A multidrug-resistant *Salmonella enterica* Typhimurium DT104 complex lineage circulating among humans and cattle in the United States lost the ability to produce pertussis-like toxin ArtAB

Laura M. Carroll[1,2,3,4], Nicolo Piacenza[5], Rachel A. Cheng[6], Martin Wiedmann[7], Claudia Guldimann[5,#]

[1]Department of Clinical Microbiology, SciLifeLab, Umeå University, Umeå, Sweden

[2]Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå University, Umeå, Sweden

[3]Umeå Centre for Microbial Research, Umeå University, Umeå, Sweden

[4]Integrated Science Lab, Umeå University, Umeå, Sweden

[5]Chair for Food Safety and Analytics, Ludwig-Maximillians-University Munich, Munich, Germany

[6]Department of Food Science and Technology, Virginia Tech, Blacksburg, Virginia, USA

[7]Department of Food Science, Cornell University, Ithaca, New York, USA

[#]Corresponding author: Claudia Guldimann, c.guldimann@lmu.de

Keywords: *Salmonella* Typhimurium, DT104, ArtAB, GogB, Gifsy-1, prophage

Repositories: Not applicable

Abstract

Salmonella enterica subspecies *enterica* serotype Typhimurium definitive type 104 (DT104) can infect both humans and animals and is often multidrug-resistant (MDR). Previous studies have indicated that, unlike most *S.* Typhimurium, the overwhelming majority of DT104 strains produce pertussis-like toxin ArtAB via prophage-encoded genes *artAB*. However, DT104 that lack *artAB* have been described on occasion. Here, we identify a MDR DT104 complex lineage circulating among humans and cattle in the United States, which lacks *artAB* (i.e., the "U.S. *artAB*-negative major clade"; $n = 42$ genomes). Unlike most other bovine- and human-associated DT104 complex strains from the U.S. ($n = 230$ total genomes), which harbor *artAB* on prophage Gifsy-1 ($n = 177$), members of the U.S. *artAB*-negative major clade lack Gifsy-1, as well as anti-inflammatory effector *gogB*. The U.S. *artAB*-negative major clade encompasses human- and cattle-associated strains isolated from ≥11 U.S. states over a twenty-year period. The clade was predicted to have lost *artAB*, Gifsy-1, and *gogB* circa 1985-1987 (95% highest posterior density interval 1979.0-1992.1). When compared to DT104 genomes from other world regions ($n = 752$ total genomes), several additional, sporadic *artAB*, Gifsy-1, and/or *gogB* loss events among clades encompassing ≤5 genomes were observed. Using phenotypic assays that simulate conditions encountered during human and/or bovine digestion, members of the U.S. *artAB*-negative major clade did not differ from closely related Gifsy-1/*artAB*/*gogB*-harboring U.S. DT104 complex strains (ANOVA raw P-value > 0.05); thus, future research is needed to elucidate the roles that *artAB*, *gogB*, and Gifsy-1 play in DT104 virulence in humans and animals.

53 **Impact Statement**

54 Multi-drug resistant (MDR) *Salmonella enterica* serotype Typhimurium definitive type 104
55 (DT104) was responsible for a global epidemic among humans and animals throughout the
56 1990s and continues to circulate worldwide. Previous studies have indicated that the vast
57 majority of DT104 produce pertussis-like toxin ArtAB via prophage-encoded *artAB*. Here, we
58 identify a DT104 complex lineage that has been circulating among cattle and humans across ≥11
59 U.S. states for over twenty years, which lacks the ability to produce ArtAB (i.e., the "U.S. *artAB*-
60 negative major clade"). The common ancestor of all U.S. *artAB*-negative major clade members
61 lost the ability to produce ArtAB in the 1980s; however, the reason for this loss-of-function
62 event within this well-established pathogen remains unclear. The role that ArtAB plays in DT104
63 virulence remains elusive, and phenotypic assays conducted here indicate that members of the
64 U.S. *artAB*-negative major clade do not have a significant advantage or disadvantage relative to
65 closely related, Gifsy-1/*artAB*/*gogB*-harboring U.S. DT104 complex strains when exposed to
66 stressors encountered during human and/or bovine digestion *in vitro*. However, ArtAB
67 heterogeneity within the DT104 complex suggests clade-specific selection for or against
68 maintenance of ArtAB. Thus, future studies querying the virulence characteristics of the U.S.
69 *artAB*-negative major clade are needed.

70

71 **Data Summary.** Supplementary Data is available under DOI 10.5281/zenodo.7688792, with
72 URL https://doi.org/10.5281/zenodo.7688792.

73

74 **INTRODUCTION**

75 Prophages, which are viruses located within the genomes of bacteria, play important roles
76 in the evolution of their microbial hosts [1-4]. In addition to possessing machinery that is
77 antagonistic to host cell survival (e.g., virion production, lysis of host cells), many prophages
78 encode accessory genes, which may provide the host with a selective advantage [1, 3, 5],
79 including stress tolerance, resistance to antimicrobials and phage, biofilm formation, increased
80 virulence, and evasion of the host immune system [1, 2, 4-8]. While they may persist within a
81 lineage through vertical transmission [5, 6, 9], prophages can undergo gain and loss events
82 within a population over time [1, 3]. Furthermore, integrated prophages can be hotspots for
83 horizontal gene transfer (HGT) and genomic recombination, allowing their bacterial hosts to
84 gain, lose, and exchange genetic information [4]. Thus, prophage-mediated HGT may confer
85 novel functions, which allow the bacterial host to survive and compete in its environment,
86 potentially contributing to the emergence of novel epidemic lineages [4, 10].

87 *Salmonella enterica* subsp. *enterica* serotype Typhimurium (*S.* Typhimurium) is among
88 the *Salmonella* serotypes most commonly isolated from human and animal salmonellosis cases
89 worldwide [11, 12] and is known to host a range of prophages within its chromosome [10]. Of
90 particular concern is *S.* Typhimurium definitive type 104 (DT104), a lineage within *S.*
91 Typhimurium that is known for its typical ampicillin-, chloramphenicol-, streptomycin-,
92 sulfonamide-, and tetracycline-resistant (ACSSuT) phenotype, although its antimicrobial
93 resistance (AMR) profile may vary [13]. Multidrug-resistant (MDR) DT104 is predicted to have
94 emerged circa 1972 [13] and rapidly disseminated around the world in the following decades
95 [13-15], culminating in a global epidemic among animals and humans in the 1990s [13-15].

96 However, despite its rapid global dissemination, DT104 does not appear to be more virulent than
97 non-DT104 *S.* Typhimurium in a classical mouse model [16].
98 In addition to its characteristic MDR phenotype, DT104 is notable for its ability to
99 produce ArtAB, a pertussis-like toxin that catalyzes ADP-ribosylation of host G proteins [17-19].
100 Treatment of various cell lines with purified ArtAB from DT104 recapitulates some of the
101 phenotypes established for pertussis toxin cytotoxicity [20-22], such as the characteristic "cell
102 clustering" phenotype in CHO-K1 cells [23], increased levels of intracellular cAMP in RAW
103 264.7 macrophage-like cells [18], and increased serum insulin levels (e.g., insulinemia); further,
104 intraperitoneal injection of purified toxin in neonatal mice was fatal [18].
105 *artAB,* which encodes ArtAB, has been detected in representative strains of at least 88
106 *Salmonella* serotypes [24], and previous studies have found that *artAB* can be encoded by
107 prophages (e.g., Gifsy-1, PhInv-1b) [17-19, 25-27]. Within *S.* Typhimurium specifically, *artAB*
108 shares a strong association with DT104 relative to other *S.* Typhimurium lineages: while
109 typically absent in most non-DT104 *S.* Typhimurium strains, the overwhelming majority of
110 DT104 possess *artAB* [18]. In DT104 specifically, *artAB* has been identified within prophage
111 Gifsy-1 (Figure 1A and Supplementary Figure S1) [10, 19, 25]. Gifsy-1 has been detected in
112 numerous *Salmonella* serotypes [28] and has been shown to harbor virulence factors [29] such as
113 *gogB*, *gipA*, and *gtgA* [29-31]; however, *artAB*-harboring Gifsy-1 has been proposed to be a
114 characteristic feature of DT104 (Figure 1A and Supplementary Figure S1) [10].
115 Despite their strong association, DT104 strains that lack *artAB* and, thus, the ability to
116 produce ArtAB toxin, have been described on occasion (referred to hereafter as "*artAB*-
117 negative" strains) [18, 32]. We identified three *artAB*-negative DT104 complex strains in a
118 previous study of *S.* Typhimurium from cattle and humans in New York State (United States,

119 U.S.) [32]. Because *artAB* tends to be prophage-encoded [17-19, 25, 26], we hypothesized that it
120 may be possible for *artAB* to be gained or lost as an *artAB*-harboring prophage integrates or
121 excises from a genome, or via HGT within an integrated prophage. However, the extent to which
122 any of these scenarios occur is unknown. Using (i) 230 human- and bovine-associated DT104
123 complex genomes collected across the U.S., plus (ii) 752 DT104 complex genomes collected
124 from a range of sources worldwide, we provide large-scale insight into the dynamics of *artAB*
125 acquisition and loss within the DT104 complex.
126
127 **METHODS**
128 **Acquisition of U.S. human- and bovine-associated DT104 complex genomic data and**
129 **metadata.** In a previous study of human- and bovine-associated *S.* Typhimurium from New
130 York State, we identified three closely related, *artAB*-negative DT104 complex genomes from
131 both humans and cattle (out of 14 total DT104 complex genomes from humans and cattle in New
132 York State) [32]. Thus, as a first evaluation of *artAB* presence and absence in the DT104
133 complex, we compiled a set of DT104 complex genomes from humans and cattle across the U.S.
134 (referred to hereafter as "Dataset 1 [U.S. bovine and human data]"; Figure 1B and
135 Supplementary Figures S2 and S3).
136 To construct Dataset 1 (U.S. bovine and human data), we first collected genomic data
137 derived from 14 human- and bovine-associated DT104 complex isolates from New York State,
138 which we had sequenced in a previous study (members of the *S.* Typhimurium Lineage III
139 cluster described in Supplementary Figures S2 and S5 of Carroll, et al.) [32]. We then aggregated
140 these 14 New York State genomes with 223 human- and bovine-associated DT104 complex
141 genomes from across the U.S., as described previously [32]. Briefly, paired-end Illumina short

142 reads associated with 223 *S.* Typhimurium genomes meeting the following criteria were
143 downloaded from the National Center for Biotechnology Information (NCBI) Sequence Read
144 Archive (SRA; https://www.ncbi.nlm.nih.gov/sra, accessed November 29, 2018), using
145 accession numbers provided by Enterobase and the SRA Toolkit version 2.9.3 [33-36]: (i)
146 genomes were serotyped as *S.* Typhimurium *in silico* using the implementation of SISTR [37] in
147 Enterobase; (ii) the country of isolation was the United States; (iii) the isolation source was
148 reported as either "Human" or "Bovine" in the "Source Niche" and "Source Type" fields in
149 Enterobase, respectively; (iv) genomes had an isolation year reported in Enterobase; (v) using
150 RhierBAPS [38], genomes were assigned to the DT104 complex, a well-supported cluster within
151 the larger bovine- and human-associated U.S. *S.* Typhimurium phylogeny, which clustered
152 among known DT104 genomes from other countries (see Supplementary Figures S2 and S5 of
153 Carroll, et al.) [32].
154 Trimmomatic version 0.36 [39] was used to trim low quality bases and Illumina adapters
155 from all read sets using the default settings for paired-end reads, and SPAdes version 3.13.0 [40]
156 was used to assemble all genomes using default settings plus the "careful" option. FastQC
157 version 0.11.5 [41] and QUAST version 4.0 [42] were used to assess the quality of each read pair
158 set and assembly, respectively, and MultiQC version 1.6 [43] was used to aggregate all FastQC
159 and QUAST results. Trimmed paired-end read sets/assemblies that were flagged by MultiQC as
160 meeting any of the following conditions were excluded: (i) Illumina adapters present after
161 trimming (n = 2), (ii) an abnormal per sequence GC content distribution (n = 3), (iii) an assembly
162 with over 200 contigs (n = 11), and (iv) a sequence quality histogram flagged as poor quality (n
163 = 2). After excluding genomes that met these conditions, a set of 219 DT104 complex genomes
164 was produced (Figure 1B, Supplementary Figures S2 and S3).

165 Finally, the 219 U.S. human- and bovine-associated DT104 complex genomes identified
166 here were supplemented with 11 U.S. bovine- and human-associated DT104 genomes from a
167 previous study [13], which did not have metadata available in Enterobase at the time and were
168 thus not included in the initial set of 219 bovine- and human-associated U.S. DT104 complex
169 genomes. Overall, the search conducted here produced a set of 230 bovine- and human-
170 associated U.S. DT104 complex genomes, which were used in subsequent steps (i.e., Dataset 1
171 [U.S. bovine and human data]; Figures 1B and 2A, Supplementary Figures S2 and S3, and
172 Supplementary Tables S1 and S2).

173 ***In silico* detection of prophage, antimicrobial resistance genes, plasmid replicons, and**
174 **virulence factors.** To identify putative prophage regions in all 230 genomes in Dataset 1 (U.S.
175 bovine and human data), each assembly was submitted to the PHASTER web server
176 (https://phaster.ca/) via the URL API [44, 45], with the "contigs" option set to "1"
177 (Supplementary Table S3). To compare prophage regions identified in Dataset 1 (U.S. bovine
178 and human data) genomes to previously described prophage in well-characterized *S.*
179 Typhimurium strains, prophage in the following *S.* Typhimurium strains were obtained from the
180 PHASTER prophage database (accessed September 18, 2020): (i) LT2 (NCBI Nucleotide
181 accession NC_003197.2), (ii) DT104 (NCBI Nucleotide accession NC_022569.1), (iii) D23580
182 (NCBI Nucleotide accession FN424405.1), and (iv) SL1344 (NCBI Nucleotide accession
183 NC_016810.1; Figure 1A and Supplementary Figure S1). All prophage regions were annotated
184 using Prokka version 1.14.6 [46], using default settings and the "Viruses" kingdom database.
185 The resulting GFF and FNA files produced by Prokka were supplied to clinker version 0.0.26
186 [47], which was used to perform pairwise alignments of all genes within prophage regions using
187 default settings.

188 ABRicate version 0.8 [48] was used to detect antimicrobial resistance (AMR) genes,
189 plasmid replicons, and virulence factors in each assembled DT104 complex genome using
190 NCBI's National Database of Antibiotic Resistant Organisms (NDARO) [49], the PlasmidFinder
191 database [50], and the Virulence Factor Database (VFDB) [51], respectively, using minimum
192 nucleotide identity and coverage thresholds of 75 and 50%, respectively (all databases accessed
193 December 10, 2020; Supplementary Table S3). The aforementioned ABRicate analyses were
194 repeated, using a minimum coverage threshold of 0% (e.g., to confirm that virulence factors
195 discussed in the manuscript were absent from genomes in which they were not initially detected).
196 Each assembled genome was additionally queried for the presence of selected virulence
197 factors, which have previously been associated with prophage in *Salmonella* [29]: (i) *artAB*
198 (NCBI Nucleotide accession AB104436.1), (ii) *gogA* (European Nucleotide Archive [ENA]
199 accession EAA7850902.1), (iii) *gtgA* (ENA accession PVI70081.1), and (iv) *gipA* (ENA
200 accession CAI93790.1). Assembled genomes were queried for selected virulence factors using
201 the command-line implementation of nucleotide BLAST (blastn) version 2.11.0 [52], using
202 default settings plus a minimum coverage threshold of 40% (Supplementary Table S3). To
203 confirm that the aforementioned genes were absent from genomes in which they were not
204 initially detected, all genomes were queried again (i) as described above, with the coverage
205 threshold lowered to 0%; and (ii) using translated nucleotide BLAST (tblastx; Supplementary
206 Tables S4 and S5). ARIBA version 2.14.6 [53] was used to further confirm *artAB* and *gogB*
207 presence/absence in all genomes with associated paired-end Illumina reads (Supplementary
208 Table S6 and Supplementary Text).
209 **Variant calling and maximum likelihood phylogeny construction within Dataset 1 (U.S.**
210 **bovine and human data).** Core single nucleotide polymorphisms (SNPs) were identified among

211 all 230 genomes within Dataset 1 (U.S. bovine and human data), using the default pipeline
212 implemented in Snippy version 4.6.0 (https://github.com/tseemann/snippy; Supplementary
213 Tables S1 and S2 and Supplementary Text) [54-67]. The closed DT104 chromosome (NCBI
214 Nucleotide accession NC_022569.1) was used as a reference, and core SNPs identified in
215 regions of the DT104 chromosome predicted to belong to phage were masked (Supplementary
216 Text). Gubbins version 2.4.1 [68] was used to identify and remove recombination events in all
217 genomes using default settings, and snp-sites was used to query the resulting recombination-free
218 alignment for core SNPs (i.e., using the "-c" option).

219 A maximum likelihood (ML) phylogeny was constructed with IQ-TREE version 1.5.4
220 [69], using (i) the resulting core SNPs as input, (ii) the optimal nucleotide substitution model
221 selected using ModelFinder [70, 71], (iii) an ascertainment bias correction to account for the use
222 of solely variant sites, and (iv) 1,000 replicates of the ultrafast bootstrap approximation
223 (Supplementary Text) [72, 73]. TempEst version 1.5.3 [74] was used to assess the temporal
224 structure of the resulting unrooted ML phylogeny, using the best-fitting root and the R^2 function
225 (R^2 = 0.33, slope = 3.05×10^{-7} substitutions/site/year, X-intercept = 1988.1). The unrooted ML
226 phylogeny was additionally rooted and time scaled using LSD2 version 1.4.2.2 [75], using tip
227 dates corresponding to the year of isolation reported for each genome (Supplementary Text). The
228 resulting rooted, time-scaled ML phylogeny was viewed using FigTree version 1.4.4 [76]
229 (Supplementary Data).

230 **Dataset 1 (U.S. bovine and human data) Bayesian time-scaled phylogeny construction.** In
231 addition to constructing a time-scaled ML phylogeny (see section "Variant calling and maximum
232 likelihood phylogeny construction within Dataset 1 [U.S. bovine and human data]" above), a
233 Bayesian approach was additionally employed to construct a time-scaled phylogeny, using a

234 subset of 146 Dataset 1 (U.S. bovine and human data) genomes (Supplementary Figure S4 and
235 Supplementary Text). All aforementioned SNP calling and ML phylogeny construction steps
236 were repeated within the 146-genome Dataset 1 (U.S. bovine and human data) subset, and the
237 resulting ML phylogeny was time-scaled using TempEst and LSD2 as described above (see
238 section "Variant calling and maximum likelihood phylogeny construction within Dataset 1 [U.S.
239 bovine and human data]" above; Supplementary Table S7, Supplementary Data, and
240 Supplementary Text).

241 BEAST2 version 2.5.1 [77, 78] was used to construct a tip-dated phylogeny, using core
242 SNPs detected among the 146-genome Dataset 1 (U.S. bovine and human data) subset as input
243 (Supplementary Text). An initial clock rate of 2.79×10^{-7} substitutions/site/year [13] was used,
244 along with an ascertainment bias correction to account for the use of solely variant sites [79].
245 bmodeltest [80] was used to infer a substitution model using Bayesian model averaging, with
246 transitions and transversions split. A relaxed lognormal molecular clock [81] and a coalescent
247 Bayesian skyline population model [82] were used, as these models have been selected as the
248 optimal clock/population model combination for DT104 previously [13] (Supplementary Text).
249 Five independent BEAST2 runs (i.e., BEAST2 runs with different random seeds) were
250 performed, using chain lengths of at least 100 million generations, sampling every 10 thousand
251 generations. LogCombiner-2 was used to aggregate the resulting log and tree files with 10% of
252 the states treated as burn-in, and TreeAnnotator-2 was used to produce a maximum clade
253 credibility (MCC) tree using Common Ancestor node heights (Supplementary Figure S5,
254 Supplementary Table S8, and Supplementary Data). The resulting phylogenies were displayed
255 and annotated using R version 4.1.2 (Figures 3 and 4, Supplementary Figures S6-S9, and
256 Supplementary Text) [83-88].

257 ***artAB* ancestral state reconstruction for Dataset 1 (U.S. bovine and human data).** To
258 estimate ancestral character states of internal nodes in the Dataset 1 (U.S. bovine and human
259 data) phylogeny as they related to *artAB* presence/absence (i.e., whether a node in the tree
260 represented an ancestor that was more likely to be *artAB*-positive or *artAB*-negative), the
261 presence or absence of *artAB* within each genome was treated as a binary state (see section "*In*
262 *silico* detection of prophage, antimicrobial resistance genes, plasmid replicons, and virulence
263 factors" above). *artAB* ancestral state reconstruction runs were performed using the BEAST2
264 time-scaled Bayesian Dataset 1 (U.S. bovine and human data) phylogeny as input (n = 146
265 genomes; see section "Dataset 1 [U.S. bovine and human data] Bayesian time-scaled phylogeny
266 construction" above; Supplementary Data and Supplementary Text). Stochastic character maps
267 were simulated on the phylogeny using the make.simmap function in the phytools version 1.0-1
268 R package [89] and the all-rates-different (ARD) model in the ape version 5.6-1 package [90,
269 91]; two root node priors were tested (Supplementary Text). The resulting phylogenies (one for
270 each root node prior) were plotted using the densityMap function in the phytools R package
271 (Supplementary Figures S10-S12, Supplementary Data).
272 **Pan-genome characterization of Dataset 1 (U.S. bovine and human data).** Prokka version
273 1.13.3 [46] was used to annotate all 230 genomes within Dataset 1 (U.S. bovine and human
274 data), using the "Bacteria" database and default settings (Supplementary Tables S1 and S2). GFF
275 files produced by Prokka were supplied as input to Panaroo version 1.2.7 [92], which was used
276 to identify core- and pan-genome orthologous gene clusters among the 230 Dataset 1 (U.S.
277 bovine and human data) genomes (Supplementary Text) [93, 94]. The LSD2 time-scaled ML
278 phylogeny for Dataset 1 (U.S. bovine and human data) (see section "Variant calling and
279 maximum likelihood phylogeny construction within Dataset 1 [U.S. bovine and human data]"

280 above) was supplied as input to Panaroo's "panaroo-img" and "panaroo-fmg" commands, which
281 were used to estimate the pan-genome size under the Infinitely Many Genes (IMG) [95, 96] and
282 Finite Many Genes (FMG) models (with 100 bootstrap replicates) [97], respectively
283 (Supplementary Figure S13).

284 Reference pan-genome coding sequences (CDS) identified by Panaroo underwent
285 functional annotation using the eggNOG-mapper version 2 webserver (http://eggnog-
286 mapper.embl.de/; accessed July 24, 2022) using default settings [98, 99]. The "table" function in
287 R was used to identify genes associated with (i) prophage Gifsy-1 presence/absence
288 (Supplementary Table S9) and (ii) clade membership (Supplementary Table S10); the
289 "fisher.test" function in R's stats package was used to conduct two-sided Fisher's exact tests, and
290 the "p.adjust" function was used to control the false discovery rate (FDR; i.e., p.adjust method =
291 "fdr") [100].

292 **Genome-wide identification of host-associated orthologous gene clusters for Dataset 1 (U.S.**
293 **bovine and human data).** The treeWAS version 1.0 R package [101] was used to identify
294 potential orthologous gene cluster-host associations among the 230 human- and bovine-
295 associated U.S. DT104 complex genomes in Dataset 1 (U.S. bovine and human data) (i.e.,
296 whether an orthologous gene cluster identified with Panaroo was human- or bovine-associated
297 while accounting for population structure; Supplementary Text). No orthologous gene clusters
298 were found to be significantly associated with isolation source via any of the treeWAS
299 association tests (FDR-corrected *P*-value > 0.10).

300 **Acquisition of global DT104 complex genomic data and metadata.** To compare the 230 U.S.
301 human- and bovine-associated DT104 complex genomes in Dataset 1 (U.S. bovine and human
302 data) to a larger set of DT104 complex genomes from numerous sources worldwide, genomic

303 data associated with the following studies were downloaded via Enterobase: (i) 243 bovine- and
304 human-associated DT104 isolates from a study of between-host transmission within Scotland
305 [102] (referred to hereafter as "Dataset 2 [Scottish bovine and human data]"; Supplementary
306 Text); (ii) 290 DT104 isolates from a variety of sources and countries from a study describing
307 the global spread of DT104 [13] (referred to hereafter as "Dataset 3 [multi-source data]"; eleven
308 of the 290 genomes were isolated from cattle and humans in the U.S. and thus had also been
309 included in Dataset 1 [U.S. bovine and human data], Figure 1B, Supplementary Figures S2 and
310 S3, Supplementary Table S1, and Supplementary Text).

311 The following datasets were aggregated to create a final set of 752 DT104 complex
312 genomes derived from numerous countries and isolation sources, which was used in subsequent
313 steps (referred to hereafter as "Dataset 4 [combined global dataset]"; Figure 1B, Supplementary
314 Figures S2 and S3, and Supplementary Tables S1 and S2): (i) Dataset 1 (U.S. bovine and human
315 data) (n = 230 DT104 complex genomes), (ii) Dataset 2 (Scottish bovine and human data) (n =
316 243 DT104 genomes), and (iii) Dataset 3 (multi-source data) (n = 290 DT104 genomes,
317 including 11 genomes that were part of Dataset 1 [U.S. bovine and human data]). QUAST
318 version 4.5 was used to assess the quality of all 752 genomes in Dataset 4 (combined global
319 dataset) (Supplementary Tables S1 and S2). Prophage, AMR genes, plasmid replicons, and
320 virulence factors were detected in all 752 Dataset 4 (combined global dataset) genomes as
321 described above (see section "*In silico* detection of prophage, antimicrobial resistance genes,
322 plasmid replicons, and virulence factors" above; Supplementary Tables S3-S5).

323 **Variant calling and maximum likelihood phylogeny construction within Dataset 4**
324 **(combined global dataset).** To identify core SNPs present in all 752 DT104 complex genomes
325 within Dataset 4 (combined global dataset), Parsnp and HarvestTools version 1.2 [103] were

326 used, as Parsnp easily scales to large data sets (Supplementary Tables S1 and S2) [103].
327 Assembled genomes were used as input for Parsnp, along with the closed DT104 chromosome as
328 a reference (NCBI Nucleotide accession NC_022569.1) and Parsnp's implementation of PhiPack
329 [104] to filter recombination.

330 Core SNPs detected among all 752 assembled genomes within Dataset 4 (combined
331 global dataset) were supplied as input to IQ-TREE version 1.5.4, which was used to construct a
332 ML phylogeny as described above; the resulting ML phylogeny was rooted and time-scaled
333 using LSD2 as described above (see section "Variant calling and maximum likelihood phylogeny
334 construction within Dataset 1 [U.S. bovine and human data]" above; Supplementary Data and
335 Supplementary Text). The resulting LSD2 time-scaled ML phylogeny was annotated using the
336 Interactive Tree of Life (iTOL) version 6 webserver (https://itol.embl.de/, accessed March 7,
337 2022; Figure 5, Supplementary Figure S14, and Supplementary Data) [105]. The LSD2 time-
338 scaled ML phylogeny for Dataset 4 (combined global dataset) was further used for *artAB*
339 presence/absence ancestral state reconstruction as described above (see section "*artAB* ancestral
340 state reconstruction for Dataset 1 [U.S. bovine and human data]" above; Supplementary Figure
341 S15).

342 **Pan-genome characterization of Dataset 4 (combined global dataset).** Pan-genome analyses
343 were carried out for Dataset 4 (combined global dataset) as described above (see section "Pan-
344 genome characterization of Dataset 1 [U.S. bovine and human data]" above; Supplementary
345 Tables S1 and S2 and Supplementary Text). The pan-genome size for Dataset 4 (combined
346 global dataset) was estimated using Panaroo's "panaroo-img" and "panaroo-fmg" commands,
347 using the LSD2 time-scaled ML phylogeny for Dataset 4 (combined global dataset) as input (see

348 section "Variant calling and maximum likelihood phylogeny construction within Dataset 4
349 [combined global dataset]" above; Supplementary Figure S13).
350 **Strain selection for phenotypic stress assays.** Phenotypic stress assays (discussed in detail in
351 the sections below) were used to compare (i) bovine- and human-associated, Gifsy-
352 1/*artAB*/*gogB*-positive U.S. DT104 complex strains to (ii) bovine- and human-associated, Gifsy-
353 1/*artAB*/*gogB*-negative U.S. DT104 complex strains. Thus, additional analyses were performed
354 to identify the most closely related, Gifsy-1/*artAB*/*gogB*-positive and -negative strains available
355 in the Cornell University Food Safety Laboratory (CUFSL) culture collection for phenotypic
356 testing (Supplementary Figures S2 and S3, Supplementary Table S2, and Supplementary Text)
357 [106-108]. Overall, the CUFSL DT104 complex genomes differed little in terms of their core and
358 pan-genome compositions (Supplementary Figure S16 and Supplementary Text).
359 Considering both (i) core- and pan-genome similarities between all 13 available CUFSL
360 DT104 complex genomes (Supplementary Figure S16 and Supplementary Table S2), as well as
361 (ii) Gifsy-1/*artAB*/*gogB* presence and absence, we selected six closely related, CUFSL DT104
362 complex strains to undergo phenotypic characterization: three Gifsy-1/*artAB*/*gogB*-positive
363 strains, and three Gifsy-1/*artAB*/*gogB*-negative strains (Figure 5, Supplementary Figure S16,
364 Supplementary Table S11, and Supplementary Text). All three Gifsy-1/*artAB*/*gogB*-negative
365 strains were members of the U.S. *artAB*-negative major clade (discussed in detail in the
366 "Results" section below; Figure 5 and Supplementary Table S11). All six selected strains had
367 been isolated from humans or cattle in New York State (U.S.) and were part of Dataset 1 (U.S.
368 bovine and human data) (Supplementary Figures S2 and S3 and Supplementary Tables S2 and
369 S11).

370 **Phenotypic assays.** Six carefully selected, human- and bovine-associated DT104 complex
371 strains from New York State (U.S.), which were available to us in the CUFSL culture collection,
372 were characterized using phenotypic assays (see section "Strain selection for phenotypic stress
373 assays" above; Supplementary Table S11). All strain stocks were maintained in CRYOBANK®
374 tubes (Mast Ltd., Reinfeld, Germany) at -80°C. Strains were streaked out from stocks on tryptic
375 soy agar (TSA; Merck KGaA, Darmstadt, Germany) and incubated overnight at 37°C. Single
376 colonies from those plates were inoculated in 5 mL of tryptic soy broth (TSB; Merck KGaA,
377 Darmstadt, Germany) and incubated for 16 - 18 h at 37°C with shaking at 200 rpm. The resulting
378 overnight cultures were diluted 1/100 into 5 mL of fresh, pre-warmed TSB, followed by
379 incubation at 37°C with shaking at 200 rpm to allow cultures to reach mid log phase (defined as
380 OD_{600} of 0.4; 1-2 x10^8 CFU/mL). These cultures were used as input into three different
381 phenotypic assays (exposure to ruminal fluid, acid stress, and bile stress; discussed in detail
382 below). Bacterial enumeration before and after stress exposure was performed by direct colony
383 counts of tilt plates according to Kühbacher et al. [109].
384 To evaluate exposure to ruminal fluid (RF), approximately 2 L of RF was acquired from
385 a Jersey cow with a ruminal fistula on each experimental day prior to the experiments (same
386 collection time was used for each experiment). The RF was immediately filtered through a
387 cellulose filter (Labsolute® Type 80, Th. Geyer GmbH& Co. KG., Renningen, Germany) to
388 remove any large debris, and the pH was measured, ranging from 7.20 to 7.62. Mid-log phase
389 cultures were prepared and inoculated into the RF at two different concentrations. One hundred
390 µl of culture suspensions were inoculated into 5mL of the RF at final concentrations of 10^8
391 (high) and 10^5 (low) CFU/mL and incubated for 1 h at 37°C without shaking with enumeration
392 by direct colony counting on XLT-4 agar (Oxoid Ltd., Basingstoke, UK) prior and after RF

393 exposure (Supplementary Table S12). The absence of *Salmonella* in the RF at the start of the
394 experiments was confirmed by plating on XLT-4 agar.

395 Acid stress resistance of the different strains at pH 3.5 with and without prior adaption
396 was tested using an adopted protocol from Horlbog et al. [110]. To carry out the acid stress
397 assay, the pH of the TSB was adjusted with hydrochloric acid solution (1M and 6 M HCL;
398 Merck KGaA, Darmstadt, Germany) immediately prior to the experiment. 1 mL aliquots of mid
399 log phase cultures were transferred to reaction tubes and centrifuged at 14,000 x g for 10 min.
400 For the non-adapted acid stress experiments, the pellets were resuspended in 1mL TSB pH 3.5
401 and incubated for 1 h at 37°C without shaking. For acid adaption, 1 mL of the same cultures
402 were pelleted, resuspended in 1 mL TSB adjusted to pH 5.5 and incubated for 1h at 37°C
403 (without shaking). Afterwards, the cultures were centrifuged again, resuspended in 1mL TSB pH
404 3.5, and incubated for 1 h at 37°C without shaking. Bacteria enumeration was performed before
405 and after the one-hour incubation at pH 3.5 (Supplementary Table S13).

406 Susceptibility to bile salts (cholic acid and deoxycholic acid in a mixture of 1:1, Bile
407 Salts No.3, Thermo Fisher Scientific Inc., Waltham, USA) was tested in two different
408 concentrations: 14.5 mmol/L corresponding to 0.6% [111] and 26.0 mmol/L corresponding to
409 1.1% [112] were chosen to represent reasonable physiological states in the duodenum. Bile salts
410 were added, and the pH of the TSB was adjusted to 5.5 (TSB-bile) immediately prior to the
411 experiment. Mid log phase cultures were centrifuged, resuspended in TSB-bile, incubated for 1h
412 37°C without shaking, and enumerated by direct colony counting prior and after bile exposure
413 (Supplementary Table S14).

414 For each stress assay, base-ten logarithmic fold change (FC) values were calculated as
415 follows: FC = log CFU/g at the start of the experiments – log CFU/g after the stress assay.

416 Analysis of Variance (ANOVA) for the interpretation of the phenotypic assays were conducted
417 using the "aov" function in R's "stats" package, with the FC values for the respective assay
418 treated as a response. Figures were designed using the ggplot2 package.
419 **Data availability.** Strain metadata, genome quality metrics, and Enterobase accession numbers
420 for all publicly available genomes queried in this study are available in Supplementary Table S1.
421 Strain metadata, genome quality metrics, CUFSL IDs [106], and NCBI BioSample accession
422 numbers [113] for the 13 New York State CUFSL DT104 complex strains queried in this study
423 (including those queried via phenotypic assays) are available in Supplementary Table S2. LSD2
424 results (for Dataset 1 [U.S. bovine and human data] and Dataset 4 [combined global dataset]) and
425 BEAST2 results (for subsets of Dataset 1 [U.S. bovine and human data]) are available as
426 Supplementary Data.
427
428 **RESULTS**
429 **Human- and bovine-associated DT104 complex strains from the U.S. harbor *artAB* on**
430 **prophage Gifsy-1.** Within the set of 230 human- and bovine-associated U.S. DT104 complex
431 genomes (i.e., Dataset 1 [U.S. bovine and human data]; Figure 2A) [32], *artAB* was present in
432 over 75% of genomes (177 of 230, 77.0%; Figure 3, Table 1, and Supplementary Figures S6-S8).
433 *artAB* presence and absence was strongly associated with the presence and absence of anti-
434 inflammatory effector *gogB* (two-sided Fisher's Exact Test [FET] raw P-value $< 2.2\times10^{-16}$,
435 infinite odds ratio [OR]), as co-occurrence was observed in all 177 *artAB*-harboring Dataset 1
436 (U.S. bovine and human data) genomes (100.0%; Figure 3, Table 1, and Supplementary Figures
437 S6-S8). Additionally, within Dataset 1 (U.S. bovine and human data), *artAB* and *gogB* presence
438 was strongly associated with the presence of prophage Gifsy-1 (NCBI Nucleotide accession

439 NC_010392.1; two-sided FET raw *P*-value < 2.2×10^{-16}, infinite OR; Figure 3, Table 1, and
440 Supplementary Figures S6-S8).

441 Subsequent investigation confirmed that, for all 177 *artAB*-harboring Dataset 1 (U.S.
442 bovine and human data) genomes, *artAB* was located within Gifsy-1 prophage regions classified
443 as "intact" via PHASTER (Table 2, Supplementary Figures S17 and S18, and Supplementary
444 Table S5). *gogB* was largely harbored within regions annotated via PHASTER as Gifsy-1 (126
445 of 180 *gogB*-harboring Dataset 1 [U.S. bovine and human data] genomes, 70.0%), although only
446 51 of these Gifsy-1 regions were annotated as intact prophage via PHASTER (28.3% of *gogB*-
447 harboring Dataset 1 [U.S. bovine and human data] genomes; Table 2, Supplementary Figure S19,
448 and Supplementary Table S5). Occasionally, among Dataset 1 (U.S. bovine and human data)
449 genomes, *gogB* was detected elsewhere in the genome: three genomes harbored *gogB* within
450 regions annotated as prophage Gifsy-2 (3 of 180 *gogB*-harboring Dataset 1 [U.S. bovine and
451 human data] genomes, 1.7%; Table 2 and Supplementary Table S5), while *gogB* was detected
452 outside of annotated prophage regions within the remaining 51 *gogB*-harboring genomes (via
453 PHASTER, 28.3% of *gogB*-harboring Dataset 1 [U.S. bovine and human data] genomes; Table 2
454 and Supplementary Table S5).

455 Only three genomes within Dataset 1 (U.S. bovine and human data) possessed an intact
456 Gifsy-1 prophage via PHASTER but did not possess *artAB* (i.e., bovine-associated
457 BOV_TYPH_Washington_2007_SRR1519881, BOV_TYPH_Minnesota_2010_SRR1089590,
458 and BOV_TYPH_Minnesota_2008_SRR1177378, 1.7% of Dataset 1 [U.S. bovine and human
459 data] genomes in which an intact Gifsy-1 was detected; Supplementary Figure S20 and
460 Supplementary Tables S3 and S4). Interestingly, all three genomes possessed *gogB*
461 (Supplementary Table S4). *gogB* was detected within an incomplete Gifsy-1 prophage region in

462 the two genomes from Minnesota, while the genome from Washington did not harbor *gogB*
463 within an annotated prophage region (via PHASTER; Tables 1 and 2, Supplementary Figure
464 S20, and Supplementary Table S5).

465 Of the 168 bovine-associated Dataset 1 (U.S. bovine and human data) genomes, 150
466 (89.3%) possessed *artAB, gogB*, and Gifsy-1, while 153 (91.1%) possessed *gogB* and Gifsy-1
467 (Figure 3, Table 1, and Supplementary Figures S6-S8). Interestingly, of 62 human-associated
468 Dataset 1 (U.S. bovine and human data) genomes, only 27 (43.5%) possessed *artAB, gogB,* and
469 Gifsy-1 (Figure 3, Table 1, and Supplementary Figures S6-S8), indicating that Gifsy-
470 1/*artAB*/*gogB* have a negative association with human-associated DT104 complex strains from
471 the U.S. (two-sided FET raw *P*-value < 4.1×10^{-12}, OR = 0.094; Table 1). However, no
472 orthologous gene clusters within the Dataset 1 (U.S. bovine and human data) pan-genome shared
473 a significant association with bovine or human host when accounting for population structure
474 (treeWAS FDR-corrected *P*-value > 0.10).

475 Overall, 90 orthologous gene clusters within the Dataset 1 (U.S. bovine and human data)
476 pan-genome were associated with Gifsy-1 presence or absence (via PHASTER; two-sided FET
477 FDR-corrected *P*-value < 0.05, Supplementary Figure S13 and Supplementary Table S9). The
478 presence and absence of 30 orthologous gene clusters shared a perfect association with Gifsy-1
479 presence and absence (via PHASTER; Supplementary Table S9). These genes were absent from
480 all Dataset 1 (U.S. bovine and human data) genomes that did not possess Gifsy-1 and were
481 present in all Dataset 1 (U.S. bovine and human data) genomes that did possess Gifsy-1 (FDR-
482 corrected *P*-value < 0.05 and OR of infinity); in addition to *gogB*, these genes included
483 numerous phage-associated proteins (Supplementary Table S9). Interestingly, genomes in which
484 PHASTER did not detect an intact Gifsy-1 prophage region tended to possess a ColRNAI

485 plasmid replicon (two-sided FET raw *P*-value < 1.0×10^{-27}; Supplementary Figures S6-S8 and
486 Supplementary Table S3).
487 **A MDR DT104 complex lineage circulating among cattle and humans across the U.S. lost**
488 ***artAB*- and *gogB*-harboring prophage Gifsy-1 in the 1980s.** To gain insight into the
489 evolutionary relationships of *artAB*-negative U.S. DT104 complex strains, a time-scaled
490 phylogeny was constructed using human- and bovine-associated U.S. DT104 complex genomes
491 (i.e., genomes within Dataset 1 [U.S. bovine and human data]; Figure 3 and Supplementary
492 Figures S6-S8). The common ancestor of the MDR U.S. bovine- and human-associated DT104
493 complex genomes included in this study was predicted to have existed circa 1975 (estimated
494 node age 1974.9, node height 95% highest posterior density [HPD] interval [1958.1, 1986.4];
495 Figure 3 and Supplementary Figures S6-S8); this is consistent with observations made in
496 previous studies [13, 114], in which DT104 was predicted to have acquired its MDR phenotype
497 in the 1970s. The mean evolutionary rate estimated for the Dataset 1 (U.S. bovine and human
498 data) genomes queried here was 1.75×10^{-7} substitutions/site/year (95% HPD interval [1.38×10^{-7},
499 2.11×10^{-7}]), which is similar to evolutionary rates estimated in previous studies of DT104
500 isolates from other world regions [13, 102] (Supplementary Figure S5, Supplementary Table S8,
501 and Supplementary Data).
502 Notably, over 75% of all Dataset 1 (U.S. bovine and human data) *artAB*-negative
503 genomes (42 of 53 Dataset 1 [U.S. bovine and human data] *artAB*-negative genomes, 79.2%)
504 were members of a single, well-supported clade (posterior probability = 1.0, referred to hereafter
505 as the "U.S. *artAB*-negative major clade"; Figure 3, Supplementary Figures S6-S8, and
506 Supplementary Table S15). In addition to lacking *artAB*, all members of the U.S. *artAB*-negative
507 major clade lacked Gifsy-1 and 50 additional genes, which were present in over half of all

508 Dataset 1 (U.S. bovine and human data) genomes not included in the U.S. *artAB*-negative major
509 clade, including *gogB*, a chitinase, and many phage-associated proteins (Figure 3, Supplementary
510 Figures S6-S8, and Supplementary Table S10).

511 Strains within the U.S. *artAB*-negative major clade were reportedly isolated between
512 1997 and 2018 (the most recent year included in this study) from at least 11 different states
513 across the U.S. (for two isolates, the U.S. state in which the strain was isolated was unknown;
514 Figures 2B and 3, Supplementary Figures S6-S8, and Supplementary Table S15). Interestingly,
515 most strains within the U.S. *artAB*-negative major clade were isolated from humans (*n* = 30 of 42
516 U.S. *artAB*-negative major clade strains, 71.4%), and nearly half of all Dataset 1 (U.S. bovine
517 and human data) genomes from human sources were members of this clade (*n* = 30 of 62 Dataset
518 1 [U.S. bovine and human data] genomes from human sources, 48.4%; Supplementary Tables S1
519 and S15). Human-associated U.S. *artAB*-negative major clade strains were reportedly isolated
520 from six U.S. states between 1997 and 2014 (Supplementary Table S15). The majority of
521 human-associated strains were isolated in Pennsylvania (*n* = 22 of 30 human-associated U.S.
522 *artAB*-negative major clade strains, 73.3%; Figure 2B and Supplementary Table S15); however,
523 Pennsylvania strains were reportedly isolated over a five-year period (i.e., from 2009 to 2014;
524 Supplementary Table S15) and showcased considerable genomic diversity (Figure 3), indicating
525 that it is highly unlikely that all human cases have an epidemiological link (i.e., they were not
526 sequenced as part of a single, point-source outbreak). Bovine strains within the U.S. *artAB*-
527 negative major clade were isolated from cattle or beef products (*n* = 12 of 42 U.S. *artAB*-
528 negative major clade genomes, 28.6%; Figure 2B and Supplementary Table S15). Much like
529 their human-associated counterparts, bovine-associated members of the U.S. *artAB*-negative
530 major clade were interspersed throughout the clade's phylogeny and varied in terms of isolation

531 date (i.e., 2004 to 2018) and geographic origin (i.e., six U.S. states; Figure 2B and
532 Supplementary Table S15).
533 Based on results of ancestral state reconstruction using *artAB* presence/absence, the loss
534 of Gifsy-1, *artAB*, *gogB*, and other Gifsy-1-associated genes among members of the U.S. *artAB*-
535 negative major clade was estimated to have occurred between 1985 and 1987 (estimated node
536 ages 1985.0 and 1987.2, node height 95% HPD intervals [1979.0, 1990.2] and [1981.7, 1992.1],
537 respectively; Figure 3 and Supplementary Figures S10-S12). This predicted loss event occurred
538 circa a predicted rapid increase in the U.S. DT104 complex effective population size in the mid-
539 to-late 1980s (Figure 4 and Supplementary Figure S9). Following this predicted rapid increase in
540 the 1980s, the U.S. DT104 complex effective population size was predicted to have increased
541 again in the mid-to-late 1990s, peaking circa 2000 (Figure 4 and Supplementary Figure S9).
542 **Loss of *artAB* and *gogB* within the global DT104 complex population occurs sporadically.**
543 The absence of Gifsy-1, *artAB*, and/or *gogB* among DT104 complex strains was not strictly a
544 U.S. phenomenon: Gifsy-1, *artAB*, and *gogB* were not detected in three and 19 genomes out of
545 (i) 243 DT104 strains isolated from cattle and humans in Scotland (referred to here as "Dataset 2
546 [Scottish bovine and human data]") [102], and (ii) 290 DT104 strains collected from numerous
547 sources around the world (referred to here as "Dataset 3 [multi-source data]") [13], respectively
548 (representing 1.2% and 6.6% of genomes in Dataset 2 [Scottish bovine and human data] and
549 Dataset 3 [multi-source data], respectively; Figure 5, Table 1, Supplementary Figures S14 and
550 S15, and Supplementary Tables S3-S5). Overall, out of 752 total DT104 complex genomes
551 queried in this study (i.e., the union of Dataset 1 [U.S. bovine and human data], Dataset 2
552 [Scottish bovine and human data], and Dataset 3 [multi-source data], referred to here as "Dataset

553 4 [combined global dataset]"), *artAB* could not be detected in 74 genomes (9.8% of 752 Dataset
554 4 [combined global dataset] genomes; Table 1 and Supplementary Tables S15 and S16).
555 The Gifsy-1/*artAB*/*gogB* loss event associated with the U.S. *artAB*-negative major clade
556 represented the single largest *artAB* loss event observed in this study (n = 42 of 752 total Dataset
557 4 [combined global dataset] genomes; Figure 5 and Supplementary Figures S14 and S15).
558 However, several additional, sporadic *artAB* loss events among clades encompassing ≤5
559 genomes were observed (Figure 5, Supplementary Figures S14 and S15, and Supplementary
560 Table S16). Overall, the 32 *artAB*-negative genomes that did not belong to the U.S. *artAB*-
561 negative major clade were isolated from (i) a variety of sources (i.e., humans, cattle, pigs, and
562 poultry), (ii) on four continents (i.e., North America, Europe, Asia, and Oceania), (iii) between
563 1992 and 2015 (Supplementary Table S16).
564 Among all 752 Dataset 4 (combined global dataset) genomes, the presence and absence
565 of *artAB* and *gogB* was correlated with that of Gifsy-1 (two-sided FET raw P-value < 2.2×10^{-16}
566 for each, ORs of 2069.8 and infinity, respectively), as well as each other (two-sided FET raw P-
567 value < 2.2×10^{-16}, infinite OR; Figure 5, Table 1, and Supplementary Figure S14). However,
568 unlike the 177 *artAB*-harboring Dataset 1 (U.S. bovine and human data) genomes queried here,
569 *artAB* was not always detected within prophage regions annotated as Gifsy-1 in other datasets
570 (i.e., genomes in Dataset 2 [Scottish bovine and human data] and Dataset 3 [multi-source data];
571 Table 2, Supplementary Figure S21, and Supplementary Table S5). Two genomes from Dataset 3
572 (multi-source data) harbored *artAB* within regions annotated as *Salmonella* phage 118970_sal3
573 (using PHASTER's nomenclature, "PHAGE_Salmon_118970_sal3_NC_031940"; Table 2 and
574 Supplementary Table S5). However, despite not being annotated by PHASTER as "Gifsy-1",

575 these prophage regions shared a high degree of sequence homology with the DT104 Gifsy-1
576 prophage (Supplementary Figure S21).
577 Notably, for over 90% of *artAB*-harboring genomes in Dataset 2 (Scottish bovine and
578 human data), *artAB* was identified outside of prophage regions annotated by PHASTER (*n* = 219
579 of 240 *artAB*-harboring genomes in Dataset 2 [Scottish bovine and human data]; Table 2 and
580 Supplementary Table S5). However, when 3 Kbp regions on either side of PHASTER prophage
581 regions were considered, 234 Dataset 2 (Scottish bovine and human data) genomes harbored
582 *artAB* within 3 Kbp of Gifsy-1 (97.5% of 240 *artAB*-harboring genomes in Dataset 2 [Scottish
583 bovine and human data]). Two and one genome harbored *artAB* within 3 Kb of other PHASTER
584 prophage regions (annotated by PHASTER as Edward_GF_2_NC_026611 and
585 PHAGE_Entero_HK630_NC_019723, respectively). For Dataset 3 (multi-source data), six
586 *artAB*-positive genomes did not harbor *artAB* within PHASTER prophage regions (Table 2 and
587 Supplementary Table S5). For one of these genomes (DK_7322994_6_swine_08-08-01), *artA*
588 and *artB* were detected on separate contigs, with *artB* present within 3 Kbp of a PHASTER
589 prophage region annotated as intact Gifsy-1; for the remaining five genomes, *artAB* was not
590 located within a 5 Kb region upstream or downstream of any PHASTER prophages (Table 2 and
591 Supplementary Table S5).
592 ***In vitro* response of U.S. DT104 complex strains to human- and bovine-associated**
593 **gastrointestinal stress factors is not correlated with the presence of *artAB*- and *gogB*-**
594 **harboring Gifsy-1.** The (i) loss of Gifsy-1/*artAB*/*gogB* associated with the U.S. *artAB*-negative
595 major clade circa a predicted rapid increase in the U.S. DT104 complex effective population
596 size, plus (ii) the over-representation of human strains in the U.S. *artAB*-negative major clade led
597 us to hypothesize that ArtAB and/or GogB production (or some other genomic element harbored

598 on Gifsy-1) may influence the dynamics of DT104 complex strains in the digestive tracts of
599 human and bovine hosts. Thus, we used phenotypic assays that simulated human and/or bovine
600 digestion-associated stress conditions to compare the phenotypes of Gifsy-1/*artAB*/*gogB*-
601 negative members of the U.S. *artAB*-negative major clade to those of the most closely related,
602 Gifsy-1/*artAB*/*gogB*-positive U.S. DT104 complex strains available (Supplementary Table S11).
603 As the first three compartments of the bovine digestive tract differ massively from that of
604 the human gut, the phenotype of Gifsy-1/*artAB*/*gogB*-positive and -negative strains was
605 investigated in fresh bovine ruminal fluid (RF) obtained from a donor cow (Supplementary Table
606 S12). DT104 complex concentrations were reduced by 3.4 log CFU (SD=0.2) when inoculated
607 into RF at a final concentration of 10^5 CFU/mL, whereas DT104 complex numbers were reduced
608 by 1.3 log CFU (SD = 0.2) when inoculated at a final concentration of 10^8 CFU/mL (Figure 6).
609 While the inoculation density did significantly affect survival (ANOVA raw *P*-value < 0.001),
610 the phenotype in RF was not associated with the presence or absence of Gifsy-1/*artAB*/*gogB*
611 (ANOVA raw *P*-value > 0.05; Figure 6).
612 The presence of Gifsy-1/*artAB*/*gogB* also did not significantly influence acid stress
613 survival at pH 3.5 (ANOVA raw *P*-value > 0.05; Figure 6 and Supplementary Table S13). While
614 prior adaptation at an intermediate pH 5.5 significantly increased survival at pH 3.5 as expected
615 (ANOVA raw *P*-value < 0.01), there was no significant difference in acid adaptation between
616 Gifsy-1/*artAB*/*gogB*-positive and -negative strains (ANOVA raw *P*-value > 0.05; Figure 6). Both
617 groups showed a concentration-dependent reduction in growth/survival at the two tested bile
618 concentrations of 0.6% and 1.1% (ANOVA raw *P* = 0.01 for the difference in fold change at the
619 two concentrations), but there was no Gifsy-1/*artAB*/*gogB*-dependent phenotype in the response

620 of DT104 complex strains to bile stress (ANOVA raw *P*-value > 0.05; Figure 6 and
621 Supplementary Table S14).
622
623 **DISCUSSION**
624 **Members of the DT104 complex largely harbor *artAB* on Gifsy-1-like prophage.** Bacterial
625 ADP-ribosylating toxins play important roles in the virulence of numerous pathogens [19, 115].
626 While the illness caused by *Salmonella enterica* is not considered to be a toxin-mediated disease
627 in the classical sense (e.g., as is the case for *Clostridium botulinum* or *Vibrio cholerae*) [19],
628 some *Salmonella* lineages are capable of producing ADP-ribosylating toxins, allowing them to
629 alter host immune responses and promote pathogenesis [17, 19, 116-118]. ArtAB is one such
630 toxin with a variable presence among *Salmonella* lineages: genes encoding ArtAB have been
631 detected in at least 88 different serotypes and are correlated with the presence of typhoid toxin
632 genes [24], although in DT104 this is not the case [17, 25]. Additionally, in the majority of these
633 serotypes, *artA* is predicted to be a pseudogene and the selective advantage of maintaining *artB*
634 appears to be related to its use as an alternative binding subunit for the typhoid toxin [19, 119].
635 A previous study of ArtAB-producing DT104 strains [18] found that ArtAB production
636 among DT104 appears to be the norm rather than the exception, as 237 of 243 strains (97.5%) in
637 the study were *artAB*-positive [18]. We observed similar findings here, as *artAB* was detected in
638 678 of 752 DT104 complex genomes (90.2%; Table 1 and Supplementary Table S4). Among
639 U.S. human- and bovine-associated DT104 complex genomes (i.e., Dataset 1 [U.S. bovine and
640 human data]), *artAB* was exclusively harbored on prophage regions identified by PHASTER as
641 intact Gifsy-1 (177 of 177 *artAB*-harboring Dataset 1 [U.S. bovine and human data] genomes,
642 100%; Table 2). It's important to note that Gifsy-1 prophage regions—as defined via

643 PHASTER—can display a significant degree of genetic heterogeneity [28]. Furthermore, Gifsy-1
644 has been shown to share homology with Gifsy-2 [120, 121], a result observed here (Figure 1A
645 and Supplementary Figures S1 and S17-S20). Thus, defining a "ground-truth" Gifsy-1 prophage
646 in the genomic era may not necessarily be straightforward, as both Gifsy-1 and Gifsy-2 were
647 originally identified in 1997 using Southern blotting [120]. Considering PHASTER results
648 (Table 2 and Supplementary Table S5), our own sequence homology observations
649 (Supplementary Figures S17-S20), and the fact that previous studies of DT104 have reportedly
650 identified *artAB* within Gifsy-1 [10, 19], we are confident that the *artAB*-harboring prophage
651 identified in Dataset 1 (U.S. bovine and human data) can be safely referred to as Gifsy-1.
652 Among genomes from other sources and/or world regions (i.e., Dataset 2 [Scottish bovine
653 and human data] and Dataset 3 [multi-source data]), *artAB* was not always detected within the
654 bounds of prophage regions annotated as Gifsy-1 (Table 2). Two *artAB*-harboring prophages that
655 were not annotated as "Gifsy-1" by PHASTER, for example, shared a high degree of sequence
656 homology with DT104 Gifsy-1 (Supplementary Figure S21), further highlighting the challenges
657 associated with genomic differentiation of homology-sharing prophages defined in the pre-
658 genomics era. However, most notably, *artAB* was frequently detected outside of annotated
659 prophage regions in Dataset 2 (Scottish bovine and human data; Table 2). When 3 Kbp regions
660 upstream and downstream of PHASTER prophage regions were considered, nearly all *artAB*-
661 harboring Dataset 2 (Scottish bovine and human data) genomes possessed *artAB* within 3 Kbp of
662 Gifsy-1. While it's possible that *artAB* is indeed harbored outside of Gifsy-1 in these genomes,
663 we suspect this is an artifact of prophage boundary prediction. Precise prediction of prophage
664 boundaries is challenging [122], and additional factors (e.g., assembly fragmentation in prophage
665 regions) can further affect boundary accuracy [45]. We thus encourage readers to interpret these

666 results with caution. Future long-read sequencing efforts will thus likely provide much-needed
667 insight into the prophage repertoire of the DT104 complex.
668 **A DT104 complex lineage isolated across multiple U.S. states for over twenty years lost its**
669 **ability to produce the ArtAB toxin and anti-inflammatory effector GogB.** Here, we observed
670 that *artAB* loss events appear sporadically throughout the DT104 complex phylogeny (Figures 3
671 and 5). Among U.S. human- and bovine-associated DT104 complex genomes (i.e., Dataset 1
672 [U.S. bovine and human data]), these loss events usually coincided with Gifsy-1 loss, although
673 not exclusively (i.e., three strains did not possess *artAB*, but possessed Gifsy-1; Figure 3 and
674 Supplementary Figure S20). As mentioned above, *in silico* prophage detection and
675 differentiation is challenging, and it is possible that HGT within a Gifsy-1-like prophage led to
676 the loss of *artAB*, rather than the complete excision of Gifsy-1 in its entirety. In this scenario, it
677 is plausible that partial Gifsy-1 remnants within these genomes would not be classified as Gifsy-
678 1 prophage elements, or they would not be detected by *in silico* prophage detection methods.
679 Regardless, we have identified 30 prophage-associated genes, which were detected in all Gifsy-
680 1-harboring genomes and absent from all Gifsy-1-negative genomes in Dataset 1 (U.S. bovine
681 and human data), indicating that numerous prophage-associated genes were lost along with
682 *artAB* and *gogB* (Supplementary Table S9).
683 Most notably, we observed a MDR DT104 complex clade circulating among cattle and
684 humans across 11 U.S. states, which lost Gifsy-1, concomitant with the ability to produce ArtAB
685 and GogB (i.e., the U.S. *artAB*-negative major clade; Figure 3). Considering (i) genomic
686 diversity observed within the U.S. *artAB*-negative major clade, along with the fact that (ii) U.S.
687 *artAB*-negative major clade members have been isolated from multiple states and sources for
688 over twenty years, it is nearly impossible that U.S. *artAB*-negative major clade genomes are the

689 result of repeated sequencing of identical or nearly identical strains (e.g., as would be the case in
690 a point-source outbreak scenario; Supplementary Table S15). When U.S. *artAB*-negative major
691 clade genomes were compared to DT104 complex genomes collected from a variety of isolation
692 sources around the world (Figure 5), we did not identify genomes from any country other than
693 the U.S. within this clade, nor did we identify genomes from any isolation source other than
694 cattle and humans (Figure 5 and Supplementary Table S15). Furthermore, we observed a high
695 proportion of human-associated strains within the U.S. *artAB*-negative major clade relative to
696 bovine-associated strains (Figure 3 and Supplementary Table S15). The (i) limited host range,
697 (ii) limited geographic range, and (iii) high human-to-bovine ratio observed for the U.S. *artAB*-
698 negative major clade in this study is likely an artifact of sampling and/or sequencing (e.g., due to
699 the large number of bovine-associated genomes included in this study relative to other animal
700 hosts, due to this study's focus on the DT104 complex in the U.S., due to geographic and/or host
701 biases in allocation of *Salmonella* sequencing resources). Future studies querying more DT104
702 complex strains from (i) non-human and non-bovine sources (e.g., other animal hosts, foods,
703 environmental sources) and (ii) countries other than the U.S. and the United Kingdom will likely
704 reveal a greater isolation source and geographic range for this clade, respectively.
705 However, it's important to note that the U.S. *artAB*-negative major clade contained
706 nearly half of all DT104 complex strains isolated from humans in the U.S. (n = 30 of 62 U.S.
707 DT104 complex genomes from human sources, 48.4%; Supplementary Tables S1 and S15). This
708 is notable, as our study included all human-associated U.S. DT104 complex genomes with
709 metadata available in Enterobase at the time. It is certainly likely that there were U.S. DT104
710 complex genomes from human sources, which were not included in our study (e.g., due to
711 missing publicly available metadata), and it is possible that there may be biases in terms of

712 metadata reporting (e.g., some laboratories may routinely provide detailed, publicly available
713 metadata for genomes that they sequence, while other laboratories may never or rarely provide
714 metadata). Thus, in order to gain further insight into potential U.S. *artAB*-negative major clade
715 host associations (or the lack thereof), it is essential that isolation source metadata are made
716 publicly available in addition to WGS data.

717 The U.S. *artAB*-negative major clade was predicted to have lost Gifsy-1/*artAB*/*gogB*
718 circa 1985-1987, circa a predicted rapid increase in the U.S. DT104 complex effective
719 population size, which occurred in the mid-to-late 1980s (Figure 4). Our results are consistent
720 with a previous study of DT104 from multiple world regions, which also identified periods of
721 dramatic population growth in the 1980s and 1990s [13]. This rapid increase in population size is
722 notable, as it coincides with the global MDR DT104 epidemic, which occurred among humans
723 and animals throughout the 1990s [13, 14, 102]. However, it is essential to note that our data do
724 not imply that *artAB*, *gogB*, or Gifsy-1 loss played a role in the emergence and subsequent global
725 spread of DT104; any potential association between the virulence and/or fitness of MDR DT104
726 and Gifsy-1/*artAB*/*gogB* loss among DT104 complex genomes is merely speculative at this
727 point. While previous studies of DT104 have shown that prophage excision and *artAB* loss occur
728 in response to DNA damage and other stressors [17, 19], future studies are needed to better
729 understand the roles that Gifsy-1, *artAB*, and *gogB* play in DT104 evolution.

730 **Members of the U.S. *artAB*-negative major clade do not have a phenotypic advantage**
731 **relative to other U.S. DT104 complex strains when exposed to ruminal fluid-, acid-, and**
732 **bile-associated stressors *in vitro.*** *Salmonella enterica* encounters numerous stressors within the
733 gastrointestinal tracts of humans and animals, including (but not limited to) low pH, low oxygen,
734 exposure to bile, and the host immune system [123-125]. Furthermore, the gastrointestinal

735 environment that *Salmonella enterica* encounters can differ between hosts; for example, the first
736 three compartments of the bovine digestive tract differ massively from those of the human gut, as
737 they essentially serve as massive microbial fermentation chambers [126]. Here, we evaluated the
738 survival of DT104 complex strains when exposed to three stressors encountered in the human
739 and/or bovine gastrointestinal tracts: (i) ruminal fluid (RF; bovine rumen), (ii) low pH (bovine
740 abomasum and human stomach), and (iii) exposure to bile (bovine and human duodenum); we
741 discuss each step in detail below.

742 In the bovine digestion process, the RF, including the complex community of ruminal
743 microbiota [127], presents an early line of defense against potential pathogens, like *Salmonella*
744 spp. In RF, the kill rate of DT104 complex strains was dependent on the inoculation density. The
745 high inoculation rate (10^8 CFU/mL) was chosen to test the ability of the ruminal microbiota to
746 efficiently kill or impede *Salmonella*. The lower inoculation rate of 10^5 CFU/mL was chosen for
747 its dynamic range to measure either growth or decrease of *Salmonella* concentration. An
748 interaction of the complex ruminal microbiota with the inoculated *Salmonella* is conceivable in
749 two ways: either the microbiota exhibit strategies to produce antimicrobial compounds against
750 *Salmonella* species [128, 129], or through competition for nutrients, e.g., iron [130]. The fact that
751 the ruminal microbiota was less effective at killing DT104 complex strains at the high
752 inoculation rate suggests that their defense mechanisms against DT104 complex strains are
753 limited and/or the system started to be overrun by the high numbers of the DT104 complex
754 strain.

755 Gastric acids in the stomach (or abomasum) are the next line of host defense, which
756 *Salmonella* must overcome during gastrointestinal passage [131]. A pH of 3.5 was selected based
757 on the following considerations: the human gastric pH varies from pH <2 in a fasted state to pH

758 >6 during meals, returning to a low pH within hours postprandially [132, 133]. Intracellular
759 pathogens like *Salmonella* spp. have adapted to survive low pH intracellularly in the phagosomes
760 of phagocytes (e.g. pH 4-6) [134, 135] and express adaptive acid tolerance that allows them to
761 tolerate pH values between 2-3 [136]. Therefore, the chosen pH 3.5 reflects a relevant
762 physiological state of the human stomach and represents a sublethal stress to *Salmonella* spp.
763 Our experiments confirmed that acid adaptation with HCl at pH 5.5 led to much higher survival
764 rates at pH 3.5. Well-known mechanisms such as decreased membrane conductivity for H^+,
765 increased proton extrusion, or changes in the cell envelope composition [136-138] could be
766 responsible for this.

767 Upon leaving the stomach, enteric pathogens are confronted with bile. Bile salts show
768 antimicrobial activity by dissolving membrane lipids, dissociating integral membrane proteins
769 [139], and lead to general cell damage by misfolding and denaturation of proteins [140, 141] and
770 DNA damage [142, 143]. *Salmonella enterica* is able to survive duodenal bile salt concentrations
771 through DNA repair mechanisms [143], multiple changes in gene expression [144], and
772 increased production of anti-oxidative enzymes [145]. Here, selected DT104 complex strains
773 were able to survive at both tested bile salt concentrations (14.5 mmol/L and 26.0 mmol/L);
774 however, no significant differences were observed between strains that harbored Gifsy-
775 1/*artAB*/*gogB* and those that did not (Figure 6).

776 In summary, the *in vitro* stress assays performed in this study aimed to mimic the
777 stressors that DT104 complex strains encounter in the gastrointestinal tracts of humans and
778 ruminants. Given the over-representation of human-associated Gifsy-1/*artAB*/*gogB*-negative
779 strains observed here, one may be tempted to speculate that Gifsy-1, *artAB*, and/or *gogB* absence
780 may confer members of the U.S. *artAB*-negative major clade with a competitive advantage in the

781 human host gastrointestinal tract; however, no Gifsy-1/*artAB*/*gogB*-dependent phenotype was
782 observed in DT104 complex strains under the tested conditions (Figure 6). Furthermore, as
783 mentioned above, the overrepresentation of human strains in this clade could merely be an
784 artifact of sampling/sequencing. Thus, it may be possible that Gifsy-1/*artAB*/*gogB* absence may
785 confer some advantage(s) to U.S. *artAB*-negative major clade strains in hosts underrepresented in
786 this study, or in environmental conditions, which were not tested in this study, including those
787 outside of the host (e.g., high osmotic pressure and competitive microbiota in manure or
788 wastewater, food safety measures like disinfectants, antimicrobials and food processing) [124].
789 However, at the present, this is merely speculation; future studies are needed to evaluate whether
790 Gifsy-1/*artAB*/*gogB* loss among members of the U.S. *artAB*-negative major clade is merely
791 coincidental or indicative of some evolutionarily advantageous phenotype.
792 **Future research is needed to understand the roles that Gifsy-1, ArtAB, and GogB play in**
793 **DT104 virulence.** The results presented here indicate that prophage-mediated ArtAB production
794 within the DT104 complex can undergo temporal changes. Most notably, we identified the U.S.
795 *artAB*-negative major clade, which lost the ability to produce ArtAB and GogB, likely due to a
796 Gifsy-1 loss event (Figure 3). However, the ecological and/or evolutionary significance of this
797 loss-of-function event remain unclear. Although phenotypic assessments have demonstrated a
798 role for DT104-encoded ArtAB in both cell culture and a mouse model [18], the true benefit of
799 this toxin in the context of human and bovine salmonellosis has not been investigated. It has
800 been previously shown that reactive oxygen species (ROS) induce production of ArtAB [23],
801 which may suggest that *artAB* is expressed in response to immune cell derived ROS.
802 Furthermore, as treatment with ArtA increases intracellular levels of cAMP in macrophage-like
803 cells [18], ArtAB may play a role in delaying *Salmonella* clearance by altering the activity of

host immune cells [19]. Hence, future studies, including in tissue culture and animal models, will be needed to determine whether *artAB* presence or absence confers a selective advantage among human- and animal-associated DT104.

AUTHOR STATEMENTS

Conflicts of interest. The authors declare that there are no conflicts of interest.

Funding information. This material is based on work supported by the National Science Foundation (NSF) Graduate Research Fellowship Program under grant no. DGE-1650441, with additional funding provided by an NSF Graduate Research Opportunities Worldwide (GROW) grant through a partnership with the Swiss National Science Foundation (SNF). This work was additionally supported by the SciLifeLab & Wallenberg Data Driven Life Science Program (grant: KAW 2020.0239).

Acknowledgments. Figure 2 was created with BioRender.com.

827 **REFERENCES**

828 1. **Bobay LM, Touchon M, Rocha EP**. Pervasive domestication of defective prophages by
829 bacteria. *Proc Natl Acad Sci U S A* 2014;111(33):12127-12132.
830 2. **Fortier LC, Sekulovic O**. Importance of prophages to evolution and virulence of
831 bacterial pathogens. *Virulence* 2013;4(5):354-365.
832 3. **Pleska M, Lang M, Refardt D, Levin BR, Guet CC**. Phage-host population dynamics
833 promotes prophage acquisition in bacteria with innate immunity. *Nat Ecol Evol* 2018;2(2):359-
834 366.
835 4. **Ramisetty BCM, Sudhakari PA**. Bacterial 'Grounded' Prophages: Hotspots for Genetic
836 Renovation and Innovation. *Front Genet* 2019;10:65.
837 5. **Owen SV, Canals R, Wenner N, Hammarlöf DL, Kröger C et al.** A window into
838 lysogeny: Revealing temperate phage biology with transcriptomics. *bioRxiv* 2019:787010.
839 6. **Haaber J, Leisner JJ, Cohn MT, Catalan-Moreno A, Nielsen JB et al.** Bacterial
840 viruses enable their host to acquire antibiotic resistance genes from neighbouring cells. *Nat*
841 *Commun* 2016;7:13333.
842 7. **Wang X, Kim Y, Ma Q, Hong SH, Pokusaeva K et al.** Cryptic prophages help bacteria
843 cope with adverse environments. *Nat Commun* 2010;1:147.
844 8. **Liu B, Knirel YA, Feng L, Perepelov AV, Senchenkova SyN et al.** Structural diversity
845 in *Salmonella* O antigens and its genetic basis. *FEMS Microbiology Reviews* 2014;38(1):56-89.
846 9. **Czajkowski R**. May the Phage be With You? Prophage-Like Elements in the Genomes
847 of Soft Rot Pectobacteriaceae: *Pectobacterium* spp. and *Dickeya* spp. *Frontiers in Microbiology*,
848 Original Research 2019;10(138).
849 10. **Hiley L, Fang NX, Micalizzi GR, Bates J**. Distribution of Gifsy-3 and of variants of
850 ST64B and Gifsy-1 prophages amongst *Salmonella enterica* Serovar Typhimurium isolates:
851 evidence that combinations of prophages promote clonality. *PLoS One* 2014;9(1):e86203.
852 11. **World Health Organization**. 2018. *Salmonella* (non-typhoidal).
853 https://www.who.int/news-room/fact-sheets/detail/salmonella-(non-typhoidal) [accessed 5 May
854 2021].
855 12. **Wang X, Biswas S, Paudyal N, Pan H, Li X et al.** Antibiotic Resistance in *Salmonella*
856 Typhimurium Isolates Recovered From the Food Chain Through National Antimicrobial
857 Resistance Monitoring System Between 1996 and 2016. *Front Microbiol* 2019;10:985.
858 13. **Leekitcharoenphon P, Hendriksen RS, Le Hello S, Weill FX, Baggesen DL et al.**
859 Global Genomic Epidemiology of *Salmonella enterica* Serovar Typhimurium DT104. *Appl*
860 *Environ Microbiol* 2016;82(8):2516-2526.
861 14. **Threlfall EJ**. Epidemic *salmonella typhimurium* DT 104--a truly international
862 multiresistant clone. *J Antimicrob Chemother* 2000;46(1):7-10.
863 15. **Helms M, Ethelberg S, Molbak K, Group DTS**. International *Salmonella* Typhimurium
864 DT104 infections, 1992-2001. *Emerg Infect Dis* 2005;11(6):859-867.
865 16. **Allen CA, Fedorka-Cray PJ, Vazquez-Torres A, Suyemoto M, Altier C et al.** *In Vitro*
866 and *In Vivo* Assessment of *Salmonella enterica* Serovar Typhimurium DT104 Virulence.
867 *Infection and Immunity* 2001;69(7):4673-4677.
868 17. **Saitoh M, Tanaka K, Nishimori K, Makino SI, Kanno T et al.** The *artAB* genes
869 encode a putative ADP-ribosyltransferase toxin homologue associated with *Salmonella enterica*
870 serovar Typhimurium DT104. *Microbiology* 2005;151(Pt 9):3089-3096.

871 18. **Tamamura Y, Tanaka K, Uchida I**. Characterization of pertussis-like toxin from
872 *Salmonella* spp. that catalyzes ADP-ribosylation of G proteins. *Sci Rep* 2017;7(1):2653.
873 19. **Cheng RA, Wiedmann M**. The ADP-Ribosylating Toxins of *Salmonella*. *Toxins (Basel)*
874 2019;11(7).
875 20. **Weiss AA, Hewlett EL**. Virulence factors of *Bordetella pertussis*. *Annu Rev Microbiol*
876 1986;40:661-686.
877 21. **Hewlett EL, Sauer KT, Myers GA, Cowell JL, Guerrant RL**. Induction of a novel
878 morphological response in Chinese hamster ovary cells by pertussis toxin. *Infection and*
879 *immunity* 1983;40(3):1198-1203.
880 22. **Carbonetti NH**. Pertussis toxin and adenylate cyclase toxin: key virulence factors of
881 *Bordetella pertussis* and cell biology tools. *Future microbiology* 2010;5(3):455-469.
882 23. **Uchida I, Ishihara R, Tanaka K, Hata E, Makino S et al.** *Salmonella enterica* serotype
883 Typhimurium DT104 ArtA-dependent modification of pertussis toxin-sensitive G proteins in the
884 presence of [32P]NAD. *Microbiology* 2009;155(Pt 11):3710-3718.
885 24. **Gaballa A, Cheng RA, Harrand AS, Cohn AR, Wiedmann M**. The Majority of
886 Typhoid Toxin-Positive *Salmonella* Serovars Encode ArtB, an Alternate Binding Subunit.
887 *mSphere* 2021;6(1).
888 25. **Ojiakor A, Gibbs RN, Chen Z, Gao X, Fowler CC**. The evolutionary diversification of
889 the *Salmonella artAB* toxin locus. *Front Microbiol* 2022;13:1016438.
890 26. **Moreno Switt AI, den Bakker HC, Cummings CA, Rodriguez-Rivera LD, Govoni G**
891 **et al.** Identification and characterization of novel *Salmonella* mobile elements involved in the
892 dissemination of genes linked to virulence and transmission. *PLoS One* 2012;7(7):e41247.
893 27. **Pang S, Octavia S, Feng L, Liu B, Reeves PR et al.** Genomic diversity and adaptation
894 of *Salmonella enterica* serovar Typhimurium from analysis of six genomes of different phage
895 types. *BMC Genomics* 2013;14:718.
896 28. **Worley J, Meng J, Allard MW, Brown EW, Timme RE**. *Salmonella enterica*
897 Phylogeny Based on Whole-Genome Sequencing Reveals Two New Clades and Novel Patterns
898 of Horizontally Acquired Genetic Elements. *MBio* 2018;9(6).
899 29. **Owen SV, Wenner N, Canals R, Makumi A, Hammarlof DL et al.** Characterization of
900 the Prophage Repertoire of African *Salmonella* Typhimurium ST313 Reveals High Levels of
901 Spontaneous Induction of Novel Phage BTP1. *Front Microbiol* 2017;8:235.
902 30. **Coombes BK, Wickham ME, Brown NF, Lemire S, Bossi L et al.** Genetic and
903 molecular analysis of GogB, a phage-encoded type III-secreted substrate in *Salmonella enterica*
904 serovar typhimurium with autonomous expression from its associated phage. *J Mol Biol*
905 2005;348(4):817-830.
906 31. **Takemura M, Haneda T, Idei H, Miki T, Okada N**. A *Salmonella* type III effector,
907 PipA, works in a different manner than the PipA family effectors GogA and GtgA. *PLoS One*
908 2021;16(3):e0248975.
909 32. **Carroll LM, Huisman JS, Wiedmann M**. Twentieth-century emergence of
910 antimicrobial resistant human- and bovine-associated *Salmonella enterica* serotype
911 Typhimurium lineages in New York State. *Sci Rep* 2020;10(1):14428.
912 33. **Zhou Z, Alikhan N-F, Mohamed K, Fan Y, Achtman M**. The user's guide to
913 comparative genomics with EnteroBase, including case studies on transmissions of micro-clades
914 of *Salmonella*, the phylogeny of ancient and modern *Yersinia pestis* genomes, and the core
915 genomic diversity of all *Escherichia*. *bioRxiv* 2019:613554.

916 34. **Alikhan NF, Zhou Z, Sergeant MJ, Achtman M**. A genomic overview of the
917 population structure of *Salmonella*. *PLoS Genet* 2018;14(4):e1007261.
918 35. **Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence**
919 **Database C**. The sequence read archive. *Nucleic Acids Res* 2011;39(Database issue):D19-21.
920 36. **Kodama Y, Shumway M, Leinonen R, International Nucleotide Sequence Database**
921 **C**. The Sequence Read Archive: explosive growth of sequencing data. *Nucleic Acids Res*
922 2012;40(Database issue):D54-56.
923 37. **Yoshida CE, Kruczkiewicz P, Laing CR, Lingohr EJ, Gannon VP et al.** The
924 *Salmonella In Silico* Typing Resource (SISTR): An Open Web-Accessible Tool for Rapidly
925 Typing and Subtyping Draft *Salmonella* Genome Assemblies. *PLoS One* 2016;11(1):e0147101.
926 38. **Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J**. RhierBAPS: An R
927 implementation of the population clustering algorithm hierBAPS. *Wellcome Open Res*
928 2018;3:93.
929 39. **Bolger AM, Lohse M, Usadel B**. Trimmomatic: a flexible trimmer for Illumina
930 sequence data. *Bioinformatics* 2014;30(15):2114-2120.
931 40. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M et al.** SPAdes: a new
932 genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*
933 2012;19(5):455-477.
934 41. **Andrews S**. FastQC: a quality control tool for high throughput sequence data.
935 https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
936 42. **Gurevich A, Saveliev V, Vyahhi N, Tesler G**. QUAST: quality assessment tool for
937 genome assemblies. *Bioinformatics* 2013;29(8):1072-1075.
938 43. **Ewels P, Magnusson M, Lundin S, Kaller M**. MultiQC: summarize analysis results for
939 multiple tools and samples in a single report. *Bioinformatics* 2016;32(19):3047-3048.
940 44. **Arndt D, Marcu A, Liang Y, Wishart DS**. PHAST, PHASTER and PHASTEST: Tools
941 for finding prophage in bacterial genomes. *Brief Bioinform* 2019;20(4):1560-1567.
942 45. **Arndt D, Grant JR, Marcu A, Sajed T, Pon A et al.** PHASTER: a better, faster version
943 of the PHAST phage search tool. *Nucleic Acids Res* 2016;44(W1):W16-21.
944 46. **Seemann T**. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*
945 2014;30(14):2068-2069.
946 47. **Gilchrist CLM, Chooi YH**. Clinker & clustermap.js: Automatic generation of gene
947 cluster comparison figures. *Bioinformatics* 2021.
948 48. **Seemann T**. ABRicate: Mass screening of contigs for antimicrobial resistance or
949 virulence genes. https://github.com/tseemann/abricate.
950 49. **Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ et al.** Validating the
951 AMRFinder Tool and Resistance Gene Database by Using Antimicrobial Resistance Genotype-
952 Phenotype Correlations in a Collection of Isolates. *Antimicrob Agents Chemother* 2019;63(11).
953 50. **Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O et al.** *In*
954 *silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence
955 typing. *Antimicrob Agents Chemother* 2014;58(7):3895-3903.
956 51. **Chen L, Yang J, Yu J, Yao Z, Sun L et al.** VFDB: a reference database for bacterial
957 virulence factors. *Nucleic Acids Res* 2005;33(Database issue):D325-328.
958 52. **Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J et al.** BLAST+:
959 architecture and applications. *BMC Bioinformatics* 2009;10:421.

960 53. **Hunt M, Mather AE, Sanchez-Buso L, Page AJ, Parkhill J et al.** ARIBA: rapid
961 antimicrobial resistance genotyping directly from sequencing reads. *Microb Genom*
962 2017;3(10):e000131.
963 54. **Li H**. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
964 *arXiv* 2013:1303.3997.
965 55. **Li H, Durbin R**. Fast and accurate short read alignment with Burrows-Wheeler
966 transform. *Bioinformatics* 2009;25(14):1754-1760.
967 56. **Li H**. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*
968 2018;34(18):3094-3100.
969 57. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J et al.** The Sequence
970 Alignment/Map format and SAMtools. *Bioinformatics* 2009;25(16):2078-2079.
971 58. **Quinlan AR, Hall IM**. BEDTools: a flexible suite of utilities for comparing genomic
972 features. *Bioinformatics* 2010;26(6):841-842.
973 59. **Quinlan AR**. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr*
974 *Protoc Bioinformatics* 2014;47:11 12 11-34.
975 60. **Li H**. A statistical framework for SNP calling, mutation discovery, association mapping
976 and population genetical parameter estimation from sequencing data. *Bioinformatics*
977 2011;27(21):2987-2993.
978 61. **Garrison E, Marth G**. Haplotype-based variant detection from short-read sequencing.
979 *arXiv* 2012:1207.3907.
980 62. **Cleary JG, Braithwaite R, Gaastra K, Hilbush BS, Inglis S et al.** Comparing Variant
981 Call Files for Performance Benchmarking of Next-Generation Sequencing Variant Calling
982 Pipelines. *bioRxiv* 2015:023754.
983 63. **Tan A, Abecasis GR, Kang HM**. Unified representation of genetic variants.
984 *Bioinformatics* 2015;31(13):2202-2204.
985 64. **Cingolani P, Platts A, Wang le L, Coon M, Nguyen T et al.** A program for annotating
986 and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of
987 *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 2012;6(2):80-92.
988 65. **Seemann T**. samclip: Filter SAM file for soft and hard clipped alignments.
989 https://github.com/tseemann/samclip.
990 66. **Li H**. Seqtk: a fast and lightweight tool for processing sequences in the FASTA or
991 FASTQ format. https://github.com/lh3/seqtk.
992 67. **Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T et al.** SNP-sites: rapid efficient
993 extraction of SNPs from multi-FASTA alignments. *Microb Genom* 2016;2(4):e000056.
994 68. **Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA et al.** Rapid phylogenetic
995 analysis of large samples of recombinant bacterial whole genome sequences using Gubbins.
996 *Nucleic Acids Res* 2015;43(3):e15.
997 69. **Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ**. IQ-TREE: a fast and effective
998 stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*
999 2015;32(1):268-274.
1000 70. **Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS**.
1001 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*
1002 2017;14(6):587-589.
1003 71. **Kimura M**. Estimation of evolutionary distances between homologous nucleotide
1004 sequences. *Proc Natl Acad Sci U S A* 1981;78(1):454-458.

1005 72. **Minh BQ, Nguyen MA, von Haeseler A**. Ultrafast approximation for phylogenetic
1006 bootstrap. *Mol Biol Evol* 2013;30(5):1188-1195.
1007 73. **Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS**. UFBoot2: Improving
1008 the Ultrafast Bootstrap Approximation. *Mol Biol Evol* 2018;35(2):518-522.
1009 74. **Rambaut A, Lam TT, Max Carvalho L, Pybus OG**. Exploring the temporal structure
1010 of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol*
1011 2016;2(1):vew007.
1012 75. **To T-H, Jung M, Lycett S, Gascuel O**. Fast Dating Using Least-Squares Criteria and
1013 Algorithms. *Systematic Biology* 2015;65(1):82-97.
1014 76. **Rambaut A**. FigTree: a graphical viewer of phylogenetic trees.
1015 http://tree.bio.ed.ac.uk/software/figtree/.
1016 77. **Bouckaert R, Heled J, Kuhnert D, Vaughan T, Wu CH et al.** BEAST 2: a software
1017 platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 2014;10(4):e1003537.
1018 78. **Bouckaert R, Vaughan TG, Barido-Sottani J, Duchene S, Fourment M et al.** BEAST
1019 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol*
1020 2019;15(4):e1006650.
1021 79. **Bouckaert R**. 2014. Correcting for constant sites in BEAST2.
1022 https://groups.google.com/forum/#!topic/beast-users/QfBHMOqImFE [accessed March 1, 2023].
1023 80. **Bouckaert RR, Drummond AJ**. bModelTest: Bayesian phylogenetic site model
1024 averaging and model comparison. *BMC Evol Biol* 2017;17(1):42.
1025 81. **Drummond AJ, Ho SY, Phillips MJ, Rambaut A**. Relaxed phylogenetics and dating
1026 with confidence. *PLoS Biol* 2006;4(5):e88.
1027 82. **Drummond AJ, Rambaut A, Shapiro B, Pybus OG**. Bayesian coalescent inference of
1028 past population dynamics from molecular sequences. *Mol Biol Evol* 2005;22(5):1185-1192.
1029 83. **R Core Team**. R: A Language and Environment for Statistical Computing. R Foundation
1030 for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
1031 84. **Wickham H**. *ggplot2: Elegant Graphics for Data Analysis*: Springer-Verlag New York;
1032 2016.
1033 85. **Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y**. ggtree: an r package for visualization
1034 and annotation of phylogenetic trees with their covariates and other associated data. *Methods in*
1035 *Ecology and Evolution* 2017;8(1):28-36.
1036 86. **Yu G, Lam TT, Zhu H, Guan Y**. Two Methods for Mapping and Visualizing
1037 Associated Data on Phylogeny Using Ggtree. *Mol Biol Evol* 2018;35(12):3041-3043.
1038 87. **R Hackathon**. phylobase: Base Package for Phylogenetic Structures and Comparative
1039 Data. https://CRAN.R-project.org/package=phylobase.
1040 88. **Yu G**. treeio: Base Classes and Functions for Phylogenetic Tree Input and Output.
1041 https://guangchuangyu.github.io/software/treeio.
1042 89. **Revell LJ**. phytools: an R package for phylogenetic comparative biology (and other
1043 things). *Methods in Ecology and Evolution* 2012;3(2):217-223.
1044 90. **Paradis E, Claude J, Strimmer K**. APE: Analyses of Phylogenetics and Evolution in R
1045 language. *Bioinformatics* 2004;20(2):289-290.
1046 91. **Paradis E, Schliep K**. ape 5.0: an environment for modern phylogenetics and
1047 evolutionary analyses in R. *Bioinformatics* 2019;35(3):526-528.
1048 92. **Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G et al.** Producing
1049 polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol* 2020;21(1):180.

1050 93. **Katoh K, Standley DM**. MAFFT multiple sequence alignment software version 7:
1051 improvements in performance and usability. *Mol Biol Evol* 2013;30(4):772-780.
1052 94. **Katoh K, Misawa K, Kuma K, Miyata T**. MAFFT: a novel method for rapid multiple
1053 sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002;30(14):3059-3066.
1054 95. **Collins RE, Higgs PG**. Testing the infinitely many genes model for the evolution of the
1055 bacterial core genome and pangenome. *Mol Biol Evol* 2012;29(11):3413-3425.
1056 96. **Baumdicker F, Hess WR, Pfaffelhuber P**. The infinitely many genes model for the
1057 distributed genome of bacteria. *Genome Biol Evol* 2012;4(4):443-456.
1058 97. **Zamani-Dahaj SA, Okasha M, Kosakowski J, Higgs PG**. Estimating the Frequency of
1059 Horizontal Gene Transfer Using Phylogenetic Models of Gene Gain and Loss. *Mol Biol Evol*
1060 2016;33(7):1843-1857.
1061 98. **Huerta-Cepas J, Szklarczyk D, Heller D, Hernandez-Plaza A, Forslund SK et al.**
1062 eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource
1063 based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 2019;47(D1):D309-D314.
1064 99. **Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J**.
1065 eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at
1066 the Metagenomic Scale. *Mol Biol Evol* 2021;38(12):5825-5829.
1067 100. **Benjamini Y, Hochberg Y**. Controlling the False Discovery Rate: A Practical and
1068 Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B*
1069 *(Methodological)* 1995;57(1):289-300.
1070 101. **Collins C, Didelot X**. A phylogenetic method to perform genome-wide association
1071 studies in microbes that accounts for population structure and recombination. *PLoS Comput Biol*
1072 2018;14(2):e1005958.
1073 102. **Mather AE, Reid SW, Maskell DJ, Parkhill J, Fookes MC et al.** Distinguishable
1074 epidemics of multidrug-resistant *Salmonella* Typhimurium DT104 in different hosts. *Science*
1075 2013;341(6153):1514-1517.
1076 103. **Treangen TJ, Ondov BD, Koren S, Phillippy AM**. The Harvest suite for rapid core-
1077 genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome*
1078 *Biol* 2014;15(11):524.
1079 104. **Bruen TC, Philippe H, Bryant D**. A simple and robust statistical test for detecting the
1080 presence of recombination. *Genetics* 2006;172(4):2665-2681.
1081 105. **Letunic I, Bork P**. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic
1082 tree display and annotation. *Nucleic Acids Res* 2021;49(W1):W293-W296.
1083 106. **Vangay P, Fugett EB, Sun Q, Wiedmann M**. Food microbe tracker: a web-based tool
1084 for storage and comparison of food-associated microbes. *J Food Prot* 2013;76(2):283-294.
1085 107. **Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S et al.** Roary: rapid large-scale
1086 prokaryote pan genome analysis. *Bioinformatics* 2015;31(22):3691-3693.
1087 108. **Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P et al.** vegan: Community
1088 Ecology Package. https://CRAN.R-project.org/package=vegan.
1089 109. **Kuhbacher A, Cossart P, Pizarro-Cerda J**. Internalization assays for *Listeria*
1090 *monocytogenes*. *Methods Mol Biol* 2014;1157:167-178.
1091 110. **Horlbog JA, Kent D, Stephan R, Guldimann C**. Surviving host - and food relevant
1092 stresses: phenotype of *L. monocytogenes* strains isolated from food and clinical sources. *Sci Rep*
1093 2018;8(1):12931.
1094 111. **Fausa O**. Duodenal bile acids after a test meal. *Scand J Gastroenterol* 1974;9(6):567-
1095 570.

1096 112. **Guariglia-Oropeza V, Orsi RH, Guldimann C, Wiedmann M, Boor KJ**. The *Listeria*
1097 *monocytogenes* Bile Stimulon under Acidic Conditions Is Characterized by Strain-Specific
1098 Patterns and the Upregulation of Motility, Cell Wall Modification Functions, and the PrfA
1099 Regulon. *Front Microbiol* 2018;9:120.
1100 113. **Barrett T, Clark K, Gevorgyan R, Gorelenkov V, Gribov E et al.** BioProject and
1101 BioSample databases at NCBI: facilitating capture and organization of metadata. *Nucleic Acids*
1102 *Res* 2012;40(Database issue):D57-63.
1103 114. **Carroll LM, Pierneef R, Mathole M, Matle I**. Genomic Characterization of Endemic
1104 and Ecdemic Non-typhoidal *Salmonella enterica* Lineages Circulating Among Animals and
1105 Animal Products in South Africa. *Front Microbiol* 2021;12:748611.
1106 115. **Simon NC, Aktories K, Barbieri JT**. Novel bacterial ADP-ribosylating toxins: structure
1107 and function. *Nat Rev Microbiol* 2014;12(9):599-611.
1108 116. **Guiney DG, Fierer J**. The Role of the *spv* Genes in *Salmonella* Pathogenesis. *Front*
1109 *Microbiol* 2011;2:129.
1110 117. **Spano S, Ugalde JE, Galan JE**. Delivery of a *Salmonella* Typhi exotoxin from a host
1111 intracellular compartment. *Cell Host Microbe* 2008;3(1):30-38.
1112 118. **Pollard DJ, Young JC, Covarelli V, Herrera-Leon S, Connor TR et al.** The Type III
1113 Secretion System Effector SeoC of *Salmonella enterica* subsp. *salamae* and *S. enterica* subsp.
1114 *arizonae* ADP-Ribosylates Src and Inhibits Opsonophagocytosis. *Infect Immun*
1115 2016;84(12):3618-3628.
1116 119. **Fowler CC, Stack G, Jiao X, Lara-Tejero M, Galan JE**. Alternate subunit assembly
1117 diversifies the function of a bacterial toxin. *Nat Commun* 2019;10(1):3684.
1118 120. **Figueroa-Bossi N, Coissac E, Netter P, Bossi L**. Unsuspected prophage-like elements in
1119 *Salmonella typhimurium*. *Mol Microbiol* 1997;25(1):161-173.
1120 121. **Figueroa-Bossi N, Bossi L**. Inducible prophages contribute to *Salmonella* virulence in
1121 mice. *Molecular Microbiology*, https://doi.org/10.1046/j.1365-2958.1999.01461.x
1122 1999;33(1):167-176.
1123 122. **Tang K, Wang W, Sun Y, Zhou Y, Wang P et al.** Prophage Tracer: precisely tracing
1124 prophages in prokaryotic genomes using overlapping split-read alignment. *Nucleic Acids Res*
1125 2021;49(22):e128.
1126 123. **Flint A, Butcher J, Stintzi A**. Stress Responses, Adaptation, and Virulence of Bacterial
1127 Pathogens During Host Gastrointestinal Colonization. *Microbiol Spectr* 2016;4(2).
1128 124. **Burgess CM, Gianotti A, Gruzdev N, Holah J, Knochel S et al.** The response of
1129 foodborne pathogens to osmotic and desiccation stresses in the food chain. *Int J Food Microbiol*
1130 2016;221:37-53.
1131 125. **Horn N, Bhunia AK**. Food-Associated Stress Primes Foodborne Pathogens for the
1132 Gastrointestinal Phase of Infection. *Front Microbiol* 2018;9:1962.
1133 126. **Hofmann RR**. Evolutionary steps of ecophysiological adaptation and diversification of
1134 ruminants: a comparative view of their digestive system. *Oecologia* 1989;78(4):443-457.
1135 127. **McCann JC, Wickersham TA, Loor JJ**. High-throughput Methods Redefine the
1136 Rumen Microbiome and Its Relationship with Nutrition and Metabolism. *Bioinform Biol Insights*
1137 2014;8:109-125.
1138 128. **Kalmokoff ML, Cyr TD, Hefford MA, Whitford MF, Teather RM**. Butyrivibriocin
1139 AR10, a new cyclic bacteriocin produced by the ruminal anaerobe *Butyrivibrio fibrisolvens*
1140 AR10: characterization of the gene and peptide. *Can J Microbiol* 2003;49(12):763-773.

1141 129. **Oyama LB, Girdwood SE, Cookson AR, Fernandez-Fuentes N, Prive F et al.** The
1142 rumen microbiome: an underexplored resource for novel antimicrobial discovery. *NPJ Biofilms*
1143 *Microbiomes* 2017;3:33.
1144 130. **Deriu E, Liu JZ, Pezeshki M, Edwards RA, Ochoa RJ et al.** Probiotic bacteria reduce
1145 *salmonella typhimurium* intestinal colonization by competing for iron. *Cell Host Microbe*
1146 2013;14(1):26-37.
1147 131. **Singh A, Barnard TG**. Adaptations in the physiological heterogeneity and viability of
1148 *Shigella dysenteriae*, *Shigella flexneri* and *Salmonella typhimurium*, after exposure to simulated
1149 gastric acid fluid. *Microb Pathog* 2017;113:378-384.
1150 132. **Dressman JB, Berardi RR, Dermentzoglou LC, Russell TL, Schmaltz SP et al.** Upper
1151 gastrointestinal (GI) pH in young, healthy men and women. *Pharm Res* 1990;7(7):756-761.
1152 133. **Russell TL, Berardi RR, Barnett JL, Dermentzoglou LC, Jarvenpaa KM et al.**
1153 Upper gastrointestinal pH in seventy-nine healthy, elderly, North American men and women.
1154 *Pharm Res* 1993;10(2):187-196.
1155 134. **Rathman M, Sjaastad MD, Falkow S**. Acidification of phagosomes containing
1156 *Salmonella typhimurium* in murine macrophages. *Infect Immun* 1996;64(7):2765-2773.
1157 135. **Smith JL**. The role of gastric acid in preventing foodborne disease and how bacteria
1158 overcome acid conditions. *J Food Prot* 2003;66(7):1292-1303.
1159 136. **Foster JW, Hall HK**. Adaptive acidification tolerance response of *Salmonella*
1160 *typhimurium*. *J Bacteriol* 1990;172(2):771-778.
1161 137. **Koutsoumanis KP, Sofos JN**. Comparative acid stress response of *Listeria*
1162 *monocytogenes*, *Escherichia coli* O157:H7 and *Salmonella* Typhimurium after habituation at
1163 different pH conditions. *Lett Appl Microbiol* 2004;38(4):321-326.
1164 138. **Torres MA, Terraf MCL, Minahk CJ, Delgado MA**. Stability of the *Salmonella*
1165 Typhimurium *rcsC11* mutant under different stress conditions. *Microbiology (Reading)*
1166 2020;166(2):157-168.
1167 139. **Hofmann AF, Hagey LR**. Bile acids: chemistry, pathochemistry, biology, pathobiology,
1168 and therapeutics. *Cell Mol Life Sci* 2008;65(16):2461-2483.
1169 140. **Begley M, Gahan CG, Hill C**. The interaction between bacteria and bile. *FEMS*
1170 *Microbiol Rev* 2005;29(4):625-651.
1171 141. **Merritt ME, Donaldson JR**. Effect of bile salts on the DNA and membrane integrity of
1172 enteric bacteria. *J Med Microbiol* 2009;58(Pt 12):1533-1541.
1173 142. **Prieto AI, Ramos-Morales F, Casadesus J**. Bile-induced DNA damage in *Salmonella*
1174 *enterica*. *Genetics* 2004;168(4):1787-1794.
1175 143. **Prieto AI, Ramos-Morales F, Casadesus J**. Repair of DNA damage induced by bile
1176 salts in *Salmonella enterica*. *Genetics* 2006;174(2):575-584.
1177 144. **Hernandez SB, Cota I, Ducret A, Aussel L, Casadesus J**. Adaptation and
1178 preadaptation of *Salmonella enterica* to Bile. *PLoS Genet* 2012;8(1):e1002459.
1179 145. **Walawalkar YD, Vaidya Y, Nayak V**. Response of *Salmonella* Typhi to bile-generated
1180 oxidative stress: implication of quorum sensing and persister cell populations. *Pathogens and*
1181 *Disease* 2016;74(8):ftw090.
1182 146. **Di Lorenzo P**. usmap: US Maps Including Alaska and Hawaii. https://CRAN.R-
1183 project.org/package=usmap.
1184

TABLES

Table 1. Presence and absence of *artAB*, *gogB*, and Gifsy-1 among four DT104 complex genome datasets queried in this study.

Data Set(s)	Host(s)	Total # of Genomes	*artAB* Present (%)[a]	*gogB* Present (%)[a]	Gifsy-1 Present (%)[b]
Dataset 1 *(U.S. bovine and human data)* [c]					
	All	230	177 (77.0%)	180 (78.3%)	180 (78.3%)
	Bovine	168	150 (89.3%)	153 (91.1%)	153 (91.1%)
	Human	62	27 (43.5%)	27 (43.5%)	27 (43.5%)
Dataset 2 *(Scottish bovine and human data)* [d]					
	All	243	240 (98.8%)	240 (98.8%)	144 (59.3%)
	Bovine	82	82 (100.0%)	82 (100.0%)	48 (58.5%)
	Human	161	158 (98.1%)	158 (98.1%)	96 (59.6%)
Dataset 3 *(multi-source data)*[e]					
	All	290	271 (93.4%)	271 (93.4%)	265 (91.4%)
Dataset 4 *(combined global dataset)*[f]					
	All	752	678 (90.2%)	681 (90.6%)	579 (77.0%)

[a] Identified using nucleotide BLAST (blastn; default settings, with no minimum identity or coverage threshold employed).
[b] Identified using the PHASTER webserver; Gifsy-1 regions annotated as "intact", "incomplete" or "questionable" were considered to be "present" in a genome.
[c] Refers to the set of 230 U.S. human- and bovine-associated DT104 complex genomes identified and aggregated here.
[d] Refers to a set of 243 Scottish human- and bovine-associated DT104 genomes sequenced and characterized previously [102].
[e] Refers to a set of 290 DT104 genomes collected from various sources around the world, which were sequenced and characterized previously [13].
[f] Refers to the union of Dataset 1 (U.S. bovine and human data), Dataset 2 (Scottish bovine and human data), and Dataset 3 (multi-source data); eleven of the 290 genomes in Dataset 3 (multi-source data) were isolated from cattle and humans in the U.S. and thus were also included in Dataset 1 (U.S. bovine and human data).
=

Table 2. Location of *artAB* and *gogB* in DT104 complex genomes within the four datasets queried in this study.

Genes[a]	Dataset (# of Genomes with Gene/ Total # of Genomes; %)	# of Genes Detected[a]						
		Within Gifsy-1[b]		Within Gifsy-2[b]		Within *Salmonella* phage 118970_sal3[b]		Outside of Annotated Prophage Regions (%)[b,c]
		Intact (%)[c]	Incomplete (%)[c]	Intact (%)[c]	Incomplete (%)[c]	Intact (%)[c]	Incomplete (%)[c]	
artAB								
	Dataset 1[d] (177/230; 77.0%)	177 (100.0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	Dataset 2[e] (240/243; 98.8%)	21 (8.8)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	219 (91.3)[h]
	Dataset 3[f] (271/290; 93.4%)	263 (97.0)	0 (0)	0 (0)	0 (0)	2 (0.7)	0 (0)	6 (2.2)[i]
	Dataset 4[g] (678/752; 90.2%)	451 (66.5)	0 (0)	0 (0)	0 (0)	2 (0.3)	0 (0)	225 (33.2)
gogB								
	Dataset 1[d] (180/230; 78.3%)	51 (28.3)	75 (41.7)	1 (0.6)	2 (1.1)	0 (0)	0 (0)	51 (28.3)
	Dataset 2[e] (240/243; 98.8%)	50 (20.8)	85 (35.4)	0 (0)	0 (0)	0 (0)	0 (0)	105 (43.8)
	Dataset 3[f] (271/290; 93.4%)	11 (4.1)	67 (24.7)	2 (0.7)	1 (0.4)	0 (0)	0 (0)	190 (70.1)
	Dataset 4[g] (681/752; 90.6%)	112 (16.4)	223 (32.7)	3 (0.4)	3 (0.4)	0 (0)	0 (0)	340 (49.9)

[a] Identified using nucleotide BLAST (blastn; default settings, with no minimum identity or coverage threshold employed).
[b] Identified using the PHASTER webserver; "intact" refers to prophage classified by PHASTER as "intact", while "incomplete" encompasses prophage classified as "incomplete" or "questionable".
[c] Percentages in parentheses were calculated relative to the "# of genomes with gene" value in the "Dataset" column used as a denominator.
[d] Refers to the set of 230 U.S. human- and bovine-associated DT104 complex genomes identified and aggregated here.
[e] Refers to a set of 243 Scottish human- and bovine-associated DT104 genomes sequenced and characterized previously [102].
[f] Refers to a set of 290 DT104 genomes collected from various sources around the world, which were sequenced and characterized previously [13].
[g] Refers to the union of Dataset 1 (U.S. bovine and human data), Dataset 2 (Scottish bovine and human data), and Dataset 3 (multi-source data); eleven of the 290 genomes in Dataset 3 (multi-source data) were isolated from cattle and humans in the U.S. and thus were also included in Dataset 1 (U.S. bovine and human data).
[h] When 3 Kbp regions on either side of PHASTER prophage regions were considered, 234 Dataset 2 (Scottish bovine and human data) genomes harbored *artAB* within 3 Kbp of Gifsy-1 (97.5% of 240 *artAB*-harboring genomes in Dataset 2 [Scottish bovine and human data]); two and one genome harbored *artAB* within 3 Kb of other PHASTER prophage regions (annotated by PHASTER as Edward_GF_2_NC_026611 and PHAGE_Entero_HK630_NC_019723, respectively).
[i] One genome (DK_7322994_6_swine_08-08-01) had *artA* and *artB* on separate contigs, with *artB* detected within 3 Kbp of an intact Gifsy-1 prophage region.

FIGURES

Figure 1. (A) Selected prophages that share homology with prophage Gifsy-1, as described in *Salmonella* Typhimurium strains (i) LT2, (ii) DT104, and (iii) D23580. Prophage regions were acquired from the PHASTER database and annotated using Prokka. clinker was used to compare prophage regions using default settings. Arrows correspond to open reading frames (ORFs), with grayscale links denoting the percent (%) amino acid identity shared between corresponding ORFs. Selected Gifsy-1 and Gifsy-2 virulence factors [29] are annotated. To view a similar plot constructed using all PHASTER prophage regions in LT2, DT104, and D23580, see Supplementary Figure S1. (B) Venn diagram showcasing the relationship between DT104 complex datasets used in this study. Numbers denote the number of genomes within a given dataset or subset of a dataset. For an extended version of this figure, see Supplementary Figure S2. For a flow chart with detailed descriptions of the datasets used in this study, see Supplementary Figure S3.

Figure 2. Geographic and source origins (i.e., human or bovine) of (A) all 230 human- and bovine-associated United States (U.S.) DT104 complex genomes queried in this study (i.e., Dataset 1 [U.S. bovine and human data]), and (B) 42 Gifsy-1/*artAB*/*gogB*-negative genomes assigned to the U.S. *artAB*-negative major clade. U.S. states shown in gray did not contribute any genomes to the respective data set. The U.S. state that contributed the most genomes to its respective data set is labeled. The figure was created using BioRender (https://biorender.com/) and the "plot_usmap" function in the usmap version 0.6.0 R package [146].

Figure 3. Bayesian time-scaled phylogeny constructed using 146 human- and bovine-associated DT104 complex genomes collected in the United States (U.S.; i.e., a subset of genomes from Dataset 1 [U.S. bovine and human data]). Tip label colors denote the isolation source reported for each genome (human or bovine in pink and blue, respectively). The heatmap to the right of the phylogeny denotes the presence and absence of (i) selected virulence factors (dark and light pink, respectively) and (ii) prophage (dark and light green, respectively). The U.S. *artAB*-negative major clade is denoted by the bright purple bar; light purple shading around the node of the U.S. *artAB*-negative major clade denotes the 95% highest posterior density (HPD) interval, in which Gifsy-1/*artAB*/*gogB* were predicted to have been lost. The phylogeny was constructed and rooted using BEAST2. Time in years is plotted along the X-axis, while branch labels correspond to posterior probabilities of branch support (selected for readability). For extended versions of this figure, see Supplementary Figures S6-S8.

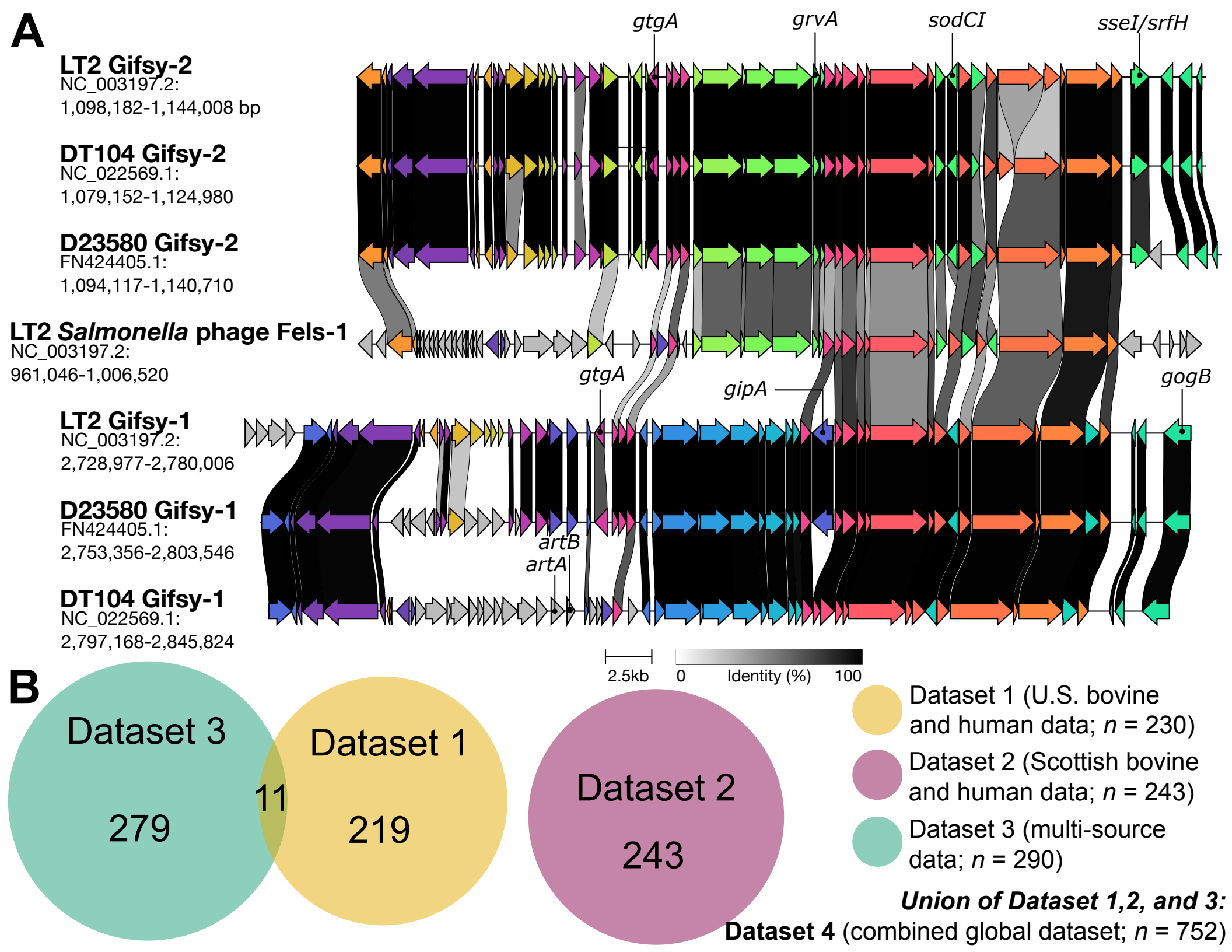
Figure 4. Coalescent Bayesian Skyline plot constructed using 146 U.S. bovine- and human-associated *S.* Typhimurium DT104 complex genomes (i.e., a subset of genomes from Dataset 1 [U.S. bovine and human data]). Effective population size and time in years are plotted along the Y- and X-axes, respectively. The median effective population size estimate is denoted by the solid black line, with upper and lower 95% highest posterior density (HPD) interval bounds denoted by gray shading. The interval shaded in light blue and bounded by dashed vertical lines denotes the time interval in which Gifsy-1/*artAB*/*gogB* were predicted to have been lost by the common ancestor of the U.S. *artAB*-negative major clade (corresponding to the years 1985.0 and 1987.2, denoted by turquoise and pink dashed lines, respectively). The dotted turquoise and pink vertical lines correspond to the 95% HPD interval lower and upper bounds for Gifsy-1/*artAB*/*gogB* loss

among members of the U.S. *artAB*-negative major clade (corresponding to the years 1979.0 and 1992.1, respectively).

Figure 5. Time-scaled maximum likelihood (ML) phylogeny constructed using 752 DT104 complex genomes (i.e., Dataset 4 [combined global dataset], the union of Dataset 1 [U.S. bovine and human data], Dataset 2 [Scottish bovine and human data], and Dataset 3 [multi-source data]). Tip label colors denote the dataset or dataset subset with which each genome is affiliated ("Dataset"; eleven genomes within the intersection of Dataset 1 [U.S. bovine and human data] and Dataset 3 [multi-source data] were colored as Dataset 1 [U.S. bovine and human data]). The heatmap encompassing the phylogeny denotes the presence and absence of selected virulence factors and intact prophage (identified via nucleotide BLAST [blastn; default settings, with no minimum identity or coverage threshold employed] and PHASTER, respectively). Bright green circles in the outermost ring surrounding the phylogeny denote genomes reportedly assigned to DT104 using phage typing ("DT104"). The ML phylogeny was constructed using IQ-TREE and rooted and time-scaled using LSD2. Branch lengths are reported in years. For an extended version of this figure, see Supplementary Figure S14.

Figure 6. Response of DT104 complex isolates (n = 6) to environmental stress factors within the context of Gifsy-1/*artAB*/*gogB* presence and absence. Base-ten logarithmic fold change (FC) was calculated as follows: FC = log CFU/g at the start of the experiments – log CFU/g after the stress assay. (A) Log FC of DT104 inoculated into ruminal fluid at high (10^8 CFU/mL; "High") or low (10^5 CFU/mL; "Low") bacterial numbers. (B) Log FC of DT104 isolates exposed to inorganic

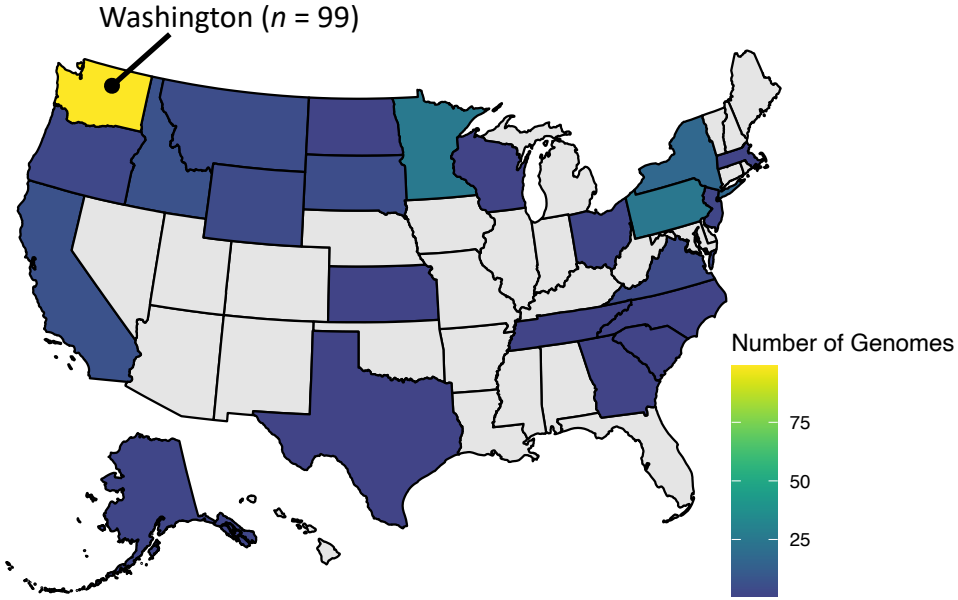
acid stress (pH 3.5) with or without a prior adaption step with an intermediate pH (pH 5.5). (C) Log FC of DT104 isolates after exposure to bile salt at two concentrations.

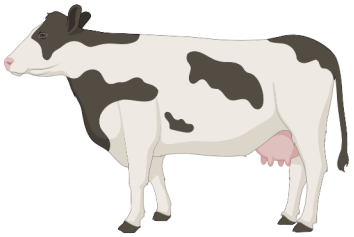

A
LT2 Gifsy-2
NC_003197.2:
1,098,182-1,144,008 bp
DT104 Gifsy-2
NC_022569.1:
1,079,152-1,124,980
D23580 Gifsy-2
FN424405.1:
1,094,117-1,140,710
LT2 Salmonella phage Fels-1
NC_003197.2:
961,046-1,006,520
LT2 Gifsy-1
NC_003197.2:
2,728,977-2,780,006
D23580 Gifsy-1
FN424405.1:
2,753,356-2,803,546
DT104 Gifsy-1
NC_022569.1:
2,797,168-2,845,824
gtgA
grvA
sodCI
sseI/srfH
gtgA
gipA
gogB
artB
artA
2.5kb
0
Identity (%)
100
B
Dataset 3
279
11
Dataset 1
219
Dataset 2
243
Dataset 1 (U.S. bovine and human data; n = 230)
Dataset 2 (Scottish bovine and human data; n = 243)
Dataset 3 (multi-source data; n = 290)
Union of Dataset 1,2, and 3:
Dataset 4 (combined global dataset; n = 752)

A.

U.S. DT104 genomes

(full data set; *n* = 230)





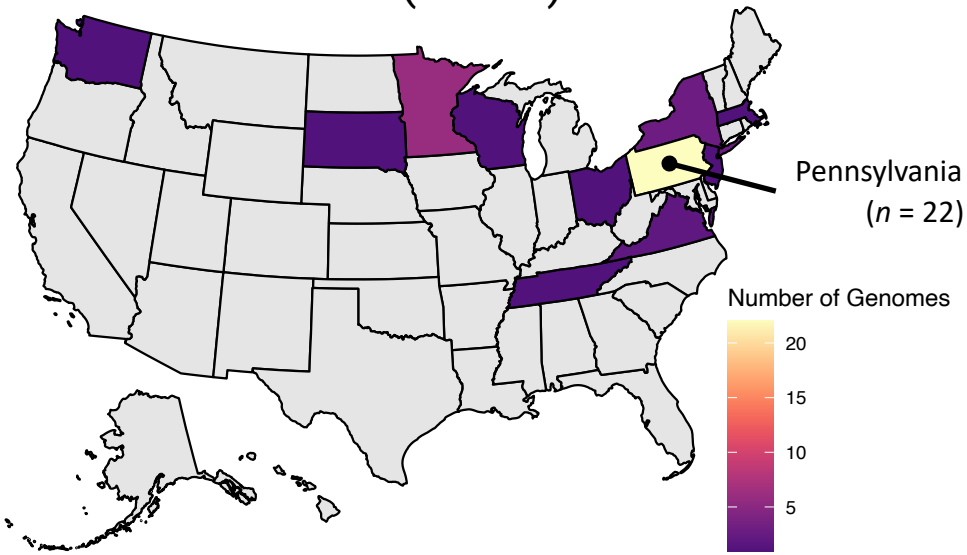
n = 62

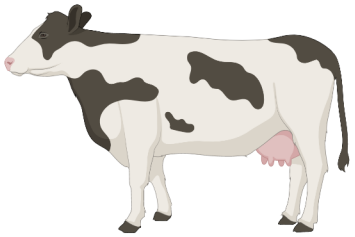
n = 168

B.

U.S. *artAB*-negative major clade

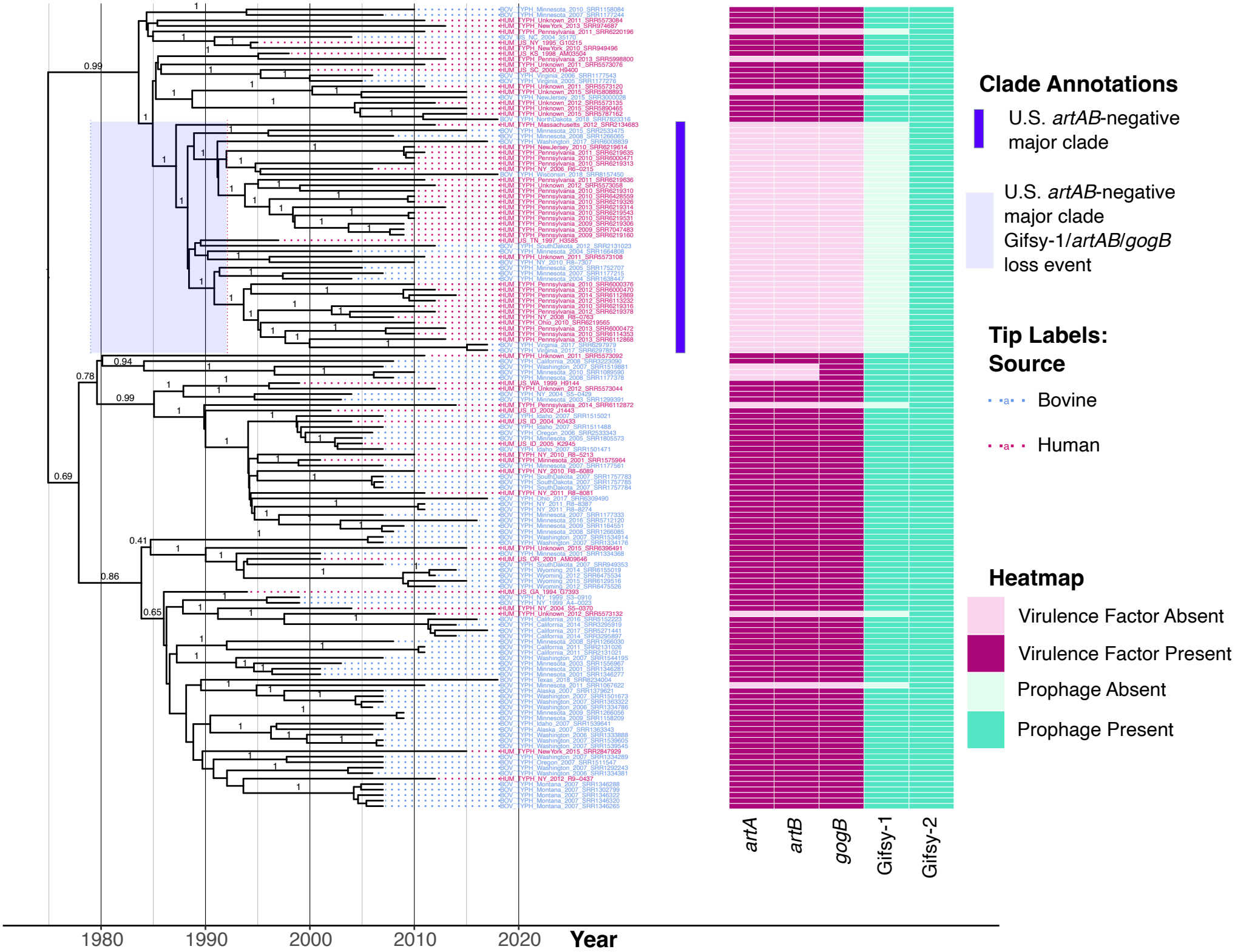
(*n* = 42)

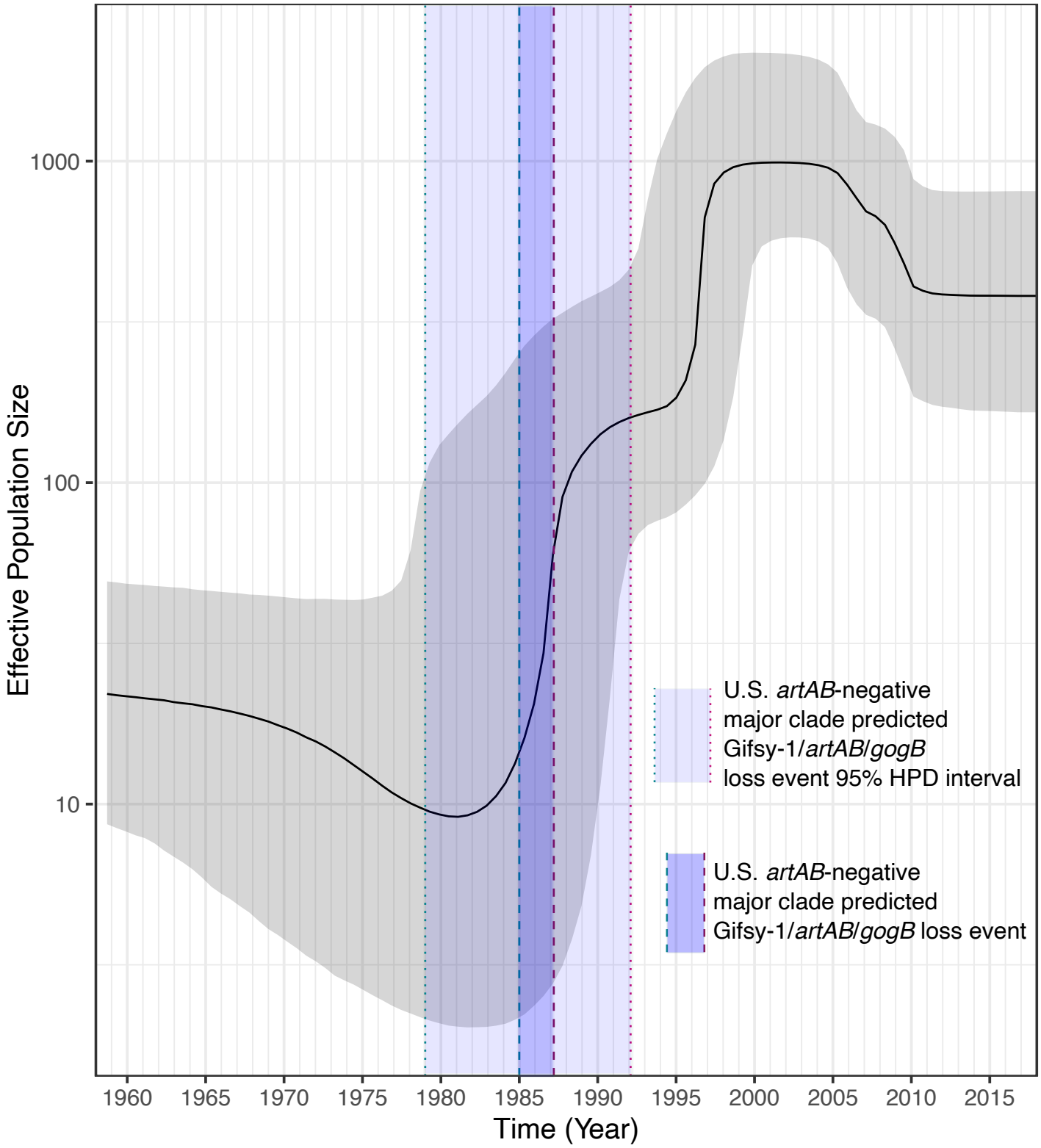


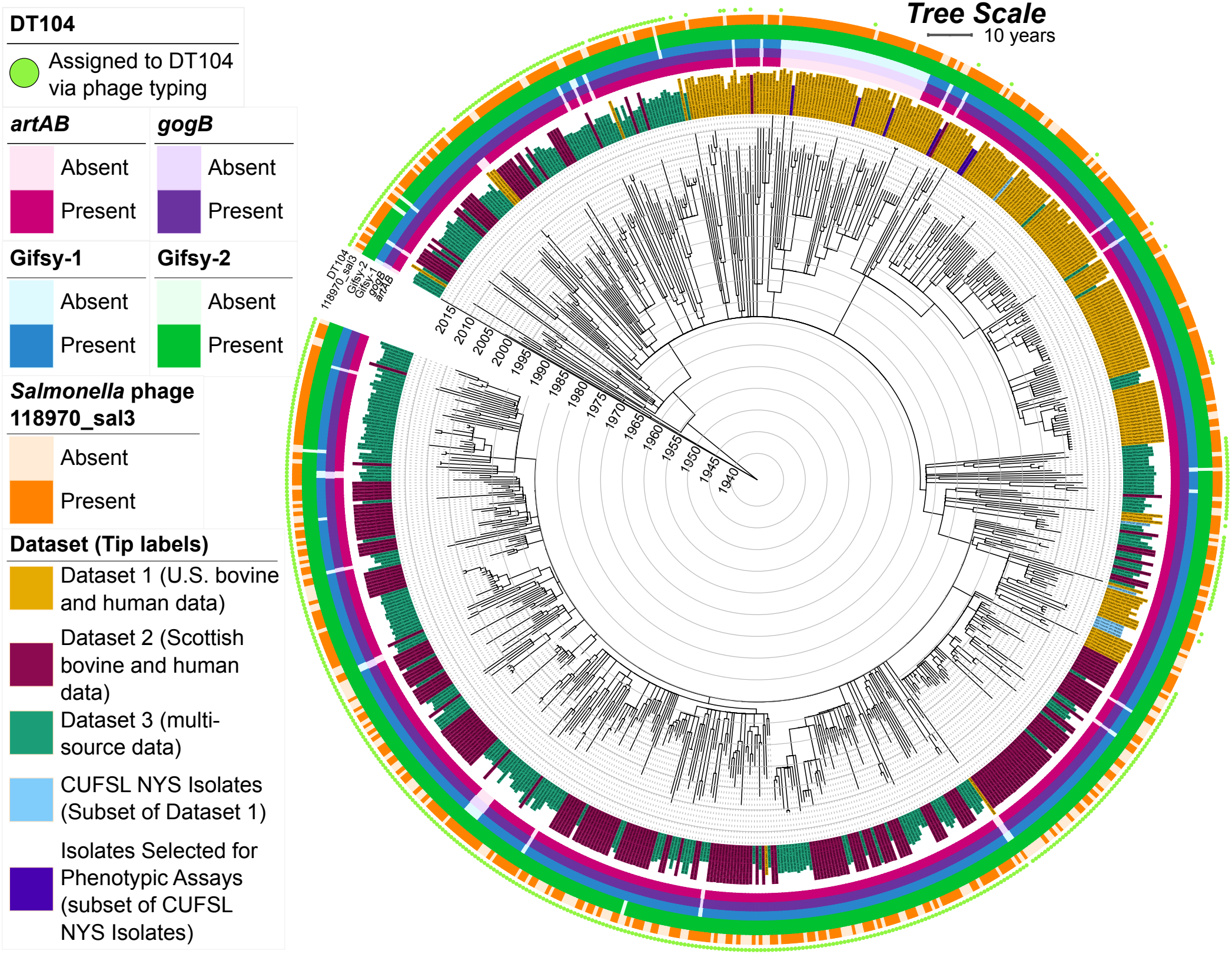


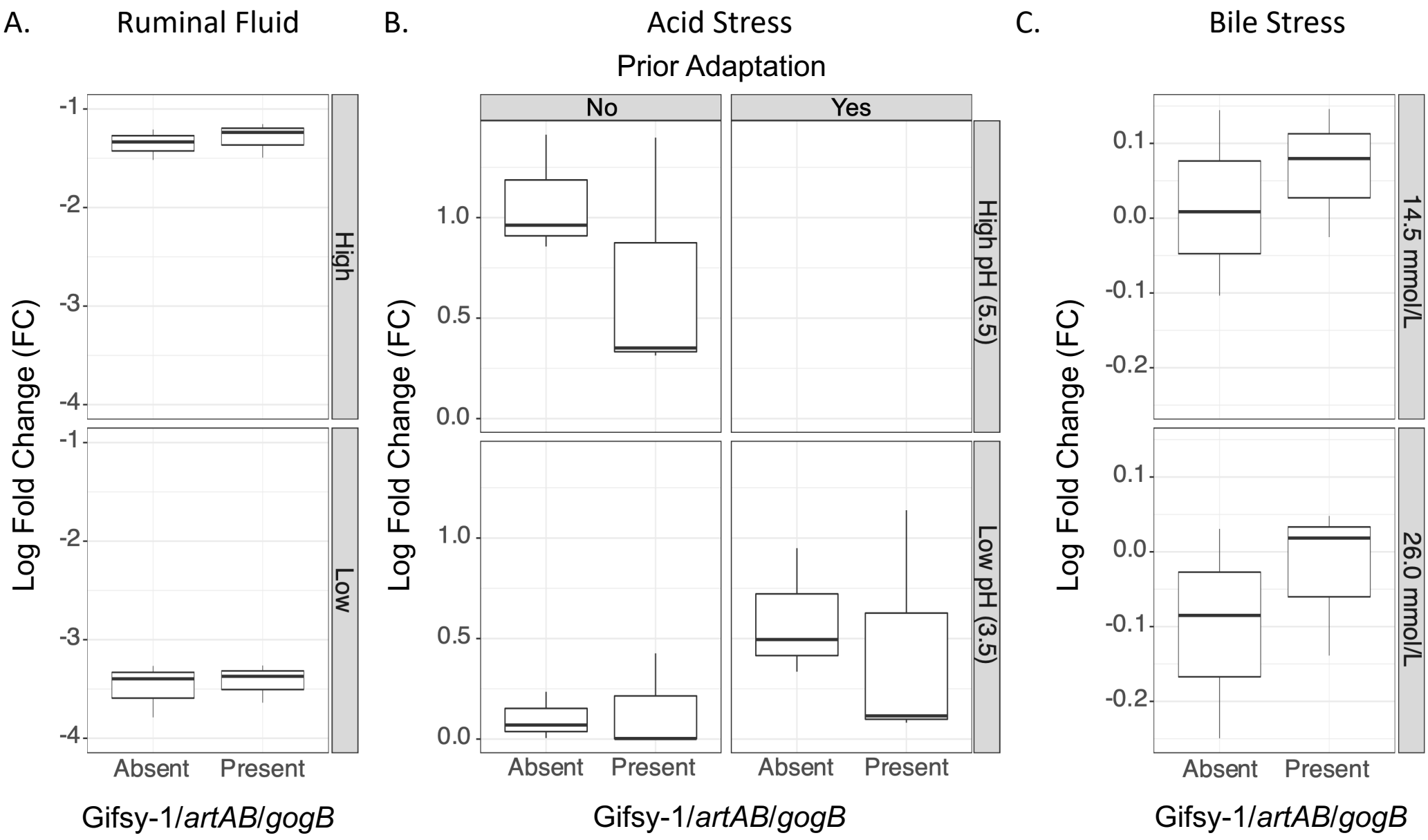
n = 30

n = 12


Clade Annotations
U.S. artAB-negative major clade
U.S. artAB-negative major clade Gifsy-1/artAB/gogB loss event
Tip Labels: Source
Bovine
Human
Heatmap
Virulence Factor Absent
Virulence Factor Present
Prophage Absent
Prophage Present
artA
artB
gogB
Gifsy-1
Gifsy-2
1980
1990
2000
2010
2020
Year


Effective Population Size
1000
100
10
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
Time (Year)
U.S. *artAB*-negative
major clade predicted
Gifsy-1/*artAB*/*gogB*
loss event 95% HPD interval
U.S. *artAB*-negative
major clade predicted
Gifsy-1/*artAB*/*gogB* loss event


DT104
Assigned to DT104 via phage typing
artAB
Absent
Present
gogB
Absent
Present
Gifsy-1
Absent
Present
Gifsy-2
Absent
Present
Salmonella phage 118970_sal3
Absent
Present
Dataset (Tip labels)
Dataset 1 (U.S. bovine and human data)
Dataset 2 (Scottish bovine and human data)
Dataset 3 (multi-source data)
CUFSL NYS Isolates (Subset of Dataset 1)
Isolates Selected for Phenotypic Assays (subset of CUFSL NYS Isolates)
Tree Scale
10 years
DT104
118970_sal3
Gifsy-2
Gifsy-1
gogB
artAB
2015
2010
2005
2000
1995
1990
1985
1980
1975
1970
1965
1960
1955
1950
1945
1940


A.
Ruminal Fluid
B.
Acid Stress
Prior Adaptation
C.
Bile Stress
Log Fold Change (FC)
High
Low
No
Yes
High pH (5.5)
Low pH (3.5)
14.5 mmol/L
26.0 mmol/L
Absent
Present
Gifsy-1/artAB/gogB