Genomic characterization of endemic and ecdemic non-typhoidal *Salmonella enterica* lineages circulating among animals and animal products in South Africa

Laura M. Carroll[a], Rian Pierneef[b], Masenyabu Mathole[c], Itumeleng Matle[c]#

[a]Structural and Computational Biology Unit, EMBL, Heidelberg, Germany

[b]Biotechnology Platform, Agricultural Research Council, Onderstepoort Veterinary Research, Onderstepoort, South Africa

[c]Bacteriology Division, Agricultural Research Council: Onderstepoort Veterinary Research, Onderstepoort, South Africa

Corresponding author: Itumeleng Matle, matlei@arc.agric.za

Running Title: South African *Salmonella enterica* genomic sequencing

ABSTRACT

Non-typhoidal *Salmonella enterica* imposes a significant burden on human and animal health in South Africa. However, very little is known about lineages circulating among animals and animal products in the country on a genomic scale. Here, we used whole-genome sequencing (WGS) to characterize 63 *Salmonella enterica* strains (n = 18, 8, 13, and 24 strains assigned to serotypes Dublin, Hadar, Enteritidis, and Typhimurium, respectively) isolated from livestock, companion animals, wildlife, and animal products in South Africa over a 60-year period. Within-serotype phylogenies were constructed using genomes sequenced in this study, as well as publicly available genomes representative of each respective serotype'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, the approaches used here identified a largely antimicrobial-susceptible, endemic lineage circulating among humans, animals, and food in South Africa, as well as a lineage that was likely recently introduced from the United States. For *S.* Hadar, multiple South African lineages harboring streptomycin and tetracycline resistance-conferring genes were identified. African *S.* Enteritidis 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 strains sequenced here were distributed across the African *S.* Typhimurium phylogeny, representing a diverse range of lineages, including numerous MDR lineages. Overall, this study provides insight into the evolution, population structure, and antimicrobial resistome composition of *Salmonella enterica* in Africa.

IMPORTANCE

Globally, *Salmonella enterica* is estimated to be responsible for more than 93 million illnesses and 150,000 deaths annually. In Africa, the burden of salmonellosis is disproportionally high; however, WGS efforts are overwhelmingly concentrated in world regions with lower salmonellosis 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. WGS data derived from non-typhoidal *Salmonella enterica* serotypes isolated from non-human sources in South Africa is extremely limited. To our knowledge, the genomes sequenced here represent the largest collection of non-typhoidal *Salmonella enterica* isolate genomes from non-human sources in South Africa to date. Furthermore, 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.

INTRODUCTION

Livestock, domestic animals, and wildlife can serve as potential reservoirs for non-typhoidal *Salmonella enterica* (1, 2). 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 (2, 3); 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 (2, 4). *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 (5-11).

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 (12, 13). 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 (14, 15). 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) (16-19).

In South Africa, the bulk of *Salmonella enterica* WGS efforts have focused on characterizing human clinical strains associated with illnesses and/or outbreaks (20-25). WGS-

based studies querying *Salmonella enterica* strains isolated from non-human sources in South Africa are limited (26), and little is known regarding which lineages are circulating among animals in the country (27). 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 the global (for *S.* Dublin and *S.* Hadar) and African (for *S.* Enteritidis and *S.* Typhimurium) *Salmonella enterica* populations. The results presented here will provide critical insights into the evolution, population structure, and AMR dynamics of *Salmonella enterica* in Africa.

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 (Supplemental Table S1). All isolates underwent *in silico* serotyping using both (i) SISTR (using its core-genome MLST [cgMLST] approach) and (ii) SeqSero2 (Supplemental Table S1); serotypes assigned using both methods were identical for all isolates (63 of 63 isolates, 100%; Supplemental Table S1). Furthermore, genomes of all isolates sequenced here clustered among publicly available *Salmonella* genomes assigned to their respective serotypes (28), 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 Supplemental Table S1). 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), unknown 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; Supplemental Table S1). 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 (Supplemental Table S1).

AMR 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 Supplemental Figures S1 and S2). 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) (29). The number of different AMR/stress response determinants detected per genome ranged from five 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;

137 46 of these (46 of 59 total unique AMR determinants, 78.0%) were present only sporadically and
138 were detected in two or fewer genomes (Figure 2).

139 In total, 17 different plasmid replicons were identified among all 63 genomes, representing
140 22 unique plasmid replicon presence/absence profiles (detected using ABRicate, the
141 PlasmidFinder database, and minimum nucleotide identity and coverage thresholds of 80 and 60%,
142 respectively; Figure 2) (30). Genomes harbored one to seven different plasmid replicons, with a
143 median of two per genome (Figure 2). Two plasmid replicons, IncFIB(S) and IncFII(S), were
144 detected in over half of all genomes sequenced here (detected in 32 and 49 of 63 genomes, 50.8%
145 and 77.8%, respectively; Figure 2). Over half of all plasmid replicons (10 of 17 unique plasmid
146 replicons; 58.8%) were detected in two or fewer genomes (Figure 2).

147 Additionally, a total of 181 different virulence factors were identified among the 63
148 genomes, with 24 unique virulence factor presence/absence profiles represented (detected using
149 ABRicate, the Virulence Factor Database [VFDB], and minimum nucleotide identity and coverage
150 thresholds of 70 and 50%, respectively; Figure 2 and Supplemental Table S2). Genomes harbored
151 146 to 171 different virulence factors, with a median of 165 (Figure 2 and Supplemental Table
152 S2). Over 75% of all unique virulence factors detected among the isolates sequenced in this study
153 were present in all genomes (137 of 181 unique virulence factors, 75.7%; Supplemental Table S2).
154 Only 13 virulence factors were detected in fewer than half of the genomes sequenced here (Figure
155 2).

156 **A largely antimicrobial-susceptible *S.* Dublin ST10 lineage circulating in South Africa**
157 **encompasses isolates from livestock, food, and human sources.** A maximum likelihood (ML)
158 phylogeny constructed using the 18 South African sequence type 10 (ST10) *S.* Dublin isolates
159 sequenced here, plus 2,784 publicly available ST10 *S.* Dublin genomes, partitioned the vast

majority of genomes (2,738 of 2,802 genomes, 97.7%) into two major *S.* Dublin ST10 clades (Figure 3 and Supplemental Figure S3), which is consistent with previous observations (31). 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 Supplemental Figure S3). 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 Supplemental Figure S3). 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 Supplemental Figure S3). Notably, multi-drug resistant (MDR) *S.* Dublin, which often possess IncA/C2 plasmids and acquired AMR determinants that confer resistance to aminoglycosides, beta-lactams, phenicols, sulfonamides, and tetracyclines (31), 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 Supplemental Figure S3). 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 Supplemental Figure S3).

All 18 South African *S.* Dublin isolates sequenced in this study belonged to *S.* Dublin Major Clade I (Figure 3 and Supplemental Figure S3); 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 (Supplemental Figure S4). 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 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 *sul1* (Figure 4). Taken together, these results indicate that a largely AMR-susceptible South African-specific *S.* Dublin lineage has been circulating among animals, foods, and humans in the country for decades.

Only one South African *S.* Dublin genome was not a member of the South African *S.* Dublin Clade within the *S.* Dublin Small Subclade (Supplemental Figure S4). 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 (Supplemental Figure S4). 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]; Supplemental Figure S4). Notably, despite clustering among MDR *S.* Dublin 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 from the United States (Supplemental Figure S4). These results indicate that a separate *S.* Dublin lineage

206 may have only recently been introduced into South Africa from North America, a hypothesis that
207 was further supported by subsequent investigation into the origin of the isolate: the poultry meat
208 from which strain FOO_2016_SouthAfrica_EasternCape_AF_7509SE-S59 was isolated had been
209 imported from North America and sold in a supermarket in South Africa's Eastern Cape province.
210 All 18 *S.* Dublin isolates sequenced in this study, as well as all seven publicly available
211 South African *S.* Dublin genomes, were members of *S.* Dublin Major Clade I (Figure 3 and
212 Supplemental Figure S3). These 25 South African genomes, 24 of which formed a well-supported
213 subclade within Major Clade I, were the only African genomes detected in *S.* Dublin Major Clade
214 I (Figure 3 and Supplemental Figures S3-S4). *S.* Dublin Major Clade II did not contain any African
215 genomes (Figure 3). However, 18 genomes from the African continent were among the few
216 genomes (i.e., 64 of 2,802 *S.* Dublin genomes, 2.3%) that fell outside of the two major *S.* Dublin
217 clades (Figure 3). These genomes were reported to have been derived from strains isolated from
218 animals, food, and humans in Ethiopia, Gambia, Nigeria, and Benin, and none harbored any
219 acquired AMR genes (Figure 3); interestingly, they clustered among human-associated genomes
220 from Asia (i.e., Taiwan), Europe (i.e., France and the United Kingdom), and North America (i.e.,
221 Canada and the United States), forming a 52-genome, well-supported clade with a common
222 ancestor dated to circa 1957 (95% CI [1142.96, 2003.00], 100% UltraFast Bootstrap Support;
223 Figure 3).
224 **South Africa harbors multiple *S.* Hadar ST33 lineages with streptomycin and tetracycline**
225 **resistance-conferring genes.** A ML phylogeny was constructed using the eight South African
226 ST33 *S.* Hadar isolates sequenced here, plus 1,561 publicly available ST33 *S.* Hadar genomes
227 (Figure 5). Notably, the majority of *S.* Hadar genomes harbored AMR genes *aph(3'')-Ib* and
228 *aph(6)-Id* (n = 1,314 and 1,347 of 1,569 *S.* Hadar 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 genomes additionally harbored *tet(A)*, which confers resistance to tetracycline (n = 1,320 of 1,569 *S.* Hadar genomes, 84.1%; Figure 5). All eight *S.* Hadar 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-conferring *aph(3'')-Ib* and *aph(6)-Id* and tetracycline resistance-conferring *tet(A)* (Figure 5).

Seven of eight *S.* Hadar 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 unknown source, were most closely related to a publicly available genome of a *S.* Hadar 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 isolates sequenced here (Figure 5). Isolated from bovine feces in 1990, this strain was most closely related to a *S.* Hadar 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 lineages have circulated in livestock populations in the country.

One largely antimicrobial-susceptible clade and one largely MDR 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 Supplemental Figure S5).

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 Supplemental Figure S5). 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 Supplemental Figure S5). None of the genomes in this clade harbored any known AMR determinants (Figure 6 and Supplemental Figure S5). 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

Supplemental Figure S5, 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 Supplemental Figure S5).

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_{TEM-1}), streptomycin (*aph(3'')-Ib, aph(6)-Id*), sulfonamides (*sul1, sul2*), chloramphenicol (*catA2*), trimethoprim (*dfrA7*), 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 MDR *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 Supplemental Text). The 24 *S.* Typhimurium strains sequenced in this study were distributed across the African *S.* Typhimurium phylogeny, representing a diverse range of lineages,

298 and eight (33.3%) possessed one or more AMR genes (Figure 7 and Supplemental Text). Notably,
299 some African *S.* Typhimurium lineages were distributed across the African continent, while others
300 were strongly associated with a particular region/country (Figure 7 and Supplemental Text). When
301 compared to genomes from a previous study of *S.* Typhimurium from New York State that have
302 been shown to be representative of the human- and bovine-associated *S.* Typhimurium population
303 in the United States as a whole (32), only five of 24 *S.* Typhimurium strains sequenced here
304 (20.8%) shared a common ancestor with one or more New York State strains after 1900
305 (Supplemental Figure S6 and Supplemental Text). This indicates that many of the strains
306 sequenced here are not closely related to *S.* Typhimurium lineages circulating among cattle and
307 humans in the United States. Below, we discuss some of these major African *S.* Typhimurium
308 clades in detail (see the Supplemental Text for discussions of additional lineages).
309 **A *S.* Typhimurium DT104-like clade emerged in Africa in the twentieth century as**
310 **antimicrobial-susceptible and later acquired MDR.** A *S.* Typhimurium strain sequenced here
311 (PIG_2002_FS_040ST-S45), isolated in 2002 from swine meat in South Africa's Free State
312 province, clustered within a 69-isolate clade, which shared a common ancestor dated circa 1884
313 (95% CI [1153.59, 1956.00]; denoted as the "Large Mixed/DT104 Clade" in Figure 7). This large
314 clade contained a mixture of human-, animal-, environmental, and food-associated isolates from
315 Senegal (n = 16), Gambia (n = 13), Tunisia (n = 10), Benin (n = 7), Ethiopia (n = 6), Morocco (n
316 = 4), DRC and South Africa (n = 3 each), Madagascar (n = 2), Algeria, Cameroon, Egypt, Kenya,
317 and Tanzania (n = 1 each). The strain isolated in this study possessed five AMR genes:
318 streptomycin resistance-conferring *aadA2*, beta lactamase bla_{CARB-2}, chloramphenicol resistance-
319 conferring *floR*, sulfonamide resistance-conferring *sul1*, and tetracycline resistance-conferring
320 *tet(G)* (Figure 7). Notably, when compared to genomes from a previous study of *S.* Typhimurium

in the United States (New York State) (32), 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]; Supplemental Figure S6).

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 (33). 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 ≈1966 and ≈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 (31, 34-38). 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,

366 for example, we identified a largely AMR-susceptible South Africa-specific clade, which has been
367 circulating among animals, foods, and humans in the country for decades. While *S.* Dublin is
368 largely considered to be a bovine-adapted serotype, human infections caused by *S.* Dublin are
369 frequently invasive and may result in severe illness and/or death (5-11, 39, 40). In South Africa,
370 invasive non-typhoidal salmonellosis is a serious public health concern: in 2019, over 25% of all
371 non-typhoidal salmonellosis cases reported to the Group for Enteric, Respiratory and Meningeal
372 disease Surveillance in South Africa (GERMS-SA) were invasive (825 and 2,437 reported invasive
373 and non-invasive non-typhoidal salmonellosis cases, respectively; 25.3%) (41). While *S.* Dublin
374 is not among the most common serotypes isolated from human clinical cases in South Africa (41),
375 routine veterinary surveillance has revealed that *S.* Dublin is frequently isolated from animal
376 sources in the country, particularly cattle (42). Further WGS efforts are needed to provide insight
377 into the evolution and between-host transmission dynamics of the endemic South African *S.*
378 Dublin lineage identified in this study.

379 In addition to the endemic South African *S.* Dublin lineage, we identified a clade of South
380 African animal- and animal product-associated *S.* Hadar ST33 strains, which clustered near the
381 root of the global *S.* Hadar ST33 phylogeny. First described as a novel *Salmonella* serotype in
382 1954 (43), *S.* Hadar was reported to have been responsible for several cases of diarrheal illness in
383 Israel (43). Reportedly, serotype *S.* Hadar was rarely isolated prior to 1971; however, in the mid-
384 1970s, *S.* Hadar quickly became the second-most common cause of human nontyphoidal
385 salmonellosis in the United Kingdom (44-46). Consistent with these observations, the South
386 African *S.* Hadar clade identified here shared a common ancestor dated circa 1962 and contained
387 strains isolated from the 1960s through 2018. In South Africa specifically, *S.* Hadar is not among
388 the top serotypes associated with human clinical cases (41); however, *S.* Hadar has been commonly

389 isolated from animals and animal associated-environments in the country for nearly two decades
390 (42, 47, 48). The results presented here indicate that an endemic *S.* Hadar ST33 lineage has been
391 circulating among animals in South Africa for over fifty years; however, future WGS efforts
392 querying *S.* Hadar strains from around the world—historical strains isolated prior to the 1970s, in
393 particular—are needed to refine estimates as to when this particular lineage emerged.

394 In addition to the *S.* Dublin ST10 and *S.* Hadar ST33 endemic South African lineages, we
395 observed that African *S.* Enteritidis ST11 could largely be partitioned into one largely
396 antimicrobial-susceptible and one largely MDR clade. South African *S.* Enteritidis ST11, including
397 those sequenced here, were confined to the largely antimicrobial-susceptible clade. These results
398 are consistent with those observed in a previous study of *S.* Enteritidis in Africa (25), in which a
399 geographically distinct MDR *S.* Enteritidis lineage was identified in Africa's Central/East regions
400 and rarely detected in South Africa. Since 2012, *S.* Enteritidis has been the serotype most
401 commonly isolated from human clinical cases in South Africa (41). Among animals, *S.* Enteritidis
402 has been one of the most frequently isolated serotypes in South Africa for decades, particularly
403 from poultry-associated sources (42, 47, 48). Our results further support that South African *S.*
404 Enteritidis, which is one of the most common *Salmonella enterica* serotypes circulating among
405 animals and humans in the country, acquires AMR only sporadically and is, on a genomic scale,
406 distinct from MDR *S.* Enteritidis lineages circulating in other regions of Africa. Collectively, our
407 study reveals that endemic lineages of several non-typhoidal *Salmonella enterica* serotypes are
408 circulating among animals and animal products in South Africa, some of which may occasionally
409 infect humans.

410 **WGS can differentiate endemic and ecdemic *Salmonella enterica* lineages.** Pathogenic bacteria
411 not previously endemic to a given geographic region can be introduced into that region through

412 the movement of humans, food, and/or animals (23, 36, 49, 50). In addition to observing
413 *Salmonella enterica* lineages that were likely endemic to South Africa, our study identified
414 numerous lineages that were likely to have been introduced into the country only recently. One *S.*
415 Dublin isolate sequenced in this study, for example, clustered among isolates from the United
416 States, indicating that this strain had been introduced into South Africa only recently. *S.* Dublin
417 from the United States has previously been shown to be distinct from *S.* Dublin strains isolated in
418 other world regions on a genomic scale (31), and the United States was one of the leading poultry
419 exporters to South Africa in 2016 (i.e., the year the ecdemic *S.* Dublin strain sequenced here was
420 isolated) (51). Our recent introduction hypothesis was further supported by metadata indicating
421 that this strain had been isolated from poultry meat imported from North America and sold in a
422 supermarket in South Africa's Eastern Cape province.

423 We observed similar results for *S.* Hadar: one *S.* Hadar strain sequenced in this study was
424 more closely related to *S.* Hadar from the United States than to its South African counterparts,
425 which all formed a clade near the global *S.* Hadar ST33 phylogeny root (44-46). Unlike the *S.*
426 Dublin strain sequenced here, which was likely introduced into South Africa from imported
427 poultry meat, it is unclear exactly how the unique *S.* Hadar lineage sequenced here was introduced
428 into the country, as its representative strain was isolated from bovine feces in 1990 and shared a
429 common ancestor circa 1982 with a canine-associated *S.* Hadar strain isolated in 1988 in the United
430 States. Future WGS efforts querying *S.* Hadar may provide insight into this lineage and its
431 emergence in South Africa.

432 The *S.* Typhimurium isolates sequenced here were distributed across the African *S.*
433 Typhimurium phylogeny, indicating that South Africa harbors numerous *S.* Typhimurium
434 lineages. Since 2012, *S.* Typhimurium has been the second-most common non-typhoidal

435 *Salmonella enterica* serotype isolated from human clinical cases in South Africa (after *S.*
436 Enteritidis) and in 2019 was the most common serotype isolated from human clinical cases in the
437 Eastern Cape province (41). *S.* Typhimurium has additionally been one of the most frequently
438 isolated serotypes from animals and wildlife in South Africa for decades, and it is frequently
439 isolated from a broad range of hosts (e.g., cattle, poultry, equine, sheep/goats, feline, rhinoceros)
440 (42, 47, 48). Interestingly, we identified a largely AMR-susceptible, primarily South African *S.*
441 Typhimurium clade, which contained isolates from produce, fish, poultry, and avian sources, and
442 one human clinical isolate from Ethiopia (referred to above as the "Vegetable ZA Clade"), which
443 was predicted to have been introduced into the country recently (i.e., after the year 2000). It is
444 unclear exactly where this lineage originated and how it was introduced into South Africa, but
445 future WGS efforts may elucidate this.

446 We additionally identified a *S.* Typhimurium clade, which contained the genomes of strains
447 assigned to phage type DT104. MDR DT104 was responsible for a global epidemic in the 1990s,
448 during which it was increasingly isolated from a broad range of animals (e.g., cattle, poultry, pigs,
449 sheep), as well as human clinical cases (52). Notably, DT104 was predicted to have emerged as
450 antimicrobial-susceptible circa 1948, later acquiring its MDR phenotype circa 1972 (33). The
451 results observed here are consistent with these findings, as the DT104 clade identified here
452 emerged as antimicrobial-susceptible circa 1939 and acquired the MDR phenotype between ≈1966
453 and ≈1980. The DT104 clade identified here spanned multiple African regions, and South African
454 DT104-like genomes were distributed across the clade, indicating that South Africa may have been
455 subjected to multiple DT104 introduction events and/or between-country transmission events;
456 however, the lack of available DT104-like genomes from South Africa (i.e., one sequenced here
457 and two publicly available genomes) and the African continent as a whole limits our ability to say

458 this conclusively. Taken together, our results further highlight the strengths of WGS in *Salmonella*
459 source tracking, both within and between countries and continents (53-55), and showcase the
460 ability of WGS-based approaches to differentiate endemic and ecdemic lineages.

461 **WGS of historical isolates from under-sequenced geographic regions can provide novel**
462 **insights into pathogen evolution and diversity.** Worldwide, *Salmonella enterica* has been
463 estimated to be responsible for more than 93 million illnesses and more than 150,000 deaths
464 annually (56). In Africa, the disease burden imposed by *Salmonella enterica* is particularly
465 significant; mortality and disability adjusted life years (DALYs) due to diarrheal disease and
466 invasive infections caused by non-typhoidal serotypes are consistently higher in Africa than in
467 other world regions (57). However, despite the disproportionally high incidence and burden of
468 salmonellosis and other foodborne illnesses, the bulk of publicly available genomic data derived
469 from *Salmonella enterica* has come from regions with lower burdens (57, 58); for example, among
470 all *Salmonella enterica* genomes in Enterobase (accessed 7 April 2021), over 80% were derived
471 from strains reported to have been isolated in North America and Europe (128,517 and 104,910
472 genomes from North America and Europe, respectively; 233,427 of 291,362 total *Salmonella*
473 *enterica* genomes).

474 Here, we used WGS to characterize 63 *Salmonella enterica* strains isolated from animals
475 and animal products in South Africa over a 60-year time span, which, to our knowledge, represents
476 the most extensive WGS-based characterization of non-human-associated non-typhoidal
477 *Salmonella enterica* in the country to date. Importantly, numerous genomes sequenced here
478 belonged to lineages that were phylogenetically distinct from those circulating in more heavily
479 sequenced/sampled regions of the world, such as North America and Europe. For example, as
480 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 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.

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 (42, 48) 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. The isolates were preserved as lyophilized and revived by inoculation into brain heart infusion (BHI) broth and incubated at 37ºC for 18-24 hours.

503 **Whole-genome sequencing.** Genomic DNA was extracted from BHI broth cultures using the
504 High Pure PCR template preparation kit (Roche, Potsdam, Germany) according to the
505 manufacturer's instructions. WGS of the isolates was performed at the Biotechnology Platform,
506 Agricultural Research Council, South Africa. DNA libraries were prepared using TruSeq and
507 Nextera DNA library preparation kits (Illumina, San Diego, CA, USA), followed by sequencing
508 on Illumina HiSeq and MiSeq instruments (Illumina, San Diego, CA, USA).
509 **Initial data processing and quality control.** Quality control, adapter removal, decontamination,
510 and error correction of the raw sequencing data was performed using BBDuk v. 37.90
511 (https://jgi.doe.gov/data-and-tools/ bbtools/bb-tools-user-guide/bbduk-guide/), and SPAdes v.
512 3.12.0 (59) was used to create a *de novo* assembly for each isolate. FastQC v. 0.11.5
513 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to assess the quality of
514 the paired-end reads associated with each isolate (n = 73 isolates total; 21, 15, 11, and 26 isolates
515 assigned to serotypes Dublin, Enteritidis, Hadar, and Typhimurium, respectively) (60), and
516 QUAST v. 4.5 (61) was used to assess the quality of the associated assembled genome
517 (Supplemental Table S1). The lineage workflow (i.e., "lineage_wf") implemented in CheckM v.
518 1.1.3 (62) was additionally used to identify potential contamination in each assembled genome,
519 as well as to assess genome completeness (Supplemental Table S1). MultiQC v. 1.8 (63) was
520 used to assess the quality of all genomes in aggregate. Several low-quality isolate genomes with
521 >5% contamination and/or <95% completeness were identified (n = 3, 2, 3, and 2 low-quality
522 isolate genomes assigned to serotypes Dublin, Enteritidis, Hadar, and Typhimurium,
523 respectively) and were thus omitted from further analysis, yielding a final set of 63 *Salmonella*
524 *enterica* genomes used in subsequent steps (Supplemental Table S1).

525 ***In silico* serotyping and multi-locus sequence typing.** All 63 assembled *Salmonella enterica*
526 genomes (see section "Initial data processing and quality control" above) underwent *in silico*
527 serotyping using the command line implementations of (i) the *Salmonella In Silico* Typing
528 Resource (SISTR) v. 1.1.1 (18) and (ii) the *k*-mer based workflow implemented in SeqSero2 v.
529 1.1.1 (17) (Supplemental Table S1). Each genome additionally underwent *in silico* seven-gene
530 MLST using mlst v. 2.9 (https://github.com/tseemann/mlst) and the seven-gene scheme available
531 for *Salmonella enterica* (--scheme 'senterica') in PubMLST (64, 65) (Supplemental Table S1).
532 **Reference-free SNP identification and phylogeny construction.** The 63 *Salmonella enterica*
533 genomes sequenced in this study were compared to 442 of the 445 *Salmonella* genomes described
534 by Worley, et al. (28) (three genomes did not have publicly available sequence read archive [SRA]
535 data at the time of access, i.e., 20 February 2019). The SRA toolkit v. 2.9.6 was used to download
536 paired-end reads for each of the 442 publicly available genomes (66, 67), which were then
537 assembled into contigs using SPAdes v. 3.8.0 (59), using *k*-mer sizes of 21, 33, 55, 77, 99, and
538 127, and the "careful" option. SNPs were identified among all 505 assembled *Salmonella* genomes
539 with kSNP3 v. 3.92 (68, 69), using the optimal *k*-mer size determined by Kchooser ($k = 19$). The
540 resulting core SNP alignment was supplied as input to IQ-TREE v. 1.5.4 (70), which was used to
541 construct a ML phylogeny using the optimal ascertainment bias-aware nucleotide substitution
542 model identified using ModelFinder (based on its Bayesian Information Criteria [BIC] value) (71)
543 and 1,000 replicates of the Ultrafast Bootstrap method (72, 73). The resulting ML phylogeny
544 (Figure 1) was annotated using FigTree v. 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/). All
545 reference-free SNP identification and ML phylogeny construction steps described above were
546 repeated to identify SNPs among the 63 *Salmonella enterica* genomes sequenced here, with
547 publicly available genomes excluded; the resulting ML phylogeny was annotated in R v. 3.6.1 (74)

548 using the bactaxR package (75) and its dependencies ggtree (76, 77), ape (78), dplyr (79),
549 phylobase (80), phytools (81), and reshape2 (82) (Figure 2 and Supplemental Table S2).
550 ***In silico* AMR determinant, plasmid replicon, and virulence factor detection.** AMR
551 determinants were identified within each of the 63 *Salmonella* genomes sequenced in this study,
552 using each of the following pipelines: (i) AMRFinderPlus v. 3.9.3 (29), (ii) ABRicate v. 1.0.1
553 (https://github.com/tseemann/abricate), and (iii) ARIBA v. 2.14.6 (83) (Figure 2 and Supplemental
554 Figure S1). For the AMRFinderPlus pipeline, Prokka v. 1.13 (84) was used to annotate each of the
555 63 assembled genomes; the resulting GFF (.gff) and FASTA (.faa and .ffn) files were used as input
556 for AMRFinderPlus, which was used to identify AMR and stress response determinants in each
557 genome, using the *Salmonella* organism option and the most recent AMRFinderPlus database
558 (database v. 2020-11-09.1, accessed 21 November 2020). For the ABRicate pipeline, AMR
559 determinants were identified in each assembled genome using the NCBI AMR database (--db ncbi;
560 accessed 19 April 2020) (29) and minimum identity and coverage thresholds of 75 (--minid 75)
561 and 50% (--mincov 50), respectively. For the ARIBA pipeline, ARIBA's getref and prepareref
562 commands were used to download and prepare the latest version of the ResFinder database
563 (accessed 14 February 2021), respectively (85). ARIBA's run command was then used to identify
564 AMR determinants in each genome, using the paired-end reads associated with each isolate as
565 input.
566 ABRicate and ARIBA were additionally used to detect plasmid replicons within each of
567 the 63 *Salmonella* genomes sequenced in this study using the PlasmidFinder database (30) (Figure
568 2 and Supplemental Figure S2). For the ABRicate pipeline, assembled genomes were used as input,
569 and plasmid replicons were detected in each genome (--db plasmidfinder; PlasmidFinder database
570 accessed 19 April 2020) using minimum identity and coverage thresholds of 80 (--minid 80) and

571 60% (--mincov 60), respectively. For the ARIBA pipeline, ARIBA's getref and prepareref
572 commands were used to download and prepare the latest version of the PlasmidFinder database
573 (accessed 14 February 2021), respectively. ARIBA's run command was then used to identify
574 plasmid replicons in each genome, using paired-end reads associated with each isolate as input.
575 ABRicate was further used to detect virulence factors in each genome, using the Virulence Factor
576 Database (VFDB; --db vfdb, accessed 19 April 2020) (86, 87), using minimum identity and
577 coverage thresholds of 70 (--minid 70) and 50% (--mincov 50), respectively (Figure 2 and
578 Supplemental Table S2).

579 **Construction of time-scaled *S.* Dublin phylogenies.** To compare the 18 *S.* Dublin isolates
580 sequenced in this study to publicly available *S.* Dublin genomes, all genomes meeting each of the
581 following conditions were downloaded via Enterobase (accessed 27 December 2020, n = 2,784;
582 Supplemental Table S3): (i) genomes were assigned to sequence type (ST) 10 (i.e., the ST to which
583 all of the *S.* Dublin isolates sequenced in this study were assigned/approximately assigned) using
584 the Achtman seven-gene MLST scheme for *Salmonella*; (ii) genomes had an exact year of isolation
585 reported in Enterobase's "Collection Year" field; (iii) genomes could be assigned to a known
586 isolation source, with "Laboratory" strains excluded, per Enterobase's "Source Niche" field; (iv)
587 genomes could be assigned to a known country of isolation, per Enterobase's "Country" field (88,
588 89). All 2,802 assembled *S.* Dublin genomes underwent *in silico* plasmid replicon and AMR
589 determinant detection using ABRicate v. 1.0.1 and the PlasmidFinder and NCBI AMR databases,
590 respectively, as described above (see section "*In silico* AMR determinant, plasmid replicon, and
591 virulence factor detection" above).

592 Parsnp and HarvestTools v. 1.2 (90) were used to identify core SNPs among all 2,802 *S.*
593 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 recombination (91). Clusters were identified within the resulting core SNP alignment using RhierBAPs v. 1.1.3 (92, 93), R v. 4.0.0, and three clustering levels. IQ-TREE v. 1.5.4 (70) 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 (71), based on its corresponding BIC value (i.e., the TVM+I model); (iv) 1,000 replicates of the UltraFast bootstrap approximation (72).

The resulting ML phylogeny was rooted and time-scaled using LSD2 v. 1.4.2.2 (94) and the following parameters: (i) tip dates corresponding to the year of isolation associated with each genome; (ii) a fixed substitution rate of 2.79×10^{-7} substitutions/site/year (i.e., the substitution rate estimated in a previous study of *S.* Typhimurium phage type DT104) (33); (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; Supplemental Figure S3); (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; Supplemental Figure S4).

Construction of time-scaled *S.* Hadar phylogeny. To compare the eight *S.* Hadar isolates sequenced in this study to publicly available *S.* Hadar genomes, all genomes meeting each of the following conditions were downloaded via Enterobase (accessed 10 January 2021, *n* = 1,562; Supplemental Table S4): (i) genomes were assigned to ST33 (i.e., the ST to which all of the *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 of isolation, per Enterobase's "Country" field (88, 89). 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 (90) 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 (91). Clusters were identified within the resulting core SNP alignment using RhierBAPs v. 1.1.3 (92, 93), R v. 4.0.0, and three clustering levels. IQ-TREE v. 1.5.4 (70) 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 (71), based on its corresponding BIC value (i.e., the K3Pu+I model) (95); (iv) 1,000 replicates of the UltraFast bootstrap approximation (72). 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 (94) and the following parameters: (i) tip dates corresponding to the year of isolation associated with each genome; (ii) a fixed substitution rate of 2.79×10^{-7} substitutions/site/year (i.e., the substitution rate estimated in a previous study of *S.* Typhimurium phage type DT104) (33); (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 4,916,780 (i.e., the length of the reference chromosome; -s 4916780). The resulting phylogeny was annotated using the bactaxR 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, all genomes meeting each of the following conditions were downloaded via Enterobase (accessed 27 December 2020, $n = 697$; Supplemental Table S5): (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

663 country of isolation within the African continent, per Enterobase's "Country" and "Continent"
664 fields, respectively (88, 89). Additionally, one isolate sequenced here was assigned to ST366, a
665 ST that differs from ST11 by a single allele (i.e., *purE*). As such, all ST366 genomes available in
666 Enterobase were additionally downloaded (n = 10), and those with known isolation years and
667 country/continents of isolation (n = 6; three isolates from South Africa, two from Zambia, and one
668 from the United Kingdom) were used in subsequent steps. All 716 assembled *S.* Enteritidis
669 genomes underwent *in silico* plasmid replicon and AMR determinant detection using ABRicate v.
670 1.0.1 and the PlasmidFinder and NCBI AMR databases, respectively, as described above (see
671 section "*In silico* AMR determinant, plasmid replicon, and virulence factor detection" above).
672 Parsnp and HarvestTools v. 1.2 (90) were used to identify core SNPs among all 716 *S.*
673 Enteritidis genomes (703 publicly available genomes, plus the 13 sequenced here), using the closed
674 chromosome of ST11 *S.* Enteritidis str. OLF-SE10-10052 (NCBI Nucleotide Accession
675 NZ_CP009092.1) as a reference and Parsnp's implementation of PhiPack to remove
676 recombination (91). Clusters were identified within the resulting core SNP alignment using
677 RhierBAPs v. 1.1.3 (92, 93), R v. 4.0.0, and three clustering levels. IQ-TREE v. 1.5.4 (70) was
678 used to construct a ML phylogeny using (i) the resulting core SNPs as input; (ii) an ascertainment
679 bias correction (to account for the use of solely variant sites), corresponding to constant sites
680 estimated using the GC content of the reference chromosome (-fconst
681 1127671,1230753,1225740,1125726); (iii) the optimal nucleotide substitution model selected
682 using ModelFinder (71), based on its corresponding BIC value (i.e., the TVM+I model); (iv) 1,000
683 replicates of the UltraFast bootstrap approximation (72).
684 The resulting ML phylogeny was rooted and time-scaled using LSD2 v. 1.4.2.2 (94) and
685 the following parameters: (i) tip dates corresponding to the year of isolation associated with each

686 genome; (ii) a fixed substitution rate of 2.20×10^{-7} substitutions/site/year (i.e., the substitution rate
687 estimated in a previous study of *S.* Enteritidis) (96); (iii) constrained mode (-c), with the root
688 estimated using constraints on all branches (-r as); (iv) variances calculated using input branch
689 lengths (-v 1); (v) 1,000 samples for calculating confidence intervals for estimated dates (-f 1000);
690 (vi) a sequence length of 4,709,890 (i.e., the length of the reference chromosome; -s 4709890).
691 The resulting phylogeny was annotated using the bactaxR package in R (Figure 6). All
692 aforementioned *S.* Enteritidis SNP calling and phylogeny construction steps were repeated to
693 construct an additional time-scaled ML phylogeny using ST11 isolates within a major clade in the
694 African *S.* Enteritidis phylogeny (referred to here as "African *S.* Enteritidis ST11 Major Clade I",
695 n = 517; Supplemental Figure S5). RhierBAPs v. 1.1.3 (92, 93) and R v. 4.0.0 were additionally
696 used to identify clusters within the resulting core SNP alignment, using three clustering levels.
697 **Construction of time-scaled *S.* Typhimurium phylogenies.** To compare the 24 *S.* Typhimurium
698 isolates sequenced in this study to publicly available *S.* Typhimurium genomes, all genomes
699 meeting each of the following conditions were downloaded via Enterobase (accessed 27 December
700 2020, n = 319; Supplemental Table S6): (i) genomes were assigned to either ST19 (the ST to which
701 23 of the 24 *S.* Typhimurium isolates sequenced in this study were assigned/approximately
702 assigned) or ST34 (the ST of the remaining isolate, which differs from ST19 by a single allele,
703 *dnaN*) using the Achtman seven-gene MLST scheme for *Salmonella*; (ii) genomes had an exact
704 year of isolation reported in Enterobase's "Collection Year" field; (iii) genomes could be assigned
705 to a known country of isolation within the African continent, per Enterobase's "Country" and
706 "Continent" fields, respectively (88, 89). All 343 assembled *S.* Typhimurium genomes underwent
707 *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 (90) 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 (91). Clusters were identified within the resulting core SNP alignment using RhierBAPs v. 1.1.3 (92, 93), R v. 4.0.0, and three clustering levels. IQ-TREE v. 1.5.4 (70) 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 (71), based on its corresponding BIC value (i.e., the TVM+I model); (iv) 1,000 replicates of the UltraFast bootstrap approximation (72).

The resulting ML phylogeny was rooted and time-scaled using LSD2 v. 1.4.2.2 (94) and the following parameters: (i) tip dates corresponding to the year of isolation associated with each genome; (ii) a fixed substitution rate of 2.79×10^{-7} substitutions/site/year (i.e., the substitution rate estimated in a previous study of *S.* Typhimurium phage type DT104) (33); (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 (32) (*n* = 111; Supplemental Figure S6).

Data availability. 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 for the *Salmonella enterica* genomes sequenced in this study are available in Supplemental Table S1. Enterobase (https://enterobase.warwick.ac.uk/) metadata for the publicly available genomes used in this study are available in Supplemental Tables S3-S6.

ACKNOWLEDGMENTS

The following organizations and individuals are acknowledged for their contributions: (i) Gauteng Department of Agriculture and Rural Development (GDRAD) for funding this project; (ii) 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; (iii) the authors are grateful to the Agricultural Research Council: Onderstepoort Veterinary Research for providing all research facilities. We also acknowledge our collaborators; EMBL.

774 **REFERENCES**

775 1. World Health Organization (WHO). 2018. *Salmonella* (non-typhoidal).
776 https://www.who.int/news-room/fact-sheets/detail/salmonella-(non-typhoidal). Accessed
777 5 May 2021.

778 2. Hoelzer K, Moreno Switt AI, Wiedmann M. 2011. Animal contact as a source of human
779 non-typhoidal salmonellosis. Vet Res 42:34.

780 3. Hald T, Aspinall W, Devleesschauwer B, Cooke R, Corrigan T, Havelaar AH, Gibb HJ,
781 Torgerson PR, Kirk MD, Angulo FJ, Lake RJ, Speybroeck N, Hoffmann S. 2016. World
782 Health Organization Estimates of the Relative Contributions of Food to the Burden of
783 Disease Due to Selected Foodborne Hazards: A Structured Expert Elicitation. PLoS One
784 11:e0145839.

785 4. CDC. 2020. Serotypes and the Importance of Serotyping *Salmonella*, *on* CDC.
786 https://www.cdc.gov/salmonella/reportspubs/salmonella-atlas/serotyping-
787 importance.html. Accessed 25 May 2021.

788 5. Rodriguez-Rivera LD, Wright EM, Siler JD, Elton M, Cummings KJ, Warnick LD,
789 Wiedmann M. 2014. Subtype analysis of *Salmonella* isolated from subclinically infected
790 dairy cattle and dairy farm environments reveals the presence of both human- and
791 bovine-associated subtypes. Vet Microbiol 170:307-16.

792 6. Uzzau S, Brown DJ, Wallis T, Rubino S, Leori G, Bernard S, Casadesus J, Platt DJ,
793 Olsen JE. 2000. Host adapted serotypes of *Salmonella enterica*. Epidemiol Infect
794 125:229-55.

795 7. Harvey RR, Friedman CR, Crim SM, Judd M, Barrett KA, Tolar B, Folster JP, Griffin
796 PM, Brown AC. 2017. Epidemiology of *Salmonella enterica* Serotype Dublin Infections
797 among Humans, United States, 1968-2013. Emerg Infect Dis 23.

798 8. Taylor DN, Bied JM, Munro JS, Feldman RA. 1982. *Salmonella* dublin infections in the
799 United States, 1979-1980. J Infect Dis 146:322-7.

800 9. Mohammed M, Le Hello S, Leekitcharoenphon P, Hendriksen R. 2017. The invasome of
801 *Salmonella* Dublin as revealed by whole genome sequencing. BMC Infect Dis 17:544.

802 10. Cummings KJ, Warnick LD, Davis MA, Eckmann K, Grohn YT, Hoelzer K, MacDonald
803 K, Root TP, Siler JD, McGuire SM, Wiedmann M, Wright EM, Zansky SM, Besser TE.
804 2012. Farm animal contact as risk factor for transmission of bovine-associated
805 *Salmonella* subtypes. Emerg Infect Dis 18:1929-36.

806 11. Rabsch W, Andrews HL, Kingsley RA, Prager R, Tschape H, Adams LG, Baumler AJ.
807 2002. *Salmonella enterica* serotype Typhimurium and its host-adapted variants. Infect
808 Immun 70:2249-55.

809 12. Ferrari RG, Rosario DKA, Cunha-Neto A, Mano SB, Figueiredo EES, Conte-Junior CA.
810 2019. Worldwide Epidemiology of *Salmonella* Serovars in Animal-Based Foods: a Meta-
811 analysis. Appl Environ Microbiol 85.

812 13. European Food Safety Authority (EFSA), European Centre for Disease Prevention and
813 Control. 2018. The European Union summary report on trends and sources of zoonoses,
814 zoonotic agents and food-borne outbreaks in 2017. EFSA Journal 16:e05500.

815 14. World Health Organization (WHO). 2014. Antimicrobial resistance: global report on
816 surveillance. World Health Organization (WHO), Geneva.

817 15. Centers for Disease Control and Prevention (CDC). 2019. Antibiotic Resistance Threats
818 in the United States, 2019. U.S. Department of Health and Human Services, CDC,
819 Atlanta, GA.

820 16. Ibrahim GM, Morin PM. 2018. *Salmonella* Serotyping Using Whole Genome
821 Sequencing. Front Microbiol 9:2993.

822 17. Zhang S, den Bakker HC, Li S, Chen J, Dinsmore BA, Lane C, Lauer AC, Fields PI,
823 Deng X. 2019. SeqSero2: Rapid and Improved *Salmonella* Serotype Determination Using
824 Whole-Genome Sequencing Data. Appl Environ Microbiol 85.

825 18. Yoshida CE, Kruczkiewicz P, Laing CR, Lingohr EJ, Gannon VP, Nash JH, Taboada EN.
826 2016. The *Salmonella In Silico* Typing Resource (SISTR): An Open Web-Accessible
827 Tool for Rapidly Typing and Subtyping Draft *Salmonella* Genome Assemblies. PLoS
828 One 11:e0147101.

829 19. Carroll LM. 2019. High-Throughput Characterization of Foodborne Pathogens using
830 Next-Generation Sequencing. ProQuest LLC, Ann Arbor, MI.

831 20. Smith AM, Tau NP, Ngomane HM, Sekwadi P, Ramalwa N, Moodley K, Govind C,
832 Khan S, Archary M, Thomas J. 2020. Whole-genome sequencing to investigate two
833 concurrent outbreaks of *Salmonella* Enteritidis in South Africa, 2018. J Med Microbiol
834 69:1303-1307.

835 21. Rule R, Said M, Mbelle N, Osei Sekyere J. 2019. Genome sequence of a clinical
836 *Salmonella* Enteritidis sequence type 11 strain from South Africa. J Glob Antimicrob
837 Resist 19:164-166.

838 22. Smith AM, Smouse SL, Tau NP, Bamford C, Moodley VM, Jacobs C, McCarthy KM,
839 Lourens A, Keddy KH, Network G-SS. 2017. Laboratory-acquired infections of
840 *Salmonella enterica* serotype Typhi in South Africa: phenotypic and genotypic analysis
841 of isolates. BMC Infect Dis 17:656.

842 23. Wong VK, Baker S, Pickard DJ, Parkhill J, Page AJ, Feasey NA, Kingsley RA, Thomson
843 NR, Keane JA, Weill FX, Edwards DJ, Hawkey J, Harris SR, Mather AE, Cain AK,
844 Hadfield J, Hart PJ, Thieu NT, Klemm EJ, Glinos DA, Breiman RF, Watson CH, Kariuki
845 S, Gordon MA, Heyderman RS, Okoro C, Jacobs J, Lunguya O, Edmunds WJ, Msefula
846 C, Chabalgoity JA, Kama M, Jenkins K, Dutta S, Marks F, Campos J, Thompson C,
847 Obaro S, MacLennan CA, Dolecek C, Keddy KH, Smith AM, Parry CM, Karkey A,
848 Mulholland EK, Campbell JI, Dongol S, Basnyat B, Dufour M, Bandaranayake D, et al.
849 2015. Phylogeographical analysis of the dominant multidrug-resistant H58 clade of
850 *Salmonella* Typhi identifies inter- and intracontinental transmission events. Nat Genet
851 47:632-9.

852 24. Park SE, Pham DT, Boinett C, Wong VK, Pak GD, Panzner U, Espinoza LMC, von
853 Kalckreuth V, Im J, Schutt-Gerowitt H, Crump JA, Breiman RF, Adu-Sarkodie Y,
854 Owusu-Dabo E, Rakotozandrindrainy R, Soura AB, Aseffa A, Gasmelseed N, Keddy
855 KH, May J, Sow AG, Aaby P, Biggs HM, Hertz JT, Montgomery JM, Cosmas L, Olack
856 B, Fields B, Sarpong N, Razafindrabe TJL, Raminosoa TM, Kabore LP, Sampo E, Teferi
857 M, Yeshitela B, El Tayeb MA, Sooka A, Meyer CG, Krumkamp R, Dekker DM, Jaeger
858 A, Poppert S, Tall A, Niang A, Bjerregaard-Andersen M, Lofberg SV, Seo HJ, Jeon HJ,
859 Deerin JF, Park J, et al. 2018. The phylogeography and incidence of multi-drug resistant
860 typhoid fever in sub-Saharan Africa. Nat Commun 9:5094.

861 25. Feasey NA, Hadfield J, Keddy KH, Dallman TJ, Jacobs J, Deng X, Wigley P, Barquist L,
862 Langridge GC, Feltwell T, Harris SR, Mather AE, Fookes M, Aslett M, Msefula C,
863 Kariuki S, Maclennan CA, Onsare RS, Weill FX, Le Hello S, Smith AM, McClelland M,
864 Desai P, Parry CM, Cheesbrough J, French N, Campos J, Chabalgoity JA, Betancor L,
865 Hopkins KL, Nair S, Humphrey TJ, Lunguya O, Cogan TA, Tapia MD, Sow SO,
866 Tennant SM, Bornstein K, Levine MM, Lacharme-Lora L, Everett DB, Kingsley RA,
867 Parkhill J, Heyderman RS, Dougan G, Gordon MA, Thomson NR. 2016. Distinct
868 *Salmonella* Enteritidis lineages associated with enterocolitis in high-income settings and
869 invasive disease in low-income settings. Nat Genet 48:1211-1217.

870 26. El Zowalaty ME, Hickman RA, Mthembu TP, Zishiri OT, El Zowalaty AE, Jarhult JD.
871 2020. Genome sequences of two *Salmonella enterica* strains (MEZSAL74 and
872 MEZSAL81) harbouring multiple antimicrobial resistance genes isolated from livestock
873 in South Africa. J Glob Antimicrob Resist 21:396-398.

874 27. Mthembu TP, Zishiri OT, El Zowalaty ME. 2021. Genomic Characterization of
875 Antimicrobial Resistance in Food Chain and Livestock-Associated *Salmonella* Species.
876 Animals (Basel) 11.

877 28. Worley J, Meng J, Allard MW, Brown EW, Timme RE. 2018. *Salmonella enterica*
878 Phylogeny Based on Whole-Genome Sequencing Reveals Two New Clades and Novel
879 Patterns of Horizontally Acquired Genetic Elements. MBio 9.

880 29. Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, Tolstoy I, Tyson GH, Zhao S,
881 Hsu CH, McDermott PF, Tadesse DA, Morales C, Simmons M, Tillman G, Wasilenko J,
882 Folster JP, Klimke W. 2019. Validating the AMRFinder Tool and Resistance Gene

883 Database by Using Antimicrobial Resistance Genotype-Phenotype Correlations in a
884 Collection of Isolates. Antimicrob Agents Chemother 63.

885 30. Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, Villa L, Moller
886 Aarestrup F, Hasman H. 2014. *In silico* detection and typing of plasmids using
887 PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother
888 58:3895-903.

889 31. Fenske GJ, Thachil A, McDonough PL, Glaser A, Scaria J. 2019. Geography Shapes the
890 Population Genomics of *Salmonella enterica* Dublin. Genome Biol Evol 11:2220-2231.

891 32. Carroll LM, Huisman JS, Wiedmann M. 2020. Twentieth-century emergence of
892 antimicrobial resistant human- and bovine-associated *Salmonella enterica* serotype
893 Typhimurium lineages in New York State. Sci Rep 10:14428.

894 33. Leekitcharoenphon P, Hendriksen RS, Le Hello S, Weill FX, Baggesen DL, Jun SR,
895 Ussery DW, Lund O, Crook DW, Wilson DJ, Aarestrup FM. 2016. Global Genomic
896 Epidemiology of *Salmonella enterica* Serovar Typhimurium DT104. Appl Environ
897 Microbiol 82:2516-26.

898 34. Strachan NJC, Rotariu O, Lopes B, MacRae M, Fairley S, Laing C, Gannon V, Allison
899 LJ, Hanson MF, Dallman T, Ashton P, Franz E, van Hoek AHAM, French NP, George T,
900 Biggs PJ, Forbes KJ. 2015. Whole Genome Sequencing demonstrates that Geographic
901 Variation of *Escherichia coli* O157 Genotypes Dominates Host Association. Scientific
902 Reports 5:14145.

903 35. Achtman M. 2008. Evolution, population structure, and phylogeography of genetically
904 monomorphic bacterial pathogens. Annu Rev Microbiol 62:53-70.

905 36. Palma F, Manfreda G, Silva M, Parisi A, Barker DOR, Taboada EN, Pasquali F, Rossi
906 M. 2018. Genome-wide identification of geographical segregated genetic markers in
907 *Salmonella enterica* serovar Typhimurium variant 4,[5],12:i. Scientific Reports 8:15251.

908 37. Carroll LM, Wiedmann M, den Bakker H, Siler J, Warchocki S, Kent D, Lyalina S, Davis
909 M, Sischo W, Besser T, Warnick LD, Pereira RV. 2017. Whole-Genome Sequencing of
910 Drug-Resistant *Salmonella enterica* Isolates from Dairy Cattle and Humans in New York
911 and Washington States Reveals Source and Geographic Associations. Appl Environ
912 Microbiol 83.

913 38. Liao J, Orsi RH, Carroll LM, Wiedmann M. 2020. Comparative genomics reveals
914 different population structures associated with host and geographic origin in
915 antimicrobial-resistant *Salmonella enterica*. Environ Microbiol doi:10.1111/1462-
916 2920.15014.

917 39. Mughini-Gras L, Pijnacker R, Duijster J, Heck M, Wit B, Veldman K, Franz E. 2020.
918 Changing epidemiology of invasive non-typhoid *Salmonella* infection: a nationwide
919 population-based registry study. Clin Microbiol Infect 26:941 e9-941 e14.

920 40. Ao TT, Feasey NA, Gordon MA, Keddy KH, Angulo FJ, Crump JA. 2015. Global
921 burden of invasive nontyphoidal *Salmonella* disease, 2010(1). Emerg Infect Dis 21.

922 41. GERMS-SA. 2019. GERMS-SA Annual Report 2019.
923 http://www.nicd.ac.za/index.php/publications/germs-annual-reports/. Accessed 29 April
924 2021.

925 42. Gelaw AK, Nthaba P, Matle I. 2018. Detection of *Salmonella* from animal sources in
926 South Africa between 2007 and 2014. J S Afr Vet Assoc 89:e1-e10.

927 43. Hirsch W, Gerichter CB, Bregman E, Lubling P, Altman G. 1954. A New *Salmonella*
928 Type (*Salmonella* hadar). Acta Medica Orientalia 13:41.

929 44. Rowe B, Hall ML, Ward LR, de Sa JD. 1980. Epidemic spread of *Salmonella* hadar in
930 England and Wales. Br Med J 280:1065-6.

931 45. O'Brien SJ. 2013. The "decline and fall" of nontyphoidal salmonella in the United
932 kingdom. Clin Infect Dis 56:705-10.

933 46. Barrow PA, Jones MA, Smith AL, Wigley P. 2012. The long view: *Salmonella*--the last
934 forty years. Avian Pathol 41:413-20.

935 47. Magwedere K, Rauff D, De Klerk G, Keddy KH, Dziva F. 2015. Incidence of
936 Nontyphoidal *Salmonella* in Food-Producing Animals, Animal Feed, and the Associated
937 Environment in South Africa, 2012-2014. Clin Infect Dis 61 Suppl 4:S283-9.

938 48. Kidanemariam A, Engelbrecht M, Picard J. 2010. Retrospective study on the incidence of
939 *Salmonella* isolations in animals in South Africa, 1996 to 2006. J S Afr Vet Assoc 81:37-
940 44.

941 49. Llarena AK, Zhang J, Vehkala M, Valimaki N, Hakkinen M, Hanninen ML, Roasto M,
942 Maesaar M, Taboada E, Barker D, Garofolo G, Camma C, Di Giannatale E, Corander J,
943 Rossi M. 2016. Monomorphic genotypes within a generalist lineage of *Campylobacter*
944 *jejuni* show signs of global dispersion. Microb Genom 2:e000088.

945 50. The HC, Thanh DP, Holt KE, Thomson NR, Baker S. 2016. The genomic signatures of
946 *Shigella* evolution, adaptation and geographical spread. Nat Rev Microbiol 14:235-50.

947 51. USDA Foreign Agricultural Service. 2020. United States Poultry Exports to South Africa
948 Continue Steady Increase, : Report Number SF2020-0024, *on* USDA Foreign

949 Agricultural Service. https://www.fas.usda.gov/data/south-africa-united-states-poultry-
950 exports-south-africa-continue-steady-increase. Accessed 29 April 2021.

951 52. Threlfall EJ. 2000. Epidemic *salmonella* typhimurium DT 104--a truly international
952 multiresistant clone. J Antimicrob Chemother 46:7-10.

953 53. Brown E, Dessai U, McGarry S, Gerner-Smidt P. 2019. Use of Whole-Genome
954 Sequencing for Food Safety and Public Health in the United States. Foodborne Pathog
955 Dis 16:441-450.

956 54. den Bakker HC, Allard MW, Bopp D, Brown EW, Fontana J, Iqbal Z, Kinney A,
957 Limberger R, Musser KA, Shudt M, Strain E, Wiedmann M, Wolfgang WJ. 2014. Rapid
958 whole-genome sequencing for surveillance of *Salmonella enterica* serovar enteritidis.
959 Emerg Infect Dis 20:1306-14.

960 55. Inns T, Ashton PM, Herrera-Leon S, Lighthill J, Foulkes S, Jombart T, Rehman Y, Fox
961 A, Dallman T, E DEP, Browning L, Coia JE, Edeghere O, Vivancos R. 2017. Prospective
962 use of whole genome sequencing (WGS) detected a multi-country outbreak of
963 *Salmonella* Enteritidis. Epidemiol Infect 145:289-298.

964 56. Majowicz SE, Musto J, Scallan E, Angulo FJ, Kirk M, O'Brien SJ, Jones TF, Fazil A,
965 Hoekstra RM, International Collaboration on Enteric Disease 'Burden of Illness S. 2010.
966 The global burden of nontyphoidal *Salmonella* gastroenteritis. Clin Infect Dis 50:882-9.

967 57. Kirk MD, Pires SM, Black RE, Caipo M, Crump JA, Devleesschauwer B, Dopfer D,
968 Fazil A, Fischer-Walker CL, Hald T, Hall AJ, Keddy KH, Lake RJ, Lanata CF,
969 Torgerson PR, Havelaar AH, Angulo FJ. 2015. World Health Organization Estimates of

970 the Global and Regional Disease Burden of 22 Foodborne Bacterial, Protozoal, and Viral
971 Diseases, 2010: A Data Synthesis. PLoS Med 12:e1001921.

972 58. Inzaule SC, Tessema SK, Kebede Y, Ogwell Ouma AE, Nkengasong JN. 2021.
973 Genomic-informed pathogen surveillance in Africa: opportunities and challenges. Lancet
974 Infect Dis doi:10.1016/S1473-3099(20)30939-7.

975 59. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM,
976 Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G,
977 Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its
978 applications to single-cell sequencing. J Comput Biol 19:455-77.

979 60. Andrews S. 2019. FastQC: a quality control tool for high throughput sequence data,
980 v0.11.8. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

981 61. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for
982 genome assemblies. Bioinformatics 29:1072-5.

983 62. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM:
984 assessing the quality of microbial genomes recovered from isolates, single cells, and
985 metagenomes. Genome Res 25:1043-55.

986 63. Ewels P, Magnusson M, Lundin S, Kaller M. 2016. MultiQC: summarize analysis results
987 for multiple tools and samples in a single report. Bioinformatics 32:3047-8.

988 64. Jolley KA, Maiden MC. 2010. BIGSdb: Scalable analysis of bacterial genome variation
989 at the population level. BMC Bioinformatics 11:595.

990 65. Jolley KA, Bray JE, Maiden MCJ. 2018. Open-access bacterial population genomics:
991 BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res
992 3:124.

993 66. Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database C.
994 2011. The sequence read archive. Nucleic Acids Res 39:D19-21.

995 67. Kodama Y, Shumway M, Leinonen R, International Nucleotide Sequence Database C.
996 2012. The Sequence Read Archive: explosive growth of sequencing data. Nucleic Acids
997 Res 40:D54-6.

998 68. Gardner SN, Hall BG. 2013. When whole-genome alignments just won't work: kSNP v2
999 software for alignment-free SNP discovery and phylogenetics of hundreds of microbial
1000 genomes. PLoS One 8:e81760.

1001 69. Gardner SN, Slezak T, Hall BG. 2015. kSNP3.0: SNP detection and phylogenetic
1002 analysis of genomes without genome alignment or reference genome. Bioinformatics
1003 31:2877-8.

1004 70. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and
1005 effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol
1006 Evol 32:268-74.

1007 71. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017.
1008 ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods
1009 14:587-589.

1010 72. Minh BQ, Nguyen MA, von Haeseler A. 2013. Ultrafast approximation for phylogenetic
1011 bootstrap. Mol Biol Evol 30:1188-95.

1012 73. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2:
1013 Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol 35:518-522.

1014 74. R Core Team. 2019. R: A Language and Environment for Statistical Computing, v3.6.1.
1015 R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

1016 75. Carroll LM, Wiedmann M, Kovac J. 2020. Proposal of a Taxonomic Nomenclature for
1017 the *Bacillus cereus* Group Which Reconciles Genomic Definitions of Bacterial Species
1018 with Clinical and Industrial Phenotypes. mBio 11.

1019 76. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017. ggtree: an r package for
1020 visualization and annotation of phylogenetic trees with their covariates and other
1021 associated data. Methods in Ecology and Evolution 8:28-36.

1022 77. Yu G, Lam TT, Zhu H, Guan Y. 2018. Two Methods for Mapping and Visualizing
1023 Associated Data on Phylogeny Using Ggtree. Mol Biol Evol 35:3041-3043.

1024 78. Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and
1025 evolutionary analyses in R. Bioinformatics 35:526-528.

1026 79. Wickham H, François R, Henry L, Müller K. 2020. dplyr: A Grammar of Data
1027 Manipulation, v0.8.5. https://CRAN.R-project.org/package=dplyr.

1028 80. R Hackathon. 2019. phylobase: Base Package for Phylogenetic Structures and
1029 Comparative Data, v0.8.6. https://CRAN.R-project.org/package=phylobase.

1030 81. Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other
1031 things). Methods in Ecology and Evolution 3:217-223.

1032 82. Wickham H. 2007. Reshaping Data with the reshape Package. 2007 21:20.

1033 83. Hunt M, Mather AE, Sanchez-Buso L, Page AJ, Parkhill J, Keane JA, Harris SR. 2017.
1034 ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads.
1035 Microb Genom 3:e000131.

1036 84. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics
1037 30:2068-9.

1038 85. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup
1039 FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. J
1040 Antimicrob Chemother 67:2640-4.

1041 86. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q. 2005. VFDB: a reference database
1042 for bacterial virulence factors. Nucleic Acids Res 33:D325-8.

1043 87. Liu B, Zheng D, Jin Q, Chen L, Yang J. 2019. VFDB 2019: a comparative pathogenomic
1044 platform with an interactive web interface. Nucleic Acids Res 47:D687-D692.

1045 88. Zhou Z, Alikhan NF, Mohamed K, Fan Y, Agama Study G, Achtman M. 2020. The
1046 EnteroBase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis*
1047 phylogeny, and *Escherichia* core genomic diversity. Genome Res 30:138-152.

1048 89. Alikhan NF, Zhou Z, Sergeant MJ, Achtman M. 2018. A genomic overview of the
1049 population structure of *Salmonella*. PLoS Genet 14:e1007261.

1050 90. Treangen TJ, Ondov BD, Koren S, Phillippy AM. 2014. The Harvest suite for rapid core-
1051 genome alignment and visualization of thousands of intraspecific microbial genomes.
1052 Genome Biol 15:524.

1053 91. Bruen TC, Philippe H, Bryant D. 2006. A simple and robust statistical test for detecting
1054 the presence of recombination. Genetics 172:2665-81.

92. Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. 2018. RhierBAPS: An R implementation of the population clustering algorithm hierBAPS. Wellcome Open Res 3:93.

93. Cheng L, Connor TR, Siren J, Aanensen DM, Corander J. 2013. Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. Mol Biol Evol 30:1224-8.

94. To T-H, Jung M, Lycett S, Gascuel O. 2015. Fast Dating Using Least-Squares Criteria and Algorithms. Systematic Biology 65:82-97.

95. Kimura M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. Proc Natl Acad Sci U S A 78:454-8.

96. Deng X, Desai PT, den Bakker HC, Mikoleit M, Tolar B, Trees E, Hendriksen RS, Frye JG, Porwollik S, Weimer BC, Wiedmann M, Weinstock GM, Fields PI, McClelland M. 2014. Genomic epidemiology of *Salmonella enterica* serotype Enteritidis based on population structure of prevalent lineages. Emerg Infect Dis 20:1481-9.

FIGURE LEGENDS

Figure 1. Maximum likelihood phylogeny constructed using core SNPs identified among 505 *Salmonella* isolate genomes. Publicly available genomes are denoted by black tip labels (n = 442), while genomes of strains isolated in conjunction with this study are denoted by tip labels colored by serotype (n = 63). The phylogeny is rooted at the midpoint with branch lengths reported in substitutions per site. Core SNPs were identified among all genomes using kSNP3, while the phylogeny was constructed and annotated using IQ-TREE and FigTree v. 1.4.4, respectively.

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

(Supplemental Table S2). The phylogeny was constructed and annotated using IQ-TREE and bactaxR/ggtree, respectively. DUBN, *S.* Dublin; ENTR, *S.* Enteritidis; HADR, *S.* Hadar; TYPH, *S.* Typhimurium.

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.

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

1123 African isolates, which encompasses 17 of the 18 *S.* Dublin isolates sequenced in this study, plus
1124 seven publicly available South African isolates. The heatmap to the right of the phylogeny
1125 denotes: (i) whether an isolate was sequenced in conjunction with this study (dark pink) or not
1126 (gray; "Study"); the presence and absence of (ii) plasmid replicons (blue) and (iii) antimicrobial
1127 resistance (AMR) determinants (orange). The phylogeny was rooted and time-scaled using
1128 LSD2, with branch lengths reported in years (X-axis). Core SNPs were identified among all
1129 genomes using Parsnp. AMR determinants were identified using ABRicate, the NCBI AMR
1130 determinant database, and minimum identity and coverage thresholds of 75 and 50%,
1131 respectively. Plasmid replicons were identified using ABRicate and the PlasmidFinder database,
1132 using minimum identity and coverage thresholds of 80 and 60%, respectively. The phylogeny
1133 was constructed and annotated using IQ-TREE and bactaxR/ggtree, respectively.
1134
1135 **Figure 5.** Maximum likelihood phylogeny constructed using core SNPs identified among 1,569
1136 *S.* Hadar genomes (1,561 publicly available genomes, plus eight sequenced here). Tip label
1137 colors denote the continent from which each strain was reported to have been isolated. The
1138 heatmap to the right of the phylogeny denotes: (i) whether an isolate was sequenced in
1139 conjunction with this study (dark pink) or not (gray; "Study"); (ii) level 1 cluster assignments
1140 obtained using RhierBAPS ("RhierBAPS"); the presence and absence of (iii) plasmid replicons
1141 (blue) and (iv) antimicrobial resistance (AMR) determinants (orange). The phylogeny was rooted
1142 and time scaled using LSD2, with branch lengths reported in years (X-axis). Core SNPs were
1143 identified among all genomes using Parsnp. AMR determinants were identified using ABRicate,
1144 the NCBI AMR determinant database, and minimum identity and coverage thresholds of 75 and
1145 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.

Figure 6. Maximum likelihood phylogeny constructed using core SNPs identified among 716 African *S.* Enteritidis genomes (703 publicly available genomes, plus 13 sequenced here). Tip label colors denote the region/country from which each strain was reported to have been isolated (based on African regions as defined by the African Union, 25 April 2021). Clade labels shown to the right of the phylogeny tip labels denote major clades discussed in the main text. 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.

Figure 7. Maximum likelihood phylogeny constructed using core SNPs identified among 343 African *S.* Typhimurium genomes (319 publicly available genomes, plus the 24 sequenced here). Tip label colors denote the region/country from which each strain was reported to have been isolated (based on African regions as defined by the African Union, 25 April 2021). Clade labels

1169 denote clades discussed in either the main manuscript or the Supplemental Text. The heatmap to
1170 the right of the phylogeny denotes: (i) whether an isolate was sequenced in conjunction with this
1171 study (dark pink) or not (gray; "Study"); (ii) level 1 cluster assignments obtained using
1172 RhierBAPS ("RhierBAPS"); the presence and absence of (iii) plasmid replicons (blue) and (iv)
1173 antimicrobial resistance (AMR) determinants (orange). The phylogeny was rooted and time-
1174 scaled using LSD2, with branch lengths reported in years (X-axis). Core SNPs were identified
1175 among all genomes using Parsnp. AMR determinants were identified using ABRicate, the NCBI
1176 AMR determinant database, and minimum identity and coverage thresholds of 75 and 50%,
1177 respectively. Plasmid replicons were identified using ABRicate and the PlasmidFinder database,
1178 using minimum identity and coverage thresholds of 80 and 60%, respectively. The phylogeny
1179 was constructed and annotated using IQ-TREE and bactaxR/ggtree, respectively.
1180

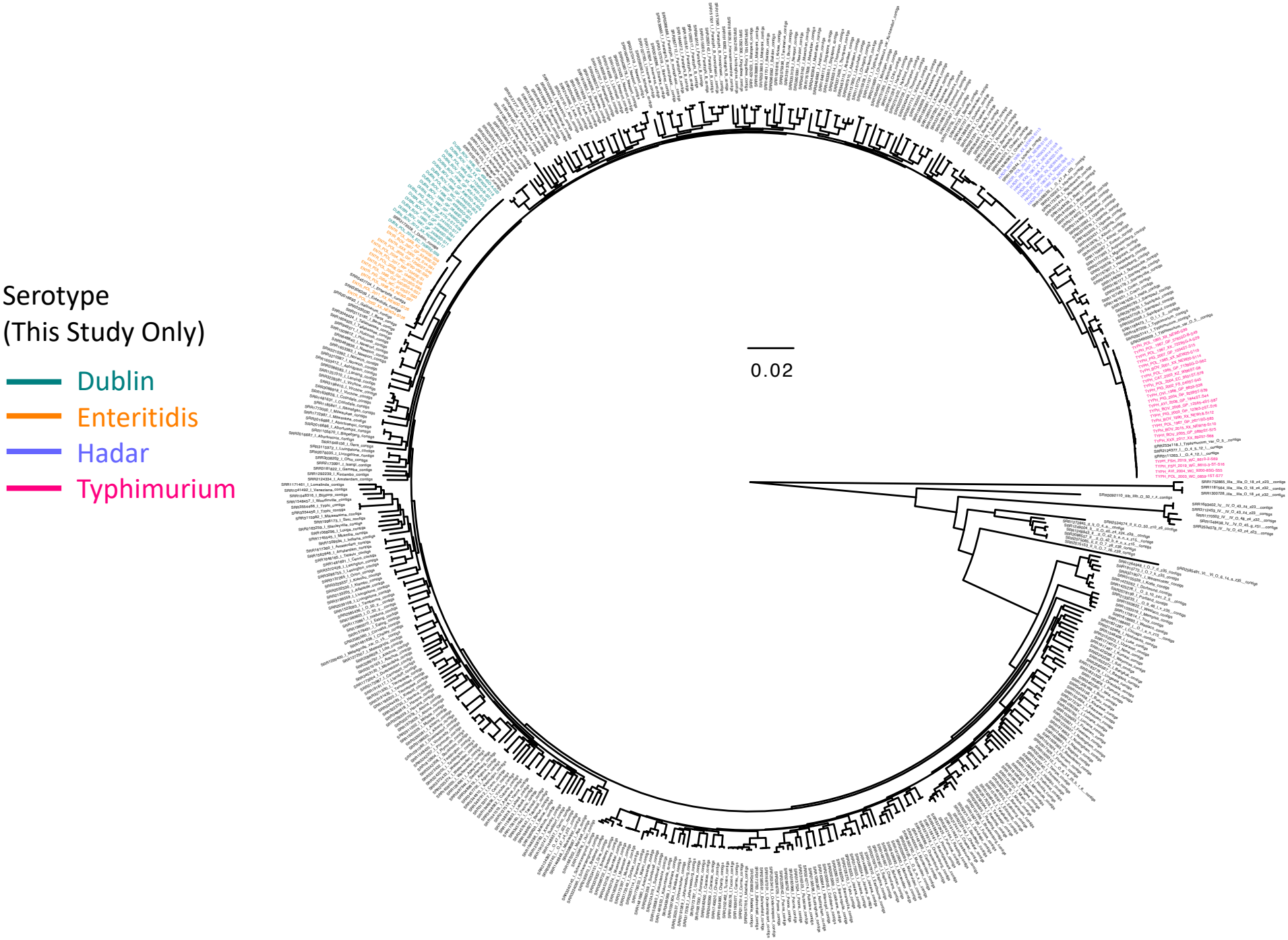

Figure 1. Maximum likelihood phylogeny constructed using core SNPs identified among 505 *Salmonella* isolate genomes. Publicly available genomes are denoted by black tip labels (n = 442), while genomes of strains isolated in conjunction with this study are denoted by tip labels colored by serotype (n = 63). The phylogeny is rooted at the midpoint with branch lengths reported in substitutions per site. Core SNPs were identified among all genomes using kSNP3, while the phylogeny was constructed and annotated using IQ-TREE and FigTree v. 1.4.4, respectively.

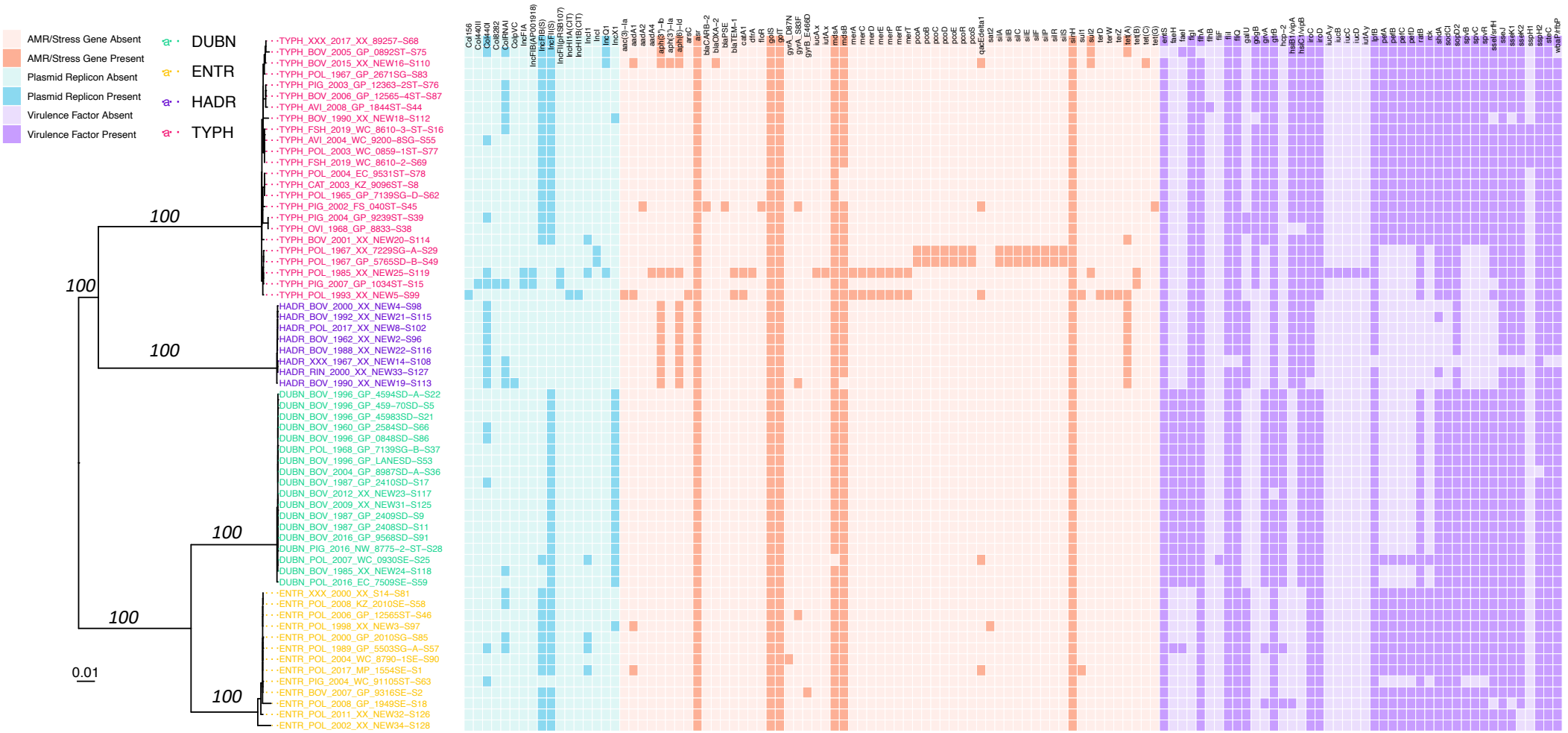

Figure 2. Maximum likelihood phylogeny constructed using core SNPs identified among the genomes of 63 *Salmonella* strains isolated in conjunction with this study. Tip label colors denote isolate serotypes, and branch labels denote ultrafast bootstrap support percentages out of 1,000 replicates (selected 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 (Supplemental Table S2). The phylogeny was constructed and annotated using IQ-TREE and bactaxR/ggtree, respectively. DUBN, *S.* Dublin; ENTR, *S.* Enteritidis; HADR, *S.* Hadar; TYPH, *S.* Typhimurium.

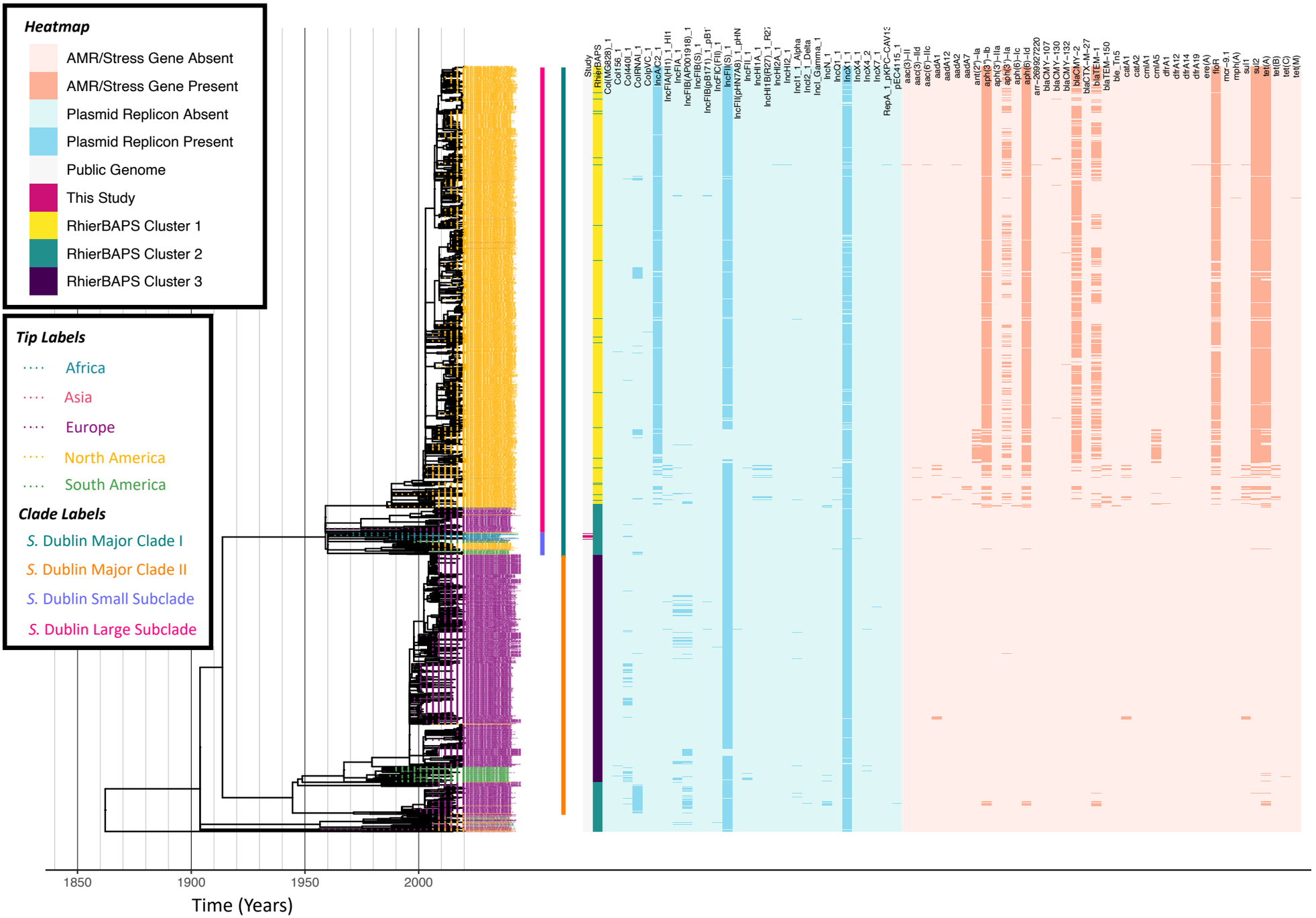

Figure 3. Maximum likelihood phylogeny constructed using core SNPs identified among 2,802 *S.* Dublin genomes (2,784 publicly available genomes, plus 18 sequenced here). Tip label colors denote the continent from which each strain was reported to have been isolated. Clade labels denote major clades assigned 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.

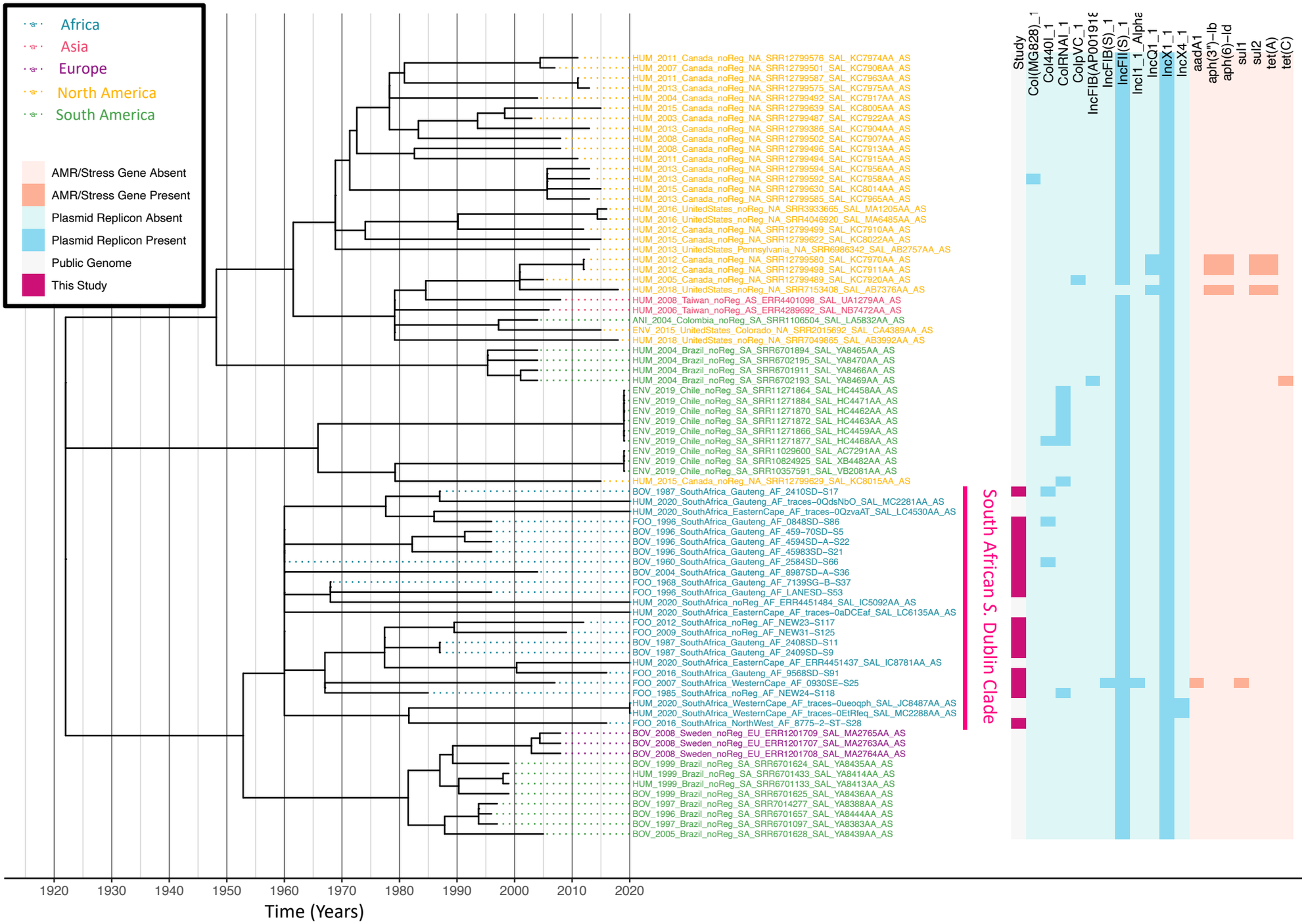

Figure 4. Maximum likelihood phylogeny constructed using core SNPs identified among 78 *S.* Dublin genomes within the *S.* Dublin Small Subclade (61 publicly available genomes, plus 17 sequenced here). Tip label colors denote the continent from which each strain was reported to have been isolated. A 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 bactaxR/ggtree, respectively.

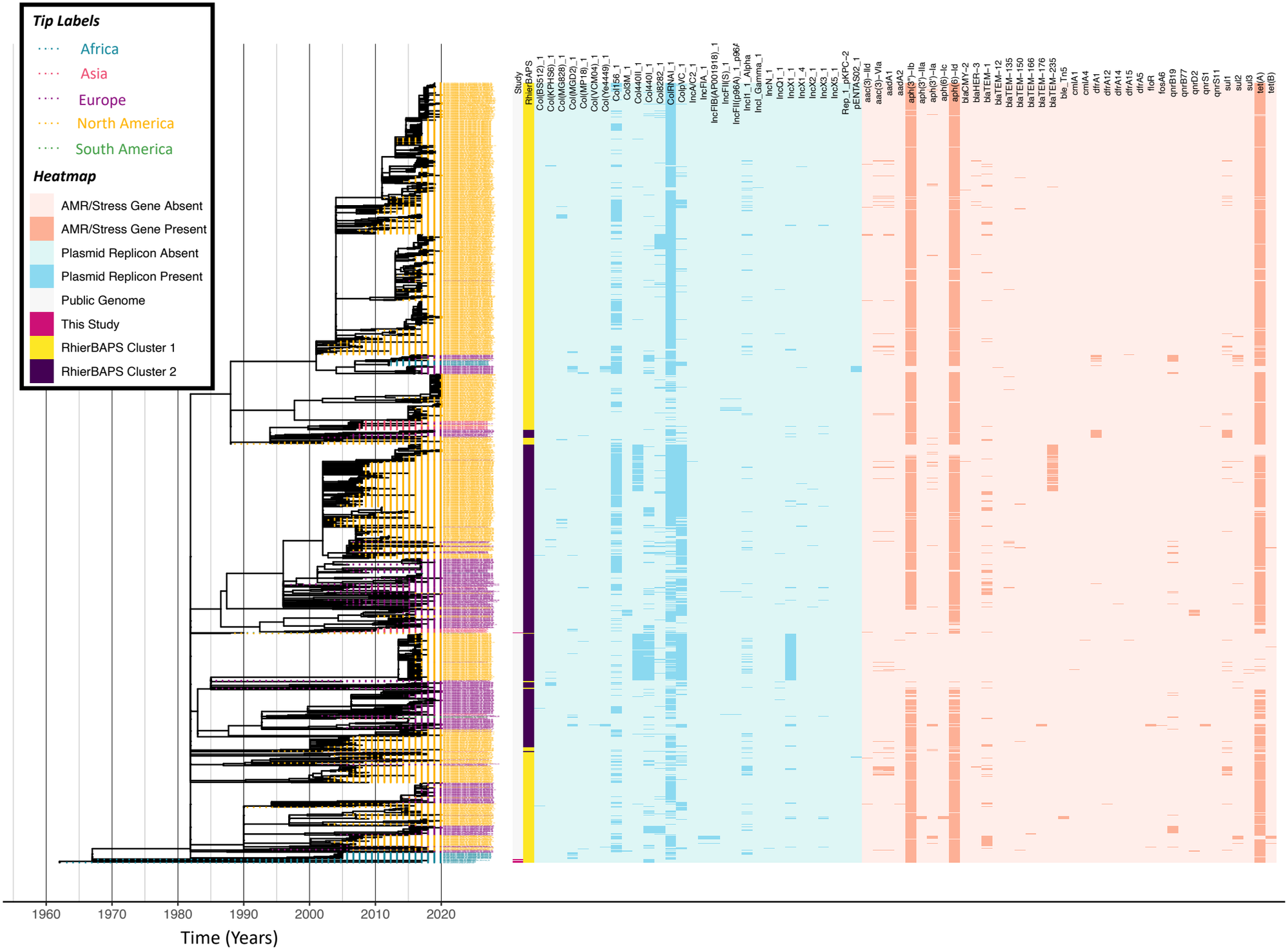

Figure 5. Maximum likelihood phylogeny constructed using core SNPs identified among 1,569 *S.* Hadar genomes (1,561 publicly available genomes, plus eight sequenced here). Tip label colors denote the continent from which each strain was reported to have been isolated. The heatmap to the right of the 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.

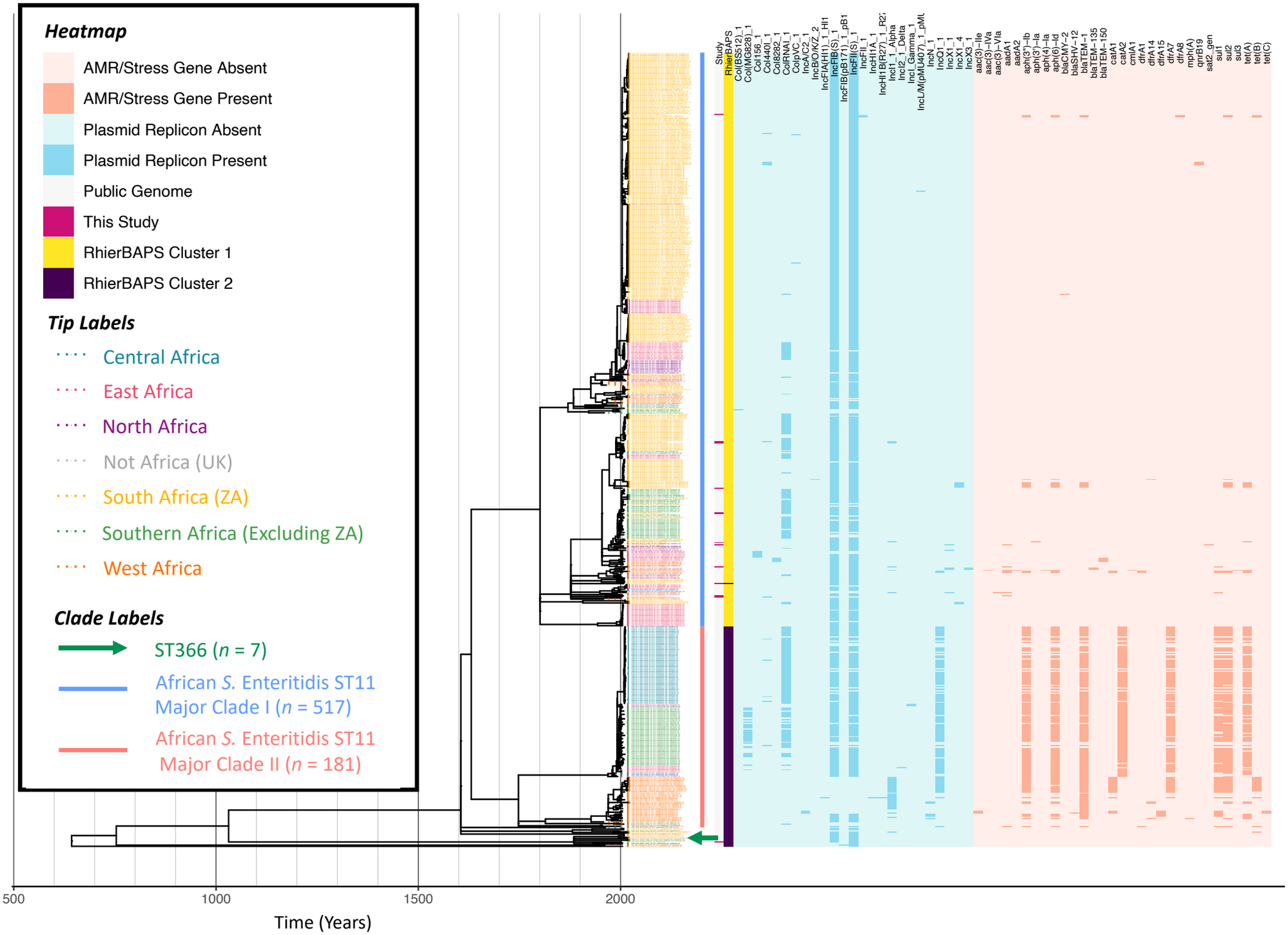

Figure 6. Maximum likelihood phylogeny constructed using core SNPs identified among 716 African *S.* Enteritidis genomes (703 publicly available genomes, plus 13 sequenced here). Tip label colors denote the region/country from which each strain was reported to have been isolated (based on African regions as defined by the African Union, 25 April 2021). Clade labels shown to the right of the phylogeny tip labels denote major clades discussed in the main text. 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.

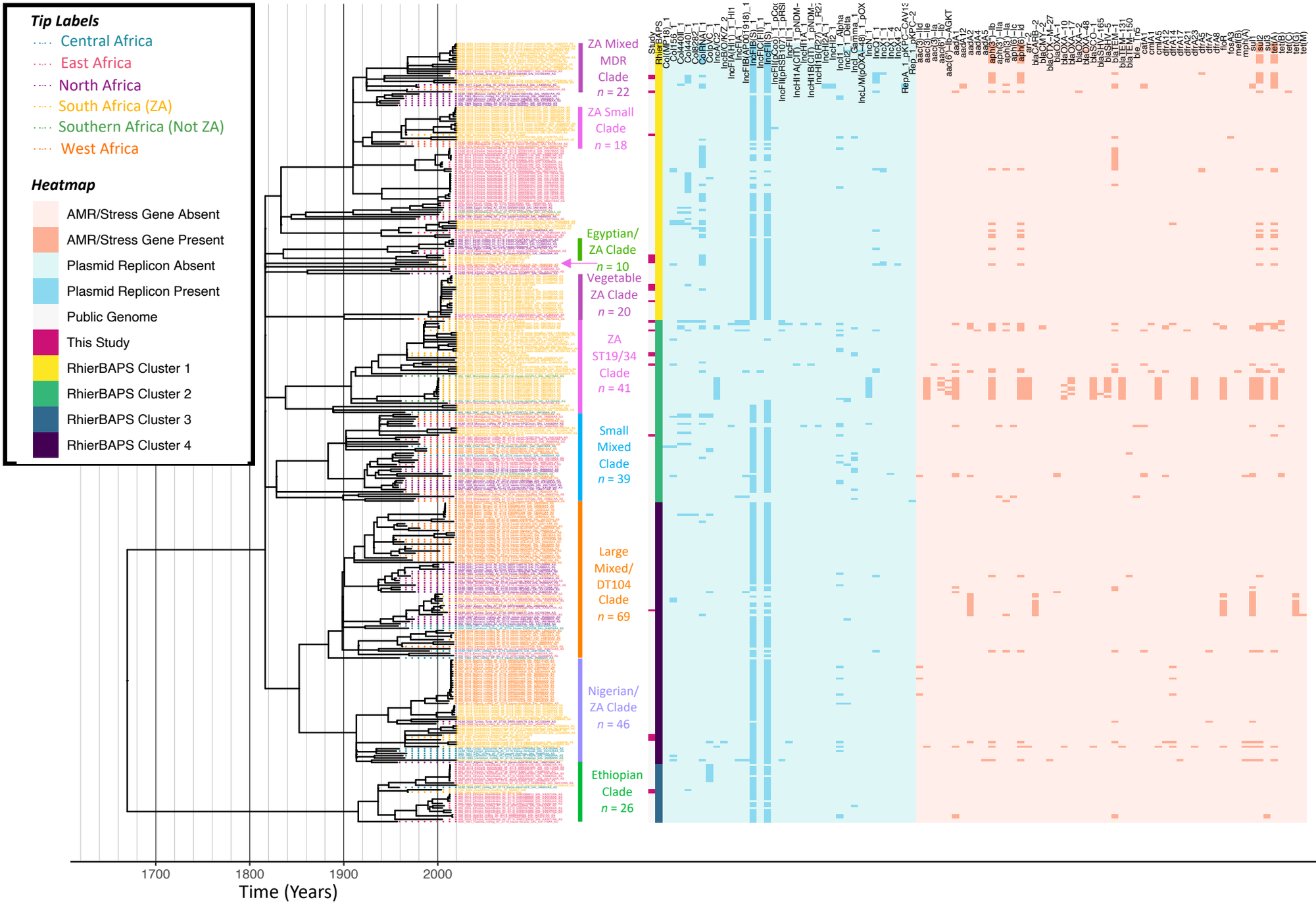

Figure 7. Maximum likelihood phylogeny constructed using core SNPs identified among 343 African *S.* Typhimurium genomes (319 publicly available genomes, plus the 24 sequenced here). Tip label colors denote the region/country from which each strain was reported to have been isolated (based on African regions as defined by the African Union, 25 April 2021). Clade labels denote clades discussed in either the main manuscript or the Supplemental Text. 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.