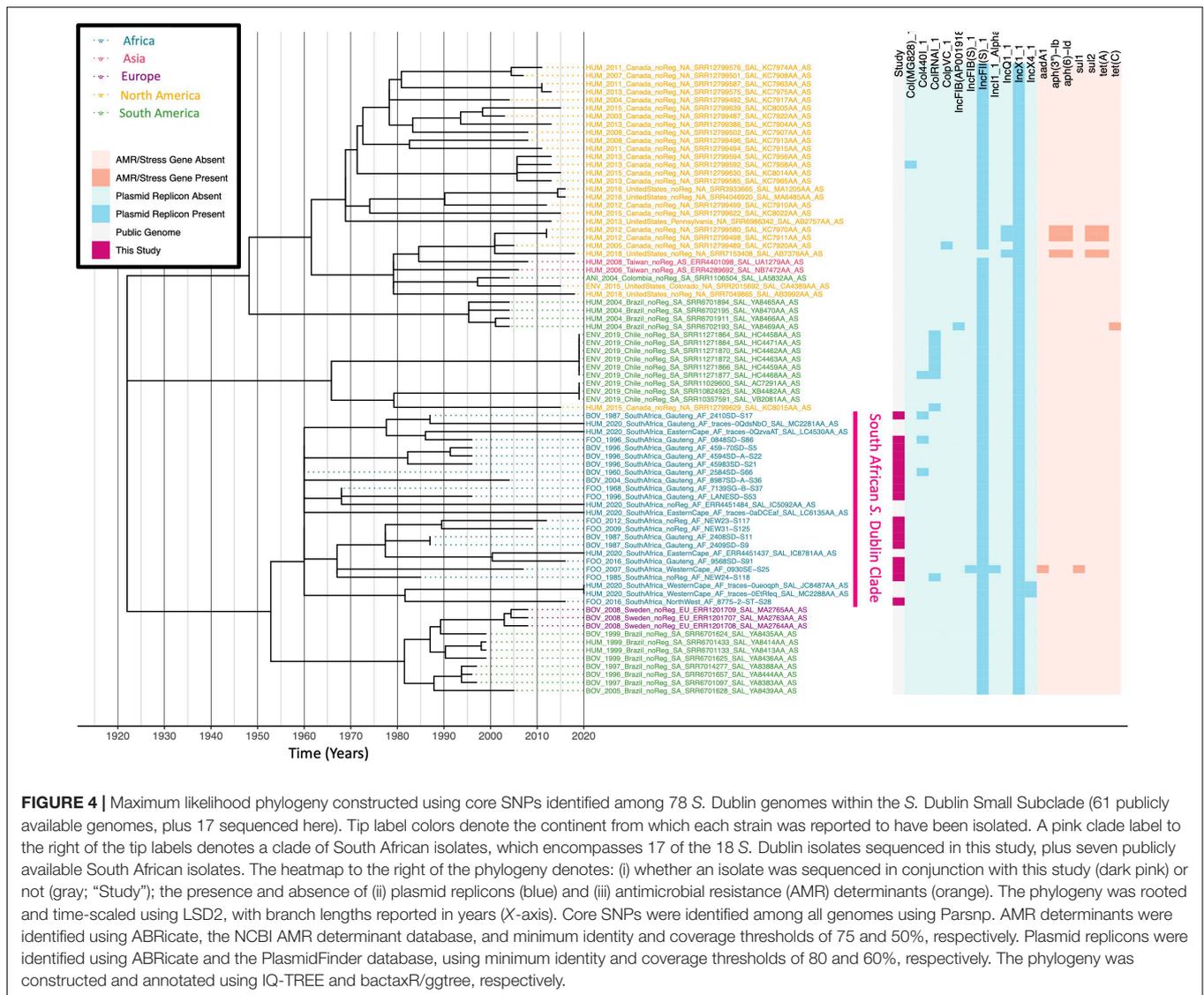
**FIGURE 3 |** Maximum likelihood phylogeny constructed using core SNPs identified among 2,802 *S. Dublin* genomes (2,784 publicly available genomes, plus 18 sequenced here). Tip label colors denote the continent from which each strain was reported to have been isolated. Clade labels denote major clades assigned in this study and are shown to the right of tip labels. The heatmap to the right of the phylogeny denotes: (i) whether an isolate was sequenced in conjunction with this study (dark pink) or not (gray; “Study”); (ii) level 1 cluster assignments obtained using RhierBAPS (“RhierBAPS”); the presence and absence of (iii) plasmid replicons (blue) and (iv) antimicrobial resistance (AMR) determinants (orange). The phylogeny was rooted and time-scaled using LSD2, with branch lengths reported in years (X-axis). Core SNPs were identified among all genomes using Parsnp. AMR determinants were identified using ABRicate, the NCBI AMR determinant database, and minimum identity and coverage thresholds of 75 and 50%, respectively. Plasmid replicons were identified using ABRicate and the PlasmidFinder database, using minimum identity and coverage thresholds of 80 and 60%, respectively. The phylogeny was constructed and annotated using IQ-TREE and bactaxR/ggtree, respectively.

of isolation, per Enterobase’s “Country” field (Alikhan et al., 2018; Zhou et al., 2020). All 1,570 assembled *S. Hadar* genomes underwent *in silico* plasmid replicon and AMR determinant detection using ABRicate v. 1.0.1 and the PlasmidFinder and NCBI AMR databases, respectively, as described above (see section “*In silico* AMR determinant, plasmid replicon, and virulence factor detection” above).

Parsnp and HarvestTools v. 1.2 (Treangen et al., 2014) were used to identify core SNPs among all 1,570 *S. Hadar* genomes (1,562 publicly available genomes, plus the eight sequenced here), using the closed chromosome of ST33 *S. Hadar* str. FDAARGOS\_313 (NCBI Nucleotide Accession NZ\_CP022069.2) as a reference and Parsnp’s implementation of PhiPack to remove recombination (Bruen et al., 2006). Clusters were identified within the resulting core SNP alignment using RhierBAPS v. 1.1.3 (Cheng et al., 2013; Tonkin-Hill et al., 2018), R v. 4.0.0, and three clustering levels. IQ-TREE v. 1.5.4 (Nguyen et al., 2015) was used to construct a ML phylogeny using (i) the resulting core SNPs as input; (ii) an ascertainment bias correction (to account for

the use of solely variant sites), corresponding to constant sites estimated using the GC content of the reference chromosome (-fconst 1179063,1283051,1279961,1174705); (iii) the optimal nucleotide substitution model selected using ModelFinder (Kalyaanamoorthy et al., 2017), based on its corresponding BIC value (i.e., the K3Pu + I model) (Kimura, 1981); (iv) 1,000 replicates of the UltraFast bootstrap approximation (Minh et al., 2013). All aforementioned SNP calling and phylogeny construction steps were repeated, with a single outlier genome from the United Kingdom (Enterobase Assembly Barcode SAL\_GB0368AA\_AS) removed, yielding a 1,569-isolate *S. Hadar* phylogeny that was used in subsequent steps.

The resulting ML phylogeny was rooted and time-scaled using LSD2 v. 1.4.2.2 (To et al., 2015) and the following parameters: (i) tip dates corresponding to the year of isolation associated with each genome; (ii) a fixed substitution rate of  $2.79 \times 10^{-7}$  substitutions/site/year (i.e., the substitution rate estimated in a previous study of *S. Typhimurium* phage type DT104) (Leekitcharoenphon et al., 2016); (iii) constrained mode



**FIGURE 4 |** Maximum likelihood phylogeny constructed using core SNPs identified among 78 *S. Dublin* genomes within the *S. Dublin* Small Subclade (61 publicly available genomes, plus 17 sequenced here). Tip label colors denote the continent from which each strain was reported to have been isolated. A pink clade label to the right of the tip labels denotes a clade of South African isolates, which encompasses 17 of the 18 *S. Dublin* isolates sequenced in this study, plus seven publicly available South African isolates. The heatmap to the right of the phylogeny denotes: (i) whether an isolate was sequenced in conjunction with this study (dark pink) or not (gray; “Study”); the presence and absence of (ii) plasmid replicons (blue) and (iii) antimicrobial resistance (AMR) determinants (orange). The phylogeny was rooted and time-scaled using LSD2, with branch lengths reported in years (*X*-axis). Core SNPs were identified among all genomes using Parsnp. AMR determinants were identified using ABRicate, the NCBI AMR determinant database, and minimum identity and coverage thresholds of 75 and 50%, respectively. Plasmid replicons were identified using ABRicate and the PlasmidFinder database, using minimum identity and coverage thresholds of 80 and 60%, respectively. The phylogeny was constructed and annotated using IQ-TREE and bactxR/ggtree, respectively.

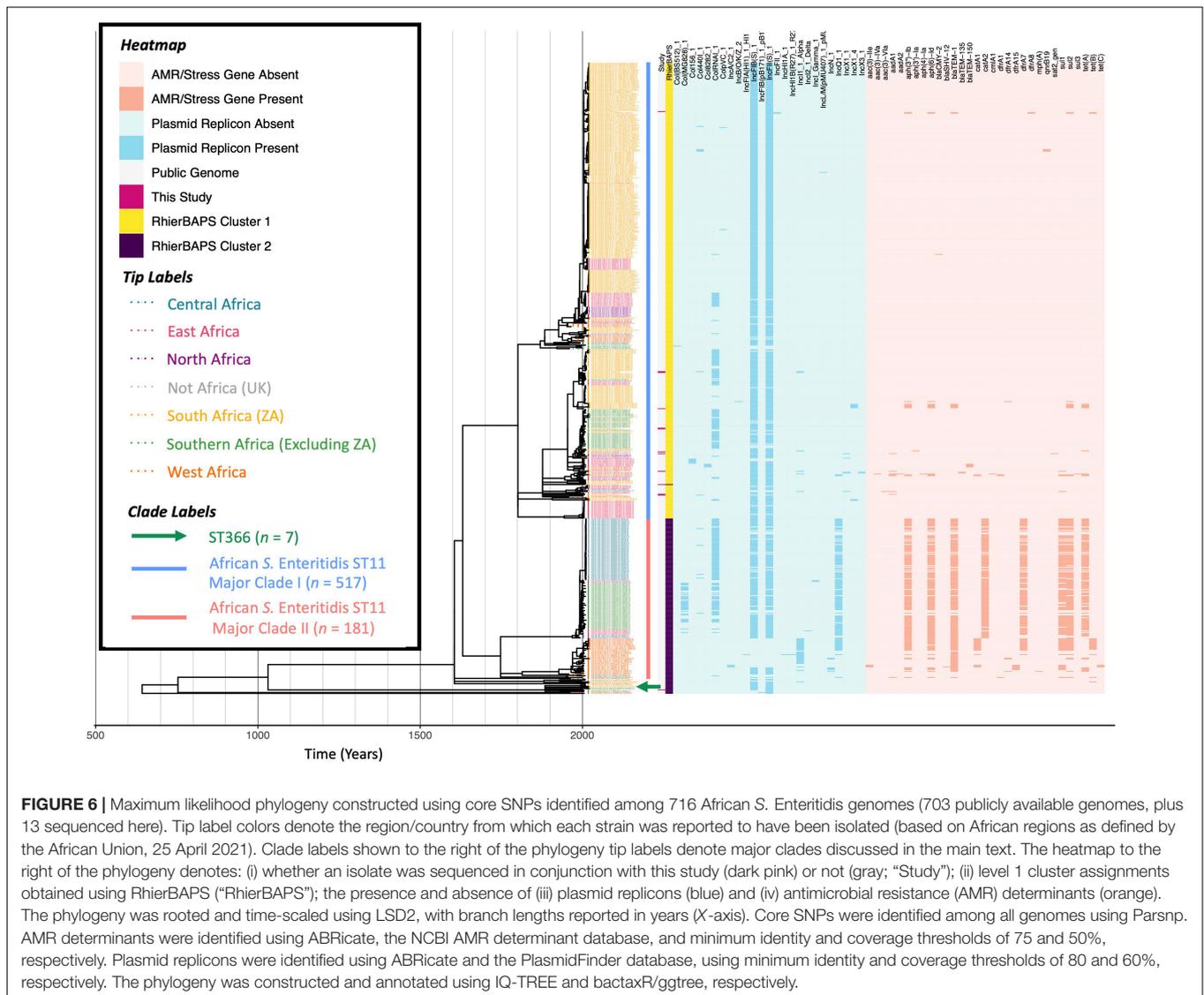
(-c), with the root estimated using constraints on all branches (-r as); (iv) variances calculated using input branch lengths (-v 1); (v) 1,000 samples for calculating confidence intervals for estimated dates (-f 1000); (vi) a sequence length of 4,916,780 (i.e., the length of the reference chromosome; -s 4916780). The resulting phylogeny was annotated using the bactxR package in R (Figure 5).

### Construction of Time-Scaled *S. Enteritidis* Phylogenies

To compare the 13 *S. Enteritidis* isolates sequenced in this study to publicly available *S. Enteritidis* genomes with known geographic origins, all genomes meeting each of the following conditions were downloaded via Enterobase (accessed 27 December 2020, *n* = 697; Supplementary Table 5): (i) genomes were assigned to ST11 (i.e., the ST to which 12 of the 13 *S. Enteritidis* isolates sequenced in this study were

assigned/approximately assigned) using the Achtman seven-gene MLST scheme for *Salmonella*; (ii) genomes had an exact year of isolation reported in Enterobase’s “Collection Year” field; (iii) genomes could be assigned to a known country of isolation within the African continent, per Enterobase’s “Country” and “Continent” fields, respectively (Alikhan et al., 2018; Zhou et al., 2020). Additionally, one isolate sequenced here was assigned to ST366, a ST that differs from ST11 by a single allele (i.e., *purE*). As such, all ST366 genomes available in Enterobase were additionally downloaded (*n* = 10), and those with known isolation years and country/continents of isolation (*n* = 6; three isolates from South Africa, two from Zambia, and one from the United Kingdom) were used in subsequent steps. All 716 assembled *S. Enteritidis* genomes underwent *in silico* plasmid replicon and AMR determinant detection using ABRicate v. 1.0.1 and the PlasmidFinder and NCBI AMR databases, respectively, as described above (see section “*In silico* AMR Determinant, Plasmid Replicon, and Virulence Factor Detection” above).





using ST11 isolates within a major clade in the African *S. Enteritidis* phylogeny (referred to here as “African *S. Enteritidis* ST11 Major Clade I,”  $n = 517$ ; **Supplementary Figure 5**). RhierBAPS v. 1.1.3 (Cheng et al., 2013; Tonkin-Hill et al., 2018) and R v. 4.0.0 were additionally used to identify clusters within the resulting core SNP alignment, using three clustering levels.

## Construction of Time-Scaled *S. Typhimurium* Phylogenies

To compare the 24 *S. Typhimurium* isolates sequenced in this study to publicly available *S. Typhimurium* genomes with known geographic origins, all genomes meeting each of the following conditions were downloaded via Enterobase (accessed 27 December 2020,  $n = 319$ ; **Supplementary Table 6**): (i) genomes were assigned to either ST19 (the ST to which 23 of the 24 *S. Typhimurium* isolates sequenced in this study were assigned/approximately assigned) or ST34 (the ST of the remaining isolate, which differs from ST19 by

a single allele, *dnaN*) using the Achtman seven-gene MLST scheme for *Salmonella*; (ii) genomes had an exact year of isolation reported in Enterobase’s “Collection Year” field; (iii) genomes could be assigned to a known country of isolation within the African continent, per Enterobase’s “Country” and “Continent” fields, respectively (Alikhan et al., 2018; Zhou et al., 2020). All 343 assembled *S. Typhimurium* genomes underwent *in silico* plasmid replicon and AMR determinant detection using ABRicate v. 1.0.1 and the PlasmidFinder and NCBI AMR databases, respectively, as described above (see section “*In silico* AMR Determinant, Plasmid Replicon, and Virulence Factor Detection” above).

Parsnp and HarvestTools v. 1.2 (Treangen et al., 2014) were used to identify core SNPs among all 343 *S. Typhimurium* genomes (319 publicly available genomes, plus the 24 sequenced here), using the closed chromosome of ST19 *S. Typhimurium* str. LT2 (NCBI Nucleotide Accession NC\_003197.2) as a reference and Parsnp’s implementation of PhiPack to remove

recombination (Brueen et al., 2006). Clusters were identified within the resulting core SNP alignment using RhierBAPs v. 1.1.3 (Cheng et al., 2013; Tonkin-Hill et al., 2018), R v. 4.0.0, and three clustering levels. IQ-TREE v. 1.5.4 (Nguyen et al., 2015) was used to construct a ML phylogeny using (i) the resulting core SNPs as input; (ii) an ascertainment bias correction (to account for the use of solely variant sites), corresponding to constant sites estimated using the GC content of the reference chromosome (-fconst 1160904,1268422,1268221,1159903); (iii) the optimal nucleotide substitution model selected using ModelFinder (Kalyaanamoorthy et al., 2017), based on its corresponding BIC value (i.e., the TVM + I model); (iv) 1,000 replicates of the UltraFast bootstrap approximation (Minh et al., 2013).

The resulting ML phylogeny was rooted and time-scaled using LSD2 v. 1.4.2.2 (To et al., 2015) and the following parameters: (i) tip dates corresponding to the year of isolation associated with each genome; (ii) a fixed substitution rate of  $2.79 \times 10^{-7}$  substitutions/site/year (i.e., the substitution rate estimated in a previous study of *S. Typhimurium* phage type DT104) (Leekitcharoenphon et al., 2016); (iii) constrained mode (-c), with the root estimated using constraints on all branches (-r as); (iv) variances calculated using input branch lengths (-v 1); (v) 1,000 samples for calculating confidence intervals for estimated dates (-f 1000); (vi) a sequence length of 4,857,450 (i.e., the length of the reference chromosome; -s 4857450). The resulting phylogeny was annotated using the bactaxR package in R (Figure 7). All aforementioned *S. Typhimurium* SNP calling and phylogeny construction steps were repeated to construct an additional time-scaled ML phylogeny, using the 24 isolates sequenced here and 87 human- and bovine-associated *S. Typhimurium* isolates from a previous study of the serotype in New York State in the United States (Carroll et al., 2020a) ( $n = 111$ ; Supplementary Figure 6).

## RESULTS

### Four Serotypes Are Represented Among the Animal-Associated South African *Salmonella enterica* Strains Sequenced Here

A total of 63 *Salmonella enterica* strains were isolated from animals and animal products in South Africa and underwent WGS (Supplementary Table 1). All isolates underwent *in silico* serotyping using both (i) SISTR (using its core-genome MLST [cgMLST] approach) and (ii) SeqSero2 (Supplementary Table 1); serotypes assigned using both methods were identical for all isolates (63 of 63 isolates, 100%; Supplementary Table 1). Furthermore, genomes of all isolates sequenced here clustered among publicly available *Salmonella* genomes assigned to their respective serotypes (Worley et al., 2018), with no observed polyphyly within serotypes among isolates sequenced here (Figure 1; note that all *S. Hadar* genomes sequenced here clustered

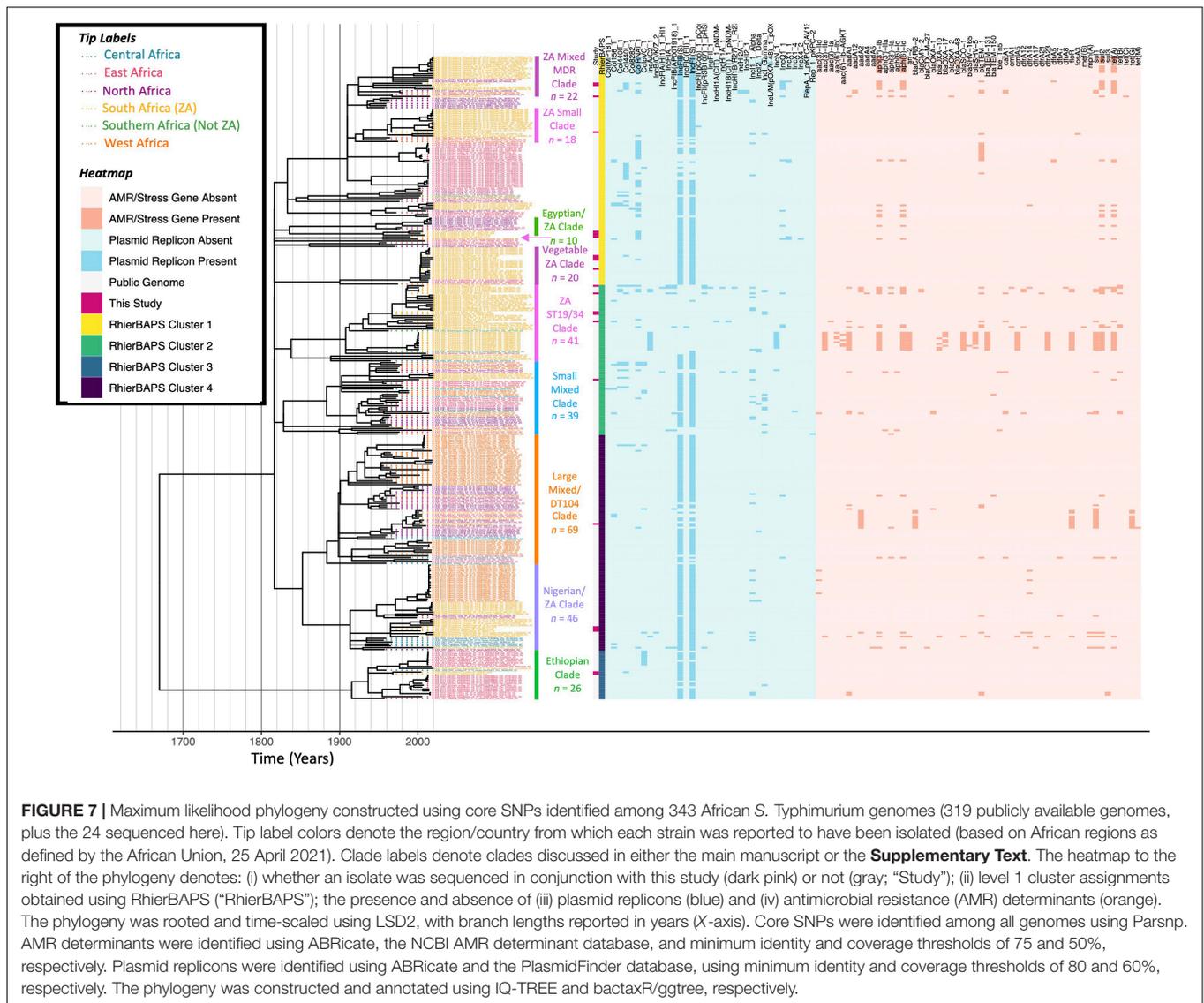
among a genome previously serotyped as *S. Istanbul*, which was serotyped as *S. Hadar in silico* using both SISTR and SeqSero2).

Four serotypes were represented among isolates sequenced in this study: *S. Dublin*, *S. Hadar*, *S. Enteritidis*, and *S. Typhimurium*, assigned to 18, 8, 13, and 24 isolates, respectively (Figure 1 and Supplementary Table 1). Strains were isolated from bovine sources (from feces, meat, or organs;  $n = 25$ ), poultry (from feces, meat, or organs;  $n = 22$ ), swine (from feces, meat, or organs;  $n = 6$ ), unspecified animal/animal product sources ( $n = 3$ ), fish (from food products;  $n = 2$ ), avian sources (feces from each of an ostrich and a pigeon;  $n = 2$ ), a rhinoceros ( $n = 1$ ), ovine sources (from feces;  $n = 1$ ), and from a cat (from feces;  $n = 1$ ; Supplementary Table 1). Strains were isolated from one of six provinces in South Africa: Gauteng ( $n = 27$ ), Western Cape ( $n = 7$ ), KwaZulu-Natal ( $n = 2$ ), Eastern Cape ( $n = 2$ ), North-West ( $n = 1$ ), Mpumalanga ( $n = 1$ ), and Free State ( $n = 1$ ); the provinces from which an additional 22 strains were isolated were unknown (Supplementary Table 1).

### Antimicrobial Resistance in *Salmonella enterica* Isolated From Animals and Animal Products in South Africa Is Acquired Sporadically

The 63 *Salmonella* genomes sequenced here underwent *in silico* AMR/stress response determinant, plasmid replicon, and virulence factor detection (Figure 2 and Supplementary Figures 1, 2). In total, 59 different AMR/stress response determinants were detected among the 63 isolates, with 18 unique AMR/stress response determinant presence/absence profiles observed (based on AMR/stress response determinants detected using AMRFinderPlus; Figure 2; Feldgarden et al., 2019). The number of different AMR/stress response determinants detected per genome ranged from 5 to 24; nearly two-thirds of all genomes sequenced in the current study (40 of 63, 63.5%) harbored six AMR/stress response determinants (the median per genome) or less (Figure 2). Six “core” AMR/stress response determinants (*asr*, *golS*, *golT*, *mdsA*, *mdsB*, *sinH*) were observed in over 90% of the isolates sequenced here (59 of 63 isolates; 93.7%), four of which were detected in all 63 isolates (*asr*, *golS*, *golT*, *sinH*; Figure 2). The remaining 53 AMR/stress response determinants were detected in less than 20% of the genomes sequenced here; 46 of these (46 of 59 total unique AMR determinants, 78.0%) were present only sporadically and were detected in two or fewer genomes (Figure 2).

In total, 17 different plasmid replicons were identified among all 63 genomes, representing 22 unique plasmid replicon presence/absence profiles (detected using ABRicate, the PlasmidFinder database, and minimum nucleotide identity and coverage thresholds of 80 and 60%, respectively; Figure 2) (Carattoli et al., 2014). Genomes harbored one to seven different plasmid replicons, with a median of two per genome (Figure 2). Two plasmid replicons, IncFIB(S) and IncFII(S), were detected in over half of all genomes sequenced here



(detected in 32 and 49 of 63 genomes, 50.8% and 77.8%, respectively; **Figure 2**). Over half of all plasmid replicons (10 of 17 unique plasmid replicons; 58.8%) were detected in two or fewer genomes (**Figure 2**).

Additionally, a total of 181 different virulence factors were identified among the 63 genomes, with 24 unique virulence factor presence/absence profiles represented (detected using ABRicate, the Virulence Factor Database [VFDB], and minimum nucleotide identity and coverage thresholds of 70 and 50%, respectively; **Figure 2** and **Supplementary Table 2**). Genomes harbored 146–171 different virulence factors, with a median of 165 (**Figure 2** and **Supplementary Table 2**). Over 75% of all unique virulence factors detected among the isolates sequenced in this study were present in all genomes (137 of 181 unique virulence factors, 75.7%; **Supplementary Table 2**). Only 13 virulence factors were detected in fewer than half of the genomes sequenced here (**Figure 2**).

## A Largely Antimicrobial-Susceptible *S. Dublin* ST10 Lineage Circulating in South Africa Encompasses Isolates From Livestock, Food, and Human Sources

A ML phylogeny constructed using the 18 South African ST10 *S. Dublin* isolates sequenced here, plus 2,784 publicly available ST10 *S. Dublin* genomes with known isolation sources and geographic origins, partitioned the vast majority of genomes (2,738 of 2,802 genomes, 97.7%) into two major *S. Dublin* ST10 clades (**Figure 3** and **Supplementary Figure 3**), which is consistent with previous observations (Fenske et al., 2019). Referred to hereafter as “*S. Dublin* Major Clade I” and “*S. Dublin* Major Clade II,” the two major clades encompassed 1,787 and 951 genomes, respectively (**Figure 3** and **Supplementary Figure 3**). While both major clades encompassed strains isolated from Asia, Europe, North America, and South America, the vast majority of North American ST10 *S. Dublin* belonged to Major Clade

I (1,641 of 1,656 *S. Dublin* ST10 strains from North America, 99.1%; **Figure 3** and **Supplementary Figure 3**). Members of Major Clade I shared a most recent common ancestor (MRCA) dated to circa 1959 (95% confidence interval [CI] of [1452.88, 1959.00]; **Figure 3** and **Supplementary Figure 3**). Notably, multi-drug resistant (MDR) *S. Dublin*, which often possess IncA/C2 plasmids and acquired AMR determinants that confer resistance to aminoglycosides, beta-lactams, phenicol, sulfonamides, and tetracyclines (Fenske et al., 2019), were almost exclusively confined to a large, primarily North American subclade within *S. Dublin* Major Clade I (referred to hereafter as the “*S. Dublin* Large Subclade”; **Figure 3** and **Supplementary Figure 3**). Conversely, members of *S. Dublin* Major Clade II shared a MRCA dated to circa 1945 (95% CI [1274.31, 1985.00]), primarily contained European isolates (893 of 951 Major Clade II genomes, 93.9%), and largely did not possess any acquired AMR determinants (**Figure 3** and **Supplementary Figure 3**).

All 18 South African *S. Dublin* isolates sequenced in this study belonged to *S. Dublin* Major Clade I (**Figure 3** and **Supplementary Figure 3**); however, 17 of the 18 isolates clustered together within a small subclade of Major Clade I (referred to hereafter as the “*S. Dublin* Small Subclade”; **Figure 4**), while the remaining isolate clustered among isolates in the *S. Dublin* Large Subclade (**Supplementary Figure 4**). Notably, within the *S. Dublin* Small Subclade, the 17 animal- and animal product-associated South African isolates sequenced here clustered among all seven publicly available *S. Dublin* ST10 genomes from South Africa, all of which were reported to have been isolated from human sources (**Figure 4**). This well-supported South African-specific *S. Dublin* lineage (referred to hereafter as the “South African *S. Dublin* Clade”), which contained animal-, animal product-, and human-associated strains isolated over a span of 60 years (i.e., 1960–2020) from the Gauteng, Eastern Cape, Western Cape, and North-West provinces, was predicted to share a common ancestor dated to circa 1960 (95% CI [1496.07, 1960.00], 98% UltraFast Bootstrap Support; **Figure 4**). Members of this South African lineage, like the *S. Dublin* Small Subclade more broadly, were largely pan-susceptible, with AMR determinants detected only sporadically; a single strain, isolated in 2007 from poultry meat in the Western Cape province (FOO\_2007\_SouthAfrica\_WesternCape\_AF\_0930SE-S25), possessed streptomycin resistance gene *aadA1* and sulfonamide resistance gene *sulI* (**Figure 4**). Taken together, these results indicate that a largely AMR-susceptible South African-specific *S. Dublin* ST10 lineage has been circulating among animals, foods, and humans in the country for decades.

Only one South African *S. Dublin* ST10 genome was not a member of the South African *S. Dublin* Clade within the *S. Dublin* Small Subclade (**Supplementary Figure 4**). This strain (i.e., FOO\_2016\_SouthAfrica\_EasternCape\_AF\_7509SE-S59), which was isolated in 2016 from poultry meat in South Africa’s Eastern Cape province, clustered among North American isolates in the *S. Dublin* Large Subclade (**Supplementary Figure 4**). This strain most closely resembled a bovine-associated strain from California isolated in 2004, and the two shared a common ancestor circa 2004 (95% CI [1973.4, 2004.00]; **Supplementary Figure 4**). Notably, despite clustering

among MDR *S. Dublin* ST10 strains from North America, neither of these strains harbored any acquired AMR genes, nor did they harbor the IncA/C2 plasmid characteristic of MDR *S. Dublin* ST10 from the United States (**Supplementary Figure 4**). These results indicate that a separate *S. Dublin* ST10 lineage may have only recently been introduced into South Africa from North America, a hypothesis that was further supported by subsequent investigation into the origin of the isolate: the poultry meat from which strain FOO\_2016\_SouthAfrica\_EasternCape\_AF\_7509SE-S59 was isolated had been imported from North America and sold in a supermarket in South Africa’s Eastern Cape province.

All 18 *S. Dublin* ST10 isolates sequenced in this study, as well as all seven publicly available South African *S. Dublin* ST10 genomes, were members of *S. Dublin* Major Clade I (**Figure 3** and **Supplementary Figure 3**). These 25 South African genomes, 24 of which formed a well-supported subclade within Major Clade I, were the only African genomes detected in *S. Dublin* Major Clade I (**Figure 3** and **Supplementary Figures 3, 4**). *S. Dublin* Major Clade II did not contain any African genomes (**Figure 3**). However, 18 genomes from the African continent were among the few ST10 genomes (i.e., 64 of 2,802 *S. Dublin* ST10 genomes, 2.3%) that fell outside of the two major *S. Dublin* ST10 clades (**Figure 3**). These genomes were reported to have been derived from strains isolated from animals, food, and humans in Ethiopia, Gambia, Nigeria, and Benin, and none harbored any acquired AMR genes (**Figure 3**); interestingly, they clustered among human-associated genomes from Asia (i.e., Taiwan), Europe (i.e., France and the United Kingdom), and North America (i.e., Canada and the United States), forming a 52-genome, well-supported clade with a common ancestor dated to circa 1957 (95% CI [1142.96, 2003.00], 100% UltraFast Bootstrap Support; **Figure 3**).

## South Africa Harbors Multiple *S. Hadar* ST33 Lineages With Streptomycin and Tetracycline Resistance-Confering Genes

A ML phylogeny was constructed using the eight South African ST33 *S. Hadar* isolates sequenced here, plus 1,561 publicly available ST33 *S. Hadar* genomes with known isolation sources and geographic origins (**Figure 5**). Notably, the majority of *S. Hadar* ST33 genomes harbored AMR genes *aph(3'')-Ib* and *aph(6)-Id* ( $n = 1,314$  and  $1,347$  of  $1,569$  *S. Hadar* ST33 genomes, 83.7 and 85.9%, respectively; **Figure 5**). Also known as *strA* and *strB*, respectively, *aph(3'')-Ib* and *aph(6)-Id* confer resistance to streptomycin. The majority of *S. Hadar* ST33 genomes additionally harbored *tet(A)*, which confers resistance to tetracycline ( $n = 1,320$  of  $1,569$  *S. Hadar* ST33 genomes, 84.1%; **Figure 5**). All eight *S. Hadar* ST33 strains sequenced in this study, which were derived from strains isolated between 1962 and 2017, were among the strains that harbored all of streptomycin resistance-confering *aph(3'')-Ib* and *aph(6)-Id* and tetracycline resistance-confering *tet(A)* (**Figure 5**).

Seven of eight *S. Hadar* ST33 genomes sequenced in this study clustered at or near the tree root, which was dated to circa 1962

(95% CI [1571.93, 1962.00]). These seven South African strains, which had been isolated between 1962 and 2017 from bovine sources (feces and meat), poultry (meat), a rhinoceros, and an unspecified animal/animal product, were most closely related to a publicly available genome of a *S. Hadar* ST33 strain isolated in 2018 from chicken in South Africa (Figure 5).

The remaining isolate sequenced in this study (i.e., BOV\_1990\_XX\_ARCZA\_NEW19-S113) was relatively distantly related to the other South African ST33 isolates sequenced here (Figure 5). Isolated from bovine feces in 1990, this strain was most closely related to a *S. Hadar* ST33 strain isolated from the spleen of a dog (*Canis lupus familiaris*) in the United States in 1988; however, these strains were relatively distant, sharing a common ancestor that existed circa 1982 (95% CI [1705.27, 1985.00]; Figure 5). While it is unclear exactly when this particular lineage was introduced into South Africa, these results indicate that multiple *S. Hadar* ST33 lineages have circulated in livestock populations in the country.

### One Largely Antimicrobial-Susceptible Clade and One Largely Multidrug-Resistant Clade Are Represented Among *S. Enteritidis* ST11 From Africa

A ML phylogeny was constructed using the 13 South African ST11 ( $n = 12$ ) and ST366 ( $n = 1$ ) *S. Enteritidis* isolates sequenced here, plus (i) 697 publicly available ST11 *S. Enteritidis* genomes of strains isolated from the African continent and (ii) all publicly available ST366 *S. Enteritidis* genomes ( $n = 10$ ; Figure 6 and Supplementary Figure 5).

Notably, one strain sequenced here (i.e., POL\_2002\_XX\_NEW34-S128), isolated from poultry meat in 2002, was assigned to ST366. Currently, there are only 15 ST366 genomes that are publicly available for download, six of which have a known collection year and isolation source and meet the quality standards used in this study (via Enterobase, accessed 18 February 2021). This can be contrasted with ST11, of which there are 50,755 publicly available genomes (via Enterobase, accessed 18 February 2021). The ST366 strain sequenced here was a member of a well-supported clade (100% UltraFast Bootstrap support), which contained eight additional publicly available genomes that shared a common ancestor dated to circa 1885 (95% CI [809.09, 2002.00]; Figure 6 and Supplementary Figure 5). In addition to the poultry-associated ST366 strain sequenced here, this clade contained all six publicly available ST366 genomes, which were all isolated from human sources in South Africa ( $n = 3$ ), Zambia ( $n = 2$ ), and the United Kingdom ( $n = 1$ ); additionally, this clade contained two ST11 genomes of strains isolated from humans in Malawi (Figure 6 and Supplementary Figure 5). None of the genomes in this clade harbored any known AMR determinants (Figure 6 and Supplementary Figure 5). Interestingly, the ST366 isolate from the United Kingdom is the only publicly available ST366 strain from outside of Africa (via Enterobase, accessed 18 February 2021), indicating that this particular ST may have a geographic association.

The remaining 12 *S. Enteritidis* strains sequenced in this study were assigned to ST11 and were confined to a large, well-supported (100% UltraFast Bootstrap support) 517-isolate clade (referred to hereafter as “African *S. Enteritidis* ST11 Major Clade I”), which shared a common ancestor dated to circa 1551–1801 (depending on the tree root/isolate set used in Figure 6 and Supplementary Figure 5, respectively; 95% CI [–639.57, 1955.0] and [738.65, 1955.00], respectively). Notably, isolates within African *S. Enteritidis* ST11 Major Clade I were largely pan-susceptible and acquired AMR determinants only sporadically; among the 12 ST11 isolates sequenced here, only three possessed AMR genes (Figure 6 and Supplementary Figure 5).

Overall, we found that the South African ST11 genomes sequenced in this study belonged to a largely antimicrobial-susceptible lineage, which showcased AMR only sporadically. This can be contrasted with a second major clade comprising 181 *S. Enteritidis* genomes (i.e., African *S. Enteritidis* ST11 Major Clade II; Figure 6); the majority of isolates in this clade were predicted to be MDR, as they possessed AMR genes conferring resistance to beta-lactams (*bla<sub>TEM-1</sub>*), streptomycin [*aph(3'')-Ib*, *aph(6)-Id*], sulfonamides (*sul1*, *sul2*), chloramphenicol (*catA2*), trimethoprim (*dhfrA7*), and tetracycline [*tet(A)*; Figure 6]. Unlike African *S. Enteritidis* ST11 Major Clade I, which encompassed 340 South African isolates, no Major Clade II isolates were found in South Africa (Figure 6); rather, African *S. Enteritidis* ST11 Major Clade II primarily included isolates from the Democratic Republic of the Congo (DRC;  $n = 72$ ) and Malawi ( $n = 55$ ), as well as from Senegal and Mali ( $n = 12$  each), Nigeria ( $n = 10$ ), Kenya ( $n = 7$ ), Burkina Faso ( $n = 5$ ), Rwanda and Guinea ( $n = 2$ ), the Central African Republic (CAR), Congo, Ivory Coast, and Madagascar ( $n = 1$  each; Figure 6).

### South Africa Harbors Numerous Antimicrobial Susceptible and Multidrug-Resistant *S. Typhimurium* Lineages

A ML phylogeny was constructed using the 24 South African ST19 ( $n = 23$ ) and ST34 ( $n = 1$ ) *S. Typhimurium* isolates sequenced here, plus publicly available *S. Typhimurium* genomes of strains isolated from the African continent assigned to (i) ST19 ( $n = 315$ ) and (ii) ST34 ( $n = 4$ ; Figure 7 and Supplementary Text). The 24 *S. Typhimurium* strains sequenced in this study were distributed across the African *S. Typhimurium* ST19/ST34 phylogeny, representing a diverse range of lineages, and eight (33.3%) possessed one or more AMR genes (Figure 7 and Supplementary Text). Notably, some African *S. Typhimurium* ST19/ST34 lineages were distributed across the African continent, while others were strongly associated with a particular region/country (Figure 7 and Supplementary Text). When compared to genomes from a previous study of *S. Typhimurium* from New York State that have been shown to be representative of the human- and bovine-associated *S. Typhimurium* population in the United States as a whole (Carroll et al., 2020a), only five of 24 *S. Typhimurium* strains sequenced here (20.8%) shared a common ancestor with one or more New York State strains after

1900 (**Supplementary Figure 6** and **Supplementary Text**). This indicates that many of the strains sequenced here are not closely related to *S. Typhimurium* lineages circulating among cattle and humans in the United States. Below, we discuss some of these major African *S. Typhimurium* ST19/ST34 clades in detail (see the **Supplementary Text** for discussions of additional lineages).

### A *S. Typhimurium* DT104-Like Clade Emerged in Africa in the Twentieth Century as Antimicrobial-Susceptible and Later Acquired Multidrug-Resistance

A *S. Typhimurium* strain sequenced here (PIG\_2002\_FS\_040ST-S45), isolated in 2002 from swine meat in South Africa's Free State province, clustered within a 69-isolate clade, which shared a common ancestor dated circa 1884 (95% CI [1153.59, 1956.00]); denoted as the "Large Mixed/DT104 Clade" in **Figure 7**). This large clade contained a mixture of human-, animal-, environmental, and food-associated isolates from Senegal ( $n = 16$ ), Gambia ( $n = 13$ ), Tunisia ( $n = 10$ ), Benin ( $n = 7$ ), Ethiopia ( $n = 6$ ), Morocco ( $n = 4$ ), DRC and South Africa ( $n = 3$  each), Madagascar ( $n = 2$ ), Algeria, Cameroon, Egypt, Kenya, and Tanzania ( $n = 1$  each). The strain isolated in this study possessed five AMR genes: streptomycin resistance-conferring *aadA2*, beta lactamase *bla<sub>CARB-2</sub>*, chloramphenicol resistance-conferring *floR*, sulfonamide resistance-conferring *sul1*, and tetracycline resistance-conferring *tet(G)* (**Figure 7**). Notably, when compared to genomes from a previous study of *S. Typhimurium* in the United States (New York State) (Carroll et al., 2020a), this isolate clustered among DT104 strains isolated from dairy cattle and humans, sharing a common ancestor dated circa 1975 (95% CI [1467.42, 1999.00]); **Supplementary Figure 6**).

Within the Large Mixed/DT104 Clade in the African *S. Typhimurium* phylogeny (**Figure 7**), the DT104-like strain sequenced here (PIG\_2002\_FS\_040ST-S45) was part of a 14-isolate subclade, which shared a common ancestor dated circa 1939 (95% CI [1623.18, 1960.00]). Notably, the four most distant members of this subclade, corresponding to strains isolated (i) in 1960 from a dog in Algeria, (ii) in 1970 and (iii) 1975 from unknown sources in Morocco, and (iv) in 1967 from a human in Morocco, were the only strains within this subclade that did not possess any AMR genes (**Figure 7**). The Moroccan strain isolated in 1975 was additionally reported to have itself been phage typed as DT104. The remaining ten genomes, which included the DT104-like strain sequenced here, clustered together, sharing a common ancestor dated circa 1980 (95% CI [1932.92, 2001.00]); **Figure 7**). Nine of these ten genomes possessed the five AMR genes listed above, which confer resistance to ampicillin, chloramphenicol, streptomycin, sulfonamides, and tetracycline [ACSSuT; one 2005 isolate from a camel in Ethiopia clustered among these genomes, but possessed only *aadA2* and *sul1*, while another strain, isolated in 2005 from poultry in Ethiopia, possessed all five AMR genes, as well as kanamycin resistance gene *aph(3')-Ia* and tetracycline resistance gene *tet(M)*; **Figure 7**). This is noteworthy, as the ACSSuT AMR profile is often seen as characteristic of MDR D104 (Leekitcharoenphon et al., 2016). Using the most parsimonious explanation for the acquisition of its MDR phenotype, the

clade of African DT104-like isolates identified here emerged as antimicrobial-susceptible circa 1939 (95% CI [1623.18, 1960.00]) and acquired the MDR phenotype between  $\approx$ 1966 and  $\approx$ 1980 (95% CI [1820.78, 2001.00]); **Figure 7**).

### An Antimicrobial-Susceptible *S. Typhimurium* Clade, Which Emerged in South Africa After 2000, Encompasses Isolates From Produce, Fish, Poultry, and Avian Sources

Four additional isolates sequenced in this study were contained within a 20-isolate clade, which shared a common ancestor dated to circa 1900 (95% CI [1220.88, 1974.00]); denoted in **Figure 7** as the "Vegetable ZA Clade"). No AMR genes were detected in any genomes within this clade, including the four strains sequenced here, which were all isolated from South Africa's Western Cape province (two strains isolated in 2019 from fish food products, one in 2004 from ostrich feces, and one in 2003 from poultry meat; **Figure 7**). Interestingly, these isolates clustered among (i) 13 publicly available genomes, all derived from food-associated strains isolated in 2015 in South Africa (i.e., 4 from cabbage, 3 from carrots, 2 from lettuce, 2 from plant salad, and one from each of spinach and red onion) and (ii) one strain isolated from human feces in Addis Ababa, Ethiopia, sharing a common ancestor dated circa 2003 (95% CI [1746.59, 2003.00]); **Figure 7**). Members of this clade additionally shared a common ancestor with a strain isolated in 2013 from swine feces in Addis Ababa, Ethiopia, which was predicted to have existed circa 1990 (95% CI [1620.12, 2003.00]); **Figure 7**). The most distant member within the clade was a strain isolated in 1974 from Burkina Faso (**Figure 7**).

## DISCUSSION

### Endemic *Salmonella enterica* Lineages Are Circulating Among Animals and Animal Products in South Africa and May Infect Humans

Geography plays an important role in shaping bacterial pathogen population structure, including that of *Salmonella enterica*, and different geographic regions may harbor their own endemic lineages (Achtman, 2008; Strachan et al., 2015; Carroll et al., 2017; Palma et al., 2018; Fenske et al., 2019; Liao et al., 2020). Here, we observed numerous endemic *Salmonella enterica* lineages circulating among animals and animal products in South Africa, some of which encompassed human clinical isolates. Within the global *S. Dublin* ST10 phylogeny, for example, we identified a largely AMR-susceptible South Africa-specific clade, which has been circulating among animals, foods, and humans in the country for decades. While *S. Dublin* is largely considered to be a bovine-adapted serotype, human infections caused by *S. Dublin* are frequently invasive and may result in severe illness and/or death (Taylor et al., 1982; Uzzau et al., 2000; Rabsch et al., 2002; Cummings et al., 2012; Rodriguez-Rivera et al., 2014; Ao et al., 2015; Harvey et al., 2017; Mohammed et al., 2017;

Mughini-Gras et al., 2020). In South Africa, invasive non-typhoidal salmonellosis is a serious public health concern: in 2019, over 25% of all non-typhoidal salmonellosis cases reported to the Group for Enteric, Respiratory and Meningeal disease Surveillance in South Africa (GERMS-SA) were invasive (825 and 2,437 reported invasive and non-invasive non-typhoidal salmonellosis cases, respectively; 25.3%) (GERMS-SA, 2019). While *S. Dublin* is not among the most common serotypes isolated from human clinical cases in South Africa (GERMS-SA, 2019), routine veterinary surveillance has revealed that *S. Dublin* is frequently isolated from animal sources in the country, particularly cattle (Gelaw et al., 2018). Further WGS efforts are needed to provide insight into the evolution and between-host transmission dynamics of the endemic South African *S. Dublin* ST10 lineage identified in this study.

In addition to the endemic South African *S. Dublin* ST10 lineage, we identified a clade of South African animal- and animal product-associated *S. Hadar* ST33 strains, which clustered near the root of the global *S. Hadar* ST33 phylogeny. First described as a novel *Salmonella* serotype in 1954 (Hirsch et al., 1954), *S. Hadar* was reported to have been responsible for several cases of diarrheal illness in Israel (Hirsch et al., 1954). Reportedly, serotype *S. Hadar* was rarely isolated prior to 1971; however, in the mid-1970s, *S. Hadar* quickly became the second-most common cause of human non-typhoidal salmonellosis in the United Kingdom (Rowe et al., 1980; Barrow et al., 2012; O'Brien, 2013). Consistent with these observations, the South African *S. Hadar* clade identified here shared a common ancestor dated circa 1962 and contained strains isolated from the 1960s through 2018. In South Africa specifically, *S. Hadar* is not among the top serotypes associated with human clinical cases (GERMS-SA, 2019); however, *S. Hadar* has been commonly isolated from animals and animal associated-environments in the country for nearly two decades (Kidānemariam et al., 2010; Magwedere et al., 2015; Gelaw et al., 2018). The results presented here indicate that an endemic *S. Hadar* ST33 lineage has been circulating among animals in South Africa for over fifty years; however, future WGS efforts querying *S. Hadar* ST33 strains from around the world—historical strains isolated prior to the 1970s, in particular—are needed to refine estimates as to when this particular lineage emerged.

In addition to the *S. Dublin* ST10 and *S. Hadar* ST33 endemic South African lineages, we observed that African *S. Enteritidis* ST11 could largely be partitioned into one largely antimicrobial-susceptible and one largely MDR clade. South African *S. Enteritidis* ST11, including those sequenced here, were confined to the largely antimicrobial-susceptible clade. These results are consistent with those observed in a previous study of *S. Enteritidis* in Africa (Feasey et al., 2016), in which a geographically distinct MDR *S. Enteritidis* lineage was identified in Africa's Central/East regions and rarely detected in South Africa. Since 2012, *S. Enteritidis* has been the serotype most commonly isolated from human clinical cases in South Africa (GERMS-SA, 2019). Among animals, *S. Enteritidis* has been one of the most frequently isolated serotypes in South Africa for decades, particularly from poultry-associated sources (Kidānemariam et al., 2010; Magwedere et al., 2015;

Gelaw et al., 2018). Our results further support that South African *S. Enteritidis*, which is one of the most common *Salmonella enterica* serotypes circulating among animals and humans in the country, acquires AMR only sporadically and is, on a genomic scale, distinct from MDR *S. Enteritidis* ST11 lineages circulating in other regions of Africa. Collectively, our study reveals that endemic lineages of several non-typhoidal *Salmonella enterica* serotypes are circulating among animals and animal products in South Africa, some of which may occasionally infect humans.

## Whole-Genome Sequencing Can Differentiate Endemic and Ecdemic *Salmonella enterica* Lineages

Pathogenic bacteria not previously endemic to a given geographic region can be introduced into that region through the movement of humans, food, and/or animals (Wong et al., 2015; Llerena et al., 2016; The et al., 2016; Palma et al., 2018). In addition to observing *Salmonella enterica* lineages that were likely endemic to South Africa, our study identified numerous lineages that were likely to have been introduced into the country only recently. One *S. Dublin* ST10 isolate sequenced in this study, for example, clustered among ST10 isolates from the United States, indicating that this strain had been introduced into South Africa only recently. *S. Dublin* from the United States has previously been shown to be distinct from *S. Dublin* strains isolated in other world regions on a genomic scale (Fenske et al., 2019), and the United States was one of the leading poultry exporters to South Africa in 2016 (i.e., the year the ecdemic *S. Dublin* ST10 strain sequenced here was isolated) (USDA Foreign Agricultural Service, 2020). Our recent introduction hypothesis was further supported by metadata indicating that this strain had been isolated from poultry meat imported from North America and sold in a supermarket in South Africa's Eastern Cape province.

We observed similar results for *S. Hadar*: one *S. Hadar* ST33 strain sequenced in this study was more closely related to *S. Hadar* ST33 from the United States than to its South African counterparts, which all formed a clade near the global *S. Hadar* ST33 phylogeny root (Rowe et al., 1980; Barrow et al., 2012; O'Brien, 2013). Unlike the *S. Dublin* strain sequenced here, which was likely introduced into South Africa from imported poultry meat, it is unclear exactly how the unique *S. Hadar* ST33 lineage sequenced here was introduced into the country, as its representative strain was isolated from bovine feces in 1990 and shared a common ancestor circa 1982 with a canine-associated *S. Hadar* ST33 strain isolated in 1988 in the United States. Future WGS efforts querying *S. Hadar* ST33 may provide insight into this lineage and its emergence in South Africa.

The *S. Typhimurium* ST19/ST34 isolates sequenced here were distributed across the African *S. Typhimurium* ST19/ST34 phylogeny, indicating that South Africa harbors numerous *S. Typhimurium* ST19/ST34 lineages. Since 2012, *S. Typhimurium* has been the second-most common non-typhoidal *Salmonella enterica* serotype isolated from human clinical cases in South Africa (after *S. Enteritidis*) and in 2019 was the most common serotype isolated from human clinical cases in the Eastern Cape province (GERMS-SA, 2019). *S. Typhimurium* has

additionally been one of the most frequently isolated serotypes from animals and wildlife in South Africa for decades, and it is frequently isolated from a broad range of hosts (e.g., cattle, poultry, equine, sheep/goats, feline, rhinoceros) (Kidanemariam et al., 2010; Magwedere et al., 2015; Gelaw et al., 2018). Interestingly, we identified a largely AMR-susceptible, primarily South African *S. Typhimurium* clade, which contained isolates from produce, fish, poultry, and avian sources, and one human clinical isolate from Ethiopia (referred to above as the “Vegetable ZA Clade”), which was predicted to have been introduced into the country recently (i.e., after the year 2000). It is unclear exactly where this lineage originated and how it was introduced into South Africa, but future WGS efforts may elucidate this.

We additionally identified a *S. Typhimurium* clade, which contained the genomes of strains assigned to phage type DT104. MDR DT104 was responsible for a global epidemic in the 1990s, during which it was increasingly isolated from a broad range of animals (e.g., cattle, poultry, pigs, sheep), as well as human clinical cases (Threlfall, 2000). Notably, DT104 was predicted to have emerged as antimicrobial-susceptible circa 1948, later acquiring its MDR phenotype circa 1972 (Leekitcharoenphon et al., 2016). The results observed here are consistent with these findings, as the DT104 clade identified here emerged as antimicrobial-susceptible circa 1939 and acquired the MDR phenotype between  $\approx$ 1966 and  $\approx$ 1980. The DT104 clade identified here spanned multiple African regions, and South African DT104-like genomes were distributed across the clade, indicating that South Africa may have been subjected to multiple DT104 introduction events and/or between-country transmission events; however, the lack of available DT104-like genomes from South Africa (i.e., one sequenced here and two publicly available genomes) and the African continent as a whole limits our ability to say this conclusively. Taken together, our results further highlight the strengths of WGS in *Salmonella* source tracking, both within and between countries and continents (Den Bakker et al., 2014; Inns et al., 2017; Brown et al., 2019), and showcase the ability of WGS-based approaches to differentiate endemic and ecemic lineages.

## Whole-Genome Sequencing of Historical Isolates From Under-Sequenced Geographic Regions Can Provide Novel Insights Into Pathogen Evolution and Diversity

Worldwide, *Salmonella enterica* has been estimated to be responsible for more than 93 million illnesses and more than 150,000 deaths annually (Majowicz et al., 2010). In Africa, the disease burden imposed by *Salmonella enterica* is particularly significant; mortality and disability adjusted life years (DALYs) due to diarrheal disease and invasive infections caused by non-typhoidal serotypes are consistently higher in Africa than in other world regions (Kirk et al., 2015). However, despite the disproportionately high incidence and burden of salmonellosis and other foodborne illnesses, the bulk of publicly available genomic data derived from *Salmonella enterica* has come from

regions with lower burdens (Kirk et al., 2015; Inzaule et al., 2021); for example, among all *Salmonella enterica* genomes in Enterobase (accessed 7 April 2021), over 80% were derived from strains reported to have been isolated in North America and Europe (128,517 and 104,910 genomes from North America and Europe, respectively; 233,427 of 291,362 total *Salmonella enterica* genomes).

Here, we used WGS to characterize 63 *Salmonella enterica* strains isolated from animals and animal products in South Africa over a 60-year time span. Importantly, numerous genomes sequenced here belonged to lineages that were phylogenetically distinct from those circulating in more heavily sequenced/sampled regions of the world, such as North America and Europe. For example, as observed here, some African *S. Dublin* ST10 isolates do not belong to the two major *S. Dublin* ST10 clades circulating primarily in North America and Europe, indicating that *S. Dublin* ST10 isolates representing clades outside of the two major North American- and European- associated clades are likely circulating in other countries around the world, including African countries outside of South Africa. Similarly, the few available *S. Enteritidis* ST366 genomes are derived from strains primarily isolated in Africa. Future WGS efforts in Africa will likely provide insight into the evolution and emergence of these lineages, as well as novel clades and those underrepresented in public databases. Overall, this study offers a glimpse into the genomics of non-typhoidal *Salmonella enterica* lineages circulating among livestock, domestic animals, wildlife, and animal products in South Africa. Future WGS-based studies querying greater numbers of isolates from animal, food, and environmental sources are needed to better understand the evolution, population structure, and AMR dynamics of this important pathogen.

## DATA AVAILABILITY STATEMENT

Illumina reads for genomes sequenced in this study are available in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under BioProject accession PRJNA727588. Metadata and genome quality statistics for the *Salmonella enterica* genomes sequenced in this study are available in **Supplementary Table 1**. Enterobase metadata for the publicly available genomes used in this study are available in **Supplementary Tables 3–6**.

## AUTHOR CONTRIBUTIONS

LC designed and carried out all computational analyses. IM conceptualized the study. RP and IM sourced the funding for the project. RP supervised the sequencing of the isolates, while MM performed all culturing work and DNA extractions. LC and IM co-wrote the manuscript, with input from all authors. All authors contributed to the article and approved the submitted version.

## FUNDING

Funding for this project was provided by the Gauteng Department of Agriculture and Rural Development (GDRAD).

## ACKNOWLEDGMENTS

The following organizations and individuals are acknowledged for their contributions: (i) the officials from the Bacteriology section of ARC: OVR (Rosina Maluleka, Palesa Nthaba, Mmatua Motau, and Lavhelesani Makhado) for the technical

support during data retrieval; (ii) the authors are grateful to the Agricultural Research Council: Onderstepoort Veterinary Research for providing all research facilities; (iii) we also acknowledge our collaborators, EMBL.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2021.748611/full#supplementary-material>

## REFERENCES

- Achtman, M. (2008). Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu. Rev. Microbiol.* 62, 53–70. doi: 10.1146/annurev.micro.62.081307.162832
- Alikhan, N. F., Zhou, Z., Sergeant, M. J., and Achtman, M. (2018). A genomic overview of the population structure of *Salmonella*. *PLoS Genet.* 14:e1007261. doi: 10.1371/journal.pgen.1007261
- Andrews, S. (2019). *FastQC: a quality control tool for high throughput sequence data*. 0.11.8 ed. Cambridge: The Babraham Institute.
- Ao, T. T., Feasey, N. A., Gordon, M. A., Keddy, K. H., Angulo, F. J., and Crump, J. A. (2015). Global burden of invasive nontyphoidal *Salmonella* disease, 2010. *Emerg. Infect. Dis.* 21, 941–949. doi: 10.3201/eid2106.140999
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Barrow, P. A., Jones, M. A., Smith, A. L., and Wigley, P. (2012). The long view: *Salmonella*—the last forty years. *Avian Pathol.* 41, 413–420. doi: 10.1080/03079457.2012.718071
- Brown, E., Dessai, U., McGarry, S., and Gerner-Smidt, P. (2019). Use of Whole-Genome Sequencing for Food Safety and Public Health in the United States. *Foodborne Pathog. Dis.* 16, 441–450. doi: 10.1089/fpd.2019.2662
- Bruen, T. C., Philippe, H., and Bryant, D. (2006). A simple and robust statistical test for detecting the presence of recombination. *Genetics* 172, 2665–2681. doi: 10.1534/genetics.105.048975
- Carattoli, A., Zankari, E., Garcia-Fernandez, A., Voldby Larsen, M., Lund, O., Villa, L., et al. (2014). *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* 58, 3895–3903. doi: 10.1128/AAC.02412-14
- Carroll, L. M. (2019). *High-Throughput Characterization of Foodborne Pathogens using Next-Generation Sequencing*. Ann Arbor, MI: ProQuest LLC.
- Carroll, L. M., Wiedmann, M., and Kovac, J. (2020b). Proposal of a Taxonomic Nomenclature for the *Bacillus cereus* Group Which Reconciles Genomic Definitions of Bacterial Species with Clinical and Industrial Phenotypes. *mBio* 11, e34–e20. doi: 10.1128/mBio.00034-20
- Carroll, L. M., Huisman, J. S., and Wiedmann, M. (2020a). Twentieth-century emergence of antimicrobial resistant human- and bovine-associated *Salmonella enterica* serotype Typhimurium lineages in New York State. *Sci. Rep.* 10:14428. doi: 10.1038/s41598-020-71344-9
- Carroll, L. M., Wiedmann, M., Den Bakker, H., Siler, J., Warchocki, S., Kent, D., et al. (2017). Whole-Genome Sequencing of Drug-Resistant *Salmonella enterica* Isolates from Dairy Cattle and Humans in New York and Washington States Reveals Source and Geographic Associations. *Appl. Environ. Microbiol.* 83, e140–e117. doi: 10.1128/AEM.00140-17
- Centers for Disease Control and Prevention (2019). *Antibiotic Resistance Threats in the United States, 2019*. Atlanta, GA: U.S. Department of Health and Human Services, CDC.
- Centers for Disease Control and Prevention (2020). *Serotypes and the Importance of Serotyping Salmonella* [Online]. Atlanta, GA: CDC.
- Chen, L., Yang, J., Yu, J., Yao, Z., Sun, L., Shen, Y., et al. (2005). VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res.* 33, D325–D328. doi: 10.1093/nar/gki008
- Cheng, L., Connor, T. R., Siren, J., Aanensen, D. M., and Corander, J. (2013). Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Mol. Biol. Evol.* 30, 1224–1228. doi: 10.1093/molbev/mst028
- Cummings, K. J., Warnick, L. D., Davis, M. A., Eckmann, K., Grohn, Y. T., Hoelzer, K., et al. (2012). Farm animal contact as risk factor for transmission of bovine-associated *Salmonella* subtypes. *Emerg. Infect. Dis.* 18, 1929–1936. doi: 10.3201/eid1812.110831
- Den Bakker, H. C., Allard, M. W., Bopp, D., Brown, E. W., Fontana, J., Iqbal, Z., et al. (2014). Rapid whole-genome sequencing for surveillance of *Salmonella enterica* serovar enteritidis. *Emerg. Infect. Dis.* 20, 1306–1314. doi: 10.3201/eid2008.131399
- Deng, X., Desai, P. T., Den Bakker, H. C., Mikoleit, M., Tolar, B., Trees, E., et al. (2014). Genomic epidemiology of *Salmonella enterica* serotype Enteritidis based on population structure of prevalent lineages. *Emerg. Infect. Dis.* 20, 1481–1489. doi: 10.3201/eid2009.131095
- El Zowalaty, M. E., Hickman, R. A., Mthembu, T. P., Zishiri, O. T., El Zowalaty, A. E., and Jarhult, J. D. (2020). Genome sequences of two *Salmonella enterica* strains (MEZSAL74 and MEZSAL81) harbouring multiple antimicrobial resistance genes isolated from livestock in South Africa. *J. Glob. Antimicrob. Resist.* 21, 396–398. doi: 10.1016/j.jgar.2020.04.001
- European Food Safety Authority, and European Centre for Disease Prevention and Control (2018). The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2017. *EFSA J.* 16:e05500. doi: 10.2903/j.efsa.2018.5500
- Ewels, P., Magnusson, M., Lundin, S., and Kaller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048. doi: 10.1093/bioinformatics/btw354
- Feasey, N. A., Hadfield, J., Keddy, K. H., Dallman, T. J., Jacobs, J., Deng, X., et al. (2016). Distinct *Salmonella* Enteritidis lineages associated with enterocolitis in high-income settings and invasive disease in low-income settings. *Nat. Genet.* 48, 1211–1217. doi: 10.1038/ng.3644
- Feldgarden, M., Brover, V., Haft, D. H., Prasad, A. B., Slotta, D. J., Tolstoy, I., et al. (2019). Validating the AMRFinder Tool and Resistance Gene Database by Using Antimicrobial Resistance Genotype-Phenotype Correlations in a Collection of Isolates. *Antimicrob. Agents Chemother.* 63, e483–e419. doi: 10.1128/AAC.00483-19
- Fenske, G. J., Thachil, A., Mcdonough, P. L., Glaser, A., and Scaria, J. (2019). Geography Shapes the Population Genomics of *Salmonella enterica* Dublin. *Genome Biol. Evol.* 11, 2220–2231. doi: 10.1093/gbe/evz158
- Ferrari, R. G., Rosario, D. K. A., Cunha-Neto, A., Mano, S. B., Figueiredo, E. E. S., and Conte-Junior, C. A. (2019). Worldwide Epidemiology of *Salmonella* Serovars in Animal-Based Foods: a Meta-analysis. *Appl. Environ. Microbiol.* 85, e591–e519. doi: 10.1128/AEM.00591-19
- Gardner, S. N., and Hall, B. G. (2013). When whole-genome alignments just won't work: kSNP v2 software for alignment-free SNP discovery and phylogenetics of hundreds of microbial genomes. *PLoS One* 8:e81760. doi: 10.1371/journal.pone.0081760
- Gardner, S. N., Slezak, T., and Hall, B. G. (2015). kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or

- reference genome. *Bioinformatics* 31, 2877–2878. doi: 10.1093/bioinformatics/btv271
- Gelaw, A. K., Nthaba, P., and Matle, I. (2018). Detection of *Salmonella* from animal sources in South Africa between 2007 and 2014. *J. S. Afr. Vet. Assoc.* 89, e1–e10. doi: 10.4102/jsava.v89i0.1643
- GERMS-SA (2019). *GERMS-SA Annual Report 2019* [Online]. Available online at: <http://www.nicd.ac.za/index.php/publications/germs-annual-reports/> (accessed April 29, 2021)
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086
- Hald, T., Aspinall, W., Devleeschauwer, B., Cooke, R., Corrigan, T., Havelaar, A. H., et al. (2016). World Health Organization Estimates of the Relative Contributions of Food to the Burden of Disease Due to Selected Foodborne Hazards: A Structured Expert Elicitation. *PLoS One* 11:e0145839. doi: 10.1371/journal.pone.0145839
- Harvey, R. R., Friedman, C. R., Crim, S. M., Judd, M., Barrett, K. A., Tolar, B., et al. (2017). Epidemiology of *Salmonella enterica* Serotype Dublin Infections among Humans, United States, 1968–2013. *Emerg. Infect. Dis.* 23, 1493–1501. doi: 10.3201/eid2309.170136
- Hirsch, W., Richter, C. B., Bregman, E., Lubling, P., and Altman, G. (1954). A New *Salmonella* Type (*Salmonella hadar*). *Acta Medica Orientalia* 13:41.
- Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281
- Hoelzer, K., Moreno Switt, A. L., and Wiedmann, M. (2011). Animal contact as a source of human non-typhoidal salmonellosis. *Vet. Res.* 42:34. doi: 10.1186/1297-9716-42-34
- Hunt, M., Mather, A. E., Sanchez-Buso, L., Page, A. J., Parkhill, J., Keane, J. A., et al. (2017). ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb. Genom.* 3:e000131. doi: 10.1099/mgen.0.000131
- Ibrahim, G. M., and Morin, P. M. (2018). *Salmonella* Serotyping Using Whole Genome Sequencing. *Front. Microbiol.* 9:2993. doi: 10.3389/fmicb.2018.02993
- Inns, T., Ashton, P. M., Herrera-Leon, S., Lighthill, J., Foulkes, S., Jombart, T., et al. (2017). Prospective use of whole genome sequencing (WGS) detected a multi-country outbreak of *Salmonella* Enteritidis. *Epidemiol. Infect.* 145, 289–298. doi: 10.1017/S0950268816001941
- Inzaule, S. C., Tessema, S. K., Kebede, Y., Ogwel Ouma, A. E., and Nkengasong, J. N. (2021). Genomic-informed pathogen surveillance in Africa: opportunities and challenges. *Lancet Infect. Dis.* 21, e281–e289. doi: 10.1016/S1473-3099(20)30939-7
- Jolley, K. A., and Maiden, M. C. (2010). BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 11:595. doi: 10.1186/1471-2105-11-595
- Jolley, K. A., Bray, J. E., and Maiden, M. C. J. (2018). Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res.* 3:124. doi: 10.12688/wellcomeopenres.14826.1
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., and Jermini, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285
- Kidanemariam, A., Engelbrecht, M., and Picard, J. (2010). Retrospective study on the incidence of *Salmonella* isolations in animals in South Africa, 1996 to 2006. *J. S. Afr. Vet. Assoc.* 81, 37–44. doi: 10.4102/jsava.v81i1.94
- Kimura, M. (1981). Estimation of evolutionary distances between homologous nucleotide sequences. *Proc. Natl. Acad. Sci. U S A.* 78, 454–458. doi: 10.1073/pnas.78.1.454
- Kirk, M. D., Pires, S. M., Black, R. E., Caipo, M., Crump, J. A., Devleeschauwer, B., et al. (2015). World Health Organization Estimates of the Global and Regional Disease Burden of 22 Foodborne Bacterial, Protozoal, and Viral Diseases, 2010: A Data Synthesis. *PLoS Med.* 12:e1001921. doi: 10.1371/journal.pmed.1001921
- Kodama, Y., Shumway, M., Leinonen, R., International Nucleotide, and Sequence Database, C. (2012). The Sequence Read Archive: explosive growth of sequencing data. *Nucleic Acids Res.* 40, D54–D56. doi: 10.1093/nar/gkr854
- Leekitcharoenphon, P., Hendriksen, R. S., Le Hello, S., Weill, F. X., Baggesen, D. L., Jun, S. R., et al. (2016). Global Genomic Epidemiology of *Salmonella enterica* Serovar Typhimurium DT104. *Appl. Environ. Microbiol.* 82, 2516–2526. doi: 10.1128/AEM.03821-15
- Leinonen, R., Sugawara, H., Shumway, M., International Nucleotide, and Sequence Database, C. (2011). The sequence read archive. *Nucleic Acids Res.* 39, D19–D21. doi: 10.1093/nar/gkq1019
- Liao, J., Orsi, R. H., Carroll, L. M., and Wiedmann, M. (2020). Comparative genomics reveals different population structures associated with host and geographic origin in antimicrobial-resistant *Salmonella enterica*. *Environ. Microbiol.* 22, 2811–2828. doi: 10.1111/1462-2920.15014
- Liu, B., Zheng, D., Jin, Q., Chen, L., and Yang, J. (2019). VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res.* 47, D687–D692. doi: 10.1093/nar/gky1080
- Llarena, A. K., Zhang, J., Vehkala, M., Valimaki, N., Hakkinen, M., Hanninen, M. L., et al. (2016). Monomorphic genotypes within a generalist lineage of *Campylobacter jejuni* show signs of global dispersion. *Microb. Genom.* 2:e000088. doi: 10.1099/mgen.0.000088
- Magwedere, K., Rauff, D., De Klerk, G., Keddy, K. H., and Dziva, F. (2015). Incidence of Nontyphoidal *Salmonella* in Food-Producing Animals, Animal Feed, and the Associated Environment in South Africa, 2012–2014. *Clin. Infect. Dis.* 61(Suppl. 4), S283–S289. doi: 10.1093/cid/civ663
- Majowicz, S. E., Musto, J., Scallan, E., Angulo, F. J., Kirk, M., O'Brien, S. J., et al. (2010). The global burden of nontyphoidal *Salmonella* gastroenteritis. *Clin. Infect. Dis.* 50, 882–889. doi: 10.1086/650733
- Minh, B. Q., Nguyen, M. A., and Von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 30, 1188–1195. doi: 10.1093/molbev/mst024
- Mohammed, M., Le Hello, S., Leekitcharoenphon, P., and Hendriksen, R. (2017). The invasiveness of *Salmonella* Dublin as revealed by whole genome sequencing. *BMC Infect. Dis.* 17:544. doi: 10.1186/s12879-017-2628-x
- Mthembu, T. P., Zishiri, O. T., and El Zowalaty, M. E. (2021). Genomic Characterization of Antimicrobial Resistance in Food Chain and Livestock-Associated *Salmonella* Species. *Animals* 11:872. doi: 10.3390/ani11030872
- Mughini-Gras, L., Pijnacker, R., Duijster, J., Heck, M., Wit, B., Veldman, K., et al. (2020). Changing epidemiology of invasive non-typhoid *Salmonella* infection: a nationwide population-based registry study. *Clin. Microbiol. Infect.* 26, e949–e941. doi: 10.1016/j.cmi.2019.11.015
- Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- O'Brien, S. J. (2013). The "decline and fall" of nontyphoidal *salmonella* in the United Kingdom. *Clin. Infect. Dis.* 56, 705–710. doi: 10.1093/cid/cis967
- Palma, F., Manfreda, G., Silva, M., Parisi, A., Barker, D. O. R., Taboada, E. N., et al. (2018). Genome-wide identification of geographical segregated genetic markers in *Salmonella enterica* serovar Typhimurium variant 4,[5],12:i:1. *Sci. Rep.* 8:15251. doi: 10.1038/s41598-018-33266-5
- Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528. doi: 10.1093/bioinformatics/bty633
- Park, S. E., Pham, D. T., Boinett, C., Wong, V. K., Pak, G. D., Panzner, U., et al. (2018). The phylogeography and incidence of multi-drug resistant typhoid fever in sub-Saharan Africa. *Nat. Commun.* 9:5094. doi: 10.1038/s41467-018-07370-z
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114
- R Core Team (2019). *"R: A Language and Environment for Statistical Computing"*, 3.6.1 Edn. Vienna: R Foundation for Statistical Computing.
- R Hackathon (2019). *"phylobase: Base Package for Phylogenetic Structures and Comparative Data"*, 0.8.6 Edn. Vienna: R Foundation for Statistical Computing.
- Rabsch, W., Andrews, H. L., Kingsley, R. A., Prager, R., Tschape, H., Adams, L. G., et al. (2002). *Salmonella enterica* serotype Typhimurium and its host-adapted variants. *Infect. Immun.* 70, 2249–2255. doi: 10.1128/IAI.70.5.2249-2255.2002
- Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3, 217–223. doi: 10.1111/j.2041-210X.2011.00169.x
- Rodriguez-Rivera, L. D., Wright, E. M., Siler, J. D., Elton, M., Cummings, K. J., Warnick, L. D., et al. (2014). Subtype analysis of *Salmonella* isolated from subclinically infected dairy cattle and dairy farm environments reveals the presence of both human- and bovine-associated subtypes. *Vet. Microbiol.* 170, 307–316. doi: 10.1016/j.vetmic.2014.02.013

- Rowe, B., Hall, M. L., Ward, L. R., and De Sa, J. D. (1980). Epidemic spread of *Salmonella* hadar in England and Wales. *Br. Med. J.* 280, 1065–1066. doi: 10.1136/bmj.280.6221.1065
- Rule, R., Said, M., Mbelle, N., and Osei Sekyere, J. (2019). Genome sequence of a clinical *Salmonella* Enteritidis sequence type 11 strain from South Africa. *J. Glob. Antimicrob. Resist.* 19, 164–166. doi: 10.1016/j.jgar.2019.09.014
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Smith, A. M., Smouse, S. L., Tau, N. P., Bamford, C., Moodley, V. M., Jacobs, C., et al. (2017). Laboratory-acquired infections of *Salmonella enterica* serotype Typhi in South Africa: phenotypic and genotypic analysis of isolates. *BMC Infect. Dis.* 17:656. doi: 10.1186/s12879-017-2757-2
- Smith, A. M., Tau, N. P., Ngomane, H. M., Sekwadi, P., Ramalwa, N., Moodley, K., et al. (2020). Whole-genome sequencing to investigate two concurrent outbreaks of *Salmonella* Enteritidis in South Africa, 2018. *J. Med. Microbiol.* 69, 1303–1307. doi: 10.1099/jmm.0.001255
- Strachan, N. J. C., Rotariu, O., Lopes, B., Macrae, M., Fairley, S., Laing, C., et al. (2015). Whole Genome Sequencing demonstrates that Geographic Variation of *Escherichia coli* O157 Genotypes Dominates Host Association. *Sci. Rep.* 5:14145. doi: 10.1038/srep14145
- Taylor, D. N., Bied, J. M., Munro, J. S., and Feldman, R. A. (1982). *Salmonella* dublin infections in the United States, 1979–1980. *J. Infect. Dis.* 146, 322–327. doi: 10.1093/infdis/146.3.322
- The, H. C., Thanh, D. P., Holt, K. E., Thomson, N. R., and Baker, S. (2016). The genomic signatures of *Shigella* evolution, adaptation and geographical spread. *Nat. Rev. Microbiol.* 14, 235–250. doi: 10.1038/nrmicro.2016.10
- Threlfall, E. J. (2000). Epidemic *salmonella* typhimurium DT 104—a truly international multiresistant clone. *J. Antimicrob. Chemother.* 46, 7–10. doi: 10.1093/jac/46.1.7
- To, T.-H., Jung, M., Lycett, S., and Gascuel, O. (2015). Fast Dating Using Least-Squares Criteria and Algorithms. *Systemat. Biol.* 65, 82–97. doi: 10.1093/sysbio/syv068
- Tonkin-Hill, G., Lees, J. A., Bentley, S. D., Frost, S. D. W., and Corander, J. (2018). RhierBAPS: An R implementation of the population clustering algorithm hierBAPS. *Wellcome Open Res.* 3:93. doi: 10.12688/wellcomeopenres.14694.1
- Treangen, T. J., Ondov, B. D., Koren, S., and Phillippy, A. M. (2014). The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* 15:524. doi: 10.1186/s13059-014-0524-x
- USDA Foreign Agricultural Service (2020). *United States Poultry Exports to South Africa Continue Steady Increase, Report Number SF2020-0024 [Online]*. Washington, DC: USDA Foreign Agricultural Service.
- Uzzau, S., Brown, D. J., Wallis, T., Rubino, S., Leori, G., Bernard, S., et al. (2000). Host adapted serotypes of *Salmonella enterica*. *Epidemiol. Infect.* 125, 229–255. doi: 10.1017/S0950268899004379
- Wickham, H. (2007). Reshaping Data with the reshape Package. *J. Statist. Softw.* 21:20. doi: 10.18637/jss.v021.i12
- Wickham, H., François, R., Henry, L., and Müller, K. (2020). "dplyr: A Grammar of Data Manipulation", 0.8.5 Edn. Vienna: R Foundation for Statistical Computing.
- Wong, V. K., Baker, S., Pickard, D. J., Parkhill, J., Page, A. J., Feasey, N. A., et al. (2015). Phylogeographical analysis of the dominant multidrug-resistant H58 clade of *Salmonella* Typhi identifies inter- and intracontinental transmission events. *Nat. Genet.* 47, 632–639. doi: 10.1038/ng.3281
- World Health Organization (2014). *Antimicrobial resistance: global report on surveillance*. Geneva: World Health Organization (WHO).
- World Health Organization (2018). *Salmonella (non-typhoidal)* [Online]. Geneva: World Health Organization (WHO).
- Worley, J., Meng, J., Allard, M. W., Brown, E. W., and Timme, R. E. (2018). *Salmonella enterica* Phylogeny Based on Whole-Genome Sequencing Reveals Two New Clades and Novel Patterns of Horizontally Acquired Genetic Elements. *MBio* 9, e2303–e2318. doi: 10.1128/mBio.02303-18
- Yoshida, C. E., Kruczkiewicz, P., Laing, C. R., Lingohr, E. J., Gannon, V. P., Nash, J. H., et al. (2016). The *Salmonella In Silico* Typing Resource (SISTR): An Open Web-Accessible Tool for Rapidly Typing and Subtyping Draft *Salmonella* Genome Assemblies. *PLoS One* 11:e0147101. doi: 10.1371/journal.pone.0147101
- Yu, G., Lam, T. T., Zhu, H., and Guan, Y. (2018). Two Methods for Mapping and Visualizing Associated Data on Phylogeny Using Ggtree. *Mol. Biol. Evol.* 35, 3041–3043. doi: 10.1093/molbev/msy194
- Yu, G., Smith, D. K., Zhu, H., Guan, Y., and Lam, T. T.-Y. (2017). ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* 8, 28–36. doi: 10.1111/2041-210X.12628
- Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., et al. (2012). Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* 67, 2640–2644. doi: 10.1093/jac/dks261
- Zhang, S., Den Bakker, H. C., Li, S., Chen, J., Dinsmore, B. A., Lane, C., et al. (2019). SeqSero2: Rapid and Improved *Salmonella* Serotype Determination Using Whole-Genome Sequencing Data. *Appl. Environ. Microbiol.* 85, e1746–e1719. doi: 10.1128/AEM.01746-19
- Zhou, Z., Alikhan, N. F., Mohamed, K., Fan, Y., Agama Study, G., and Achtman, M. (2020). The Enterobase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny, and *Escherichia coli* core genomic diversity. *Genome Res.* 30, 138–152. doi: 10.1101/gr.251678.119

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

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

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