Open access • Posted Content • DOI:10.1101/2020.10.26.356493

High-resolution genomic comparisons within Salmonella enterica serotypes derived from beef feedlot cattle: parsing the roles of cattle source, pen, animal, sample type and production period — Source link

Gizem Levent, Ashlynn Schlochtermeier, Samuel E. Ives, Keri N. Norman ...+6 more authors

Institutions: Texas A&M University, West Texas A&M University, Texas Tech University, United States Department of Agriculture ...+1 more institutions

Published on: 27 Oct 2020 - bioRxiv (Cold Spring Harbor Laboratory)

Topics: Salmonella, Beef cattle, Salmonella enterica and Population

High-resolution genomic comparisons within *Salmonella enterica* serotypes derived from beef feedlot cattle: parsing the roles of cattle source, pen, animal, sample type and production period

Gizem Levent[a], Ashlynn Schlochtermeier[b], Samuel E. Ives[b], Keri N. Norman[c], Sara D. Lawhon[a], Guy H. Loneragan[d], Robin C. Anderson[e], Javier Vinasco[a], Henk C. den Bakker[f], H. Morgan Scott[a#]

[a]Department of Veterinary Pathobiology, Texas A&M University, College Station, Texas, USA; [b]Department of Agricultural Sciences, West Texas A&M University, Canyon, Texas, USA; [c]Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, Texas, USA; [d]School of Veterinary Medicine, Texas Tech University, Amarillo, Texas, USA; [e]United States Department of Agriculture, Agricultural Research Service, College Station, Texas, USA; [f]Center for Food Safety, University of Georgia, Griffin, Georgia, USA.

#Address correspondence to H. M. Scott, Department of Veterinary Pathobiology, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, Texas, USA 77843-4467. E-mail: hmscott@cvm.tamu.edu

Short title: Genomic comparisons of cattle-origin *Salmonella*

Keywords: *Salmonella enterica*, beef cattle, population structure, feedlot, genomic comparisons, core-genome MLST, phylogenetic analysis

Abstract

Salmonella enterica is a major foodborne pathogen, and contaminated beef products have been identified as the primary source of *Salmonella*-related outbreaks. Pathogenicity and antibiotic resistance of *Salmonella* are highly serotype- and subpopulation-specific, which makes it essential to understand high-resolution *Salmonella* population dynamics in cattle. Time of year, source of cattle, pen, and sample type(i.e., feces, hide or lymph nodes) have previously been identified as important factors influencing the serotype distribution of *Salmonella* (e.g., Anatum, Lubbock, Cerro, Montevideo, Kentucky, Newport, and Norwich) that were isolated from a longitudinal sampling design in a research feedlot. In this study, we performed high-resolution genomic comparisons of *Salmonella* isolates within each serotype using both single-nucleotide polymorphism (SNP)-based maximum likelihood phylogeny and hierarchical clustering of core-genome multi-locus sequence typing. The importance of the aforementioned features on clonal *Salmonella* expansion was further explored using a supervised machine learning algorithm. In addition, we identified and compared the resistance genes, plasmids, and pathogenicity island profiles of the isolates within each sub-population. Our findings indicate that clonal expansion of *Salmonella* strains in cattle was mainly influenced by the randomization of block and pen, as well as the origin/source of the cattle; that is, regardless of sampling time and sample type (i.e., feces, lymph node or hide). Further research is needed concerning the role of the feedlot pen environment prior to cattle placement to better understand carry-over contributions of existing strains of *Salmonella* and their bacteriophages.

Importance

Salmonella serotypes isolated from outbreaks in humans can also be found in beef cattle and feedlots. Virulence factors and antibiotic resistance are among the primary defense mechanisms of *Salmonella*, and are often associated with clonal expansion. This makes understanding the subpopulation dynamics of *Salmonella* in cattle critical for effective mitigation. There remains a gap in the literature concerning subpopulation dynamics within *Salmonella* serotypes in feedlot cattle from the beginning of feeding up until slaughter. Here, we explore *Salmonella* population dynamics within each serotype using core genome phylogeny and hierarchical classifications. We used machine-learning to quantitatively parse the relative importance of both hierarchical and longitudinal clustering among cattle host samples. Our results reveal that *Salmonella* populations in cattle are highly clonal over a 6-month study period, and that clonal dissemination of *Salmonella* in cattle is mainly influenced spatially by experimental block and pen, as well by the geographical origin of the cattle.

Introduction

Globally, *Salmonella enterica* subsp. *enterica* is among the leading foodborne bacterial pathogens and is responsible for negative effects both on public health and on the economy (1). While there are over 2,500 identified *Salmonella* serotypes, only a minor set (~100) of these serotypes are commonly identified as the source of human *Salmonella* infections and outbreaks (2). Each year, approximately 93.8 million human cases occur worldwide due to *Salmonella*, resulting in 155,000 deaths (3). Human infections caused by pathogenic *Salmonella* strains often result in mild gastroenteritis and do not require antibiotic treatment. At-risk individuals, such as infants, the elderly, or immunocompromised patients, may require the use of antibiotics to counter the pathogen's invasive and adverse health effects. Over the last decade, high levels of hospitalization and increased mortality rates associated with multidrug-resistant (MDR) *Salmonella* serotypes (4, 5) have, indeed, become a serious public health concern (6, 7).

The majority of *Salmonella* outbreaks are related to the consumption of contaminated food products (3). *Salmonella*-contaminated beef products are a major culprit in foodborne S*almonella* outbreaks (8). *Salmonella* contamination of carcass surfaces can occur during slaughter evisceration and de-hiding processes due to direct contact with cattle feces and aerosolization of fecal material from hides, respectively. Lymph nodes harboring *Salmonella* embedded in fat tissue can also contaminate final beef products (e.g., ground beef) as fat trim (9).

Since the gastrointestinal tract of cattle is known to be a natural niche for a number of *Salmonella* serotypes, those that can cause human infections through the ingestion of contaminated beef products are commonly found not only in cattle but also in their feedlot pen environments (10-15).*Salmonella* serotypes Typhimurium, Newport, Enteritidis, Montevideo, Anatum, Cerro, Kentucky, and Mbandaka are commonly found in cattle, beef products, and in cattle production environments (8, 11, 14-18). Among these, Typhimurium and Newport are the serotypes most commonly derived from human cases of *Salmonella*, followed next by Anatum and Montevideo (16, 19, 20). Cattle-origin *S*. Typhimurium and Newport are also serotypes commonly associated with a phenotypic MDR profile (8, 11, 14-17).

In an effort to minimize public-health risks of *Salmonella* originating from cattle, multiple teams of researchers have previously explored serotype-level dynamics of *Salmonella* in healthy cattle among samples taken from feces, lymph nodes and hides, as well as from feedlot pen environments. These teams have documented not only a shift of the serotype-level population as influenced by geographical, ecological, and seasonal (or temporal) changes, but also an elevated association between specific antibiotic resistance profiles found in specific serotypes (10, 15, 21-25)

Pathogenicity (virulence) and antibiotic resistance are two of the major adaptation, survival, and defense mechanisms of *Salmonella.* These mechanisms are located on bacterial chromosomes or plasmids, and are considered among either the core- or accessory-genomes of *Salmonella* (26-32). The majority of the core genome and almost half of the accessory genes of *Salmonella* are serotype-specific (29, 33). For

instance, plasmidal defense mechanisms are mainly responsible for the transmission of virulent and antibiotic-resistant strains among humans, animals, and the environment, and are often serotype-specific (34, 35). The increasing use of whole-genome sequencing and the development of allele- and SNP-based techniques used for high-resolution genomic comparisons of *Salmonella* (36-39) recently revealed that pathogenicity and antibiotic resistance mechanisms in *Salmonella* were associated with genetic diversity of clonal groups belonging to the same sequence type, or else to lineages within the same serotype (30, 33, 40-42).

Several genomic epidemiology studies have investigated potential drivers of clonal expansion of outbreak-related serotypes, such as *S*. Typhimurium, Enteritidis, Newport, Dublin, Kentucky, Montevideo, and Cerro, all of which are highly prevalent in cattle, by using genomic comparisons of the isolates derived from various host species, food sources, environments and geographical locations (29, 30, 34, 36, 40-54). A subset of these studies demonstrated that subpopulation-level genetic variations were strongly influenced by time and varied by geographical location at global- (36, 40, 41, 49, 50), country- (49, 51), and state-levels (29, 30, 44, 52). In addition to geographical location, a group of studies also showed distinct host- and source-specific macroevolutionary patterns that were observed across human-, animal-, food-, and environment-origin *Salmonella* isolates (29, 30, 34, 40, 42-46, 53). The genetic variation observed among the geographical location, host and environment were attributed mainly to the alterations observed in the prophage regions (43, 45, 54), the absence or

123 presence of pathogenicity-related genes (36, 42-44, 46, 52, 54) as well as antibiotic
124 resistance genes (41, 44, 50).

125 In the literature, there remains a gap in studies of high-resolution population
126 dynamics of *Salmonella* serotypes in cattle from the beginning of the feeding period until
127 slaughter; in case of the latter, this represents the most critical period for mitigating
128 *Salmonella* impacts on public health. Notably, the temporal evolutionary changes
129 observed within serotype and the potential factors that play a role in the
130 microevolutionary patterns of *Salmonella* populations in cattle are yet to be discovered
131 and thoroughly understood (4).

132 Previous work by our group measured the effects of single-dose metaphylactic
133 antibiotic treatments on *Salmonella* prevalence in beef feedlot cattle feces, hides, and
134 lymph nodes, and reported no treatment effects on the prevalence or phenotypic
135 antibiotic resistance profiles of *Salmonella* (25). However, the serotype-level distribution
136 of *Salmonella* (Anatum, Lubbock, Cerro, Montevideo, Kentucky, Newport, and Norwich)
137 that corresponded to a single sequence type was found to be mostly influenced by the
138 geographical origin of the cattle and the experimental blocks of pens that housed them
139 through the study period.

140 Our primary objective in the present analysis was to investigate the temporal
141 microevolutionary patterns of *Salmonella* within each serotype to quantify the relative
142 importance of cattle source and experimental allocation via blocks and pens over a
143 study period of several months. We also explored the specific role of sample types (i.e.,
144 feces, lymph nodes, hides) by using in each case the core-genome SNP-based

maximum-likelihood phylogeny and hierarchical clusters of core-genome multilocus sequence types (HierCC)analyses to explore the multiple levels of resolution of core-genomic relationships among the subpopulations within the *Salmonella* serotypes. We further compared the HierCC-levels and cluster numbers with a publicly available *Salmonella* whole-genome database that includes human clinical disease isolates (EnteroBase Platform http://enterobase.warwick.ac.uk/species/index/senterica) to explore the potential public health risks of the *Salmonella* subpopulations identified in our study (38, 55). Secondarily, using a supervised machine-learning algorithm, we explored the relative importance of each experimental component in contributing towards clonal *Salmonella* expansion within each serotype. Lastly, we investigated the presence and distribution of antibiotic resistance genes, along with plasmids and pathogenicity islands, within the entire *Salmonella* population isolated during the study.

Materials and Methods

This study was reviewed for animal use and biosafety and approved by the West Texas A&M University/Cooperative Research, Educational and Extension Team Institutional Animal Care and Use Committee (Protocol no. 05-09-15) and the Texas A&M University Institutional Biosafety Committee (IBC2017-049), respectively.

The experimental design of the study was previously described by Levent et al. (2019) (25). Briefly, a total of 134 cross-bred beef cattle was purchased from two different sources (35 cattle from Source 1 [Hereford, Texas], and 99 cattle from Source

2 [Abilene, Texas], located 480 km apart from each other) in west Texas, USA. Cattle were shipped to a research feedlot operated by West Texas A&M University and located near Canyon, Texas. Upon arrival to the feedlot, cattle were source- and weight-blocked and allocated into four blocks (1-4); within each block, cattle were randomized into three pens to control for possible cattle source- and host-related confounders. Later, cattle pens (with 11-12 steers) in each block were randomly assigned to receive: 1) ceftiofur, 2) tulathromycin or, 3) remain as a control group without any antibiotic treatment, respectively. For each treatment pen, all but one (or two) of the cattle were administered subcutaneously a single dose one-time injection of either ceftiofur crystalline-free acid (Excede®, Zoetis Inc., Kalamazoo, MI, USA) at 6.6 mg/kg or else tulathromycin (Draxxin®, Zoetis Inc., Kalamazoo, MI, USA) at 2.5 mg/kg. Staff involved in this study were blinded as to treatment to avoid possible selection bias. From time of antibiotic treatment (Day 0), until the time of slaughter (between Days 99 -141), cattle were fed and raised without any additional antibiotic supplements. Starting from Day 99, cattle that had reached the desired body weight for slaughter were sent to the slaughterhouse as experimental blocks until Day 141, which was the end of the study. Fecal samples (*per rectum*) were collected via sterile obstetric gloves immediately prior to the antibiotic treatment (Day 0), a week after the injection (Day 7), and on multiple additional days up until slaughter. At slaughter, one final fecal sample and an additional hide swab, rubbed from a one-m^2 cranial ventral (brisket) area, were collected from each steer. In addition, two peripheral sub-iliac lymph nodes were collected from each hot carcass during slaughter at a local federally inspected processing plant. Study design details relating to allocation of the 134 cattle purchased from two different

189 locations (sources) to the experimental blocks and pens and the sample types
190 corresponding to each of the sampling days are provided in Table 1.

191 *Salmonella* isolates derived from fecal samples during the early feeding period
192 (Day 0 and 7) and fecal, sub-iliac lymph node, and hide swab samples obtained at
193 slaughter included for analysis in this study are previously reported in detail by Levent et
194 al. (2019). A total of 399 isolates was previously isolated, sequenced, serotyped and
195 sequence typed (using legacy 7-gene MLST [multi-locus sequence type]), and were
196 identified as Lubbock ST413 (n =136), Anatum ST64 (n=113), Montevideo ST138
197 (n=68), Cerro ST138 (n=64), Kentucky ST152 (n=11), Newport ST118 (n=6), and
198 Norwich ST2119 (n=1).

199 The methods related to *Salmonella* isolation, DNA extraction, whole-genome
200 sequencing, genome quality, genome assembly, and *in silico* tools used for legacy 7-
201 gene MLST and serotyping are previously published by Levent et al., 2019 (25). Paired-
202 end raw sequencing reads and assemblies of 399 *Salmonella* isolates have previously
203 been deposited in the NCBI database under BioProject number PRJNA521731.
204 BioSample numbers of each assembly, and the metadata related to each isolate are
205 listed in Supp. Dataset 1.

206

207 Phylogenetic analyses

208 *Salmonella* strains restricted to those from serotypes Anatum, Cerro,
209 Montevideo, and Lubbock and isolated from multiple sample types arising from

individual cattle throughout the study were included in the core-genome SNP analysis using maximum-likelihood phylogeny. This was performed to obtain a high degree of resolution of the population structure within each serotype; importantly, all isolates of one serotype were identified as being from the same legacy 7-gene MLST group (39). Cattle source, block, pen, day, and sample type distribution of the isolates included in the SNP-based phylogenetic analysis from each of the four serotypes are presented in Table 2. Kentucky, Newport, and Norwich isolates were excluded from phylogenetic analysis due to small sample sizes (n<12).

Reference genome selection

In order to perform core-genome-level SNP analysis, the complete reference genomes at the closest genomic distance for each serotype were selected using the Similar Genome Finder Service on PATRIC (the Pathosystems Resource Integration Center: available at https://www.patricbrc.org/app/GenomeDistance). The selection criteria were kept at default threshold values (maximum hit value of 50, *P* value of 1, and Mash/MinHash distance [estimating the distance based on the rate of sequence mutation] value of 0.05) (56, 57). The selected reference genomes were further screened for prophage regions using PHASTER (Phage Search Tool Enhanced Release, available at http://phaster.ca/) (58). The complete and questionable prophage regions were further masked using BEDTools v.2.18 (59) and were carried to the next step; that is, alignment of sequencing reads from the isolates for phylogenetic analyses.

231

232 *Variant calling, tree inference and tree visualization*

233 Reference alignment and variant-calling processes for those isolates that shared
234 the same serotype were performed using the McOutbryk SNP calling pipeline (available
235 under the Massachusetts Institute of Technology license at
236 https://github.com/hcdenbakker/McOutbryk) using raw (FASTQ) sequencing short reads
237 with a minimum sequencing depth of 24X and a reference genome (60). This pipeline
238 uses the McCortex tool (61) to build graphs of a reference sequence along with the data
239 of the genomes to be queried for SNP calling. The SNP-calling stage consists of two
240 phases: (Phase 1) an initial phase, which consists of a comparison of the reference
241 graph and each query genome to construct a list of putative variable sites within the
242 population, and (Phase 2) a final SNP-calling phase, which calls the allele for each
243 putative SNP site found in Phase 1. While the SNP calling is done *de novo* at the initial
244 step, the pipeline uses BWA (Burrow-Wheeler aligner)-mem (62) at the final step to
245 place the SNP sites in relation to the reference sequence. In addition to BWA, the
246 pipeline relies on VCFtools (63), and vcflib (available under the MIT license at
247 https://github.com/vcflib/vcflib) for VCF (Variant Call Format) file manipulation. K-mer
248 size 33 was set for variant analysis, and highly divergent isolates (number of SNPs >
249 5000) were excluded from the analysis by default. The matrices containing SNP sites
250 were later evaluated for the best-fit nucleotide substitution model using IQ-tree v.1.6.10
251 with ascertainment bias correction (ASC) option to construct a maximum-likelihood
252 phylogeny (64). Later, the phylogenetic trees were inferred using the selected

nucleotide submission model, including bootstrap estimates obtained with 1,000 iterations (64). The resulting trees were visualized and annotated with cattle source, pen, block, day and sample type as features of the isolates using the interactive tree of life (iTOL v.4) (65). The number of SNPs corresponding to the tree scale and the strong branch support values (bootstrap values of 80%-100%) of the clades were presented on each resulting phylogenetic tree.

Hierarchical clustering of cgMLST and feature importance analyses

Hierarchical clustering of the cgMLST (core-genome multilocus sequence typing, also known as HierCC) cluster number of the isolates – including all serotypes – were determined at multiple levels of HierCC clusters using the EnteroBase platform (based on cgMLST V2 + HierCC V1) to obtain high resolution genomic-based comparisons via HierCC (38, 55). The HierCC cluster numbers observed for each of the isolates derived from our study were also compared with the publicly available whole-genomes in the *Salmonella* database provided by EnteroBase to explore the public health-related significance of our findings at the HC0-,HC2-, HC5-, HC10- and HC50-levels (38).

Furthermore, the importance of isolate features such as, cattle source (1 and 2), block (1, 2, 3, and 4), day (0, 7, and 122 [collapsed to the average day from slaughter days 99, 120, 134, and 141]) and sample type (fecal, lymph node, and hide) for the prediction of the indistinguishable isolates (i.e., those that were highly clonal and designated to the same HC0 cluster; that is, with no cgMLST allele difference) were

explored for each serotype population using the Orange v.3.26.0 data-mining toolbox (66). The supervised machine learning algorithm Random Forest (67, 68) was used to build a set of (n=10) decision trees for the clonal groups using the features listed above, and the resulting trees were explored using Pythagorean Forest (69) algorithm. The best tree that that requires only a few attributes to split the branches was selected to evaluate the feature importance on the tree. Importance of the features was ranked from the most important to the least important based on the values predicted by the ReliefF algorithm (70, 71).

Characterization of antibiotic resistance genes, plasmids, and pathogenicity islands

In silico plasmid and antibiotic resistance gene detection and identification of 399 assembled genome sequences were performed using ABRicate v.0.8.7 (https://github.com/tseemann/abricate) using the ResFinder (72) and the PlasmidFinder (73) databases as well as the SPIFinder (74) database that were obtained from the Center for Genomic Epidemiology server (available online at https://bitbucket.org/genomicepidemiology/spifinder_db/src/master/).

Results

Population structure of *Salmonella*

293 *Salmonella* Anatum, Cerro, Montevideo, and Lubbock isolates collected from
294 individual steers throughout the study period were included in core-genome SNP-based
295 phylogenetic analyses. Based on the resulting trees, the highest number of core-
296 genome SNPs (n=84) was observed among Lubbock isolates, followed by Anatum (66
297 SNPs), Cerro (17 SNPs) and Montevideo (12 SNPs); importantly, each demonstrated
298 distinct cattle source-, block-, pen-, and day-related patterns.

299 Population-based HierCC analysis showed that there were no more than 100
300 cgMLST allelic distance between *Salmonella* genomes of each serotype. Based on
301 HierCC designations, the highest level of cluster variation across all sample features
302 and within the constraints of a 122-day study was observed in Anatum (HC100),
303 Lubbock (HC10) and then Cerro (HC10) isolates, followed by Montevideo (HC5),
304 Kentucky (HC5), and Newport (HC0). Details concerning the multi-level HierCC cluster
305 numbers for the isolates found in this study are presented in Supp. Dataset 1 and Fig.
306 S1.

307 As of August 21, 2020, there were 7,149 publicly available *Salmonella* genomes
308 in EnteroBase with an allelic distance equal or less than 50 (HC50) found to be no; of
309 those, 6,934 genomes were found in the same HC50 cluster of serotypes, followed by
310 145 genomes in HC20, 65 genomes in HC10, and 4 genomes in the same HC5 cluster.
311 A single isolate was found in the HC2 cluster suggesting nearly complete homology.
312 Detailed information related to the HierCC cluster numbers, along with the geographical
313 region of origin, source type and other *Salmonella* genome-related data were
314 downloaded from EnteroBase and provided in Supp. Dataset 2. Technical information

315 related to the HierCC clustering method is also provided in the EnteroBase (available at
316 https://enterobase.readthedocs.io/en/latest/features/clustering.html accessed on
317 10/20/2020)

318 Similarly, serotype-specific population structures of *Salmonella* isolates are
319 presented in further detail within specific sections immediately below.

320

321 *S.* Lubbock

322 The complete genome of a single previously reported Lubbock isolate (GenBank:
323 CP032814.1) was found to be the closest genome match to those Lubbock isolates
324 derived from our study. A total of 11 prophage regions were detected and masked in
325 this reference genome (Table S1). The phylogenetic tree of 136 Lubbock isolates was
326 inferred using the K2P+ASC model, which was the best nucleotide substitution model
327 according to the BIC value. The resulting tree had two clades (Clade I and II) and well-
328 supported subclades (bootstrap values of > 80%) with distinct patterns for source, block,
329 pen, and day (Fig. 1). While the isolates from Block 4 were commonly observed in
330 Clade I, isolates from Block 1 (Source 1) were only observed in Clade II. Lubbock
331 isolates derived from the early feeding period (Days 0 and 7) were only observed in
332 Clade II; that is to say, there was no early feeding period isolates observed in Clade I.
333 The phylogeny of Lubbock isolates yielded the highest total number of SNPs (n = 84)
334 when compared to the total number of SNPs obtained from other serotypes included in

this study. There were no sample type-related (feces, lymph node and hide) phylogenetic differences observed among the Lubbock isolates derived at slaughter.

The cgMLST allelic distance measured among isolates of Lubbock (n=136) found in this study was 10 or less. The most prevalent clonal group was identified in the HC0_214521 cluster (n=31) and this clonal group was only observed at slaughter, and only among pens that were from Source 2; further, 26 of the isolates in this group were from Block 4. Another clonal group of the HC0_214545 cluster (n=28) was observed only in Block 2 and only at slaughter. Interestingly, isolates in Cluster HC0_215282 (n=9) were only observed in cattle in pens from Block 1 at slaughter and identified once again from Block 2 at slaughter. Isolates that were identified in Cluster HC0_214569 (n=2) were only observed in Pen 9 during the early feeding period (Days 0 and 7) and were not identified later during the study period (Fig. S1a). There were no unique patterns observed among different sample types (i.e., fecal, lymph node, and hide). The important features contributing most to the clonality of Lubbock were ordered from highest to lowest as experimental block, pen, source, day, and lastly sample type based on their ReliefF scores (Table S2) which likewise confirmed the phylogenetic findings.

As of August 21, 2020, Lubbock isolated from this study were most closely associated with two Texas origin dairy cattle isolates found in EnteroBase at HC2 and HC5 levels, respectively. An additional two Texas origin isolates (beef cattle and beef product origin) were in the same HC10 cluster. At the HC20-level, there were 139 isolates which were from the USA – and mainly from the states of Texas and Kansas.

Isolates in this cluster were derived mainly from cattle liver abscess and beef products; importantly, no human origin isolate was reported in this cluster.

S. Anatum

The genome of a previously isolated Anatum strain (GenBank: CP007483.2) was selected as the reference genome; prophage regions (n=5, Table S1**)** were masked from this reference genome. One isolate in this study (BioSample ID SAMN10910080) showed high divergence (> 5000 SNP differences as compared to the reference genome) was unable to be genotyped using the McCortex genotyping algorithm; therefore, this isolate was excluded from further analysis. As a result, 112 isolates were carried forward for SNP analysis. The best-fit nucleotide substitution model was selected as K3P+ASC according to the lowest BIC value. The resulting tree contained two distinct clades (Clade I and II)containing well-supported subclades (bootstrap values of >80%) that showed distinct patterns relating to source, day, experimental block, and pen. While isolates derived at slaughter from cattle in Pens 7, 8, and 9 (Source 1, Block 1) were observed only in Clade I, isolates from cattle in Pens 55, 56, 57, and 58 (Source 2) were observed only in Clade II. When sample day-related phylogenetic relationships were explored, no isolates were observed in Clade I from Day 7 and no other sample day-related patterns were apparent for either clade (Fig. 2).

Considered in their entirety, Anatum isolates from this study did not show a cgMLST allelic distance of more than ten , except one isolate (same isolate that was

377 excluded from the SNP analysis) that showed a large cgMLST allelic distance (50 or
378 less) compared to the rest of the isolates at HC50 cluster (Fig. S1b). The most prevalent
379 clonal group of Anatum was HC0_214518 (n=34), which was only observed in Blocks 3
380 and 4 and from Day 0 until the end of the study (Day 141). The second most prevalent
381 (n=13) clonal group was HC0_214565, which was only observed in Pen 54 at slaughter.
382 The clonal group of HC0_215273 (n=11) was only observed in Pen 52 during the
383 feeding period on Day 0 (n=1) and Day 7 (n=1); at slaughter, this clonal group had
384 become prevalent in the rest of the pens from Source 2 (n=9). This cluster was never
385 identified in cattle from Source 1 during the study period. In contrast, the HC0_215308
386 clonal group (n=8) was only identified at slaughter and only in cattle from Source 1 (Fig.
387 S1b). Based on the ReliefF scores, experimental block was the most important feature
388 for explaining Anatum clonal differences, followed by pen, source, day, and sample type
389 (Table S2).

390 The Anatum population identified in this study did not cluster with any public
391 *Salmonella* genomes until the HC50-level. As of August 21, 2020, there were 209
392 *Salmonella* genomes that shared the same clustering at HC50. These genomes mainly
393 originated from Texas and Kansas, USA and were derived from cattle and beef product
394 samples. Of interest, the single Anatum isolate that was highly divergent when
395 compared to the rest of the Anatum population was found to be in the same HC2 cluster
396 as a single equine isolate derived from Texas in 2015.

397

398 *S.* Montevideo

A previously sequenced *S.* Montevideo (GenBank: CP032816.1) strain was the genetically closest genome to the Montevideo isolates derived from our study. Prophage regions (n=4) masked in the reference genome are presented in Table S1**.** The tree was inferred with a K2P+ASC model (according to the lowest BIC value) and resulted in two clades (Clade I and II) for 68 Montevideo isolates. Clade I contained only a single distinct hide origin isolate derived from a steer located in Pen 52. The remaining isolates all were in Clade II. There was a subclade (a) within Clade II (Fig. 3) that harbored only isolates derived at slaughter age though from all sample types. The majority of isolates in this subclade was from Source 1, especially from Pens 8 and 9. The remainder of the subclade was consistently populated with isolates derived from feces sampled during the early feeding period, from a variety of pens. Montevideo phylogeny revealed a highly conservative tree with distinct day- (or sample type), and pen-related patterns and resulted in the lowest number of SNPs (n=12) when compared to the other serotypes.

Montevideo isolates (n=68) identified in this study were highly clonal and isolates were no more than two cgMLST allelic distance values removed from each other. The most prevalent HierCC clonal group was HC0_214530 (n=30), which was observed only in the early feeding period (Day 0 and 7) in all pens (except Pen 52) and was derived from fecal samples of cattle that are from both sources. The second most prevalent clonal group was HC0_215274 (n=22), which was only observed among pens from Block 1 (Source 1) at slaughter, though it was found across all types of samples (Fig. S1c). Sampling day was the most important feature with the highest ReliefF score for

clonal clustering of Montevideo, followed by sample type, pen, source, and experimental block (Table S2).

Montevideo isolates from this study were found in the HC20_9091 cluster along with a single isolate derived from swine feces in Texas. Besides that, isolate, there were no other similar isolates observed in the same cluster until the HC50-level. At HC50, a total of 1,543 isolates was found to cluster with our isolates, and the majority of those isolates originated from the USA (mainly, the states of Texas and California). Isolates in this cluster were primarily derived from beef products, followed by other cattle sources, including the feedlot environment and human clinical samples.

S. Cerro

The Cerro population identified in this study was never observed in fecal samples from the early feeding period; thus, the analysis was restricted to slaughter age samples. The complete genome of a Cerro strain (GenBank: CP008925.1) was selected as the reference genome, and five prophage regions (Table S1) were masked before SNP analysis. The phylogenetic tree of 64 Cerro isolates was inferred using TIMe+ASC nucleotide substitution model according to the lowest BIC value. The resulting phylogenetic tree revealed two separate clades (Clade I and II). Among those, Clade I was well-supported (bootstrap value of > 80%) presenting distinct source, slaughter sampling day (collinear with experimental block), and pen patterns (Fig. 4). Isolates derived from cattle in Pens 7, 8 and 9 from Block 1 (Source 1 and slaughter Day 134)

and Block 4 (Source 2, slaughter Day 99) were found only in Clade I. Clade II harbored only the isolates from Source 2 and mainly were from Pens 55 and 56 of Block 3 (slaughter Day 120). There were no obvious sample type-related phylogenetic patterns observed among Cerro isolates.

A total of three clonal HC0 clusters was identified with an cgMLST allelic distance of five or less. The most prevalent clonal HierCC cluster was HC0_214520 (n=23) which was only found in cattle from Source 2, and 22 of these were arose from Pens 55 and 56 (Block 3); the remaining isolate was derived from Pen 52. The second most prevalent clonal group was HC0_215280 consisting of 22 isolates; of those, 19 were derived from Block 1 (Source 1) while the remaining three isolates were from pens housing cattle from Source 2 (Fig. S1d). There were three isolates in the third clonal group of HC0_214566; all arose from samples taken from Pen 59 at slaughter. This group (HC0_214566) of isolates also differed from the rest of the isolates at the HC5 level (i.e., no more than cgMLST allelic distance of five). Based on ReliefF scores, the most important feature of clonal Cerro clustering was experimental block (collinear with slaughter day) followed by source, pen, and sample type (Table S2).

Similar to Anatum isolates, Cerro isolates were not found to be in the same cluster as any other public *Salmonella* genomes until the level of HC50. At the HC50 cluster, as of August 21, 2020 there were 283 matching genomes (HC50_996) and the majority of these were sequences from isolates collected in the United States and a few other countries (e.g., Taiwan and the United Kingdom). The isolates from the USA

463 mainly were from Texas, followed by California. Cerro genomes at the HC50 cluster
464 level were mostly associated with beef products, cattle, and humans.

465

466 Other serotypes

467 Kentucky (n=11) isolates were derived only from early feeding period (Days 0
468 and 7) fecal samples of cattle purchased from Source 1 and were highly clonal; of
469 these, 9 isolates were identified as being in a single highly clonal group (HC0_214531)
470 as shown in Fig. S1e. Kentucky isolates were not closely related to any other isolated
471 Kentucky genomes until the HC50-level. At this level (HC50_16), a total of 4,521
472 Kentucky genomes were matched with the HierCC inquiry, and mainly originated from
473 the USA; of those, the majority of isolates were derived from California, followed by
474 Georgia and then Texas. A majority of Kentucky genomes at the HC50 clustering level
475 were associated with poultry, and rarely with either cattle or human isolates.

476 Newport (n=6) isolates were derived only from fecal samples and lymph nodes of
477 cattle located in Pen 53 and only at slaughter age. These isolates were
478 indistinguishable from each other and belonged to a single clonal group (HC0_215270)
479 as also shown in Fig. S1f. Newport isolates derived in our study were no more than
480 cgMLST allelic distance of 20 (clustered at HC20) from two *Salmonella* genomes
481 derived from human clinical specimens in the USA. The next HierCC cluster match was
482 observed at HC50-level with 284 isolates, the majority of those were from Arizona,

followed by California and Texas. Isolates from this cluster were primarily isolated from human, seeds and vegetables and rarely found in cattle related sources.

Only a single Norwich isolate was identified at slaughter from a fecal sample of a steer located in Pen 52. This isolate was no more than five cgMLST allelic distance to four human isolates from the USA. In addition, there was a total of 72 human or outbreak origin Norwich genomes that were observed in the same HC10 cluster. There was no other public *Salmonella* genome found in the HC20 or HC50 cluster of the single Norwich isolate.

Antibiotic resistance, plasmids, and pathogenicity island profiles

Among 399 *Salmonella* isolates screened for antibiotic resistance genes and plasmids. All Montevideo isolates (except for one) harbored a tetracycline resistance gene *tet*(A), all of which were located on an IncN (IncN_1) plasmid-related contig. A subset of isolates (n=62) carried an IncI1 (IncI_1_Alpha) plasmid, which was not serotype-specific and was found only among a subset of Anatum, Lubbock and Montevideo isolates.

A total of ten *Salmonella* pathogenicity islands (SPIs) were identified across these same 399 isolates. Of those, SPI1, SPI2, SPI5, and SPI9 were found in all serotypes with few exceptions; on the other hand, C63PI, SPI3, SPI8, SPI13, and SPI14 were found to be highly serotype-specific. The SPI4 mainly varied across the Cerro and Lubbock subpopulation of isolates, regardless of the clonal distribution of these isolates

and of any other features such as pen, day, or sample type. Antibiotic resistance genes, plasmids and SPIs that were identified within each serotype are provided in Table 3. No distinct plasmidal or pathogenicity island profiles were found to be related to a specific clonal group (Supp. Dataset 1).

Discussion

Our previously published randomized controlled longitudinal feedlot trial was initially designed to explore the effects of single-dose antibiotic metaphylaxis for bovine respiratory disease on antibiotic resistance and the prevalence of *Salmonella* in cattle feces during the feeding period, and in feces, sub-iliac lymph nodes and on hide surfaces at slaughter. The results of our previous study reflected no significant differences between the prevalence of *Salmonella* in cattle that received either antibiotic treatment (ceftiofur or tulathromycin) and those in the control group that received neither. What varied instead, was the distinct serotype distribution patterns observed across sampling days and sample types among the 12 pens, four experimental blocks, and the two sources the cattle were purchased from, which merited further investigation into how the subpopulations within the serotypes were influenced when a higher resolution was obtained. The isolates identified as Anatum, Lubbock, Cerro, Montevideo, Kentucky, and Newport serotypes each appeared in their own same legacy 7-gene MLST groups, which initially suggested clonal dissemination of *Salmonella* within the cohorts of cattle throughout the feeding period and at slaughter (25).

However, initial examinations of the phylogenies of these longitudinally sampled isolates suggested that simplistic reasoning was likely wrong.

In this follow-up study, we further explored the clonal distribution and evidence for temporal and geographical dissemination of *Salmonella* isolates by performing core-genomic comparisons with increased resolution using core-genome SNP-based maximum likelihood phylogeny and allele-based approach using HierCC method. HierCC analyses directly supported the clonal distributions that were observed on the phylogenetic trees. The latter analysis method was used specifically to identify clonal groups, predict the importance of the study design-related features as well as the genomic distance between the isolates derived from this study and the isolates derived from clinical human cases. HierCC methods have been shown to be highly reliable in predicting legacy 7-gene MLST groups (HC900), endemic persistence of strains (HC100 and HC200), as well as the clonal groups of epidemic outbreak strain (HC2, HC5, and HC10) clusters based on distances observed among cgMLST alleles (38).

Overall, the results we obtained from SNP- and allele-based distance analyses showed that experimental blocks – wherein 11-12 cattle randomized to three adjacent pens were geographically co-located – and the geographical origin of the cattle (source) were the most consistently predictive factors explaining clonal clustering and expansion observed in the *Salmonella* population. These two factors were followed by sampling day and sample type, though of lesser importance.

Salmonella Cerro (legacy 7-gene MLST ST367) isolates were identified in cattle from all experimental blocks; however, this serotype was only isolated from samples

547 collected when cattle were eligible to be sent to slaughter between Days 99 and 141
548 (specific slaughter sampling dates by experimental block are listed in Table 1). Our
549 findings show that clonal clustering/expansion of the Cerro population was most
550 influenced by experimental block (or, by slaughter day due to collinearity), followed by
551 source, pen, and sample type. However, it seems most likely that the clonal effects
552 observed are attributable to block rather than sampling day. Relatively short intervals
553 between sampling days at slaughter eligibility underlies this hypothesis, which will also
554 have applied to the other serotypes that were derived during the study within the
555 constrained period of slaughter, or where the clonal expansion of *Salmonella* seemed
556 affected less by day than by block. While the Cerro population derived from our study
557 were genetically clustered with other publicly available genomes of the isolates
558 originated from Texas, the beef-, human-, and cattle-origin Cerro population showed a
559 strong endemic persistence of the outbreak-related strains and had no more than 50
560 cgMLST allelic distance from the isolates derived from our study. Preceding our study,
561 the work of Kovac et al. (2017) identified geographical location as a potential driver of
562 the clonal expansion of the Cerro population (ST367) derived from dairy cattle located
563 either in Texas or New York state (52).

564 Prior to this present study, the potential drivers of the clonal clustering/expansion
565 of Lubbock and Anatum populations in cattle had not been explored. Our results
566 indicate that experimental block, pen, and cattle source are the most important features
567 for explaining clonal relatedness of Lubbock (ST413) and Anatum (ST64) populations in
568 fed cattle during a study period such as ours. In addition, the HierCC results of our

569 study suggest that dissemination of Lubbock and Anatum related clones has been
570 restricted primarily to Texas and Kansas, U.S. states where these two serotypes mainly
571 were isolated from cattle and beef products. However, neither set of serotypes clonal
572 groups have been associated with human isolates at the HC50-level. An emerging
573 serotype first reported in 2015, Lubbock was first identified in a sub-iliac lymph node of
574 a harvest-ready steer located in Texas. The study of Bugarel et al. (2015) suggested the
575 novel serotype evolved from its ancestral serotype Mbandaka through recombination
576 events (75). As of yet, no Lubbock strain was reported as a source of human clinical
577 *Salmonella* outbreaks. Thus, the possible public health outcome of the Lubbock
578 population remains unknown.

579 In contrast to *S.* Anatum, Lubbock and Cerro, the clonal clustering/expansion of
580 the Montevideo (legacy 7-gene MLST ST138) population was influenced mainly by
581 sampling day, sample type and pen; in contrast to the other serotypes, experimental
582 block, and cattle had much less influence on their apparent dissemination. This
583 difference was perhaps not surprising since the majority of Montevideo strains were
584 identified in fecal samples initially from the beginning of the early feeding period and
585 across all pens from both sources or mainly from Source-1 at slaughter time on day 134
586 (Fig. 3). While our machine-learning approach suggests that features related to
587 sampling day and sample type had more bearing on Montevideo clonal
588 clustering/expansion, it also is possible that our study design might have masked the
589 pen-, block-, and source-related influence. The well-supported (bootstrap value of
590 >80%) subclade (a) within the Clade I observed in the phylogenetic tree of Montevideo

591 also reinforced that there was an even greater effect of pen when multiple sample types
592 were accounted for (Fig. 3). In addition, sample type has not earlier been reported as
593 being associated with clonal clustering/expansion of *Salmonella* subpopulations
594 reported (42, 51). Typhimurium and Derby isolates derived from swine feces, lymph
595 nodes and carcasses in farms located in the state of North Carolina have yielded, for
596 example, no phylogenetic relations when comparing among the sample types from
597 which isolates were derived (42). The Montevideo (legacy 7-gene MLST ST138)
598 population in our study displayed more similarities with previously recorded cattle-origin
599 isolates than with human-origin isolates at the HC50-level. The research of Nguyen et
600 al. (2018), likewise, revealed with a few exceptions that Montevideo legacy 7-gene
601 MLST ST138 strains were related mainly to cattle origin isolates and were located in
602 different clades than human outbreak origin isolates which mainly were from ST316
603 (43).

604 We also identified 11 highly clonal Kentucky isolates from the early feeding
605 period (Day 0 and 7); importantly, these were identified only from those cattle
606 purchased from Source 1. We believe that Kentucky was the serotype most likely to
607 have pre-existed in a group of cattle prior to placement in the feedlot. Further, this helps
608 to explain the apparent greater effect of source and sampling day than of experimental
609 block and pen (geographical and environmental surrogates) on the clonal
610 clustering/expansion of the Kentucky population. In addition, a total of six Newport
611 isolates were identified, all of which were found to be highly clonal and originated from
612 either a lymph node (n=5) or fecal sample (n=1) of cattle housed in the same pen (Pen

613 53). This population exhibited no more than 20 cgMLST allelic distance from human-
614 and outbreak-related isolates. Although these isolates had no antibiotic resistant genes,
615 the presence of the highly clonal Newport population in the lymph nodes poses a
616 potential risk for contamination of ground beef. This is a tangible public health risk, since
617 a recent *Salmonella* outbreak similarly originated in a contaminated ground-beef
618 product, was traced to a pan-susceptible Newport isolate in 2018, and resulted in 403
619 reported cases in 30 states causing 117 hospitalizations (76). Besides Newport, our
620 study identified a single Norwich strain isolated from a cattle fecal sample at slaughter.
621 This isolate was found to be related to 70 human and outbreak origin isolates at the
622 HC10-level and 4 out of 70 came from the same cluster at the HC5-level which is
623 reveals this strain to be a clonal epidemic outbreak strain (38).

624 Our data did not reflect any SPIs specific to a certain clonal group of *Salmonella*.
625 The exception was the presence of SPI4, which varied within both the Cerro and
626 Lubbock populations. The SPI4 is known to carry virulence genes responsible for
627 epithelial adhesion and host invasion (77). When we examined the absence or
628 presence of SPI4 regions at the population level, for instance, patterns related to
629 individual cattle seemed more consistent compared to those related to pen, block,
630 source, and day. Over the entire feeding period, we identified only two types of plasmids
631 (IncN and IncI1) – both, incompatibility group plasmids previously reported to harbor
632 beta-lactam class resistance genes in *Salmonella* isolated from farm animals (78). Even
633 though the selection pressure of ceftriaxone metaphylaxis was applied to one group of
634 cattle on day 0, none of the isolates derived bore IncI1 and IncN plasmids harboring *bla*

635 genes. The presence of the IncI1_alpha plasmid was related neither to a particular
636 clonal group nor to a particular serotype. IncI plasmids can, in fact, be found in a wide
637 range of *Salmonella* serotypes and may or may not carry antibiotic resistance and
638 virulence genes; however, the exact function of IncI1 plasmid remains unknown (79).

639 We found an IncN_1 plasmid and *tet*(A) gene associated with all of our
640 Montevideo isolates. Cattle-origin Montevideo isolates were previously found to be
641 related to an IncN_1 plasmid harboring *tet*(A) genes (28); however, no sequence type
642 (legacy 7-gene MLST) of the Montevideo isolates in that study were reported. Another
643 study – Nguyen et al. (2018) – examined a cattle origin Montevideo (legacy 7-gene
644 MLST ST138) population, but did not report any IncN plasmids and tetracycline
645 resistance genes (43), suggesting that the IncN plasmid and *tet*(A) gene presence in
646 Montevideo ST138 may be related to the particular feedlot environment or the
647 geographical origin of the cattle. We have previously described the antimicrobial
648 susceptibility phenotype of these same isolates (25) and showed that the Montevideo
649 isolates that carried the *tet*(A) gene were also found to be phenotypically resistant to
650 tetracycline. However, based on sequencing data, 21 isolates that were previously
651 identified as phenotypically resistant to streptomycin did not bear any such streptomycin
652 (aminoglycoside) resistance genes, such as of the *str*, *aad*, *aph,* or *arm* families. This
653 inconsistency between the genotypic and phenotypic streptomycin resistance could
654 possibly be ascribed to the uncertainty of the MIC breakpoints that have been
655 determined for phenotypic streptomycin resistance (80).

Our results showed no antibiotic resistance, plasmid, and pathogenicity island profile association among varied *Salmonella* subpopulations derived from individual cattle, pens, blocks, and sources during the study period. However, the prophage regions of the isolates that are expected to usually vary by host and environment was not explored in our study. Beside the lack of prophage profile findings, another lack of finding was the absence of samples from the environment (e.g., pen floor, feedlot dust) to measure the environmental contribution on *Salmonella* dynamics in feedlot cattle.

We employed a supervised machine-learning algorithm to estimate the relative importance of features such as source, experimental block, pen, sampling day, and sample type for explaining clonal clustering/expansion of *Salmonella* serotypes. We used a Random Forest algorithm with ReliefF-based feature selection to rank the prediction power of these features by their potential to forecast clonal outcomes (68, 71). While the outcomes of the feature importance were consistent with the core-genomic subpopulation-related analysis performed in this case, the interpretation of results in cases where machine-learning methods are used needs to be handled with care; specifically, with an eye to prioritizing the qualitative roles of biology and study design over numeric predictions. Our study suggests that the application of machine-learning tools, when used for the prediction of feature importance on clonal clustering and expansion, can be a promising approach. However, used in studies such as ours, with more limited sample size numbers, makes interpretation in relation to the biology and study design features even more important.

677 In summary, this was the first study of its kind to attempt to monitor and measure
678 the microevolutionary progress of *Salmonella* subpopulations over time among various
679 sample types collected from a group of beef cattle during the feeding period and at
680 slaughter. We aimed to explore the potential initiators and drivers of *Salmonella*
681 population dynamics as observed in feces, lymph nodes, and hide samples, all of which
682 are among potential contaminants of beef products at slaughter. Our research findings
683 stress the importance of cattle origin, which impacts background microbiota on arrival
684 and also the starting feedlot environment into which these same cattle are placed
685 (including the geospatial placement of cattle and their pens). In total, these two features
686 are the main drivers of initial shared *Salmonella* clonal profiles that later morph in
687 response to microevolutionary pressures. In order to combat antibiotic resistance and
688 reduce the pathogenicity of *Salmonella* found in final beef products, more research is
689 needed to explore the feedlot management and environmental effects on clonal
690 *Salmonella* clustering and expansion, including pre-existing soil and manure pack
691 microbiota and dust contribution as well as the bacteriophages that are in the feedlot
692 environment.

693

694 **Acknowledgments**

695 We acknowledge the H. M. Scott laboratory graduate and undergraduate students for
696 assisting with sample collection and microbiological processing. This study was funded
697 by the National Cattlemen's Beef Association, a contractor to the Beef Checkoff (no.
698 22615).

699 **References**

700 1. Ao TT, Feasey NA, Gordon MA, Keddy KH, Angulo FJ, Crump JA. 2015. Global
701 burden of invasive nontyphoidal *Salmonella* disease, 2010(1). Emerg Infect Dis
702 21.

703 2. CDC. 2015. Serotypes and the importance of serotyping *Salmonella*. Centers for
704 Disease Control and Prevention, U.S. Department of Health and Human
705 Services, Atlanta, GA.

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

710 4. Hong S, Rovira A, Davies P, Ahlstrom C, Muellner P, Rendahl A, Olsen K,
711 Bender JB, Wells S, Perez A, Alvarez J. 2016. Serotypes and antimicrobial
712 resistance in *Salmonella enterica* recovered from clinical samples from cattle and
713 swine in Minnesota, 2006 to 2015. PLoS One 11:e0168016.

714 5. McDermott PF, Zhao S, Tate H. 2018. Antimicrobial resistance in Nontyphoidal
715 *Salmonella*. Microbiol Spectr 6.

716 6. CDC. 2013. Antibiotic resistance threats in the United States, 2013. Centers for
717 Disease Control and Prevention, U.S. Department of Health and Human
718 Services, Atlanta, GA.

719 7. CDC. 2019. Antibiotic resistance threats in the United States, 2019. Centers for
720 Disease Control and Prevention, U.S. Department of Health and Human
721 Services, Atlanta, GA.

722 8. Laufer AS, Grass J, Holt K, Whichard JM, Griffin PM, Gould LH. 2015. Outbreaks
723 of *Salmonella* infections attributed to beef -United States, 1973-2011. Epidemiol
724 Infect 143:2003-13.

725 9. Arthur TM, Brichta-Harhay DM, Bosilevac JM, Guerini MN, Kalchayanand N,
726 Wells JE, Shackelford SD, Wheeler TL, Koohmaraie M. 2008. Prevalence and
727 characterization of *Salmonella* in bovine lymph nodes potentially destined for use
728 in ground beef. J Food Prot 71:1685-8.

729 10. USDA-APHIS. 2001. Info sheet- *Salmonella* in United States feedlots. United
730 States Department of Agriculture, Animal and Plant Health Inspection Service,
731 Fort Collins, CO.

11. USDA-APHIS. 2014. Info sheet- *Salmonella* in U.S. Cattle Feedlots. United States Department of Agriculture, Animal and Plant Health Inspection Service, National Animal Health Monitoring System, Fort Collins, CO.

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

13. FSIS. 2017. Serotypes profile of *Salmonella* isolates from meat and poultry products January 1998 through December 2014. Food Safety and Inspection Service, United States Department of Agriculture, Washington, D.C.

14. Afema JA, Mather AE, Sischo WM. 2015. Antimicrobial resistance profiles and diversity in *Salmonella* from Humans and Cattle, 2004-2011. Zoonoses Public Health 62:506-17.

15. Gutema FD, Agga GE, Abdi RD, De Zutter L, Duchateau L, Gabriel S. 2019. Prevalence and serotype diversity of *Salmonella* in apparently healthy cattle: systematic review and meta-analysis of published studies, 2000-2017. Front Vet Sci 6:102.

16. CDC-NARMS. 2018. Human Isolates Surveillance Report for 2015. National Antimicrobial Resistance Monitoring System, Centers for Disease Control and Prevention, Atlanta, GA.

17. FSIS. 2016. Serotypes profile of *Salmonella* isolates from meat and poultry products from January 1998 through December 2014. Food Safety And Inspection Service. USDA.

18. Carlson JC, Hyatt DR, Ellis JW, Pipkin DR, Mangan AM, Russell M, Bolte DS, Engeman RM, DeLiberto TJ, Linz GM. 2015. Mechanisms of antimicrobial resistant *Salmonella enterica* transmission associated with starling-livestock interactions. Vet Microbiol 179:60-8.

19. Harhay DM, Bono JL, Smith TP, Fields PI, Dinsmore BA, Santovenia M, Kelley CM, Wang R, Harhay GP. 2016. Complete closed genome sequences of *Salmonella enterica* subsp. enterica serotypes Anatum, Montevideo, Typhimurium, and Newport, isolated from beef, cattle, and humans. Genome Announc 4.

20. CDC. 2013. An Atlas of *Salmonella* in the United States, 1968-2011. Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, Atlanta, GA.

21. Kunze DJ, Loneragan GH, Platt TM, Miller MF, Besser TE, Koohmaraie M, Stephens T, Brashears MM. 2008. *Salmonella enterica* burden in harvest-ready

769 cattle populations from the Southern High Plains of the United States. Appl
770 Environ Microbiol 74:345-51.

771 22. Gragg SE, Loneragan GH, Nightingale KK, Brichta-Harhay DM, Ruiz H, Elder JR,
772 Garcia LG, Miller MF, Echeverry A, Ramirez Porras RG, Brashears MM. 2013.
773 Substantial within-animal diversity of *Salmonella* isolates from lymph nodes,
774 feces, and hides of cattle at slaughter. Appl Environ Microbiol 79:4744-50.

775 23. Webb HE, Brichta-Harhay DM, Brashears MM, Nightingale KK, Arthur TM,
776 Bosilevac JM, Kalchayanand N, Schmidt JW, Wang R, Granier SA, Brown TR,
777 Edrington TS, Shackelford SD, Wheeler TL, Loneragan GH. 2017. *Salmonella* in
778 peripheral lymph nodes of healthy cattle at slaughter. Front Microbiol 8:2214.

779 24. Purdy CW, Straus DC, Clark RN. 2004. Diversity of *Salmonella* serovars in
780 feedyard and nonfeedyard playas of the Southern High Plains in the summer and
781 winter. Am J Vet Res 65:40-4.

782 25. Levent G, Schlochtermeier A, Ives SE, Norman KN, Lawhon SD, Loneragan GH,
783 Anderson RC, Vinasco J, Scott HM. 2019. Population dynamics of *Salmonella*
784 *enterica* within beef cattle cohorts followed from single-dose metaphylactic
785 antibiotic treatment until slaughter. Appl Environ Microbiol 85:e01386-19.

786 26. Folster JP, Grass JE, Bicknese A, Taylor J, Friedman CR, Whichard JM. 2017.
787 Characterization of resistance genes and plasmids from outbreaks and illness
788 clusters caused by *Salmonella* resistant to ceftriaxone in the United States, 2011-
789 2012. Microb Drug Resist 23:188-193.

790 27. Karp BE, Campbell D, Chen JC, Folster JP, Friedman CR. 2018. Plasmid-
791 mediated quinolone resistance in human non-typhoidal *Salmonella* infections: An
792 emerging public health problem in the United States. Zoonoses Public Health
793 65:838-849.

794 28. McMillan EA, Gupta SK, Williams LE, Jove T, Hiott LM, Woodley TA, Barrett JB,
795 Jackson CR, Wasilenko JL, Simmons M, Tillman GE, McClelland M, Frye JG.
796 2019. Antimicrobial resistance genes, cassettes, and plasmids present in
797 *Salmonella enterica* associated with United States food animals. Front Microbiol
798 10:832.

799 29. Liao J, Orsi RH, Carroll LM, Kovac J, Ou H, Zhang H, Wiedmann M. 2019.
800 Serotype-specific evolutionary patterns of antimicrobial-resistant *Salmonella*
801 *enterica*. BMC Evol Biol 19:132.

802 30. Carroll LM, Wiedmann M, den Bakker H, Siler J, Warchocki S, Kent D, Lyalina S,
803 Davis M, Sischo W, Besser T, Warnick LD, Pereira RV. 2017. Whole-genome
804 sequencing of drug-resistant *Salmonella enterica* isolates from dairy cattle and
805 humans in New York and Washington States reveals source and geographic
806 associations. Appl Environ Microbiol 83.

807 31. Jones TF, Ingram LA, Cieslak PR, Vugia DJ, Tobin-D'Angelo M, Hurd S, Medus
808 C, Cronquist A, Angulo FJ. 2008. Salmonellosis outcomes differ substantially by
809 serotype. J Infect Dis 198:109-14.

810 32. Fierer J, Guiney DG. 2001. Diverse virulence traits underlying different clinical
811 outcomes of *Salmonella* infection. J Clin Invest 107:775-80.

812 33. den Bakker HC, Moreno Switt AI, Govoni G, Cummings CA, Ranieri ML,
813 Degoricija L, Hoelzer K, Rodriguez-Rivera LD, Brown S, Bolchacova E, Furtado
814 MR, Wiedmann M. 2011. Genome sequencing reveals diversification of virulence
815 factor content and possible host adaptation in distinct subpopulations of
816 *Salmonella enterica*. BMC Genomics 12:425.

817 34. Hong YP, Wang YW, Huang IH, Liao YC, Kuo HC, Liu YY, Tu YH, Chen BH, Liao
818 YS, Chiou CS. 2018. Genetic relationships among multidrug-resistant *Salmonella*
819 *enterica* serovar Typhimurium strains from humans and animals. Antimicrob
820 Agents Chemother 62.

821 35. Jajere SM. 2019. A review of *Salmonella enterica* with particular focus on the
822 pathogenicity and virulence factors, host specificity and antimicrobial resistance
823 including multidrug resistance. Vet World 12:504-521.

824 36. Cao G, Meng J, Strain E, Stones R, Pettengill J, Zhao S, McDermott P, Brown E,
825 Allard M. 2013. Phylogenetics and differentiation of *Salmonella* Newport lineages
826 by whole genome sequencing. PLoS One 8:e55687.

827 37. Coipan CE, Dallman TJ, Brown D, Hartman H, van der Voort M, van den Berg
828 RR, Palm D, Kotila S, van Wijk T, Franz E. 2020. Concordance of SNP- and
829 allele-based typing workflows in the context of a large-scale international
830 *Salmonella* Enteritidis outbreak investigation. Microb Genom
831 doi:10.1099/mgen.0.000318.

832 38. Zhou Z, Alikhan NF, Mohamed K, Fan Y, Agama Study G, Achtman M. 2020.
833 The EnteroBase user's guide, with case studies on *Salmonella* transmissions,
834 *Yersinia pestis* phylogeny, and *Escherichia* core genomic diversity. Genome Res
835 30:138-152.

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

838 40. Sangal V, Harbottle H, Mazzoni CJ, Helmuth R, Guerra B, Didelot X, Paglietti B,
839 Rabsch W, Brisse S, Weill FX, Roumagnac P, Achtman M. 2010. Evolution and
840 population structure of *Salmonella enterica* serovar Newport. J Bacteriol
841 192:6465-76.

842 41. Leekitcharoenphon P, Hendriksen RS, Le Hello S, Weill FX, Baggesen DL, Jun
843 SR, Ussery DW, Lund O, Crook DW, Wilson DJ, Aarestrup FM. 2016. Global

genomic epidemiology of *Salmonella enterica* serovar Typhimurium DT104. Appl Environ Microbiol 82:2516-26.

42. Pornsukarom S, van Vliet AHM, Thakur S. 2018. Whole genome sequencing analysis of multiple *Salmonella* serovars provides insights into phylogenetic relatedness, antimicrobial resistance, and virulence markers across humans, food animals and agriculture environmental sources. BMC Genomics 19:801.

43. Nguyen SV, Harhay DM, Bono JL, Smith TPL, Fields PI, Dinsmore BA, Santovenia M, Wang R, Bosilevac JM, Harhay GP. 2018. Comparative genomics of *Salmonella enterica* serovar Montevideo reveals lineage-specific gene differences that may influence ecological niche association. Microb Genom 4.

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

45. Gymoese P, Kiil K, Torpdahl M, Osterlund MT, Sorensen G, Olsen JE, Nielsen EM, Litrup E. 2019. WGS based study of the population structure of *Salmonella enterica* serovar Infantis. BMC Genomics 20:870.

46. Sevellec Y, Vignaud ML, Granier SA, Lailler R, Feurer C, Le Hello S, Mistou MY, Cadel-Six S. 2018. Polyphyletic nature of S*almonella enterica* serotype Derby and lineage-specific host-association revealed by genome-wide analysis. Front Microbiol 9:891.

47. Leekitcharoenphon P, Lukjancenko O, Friis C, Aarestrup FM, Ussery DW. 2012. Genomic variation in *Salmonella enterica* core genes for epidemiological typing. BMC Genomics 13:88.

48. Worley J, Meng J, Allard MW, Brown EW, Timme RE. 2018. *Salmonella enterica* phylogeny based on whole-genome sequencing reveals two new clades and novel patterns of horizontally acquired genetic elements. MBio 9.

49. Fenske GJ, Thachil A, McDonough PL, Glaser A, Scaria J. 2019. Geography shapes the population genomics of *Salmonella enterica* Dublin. Genome Biol Evol 11:2220-2231.

50. Hawkey J, Le Hello S, Doublet B, Granier SA, Hendriksen RS, Fricke WF, Ceyssens PJ, Gomart C, Billman-Jacobe H, Holt KE, Weill FX. 2019. Global phylogenomics of multidrug-resistant *Salmonella enterica* serotype Kentucky ST198. Microb Genom 5.

51. 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.

52. Kovac J, Cummings KJ, Rodriguez-Rivera LD, Carroll LM, Thachil A, Wiedmann M. 2017. Temporal genomic phylogeny reconstruction indicates a geospatial transmission path of *Salmonella* Cerro in the United States and a clade-specific loss of hydrogen sulfide production. Front Microbiol 8:737.

53. Allard MW, Luo Y, Strain E, Li C, Keys CE, Son I, Stones R, Musser SM, Brown EW. 2012. High resolution clustering of *Salmonella enterica* serovar Montevideo strains using a next-generation sequencing approach. BMC Genomics 13:32.

54. Zheng J, Luo Y, Reed E, Bell R, Brown EW, Hoffmann M. 2017. Whole-genome comparative analysis of *Salmonella enterica* serovar Newport strains reveals lineage-specific divergence. Genome Biol Evol 9:1047-1050.

55. Achtman M, Zhou Z, Alikhan N, Tyne W, Parkhill J, Cormican M, Chiou C, Torpdahl M, Litrup E, Prendergast D, Moore J, Strain S, Kornschober C, Meinersmann R, Uesbeck A, Weill F, Coffey A, Andrews-Polymenis H, Curtiss 3rd R, Fanning S. 2020. Genomic diversity of Salmonella enterica -The UoWUCC 10K genomes project [version 1; peer review: awaiting peer review]. Wellcome Open Research 5.

56. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. 2016. Mash: Fast genome and metagenome distance estimation using MinHash. Genome Biol 17:132.

57. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T, Gabbard JL, Gerdes S, Henry CS, Kenyon RW, Machi D, Mao C, Nordberg EK, Olsen GJ, Murphy-Olson DE, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Vonstein V, Warren A, Xia F, Yoo H, Stevens RL. 2017. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. Nucleic Acids Res 45:D535-D542.

58. Arndt D, Marcu A, Liang Y, Wishart DS. 2017. PHAST, PHASTER and PHASTEST: Tools for finding prophage in bacterial genomes. Brief Bioinform doi:10.1093/bib/bbx121.

59. Quinlan AR, Hall IM. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26:841-2.

60. Burnett J, Wu ST, den Bakker HC, Cook PW, Veenhuizen DR, Hammons SR, Singh M, Oliver HF. 2020. Listeria monocytogenes is prevalent in retail produce environments but *Salmonella enterica* is rare. Food Control 113:107173.

61. Turner I, Garimella KV, Iqbal Z, McVean G. 2018. Integrating long-range connectivity information into de Bruijn graphs. Bioinformatics 34:2556-2565.

62. Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589-95.

63. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Genomes Project Analysis G. 2011. The variant call format and VCFtools. Bioinformatics 27:2156-8.

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

65. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:W256-W259.

66. Demšar J, Curk T, Erjavec A, Gorup Č, Hočevar T, Milutinovič M, Možina M, Polajnar M, Toplak M, Starič A, Štajdohar M, Umek L, Žagar L, Žbontar J, Žitnik M, Zupan B. 2013. Orange: data mining toolbox in python. J Mach Learn Res 14:2349–2353.

67. Breiman L. 2001. Random Forests. Machine Learning 45:5-32.

68. Stephan J, Stegle O, Beyer A. 2015. A random forest approach to capture genetic effects in the presence of population structure. Nat Commun 6:7432.

69. Beck F, Burch M, Munz T, Silvestro LD, Weiskopf D. 2014. Generalized Pythagoras Trees for visualizing hierarchies. 2014 International Conference on Information Visualization Theory and Applications (IVAPP):17-28.

70. Kononenko I, Šimec E, Robnik-Šikonja M. 1997. Overcoming the myopia of inductive learning algorithms with RELIEFF. Appl Intell 7:39-55.

71. Urbanowicz RJ, Olson RS, Schmitt P, Meeker M, Moore JH. 2018. Benchmarking relief-based feature selection methods for bioinformatics data mining. J Biomed Inform 85:168-188.

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

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

951 74. Roer L, Hendriksen RS, Leekitcharoenphon P, Lukjancenko O, Kaas RS,
952 Hasman H, Aarestrup FM. 2016. Is the evolution of *Salmonella enterica* subsp.
953 *enterica* linked to restriction-modification systems? mSystems 1.

954 75. Bugarel M, den Bakker HC, Nightingale KK, Brichta-Harhay DM, Edrington TS,
955 Loneragan GH. 2015. Two draft genome sequences of a new serovar of
956 *Salmonella enterica*, serovar Lubbock. Genome Announc 3:e00215-15.

957 76. CDC. 2019. Outbreak of *Salmonella* infections linked to ground beef: Final
958 update. Centers for Disease Control and Prevention, U.S. Department of Health
959 and Human Services, Atlanta, GA.

960 77. Gerlach RG, Jackel D, Geymeier N, Hensel M. 2007. Salmonella pathogenicity
961 island 4-mediated adhesion is coregulated with invasion genes in Salmonella
962 enterica. Infect Immun 75:4697-709.

963 78. Carattoli A. 2009. Resistance plasmid families in Enterobacteriaceae. Antimicrob
964 Agents Chemother 53:2227-38.

965 79. Kaldhone PR, Carlton A, Aljahdali N, Khajanchi BK, Sanad YM, Han J, Deck J,
966 Ricke SC, Foley SL. 2019. Evaluation of incompatibility group I1 (IncI1) plasmid-
967 containing *Salmonella enterica* and assessment of the plasmids in bacteriocin
968 production and biofilm development. Front Vet Sci 6:298.

969 80. Tyson GH, Li C, Ayers S, McDermott PF, Zhao S. 2016. Using whole-genome
970 sequencing to determine appropriate streptomycin epidemiological cutoffs for
971 *Salmonella* and *Escherichia coli*. FEMS Microbiol Lett 363.

972

973

974 Table 1. Study design representing number of cattle distributed among source, block,
975 and pen along with the sample types collected each sampling day.

Source	Block	Pen identifier	Treatment	Number of cattle per pen	Feeding period (feces only)	Slaughter period (feces, lymph node, hide)
Hereford, TX (Source -1)	1	7	Tul	12	Day 0 and 7	Day 134
		8	Cont	11		
		9	Cef	12		
Abilene, TX (Source-2)	2	51	Tul	11		Day 141
		52	Cont	11		
		53	Cef	11		
	3	54	Cont	11		Day 120
		55	Cef	11		
		56	Tul	11		
	4	57	Cef	11		Day 99
		58	Tul	11		
		59	Cont	11		

976

977

978 Cattle in Pens 7, 8 and 9 were from Source 1 and were located next to one another in
979 the feedlot. Cattle from Source 2 were located next to one another in Pens 51, 52, 53,
980 54, 55, 56, 57, 58, and 59. A schematic of the pen locations in the feedlot was published
981 previously in Levent et al. (2019) (25). Pens within each block (see Table) were
982 randomly assigned for cattle to receive either ceftiofur (Cef) or tulathromycin (Tul) on
983 Day 0 prior to first fecal sample collection, or else to remain as control (Cont) animals.

984

985 Table 2. Number of isolates selected for phylogenetic analysis by serotype, source,

986 block, pen, day, and sample type.

987

Anatum (ST64)

			Day 0 fecal	Day 7 fecal	Terminal fecal	Lymph node	Hide
Source 1	Block 1	Pen-7	3	0	6	3	0
		Pen-8	0	0	1	0	0
		Pen-9	0	0	0	0	1
	Block 2	Pen-51	1	0	0	1	0
		Pen-52	3	1	3	1	3
		Pen-53	0	0	4	3	0
Source 2	Block 3	Pen-54	1	1	4	10	8
		Pen-55	2	1	1	7	1
		Pen-56	1	1	5	8	0
	Block 4	Pen-57	0	3	6	3	0
		Pen-58	1	7	0	0	1
		Pen-59	0	2	1	3	0

Lubbock(ST413)

			Day 0 fecal	Day 7 fecal	Terminal fecal	Lymph node	Hide
Source 1	Block 1	Pen-7	0	0	0	3	0
		Pen-8	0	0	4	3	4
		Pen-9	2	2	0	1	1
	Block 2	Pen-51	0	0	10	6	8
		Pen-52	4	0	4	4	5
		Pen-53	0	0	4	2	9
Source 2	Block 3	Pen-54	0	0	0	0	0
		Pen-55	1	0	0	0	0
		Pen-56	0	0	2	0	0
	Block 4	Pen-57	0	0	3	0	11
		Pen-58	1	0	10	5	9
		Pen-59	1	0	5	1	10

Cerro (ST367)

			Day 0 fecal	Day 7 fecal	Terminal fecal	Lymph node	Hide
Source 1	Block 1	Pen-7	0	0	4	4	7
		Pen-8	0	0	0	3	2
		Pen-9	0	0	0	0	3
	Block 2	Pen-51	0	0	0	0	1
		Pen-52	0	0	1	0	0
		Pen-53	0	0	0	0	0
Source 2	Block 3	Pen-54	0	0	0	0	0
		Pen-55	0	0	8	2	9
		Pen-56	0	0	3	0	11
	Block 4	Pen-57	0	0	0	0	0
		Pen-58	0	0	0	0	1
		Pen-59	0	0	3	1	1

Montevideo (ST138)

			Day 0 fecal	Day 7 fecal	Terminal fecal	Lymph node	Hide
Source 1	Block 1	Pen-7	6	3	0	2	0
		Pen-8	2	0	1	5	0
		Pen-9	1	0	10	10	4
	Block 2	Pen-51	3	0	0	0	0
		Pen-52	0	0	0	0	2
		Pen-53	4	0	0	0	0
Source 2	Block 3	Pen-54	7	0	0	0	0
		Pen-55	3	0	0	0	0
		Pen-56	1	1	0	0	0
	Block 4	Pen-57	1	0	0	0	0
		Pen-58	0	0	0	0	0
		Pen-59	2	0	0	0	0

988

989

990 Table 3. Distribution of isolates harboring antibiotic resistance genes, plasmids, and
991 pathogenicity islands by serotype.

992

		Anatum (n=113)	Cerro (n=64)	Kentucky (n=11)	Lubbock (n=136)	Montevideo (n=68)	Newport (n=6)	Norwich (n=1)
Resistance genes*	*tet*(A)_6	0	0	0	0	67	0	0
Plasmids	IncN_1	0	0	0	0	67	0	0
	IncI1_1_Alpha	16	3	0	43	2	0	0
Pathogenicity islands	C63PI	0	0	0	0	0	6	1
	SPI-1	113	64	11	136	68	6	1
	SPI-2	111	64	11	136	68	6	1
	SPI-3	113	63	11	0	68	6	1
	SPI-4	112	54	10	104	68	6	1
	SPI-5	113	64	11	136	68	6	1
	SPI-8	0	0	11	0	0	0	0
	SPI-9	113	64	11	136	68	6	1
	SPI-13	113	0	0	0	68	6	1
	SPI-14	113	0	0	0	68	6	1

993 *Cryptic aminoglycoside gene *aac(6')-Iaa_1* was in all *Salmonella* isolates.

994

995 Resistance genes were identified using the ResFinder Database updated on 06-03-
996 2020 with the minimum threshold set for 97% for ID and 97% for coverage match.
997 Plasmids were identified through the PlasmidFinder database updated on 06-03-2020
998 with the minimum threshold set for 97% for ID and 80% for coverage match.
999 Pathogenicity islands were identified through the SPIFinder database updated on 07-
1000 24-2017 with the minimum threshold set for 97% for ID% and 60% for coverage match.
1001

Figure 1**.** Phylogenetic analysis of 136 *Salmonella* Lubbock isolates, based on maximum-likelihood analysis of 84 SNP sites; of these, 36 were identified as parsimony informative and the remaining 48 were singleton sites.

The tree scale (0.0001) was calculated to be equivalent to approximately 2.8 nucleotide substitutions per site. Full tree is presented in a miniaturized boxed view; expanded here, Clade I is in the middle and Clade II is on the right. Those branches with bootstrap support values of 80-100% are represented with a grey circle located in the middle of the corresponding branch and sized proportional to the given support values. Pen (first column) and cattle source (second column) colors are presented in the legends. Sampling days are presented in the next six columns. Sample types are indicated in the last three columns.

Figure 2**.** Phylogenetic analysis of 112 *Salmonella* Anatum isolates based on maximum-likelihood analysis of 66 SNP sites; of these, 36 were identified as parsimony informative and the remaining 30 were singleton sites.

The tree scale (0.0001) shows approximately 2.2 nucleotide substitutions per site. Those branches with bootstrap support values of 80-100% are presented with a grey circle located in the middle of the corresponding branch and sized proportional to the given support values. Pen (first column) and cattle source (second column) colors are presented in the legends. Sampling days are presented in the next six columns. Sample types are indicated in the last three columns.

1023

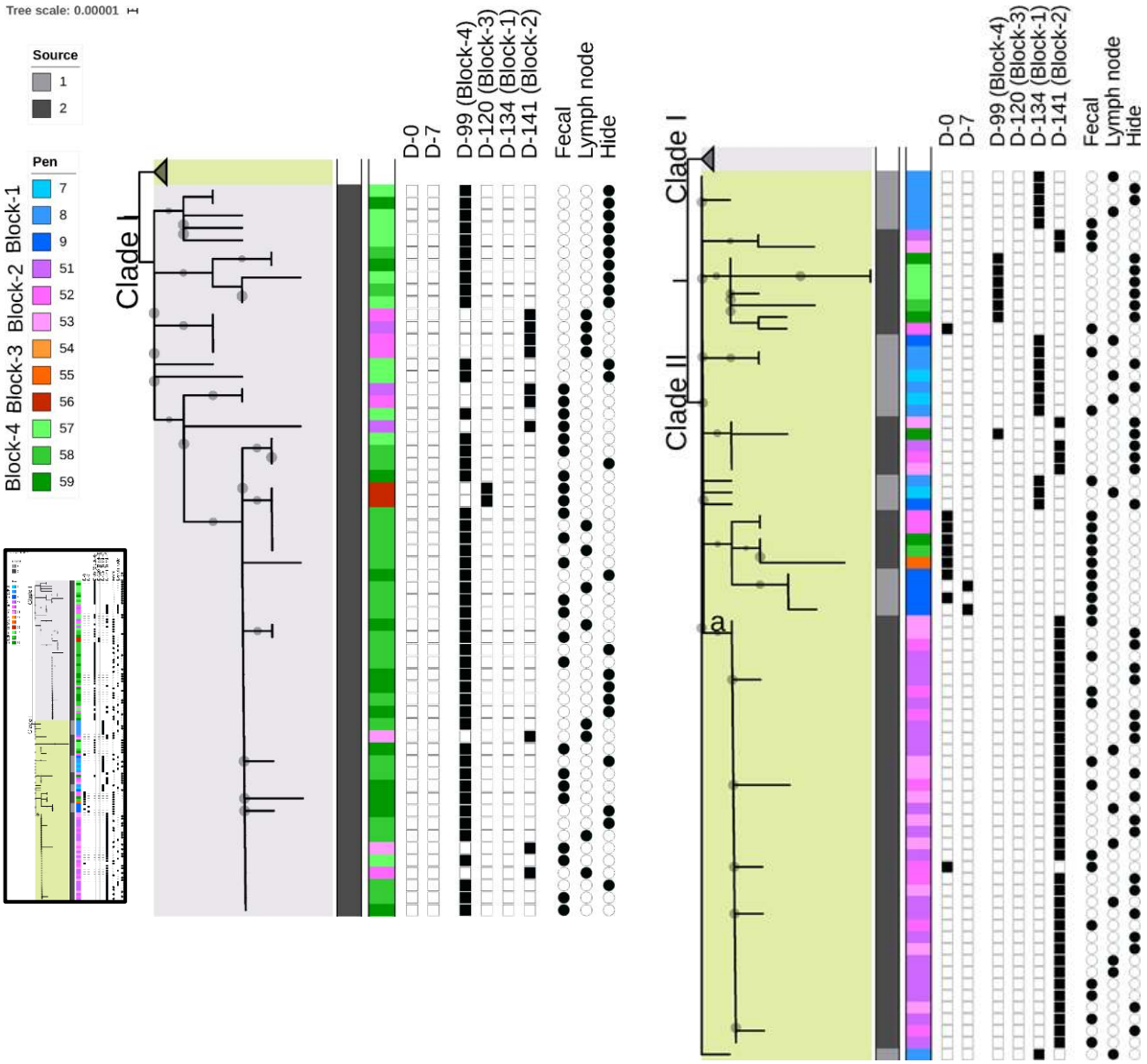
1024 Figure 3**.** Phylogenetic analysis of 68 *Salmonella* Montevideo isolates based on
1025 maximum-likelihood analysis of 12 SNP sites; of these, one was identified as parsimony
1026 informative and the remaining 11 were singleton sites.

1027 The tree scale of 0.0001 is equivalent to approx. 0.4 nucleotide substitutions per site.
1028 Those branches with bootstrap support values of 80-100% are presented with a grey
1029 circle located in the middle of the corresponding branch and sized proportional to the
1030 given support values. Pen (first column) and cattle source (second column) colors are
1031 presented in the legends. Sampling days are presented in the next six columns. Sample
1032 types are indicated in the last three columns.

1033

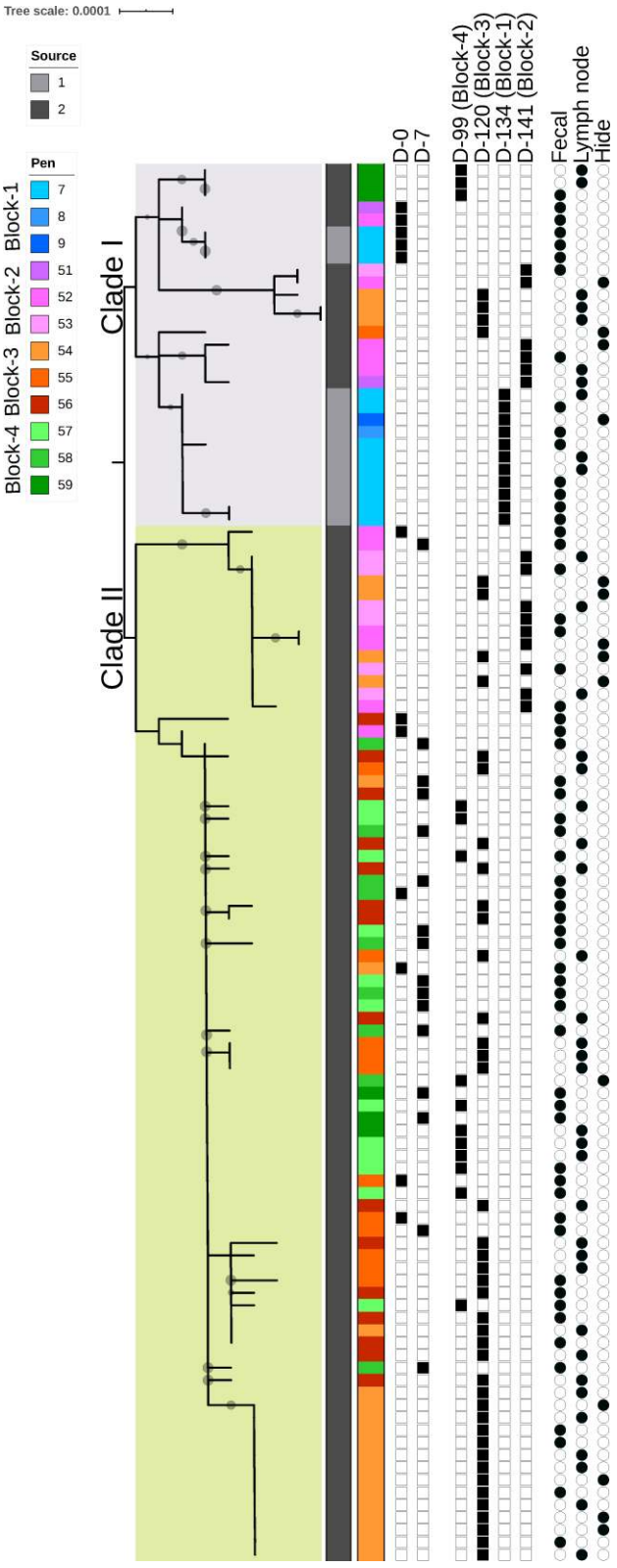
1034 Figure 4**.** Phylogenetic analysis of 64 *Salmonella* Cerro isolates based on maximum-
1035 likelihood analysis of 17 SNP sites; of these, 9 were identified as parsimony informative
1036 and the remaining 8 were singleton sites.

1037 The tree scale (0.0001) shows approximately 0.4 nucleotide substitutions per site. S.
1038 Cerro isolates were derived only at slaughter age in this study. Those branches with
1039 bootstrap support values of 80-100% are presented with a grey circle located in the
1040 middle of the corresponding branch and sized proportional to the given support values.
1041 Pen (first column) and cattle source (second column) colors are presented in the
1042 legends. Sampling days are presented in the next four columns. Sample types are
1043 indicated in the last three columns

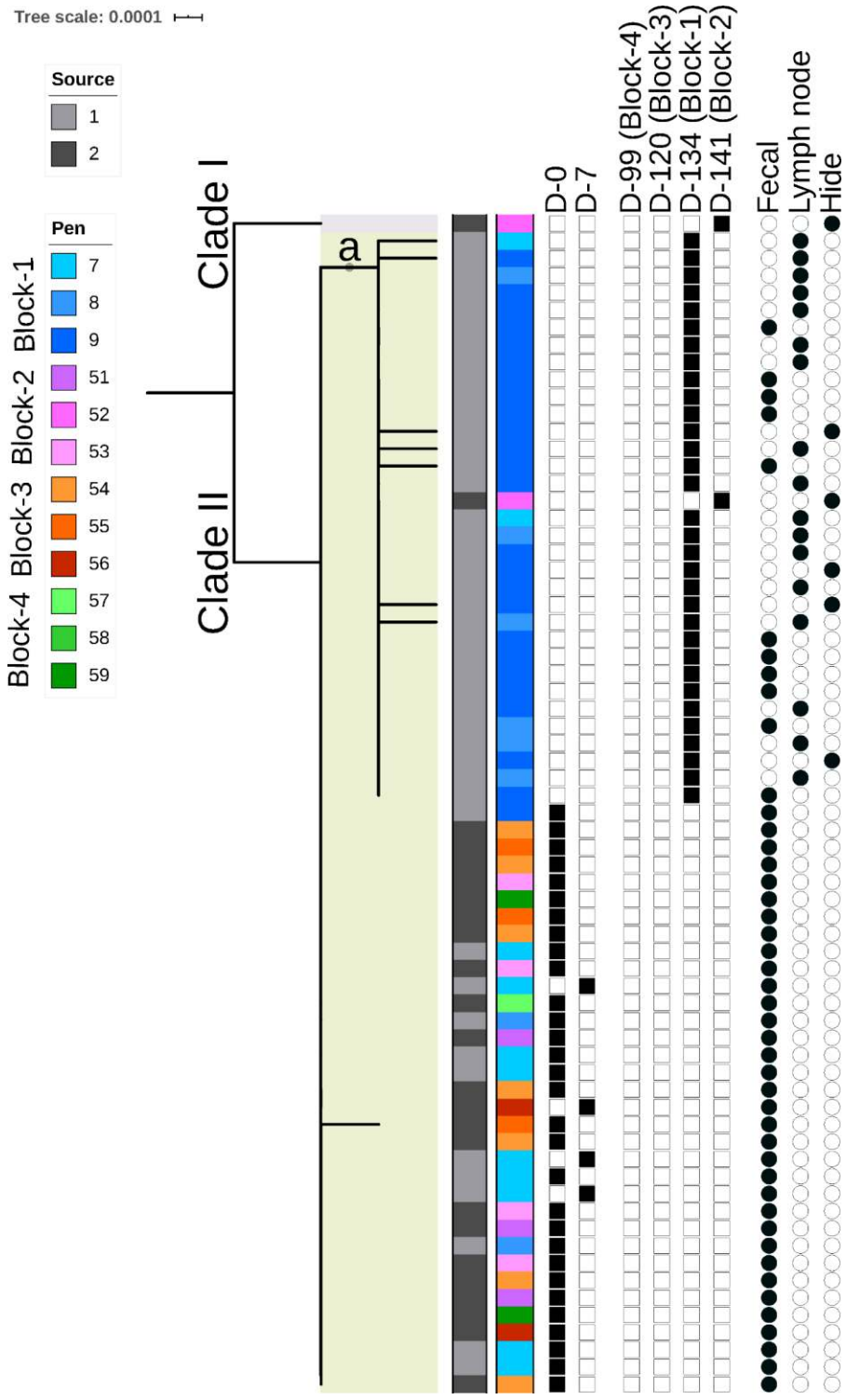


1044

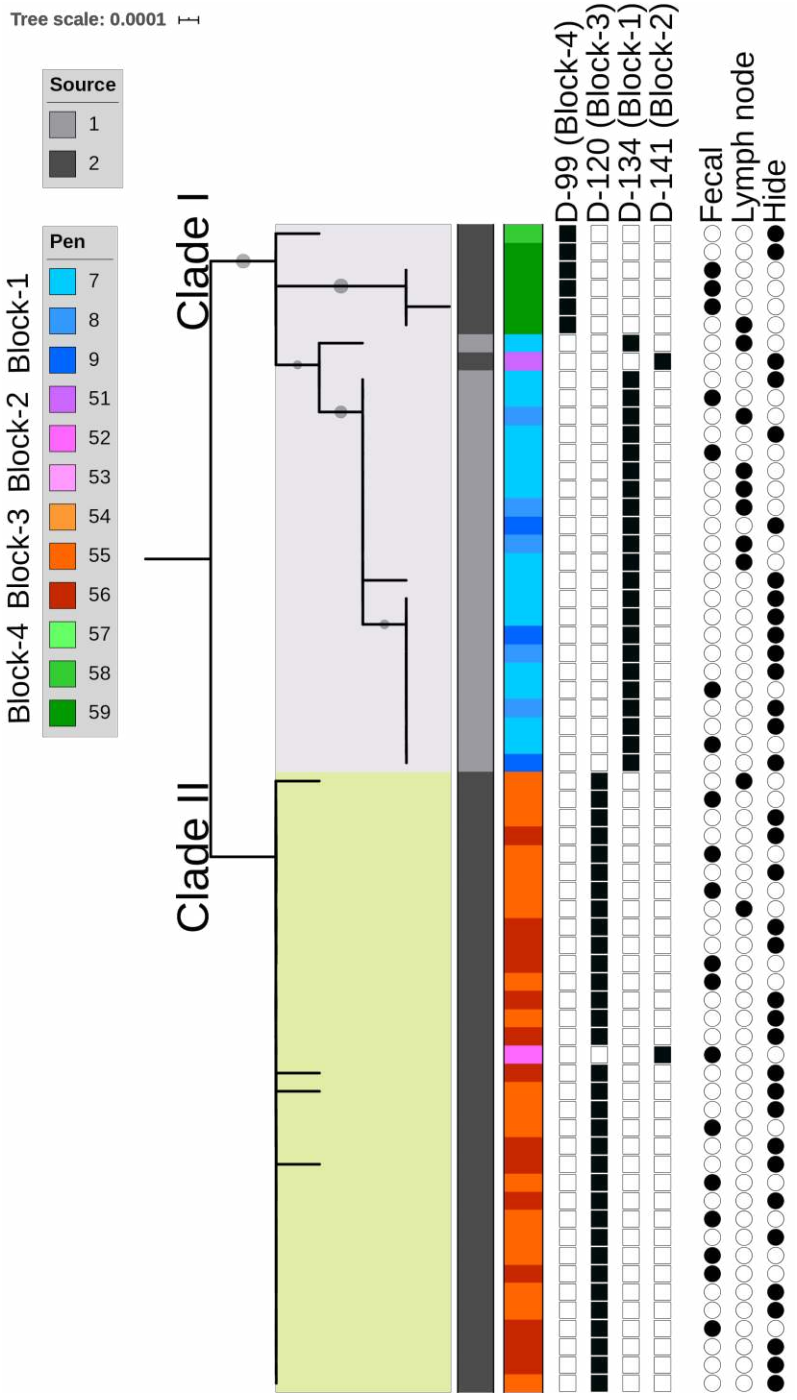
1045



1046



1047



1048