A multidrug-resistant *Salmonella enterica* serotype Typhimurium DT104 lineage circulating among humans and cattle in the United States lost the ability to produce pertussis-like toxin ArtAB in close temporal proximity to the global DT104 epidemic

Laura M. Carroll[1,*], Nicolo Piacenza[2], Rachel A. Cheng[1], Martin Wiedmann[1], Claudia Guldimann[2,#]

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

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

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

*Current address: Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany

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

Repositories: Not applicable

Abstract

Salmonella enterica subspecies *enterica* serotype Typhimurium phage type DT104 (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 a pertussis-like toxin, ArtAB, via prophage-encoded *artAB*; however, DT104 that lack *artAB* have been described on occasion. Here, we identify a MDR DT104 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 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 was predicted to have lost *artAB*, Gifsy-1, and *gogB* circa 1985-1987 (95% highest posterior density interval 1979.0-1992.1), in close temporal proximity to a predicted rapid increase in the U.S. DT104 effective population size (circa 1983-1989). 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. In 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 strains (ANOVA raw $P > 0.05$); thus, future research is needed to elucidate the roles that *artAB*, *gogB*, and Gifsy-1 play in DT104 virulence in humans and animals.

Impact Statement

Multi-drug resistant (MDR) *Salmonella enterica* serotype Typhimurium phage type DT104 (DT104) was responsible for a global epidemic among humans and animals throughout the 1990s and continues to circulate worldwide. Previous studies have indicated that the vast majority of DT104 produce a pertussis-like toxin, ArtAB, via prophage-encoded *artAB*. However, here we identify a DT104 lineage circulating among cattle and humans across ≥11 U.S. states, which lacks the ability to produce ArtAB (i.e., the "U.S. *artAB*-negative major clade"). The common ancestor of all U.S. *artAB*-negative major clade members lost the ability to produce ArtAB in close temporal proximity to the global MDR DT104 epidemic; however, the reason for this loss-of-function event within this well-established pathogen remains unclear. The role that ArtAB plays in DT104 virulence remains elusive, and phenotypic assays conducted here indicate that members of the U.S. *artAB*-negative major clade do not have a significant advantage or disadvantage relative to closely related Gifsy-1/*artAB*/*gogB*-harboring U.S. DT104 strains when exposed to stressors encountered during human and/or bovine digestion *in vitro*. However, ArtAB heterogeneity among DT104 suggest clade-specific selection for or against maintenance of ArtAB. Thus, future studies querying the virulence potential of the U.S. *artAB*-negative major clade are needed.

Data Summary. The authors confirm all supporting data, code and protocols have been provided within the article or through supplementary data files.

INTRODUCTION

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

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

However, despite its rapid global dissemination, DT104 does not appear to be more virulent than non-DT104 *S.* Typhimurium in a classical mouse model [16].

In addition to its characteristic MDR phenotype, DT104 is notable for its ability to produce ArtAB, a pertussis-like toxin that catalyzes ADP-ribosylation of host G proteins [17-19]. Treatment of various cell lines with purified ArtAB from DT104 recapitulates some of the phenotypes established for pertussis toxin cytotoxicity [20-22], such as the characteristic "cell clustering" phenotype in CHO-K1 cells [23], increased levels of intracellular cAMP in RAW 264.7 macrophage-like cells [18], and increased serum insulin levels (e.g., insulinemia); further, intraperitoneal injection of purified toxin in neonatal mice was fatal [18]. *artAB*, which encodes ArtAB, shares a strong association with DT104 relative to other *S.* Typhimurium lineages, as the overwhelming majority of DT104 possess *artAB* [18]. However, occasionally, DT104 strains that lack *artAB* and, thus, the ability to produce ArtAB toxin, have been described [18]; *artAB* is located on prophages within the DT104 genome [18, 19], indicating that it may be possible for *artAB* to be lost or gained as a prophage integrates or excises from a genome, or via HGT within an integrated prophage.

In a previous study of *S.* Typhimurium isolated from dairy cattle and human clinical cases in New York State (United States; U.S.), we identified three bovine- and human-associated DT104 strains, which did not possess *artAB* (referred to hereafter as "*artAB*-negative" strains) [24]. Interestingly, these three *artAB*-negative strains were closely related and formed a clade within the largely *artAB*-positive New York State DT104 phylogeny [24]. Here, we investigate this lineage further: using (i) 230 human- and bovine-associated DT104 genomes collected across the U.S., plus (ii) 752 DT104 genomes collected from numerous sources worldwide, we identify a major DT104 lineage circulating among cattle and humans across the U.S., which lost

artAB, as well as a co-occurring anti-inflammatory effector encoded by *gogB*, via a Gifsy-1 prophage loss event that occurred in close temporal proximity to the global MDR DT104 epidemic.

METHODS

Acquisition of U.S. DT104 genomic data and metadata. Genomic data from human- and bovine-associated DT104 isolates from the U.S. were acquired as described previously [24]. Briefly, paired-end Illumina short reads associated with 223 *S.* Typhimurium genomes meeting the following criteria were downloaded via Enterobase (accessed November 29, 2018) [25, 26] and the Sequence Read Archive (SRA) Toolkit version 2.9.3 [27, 28]: (i) genomes were serotyped as *S.* Typhimurium *in silico* using the implementation of SISTR [29] in Enterobase; (ii) the country of isolation was the United States; (iii) the isolation source was reported as either "Human" or "Bovine" in the "Source Niche" and "Source Type" fields in Enterobase, respectively; (iv) genomes had an isolation year reported in Enterobase; (v) genomes were assigned to the same well-supported cluster within the larger bovine- and human-associated U.S. *S.* Typhimurium phylogeny using RhierBAPS [30] and clustered among known DT104 genomes from other countries (see Supplementary Figures S2 and S5 of Carroll, et al.) [24]. These genomes were supplemented with an additional 14 *S.* Typhimurium genomes from bovine and human sources in New York State (U.S.) that belonged to the same DT104 cluster (members of the *S.* Typhimurium Lineage III cluster described in Supplementary Figures S2 and S5 of Carroll, et al.) [24]. Trimmomatic version 0.36 [31] was used to trim low quality bases and Illumina adapters from all read sets using the default settings for paired-end reads, and SPAdes version 3.13.0 [32] was used to assemble all genomes using default settings plus the "careful"

option. FastQC version 0.11.5 [33] and QUAST version 4.0 [34] were used to assess the quality of each read pair set and assembly, respectively, and MultiQC version 1.6 [35] was used to aggregate all FastQC and QUAST results. Trimmed paired-end read sets/assemblies that were flagged by MultiQC as meeting any of the following conditions were excluded: (i) Illumina adapters present after trimming (n = 2), (ii) an abnormal per sequence GC content distribution (n = 3), (iii) an assembly with over 200 contigs (n = 11), and (iv) a sequence quality histogram flagged as poor quality (n = 2). After excluding genomes that met these conditions, a set of 219 DT104 genomes was produced for use in subsequent steps (Supplementary Tables S1 and S2).

Acquisition of global DT104 genomic data and metadata. To contextualize the 219 U.S. bovine- and human-associated DT104 genomes identified via the initial genome search (see section "Acquisition of U.S. DT104 genomic data and metadata" above) within the larger global DT104 population, genomic data associated with the following studies were downloaded via Enterobase: (i) Illumina reads associated with 243 bovine- and human-associated DT104 isolates from a study of between-host transmission within Scotland [36] (referred to hereafter as the "Scottish DT104" data set); genomes were pre-processed and assembled as described above (see section "Acquisition of U.S. DT104 genomic data and metadata" above); (ii) assembled genomes associated with 290 DT104 isolates from a variety of sources and countries from a study describing the global spread of DT104 [13] (referred to hereafter as the "global DT104" data set; Supplementary Table S1). Overall, these searches resulted in two data sets, which were used in subsequent steps: (i) a 230-genome human- and bovine-associated U.S. DT104 data set (i.e., 219 genomes identified in this study, plus 11 additional U.S. bovine- and human-associated DT104 strains from the 290-genome "global DT104" data set, which did not have metadata available in Enterobase at the time and were thus not included in the initial set of bovine- and human-

associated U.S. DT104 genomes; see section “Acquisition of U.S. DT104 genomic data and metadata” above); (ii) a 752-genome data set, composed of genomes from all three data sets (i.e., 219 genomes identified in this study, 243 Scottish DT104 genomes, and 290 global DT104 genomes, referred to hereafter as the “combined DT104” data set; Supplementary Tables S1 and S2). QUAST version 4.5 was used to assess the quality of all 752 genomes (Supplementary Tables S1 and S2).

***In silico* detection of antimicrobial resistance genes, plasmid replicons, virulence factors, and prophage.** To identify putative prophage regions in all 752 genomes, each assembly was submitted to the PHASTER web server via the URL API [37, 38] with the “contigs” option set to “1”. ABRicate version 0.8 [39] was used to detect antimicrobial resistance (AMR) genes, plasmid replicons, and virulence factors in each assembled DT104 genome using NCBI’s National Database of Antibiotic Resistant Organisms (NDARO) [40], the PlasmidFinder database [41], and the Virulence Factor Database (VFDB) [42], respectively, using minimum nucleotide identity and coverage thresholds of 75 and 50%, respectively (all databases accessed December 10, 2020; Supplementary Table S3). The aforementioned ABRicate analyses were repeated, using a minimum coverage threshold of 0% (e.g., to confirm that virulence factors discussed in the manuscript were absent from genomes in which they were not initially detected).

Each assembled genome was additionally queried for the presence of selected virulence factors, which have previously been associated with prophage in *Salmonella* [43]: (i) *artAB* (NCBI Nucleotide Accession AB104436.1), (ii) *gogA* (European Nucleotide Archive [ENA] Accession EAA7850902.1), (iii) *gtgA* (ENA Accession PVI70081.1), and (iv) *gipA* (ENA Accession CAI93790.1). Assembled genomes were queried for selected virulence factors using the command-line implementation of nucleotide BLAST (blastn) version 2.11.0 [44], using

default settings plus a minimum coverage threshold of 40% (Supplementary Table S3). To confirm that the aforementioned genes were absent from genomes in which they were not initially detected, all genomes were queried again (i) as described above, with the coverage threshold lowered to 0%; (ii) using translated nucleotide BLAST (tblastx; e.g., to confirm that all genomes in the U.S. *artAB*-negative major clade did not possess remote *artAB* and *gogB* homologs; Supplementary Tables S4 and S5).

Variant calling and maximum likelihood phylogeny construction within the U.S. DT104 data set. Core SNPs were identified among genomes within the 230-genome human- and bovine-associated U.S. DT104 data set using the default pipeline implemented in Snippy version 4.6.0 [45] and the following dependencies: BWA version 0.7.17-r1188 [46, 47], Minimap2 version 2.23-r1111 [48], SAMtools version 1.14 [49], BEDtools version 2.30.0 [50, 51], BCFtools version 1.14 [52], FreeBayes version 1.3.2-dirty [53], vcflib version 1.0.0-rc0-349-g45c6-dirty [54], vt version 0.5 [55], SnpEff version 5.0e [56], samclip version 0.4.0 [57], seqtk version 1.3-r106 [58], and snp-sites version 2.5.1 [59]. For the 219 U.S. genomes initially identified in this study, the trimmed Illumina paired-end reads associated with each genome were treated as input; for the remaining genomes, the assembled contigs were used as input (see sections “Acquisition of U.S. DT104 genomic data and metadata” and “Acquisition of global DT104 genomic data and metadata” above; Supplementary Tables S1 and S2). The closed DT104 chromosome (NCBI Nucleotide Accession NC_022569.1) was treated as a reference. Core SNPs identified in regions of the DT104 chromosome predicted to belong to phage were masked (see section “*In silico* detection of antimicrobial resistance genes, plasmid replicons, virulence factors, and prophage” above). Gubbins version 2.4.1 [60] was used to identify and

remove recombination events among all genomes using default settings, and snp-sites was used to query the resulting recombination-free alignment for core SNPs (i.e., using the "-c" option).

A maximum likelihood (ML) phylogeny was constructed with IQ-TREE version 1.5.4 [61], using (i) the resulting core SNPs as input, (ii) the optimal nucleotide substitution model determined using Bayesian information criteria (BIC) values produced with ModelFinder [62] (i.e., the K3Pu+I model) [63], (iii) an ascertainment bias correction to account for the use of solely variant sites (corresponding to constant sites identified relative to the DT104 reference chromosome; -fconst 1092869,1195194,1193287,1094079), and (iv) 1,000 replicates of the ultrafast bootstrap approximation [64, 65].

TempEst version 1.5.3 [66] was used to assess the temporal structure of the resulting unrooted ML phylogeny, using the best-fitting root and the R^2 function (R^2 = 0.33, slope = 3.05×10^{-7} substitutions/site/year, X-intercept = 1988.1). The unrooted ML phylogeny was additionally rooted and time scaled using LSD2 version 1.4.2.2 [67] and the following parameters: (i) tip dates corresponding to the year of isolation associated with each genome; (ii) an estimated substitution rate; (iii) constrained mode (-c), with the root estimated using constraints on all branches (-r as); (iv) variances calculated using input branch lengths (-v 1); (v) 1,000 samples for calculating confidence intervals for estimated dates (-f 1000); (vi) a sequence length of 4,500,000. The resulting rooted, time-scaled ML phylogeny was viewed using FigTree version 1.4.4 [68] (Supplementary Data).

Variant calling and maximum likelihood phylogeny construction within the combined DT104 data set. Parsnp and HarvestTools version 1.2 [69] were used to detect core SNPs among all 752 assembled DT104 genomes within the combined DT104 data set (see section "Acquisition of global DT104 genomic data and metadata" above; Supplementary Tables S1 and

S2), using the closed DT104 chromosome as a reference (NCBI Nucleotide Accession NC_022569.1) and Parsnp's implementation of PhiPack [70] to filter recombination. Core SNPs detected among all 752 assembled genomes were supplied as input to IQ-TREE version 1.5.4, which was used to construct a ML phylogeny as described above (the corresponding ascertainment bias correction here was "-fconst 1181208,1285673,1280769,1179580"; see section "Variant calling and maximum likelihood phylogeny construction within the U.S. DT104 data set" above). The resulting ML phylogeny was rooted and time-scaled using LSD2 as described above (see section "Variant calling and maximum likelihood phylogeny construction within the U.S. DT104 data set" above). A range of 1900-2017 was supplied for four genomes, which were part of the global DT104 data set, but did not have a reported year of isolation. The resulting time-scaled ML phylogeny was annotated using the Interactive Tree of Life (iTOL) version 6 webserver (https://itol.embl.de/; accessed March 7, 2022) [71].

U.S. DT104 Bayesian time-scaled phylogeny construction. Due to the overrepresentation of genomes of DT104 strains reportedly isolated in 2007 from bovine sources in Washington state within the 230-genome human- and bovine-associated U.S. DT104 data set (Figure 1 and Supplementary Table S1), all aforementioned SNP calling and phylogeny construction steps were repeated among genome sets downsampled to (i) 25, (ii) 10, and (iii) 5 randomly selected bovine DT104 genomes collected in Washington state in 2007 (n = 161, 146, and 141 total genomes in each downsampled genome set, respectively; see section "Variant calling and maximum likelihood phylogeny construction within the U.S. DT104 data set" above, Supplementary Data). For each of the three downsampled U.S. DT104 data sets, BEAST2 version 2.5.1 [72, 73] was used to construct a tip-dated phylogeny using core SNPs detected among the genomes within the respective data set as input (see section "Variant calling and

maximum likelihood phylogeny construction within the U.S. DT104 data set" above), an initial clock rate of 2.79×10^{-7} substitutions/site/year [13], and an ascertainment bias correction to account for the use of solely variant sites [74]. bmodeltest [75] was used to infer a substitution model using Bayesian model averaging, with transitions and transversions split. A relaxed lognormal molecular clock [76] and a coalescent Bayesian skyline population model [77] were used, as these models have been selected as the optimal clock/population model combination for DT104 previously [13]. A log-normal distribution with a mean of 4.6×10^{-7} and standard deviation of 1 (median of 2.79×10^{-7}) was used as the prior on the uncorrelated log-normal relaxed molecular clock mean rate parameter (ucld.mean; Supplementary Data).

For each of the three downsampled U.S. DT104 data sets, five independent BEAST2 runs were performed, using chain lengths of at least 100 million generations, sampling every 10 thousand generations. For each downsampled data set, LogCombiner-2 was used to aggregate the resulting log and tree files with 10% of the states treated as burn-in, and TreeAnnotator-2 was used to produce a maximum clade credibility (MCC) tree using Common Ancestor node heights (Supplementary Data). The resulting phylogenies was displayed and annotated using R version 4.1.2 [78] and the following packages: ggplot2 version 3.3.5 [79], ggtree version 3.2.1 [80, 81], phylobase version 0.8.10 [82], and treeio version 1.18.1 [83].

All three downsampled U.S. DT104 data sets resulted in similar BEAST2 parameter estimates (Supplementary Figure S1, Supplementary Table S6, and Supplementary Data). Thus, the final Bayesian time-scaled phylogeny and associated parameter estimates reported in the main manuscript correspond to those obtained using the U.S. DT104 data set, which was downsampled to 10 randomly selected bovine DT104 genomes collected in Washington state in 2007 (n = 146 genomes, Figures 2 and 3). Results are available for the U.S. DT104 data sets

downsampled to 25 and 5 bovine DT104 genomes collected in Washington state in 2007 (Supplementary Figures S2-S5, Supplementary Table S6, and Supplementary Data).

***artAB* ancestral state reconstruction.** To estimate ancestral character states of internal nodes in the (i) U.S. DT104 and (ii) combined DT104 data set phylogenies as they related to *artAB* presence/absence (i.e., whether a node in the tree represented an ancestor that was more likely to be *artAB*-positive or *artAB*-negative), the presence or absence of *artAB* within each genome was treated as a binary state. The following phylogenies were each used as input: (i) the BEAST2 time-scaled Bayesian U.S. DT104 phylogenies (n = 161, 146, and 141 total genomes in each downsampled data set; see section "U.S. DT104 Bayesian time-scaled phylogeny construction" above); (ii) the LSD2 time-scaled ML combined DT104 data set phylogeny (n = 752; see section "Variant calling and maximum likelihood phylogeny construction within the combined DT104 data set" above). Stochastic character maps were simulated on each phylogeny using the make.simmap function in the phytools version 1.0-1 R package [84] and the all-rates-different (ARD) model in the ape version 5.6-1 package. For each phylogeny, either (i) equal root node prior probabilities for *artAB*-positive and *artAB*-negative states (i.e., P(*artAB* present) = P(*artAB*-absent) = 0.5), or (ii) estimated root node prior probabilities for *artAB*-positive and *artAB*-negative states obtained using the make.simmap function were used. For each root node prior/phylogeny combination (eight total combinations of two root node priors and four phylogenies), an empirical Bayes approach was used, in which a continuous-time reversible Markov model was fitted, followed by 10,000 simulations of stochastic character histories using the fitted model and tree tip states. The resulting phylogenies were plotted using the densityMap function in the phytools R package. For the U.S. DT104 data set, the final ancestral state results reported in the main manuscript correspond to those obtained using the U.S. DT104 data set,

which was downsampled to 10 randomly selected bovine DT104 genomes collected in Washington state in 2007 (*n* = 146 genomes, Figure 2). Results are available for the U.S. DT104 data sets downsampled to 25 and 5 bovine DT104 genomes collected in Washington state in 2007 (Supplementary Figures S6-S10 and Supplementary Data).

Pan-genome characterization. Prokka version 1.13.3 [85] was used to annotate all 752 DT104 genomes, using the "Bacteria" database (see section "Genomic comparison of U.S. DT104 to the global DT104 population" above; Supplementary Tables S1 and S2). GFF files produced by Prokka were supplied as input to Panaroo version 1.2.7 [86], which was used to identify core- and pan-genome orthologous gene clusters among (i) the 230 U.S. DT104 genomes and (ii) all 752 DT104 genomes in the combined data set, using the following parameters: (i) "strict" mode (--clean-mode strict); (ii) MAFFT as the sequence aligner (--aligner mafft) [87, 88]; (iii) a core genome threshold of 98% (i.e., genes present in at least 98% of genomes were considered to be core genes; --core_threshold 0.98); (iv) a protein family sequence identity threshold of 70% (-f 0.7, the default). The LSD2 time-scaled ML phylogenies for the (i) 230-genome U.S. DT104 and (ii) combined DT104 data sets (see sections "Variant calling and maximum likelihood phylogeny construction within the U.S. DT104 data set" and "Variant calling and maximum likelihood phylogeny construction within the combined DT104 data set" above) were supplied as input to Panaroo's "panaroo-img" and "panaroo-fmg" commands, which were used to estimate the pan-genome size under the Infinitely Many Genes (IMG) [89, 90] and Finite Many Genes (FMG) models (with 100 bootstrap replicates) [91], respectively (Supplementary Figure S11). Reference pan-genome coding sequences (CDS) identified by Panaroo for both the (i) U.S. and (ii) combined DT104 data sets underwent functional annotation using the eggNOG-mapper version 2 webserver (http://eggnog-mapper.embl.de/; accessed July 24, 2022) using default settings [92,

93]. Among genomes within the U.S. DT104 data set (n = 230), the “table” function in R was used to identify genes associated with (i) Gifsy-1 presence/absence (Supplementary Table S7) and (ii) major clade membership (Supplementary Table S8); the “fisher.test” function in R’s stats package was used to conduct two-sided Fisher’s exact tests, and the “p.adjust” function was used to control the false discovery rate (i.e., p.adjust method = “fdr”) [94].

The treewas version 1.0 R package [95] was additionally used to identify potential gene-host associations among the 230-genome U.S. DT104 data set (i.e., whether a gene identified with Panaroo was human- or bovine-associated while accounting for population structure), using the following: (i) the isolation source treated as a discrete phenotype (i.e., a vector of “human” or “bovine”, supplied to the treeWAS function’s “phen” argument; phen.type = “discrete”); (ii) unique gene presence/absence profiles of genes detected in ≥10 and ≤220 of 230 total U.S. DT104 genomes, treated as the genotypes to test (supplied to the treeWAS function’s “snps” argument); (iii) the time-scaled ML phylogeny constructed using LSD2 for the treeWAS function’s “tree” argument (see section “Variant calling and maximum likelihood phylogeny construction within the U.S. DT104 data set” above); (iv) the number of simulated loci for estimating the null distribution set to five million (i.e., n.snps.sim = 5000000); (v) ancestral state reconstruction performed using ML methods (i.e., snps.reconstruction = “ML”, snps.sim.reconstruction = “ML”, and phen.reconstruction = “ML”); (vi) a P-value significance threshold of 0.1, after controlling the FDR (p.value.correct = "fdr"). The analysis was re-run, using parsimony approaches in place of ML approaches for ancestral state reconstruction. Regardless of approach, no genes were found to be significantly associated with isolation source via any of the treeWAS association tests (FDR-corrected $P > 0.1$).

Strain selection for phenotypic stress assays. Thirteen DT104 strains isolated from a previous study of *S.* Typhimurium in New York State [24], which were each available in the Cornell University Food Safety Laboratory (CUFSL) culture collection [96], were additionally characterized separately so that Gifsy-1/*artAB*/*gogB*-positive and -negative DT104 strains from bovine and human sources in the U.S. could be selected to undergo phenotypic characterization (Supplementary Table S9). Parsnp and HarvestTools version 1.2 [69] were used to identify core SNPs among all 13 assembled genomes, using the closed DT104 chromosome as a reference (NCBI Nucleotide Accession NC_022569.1) and Parsnp's implementation of PhiPack [70] to remove recombination. IQ-TREE version 1.5.4 was used to construct a ML phylogeny, using (i) the resulting core SNPs as input, (ii) an ascertainment bias correction, based on the GC content of the DT104 reference chromosome (-fconst 1182070,1287912,1283169,1180480), (iii) the optimal nucleotide substitution model (-m MFP), selected using ModelFinder (i.e., the TIM+I model) , and (iv) 1,000 replicates of the ultrafast bootstrap approximation (-bb 1000). Prokka version 1.13 was used to annotate each genome (using the "Bacteria" database), and the resulting GFF files were supplied to Roary version 3.13.0 [97], which was used to identify orthologous gene clusters among the 13 DT104 genomes (using default thresholds, e.g., 95% BLASTP identity).

The New York State DT104 isolates differed little in terms of their core and pan-genome compositions (Supplementary Figure S12 and Supplementary Table S9). A total of 336 core SNPs were identified among the 13 DT104 genomes; pairwise core SNP distances between all 13 genomes ranged from 12-113 core SNPs (median and mean of 85 and 80.8 core SNPs, respectively, calculated using the "dist.gene" function in the ape R package). Based on gene presence/absence of pan-genome elements, the Jaccard distance between all 13 genomes ranged

from 0.0036-0.0820 (median and mean of 0.0359 and 0.0380, respectively; calculated in R using the "vegdist" function in vegan version 2.5-7) [98].

All available Gifsy-1/*artAB*/*gogB*-negative strains in the CUFSL culture collection were selected to undergo phenotypic testing (n = 3, two human isolates and one bovine isolate; Supplementary Table S9); all three strains were members of the U.S. *artAB*-negative major clade (discussed in detail below). Considering both core and pan-genome distances relative to all three available Gifsy-1/*artAB*/*gogB*-negative strains, three Gifsy-1/*artAB*/*gogB*-positive DT104 strains were additionally selected to undergo phenotypic testing (one from human and two from bovine sources; Supplementary Figure S12 and Supplementary Table S9). The three selected Gifsy-1/*artAB*/*gogB*-positive DT104 strains differed from the three available Gifsy-1/*artAB*/*gogB*-negative strains by (i) 64-83 (HUM_TYPH_NY_04_S5_0370), 74-93 (BOV_TYPH_NY_99_A4_0023), and 65-84 (BOV_TYPH_NY_99_S3_0910) core SNPs and (ii) Jaccard distances (based on pan-genome element presence/absence) of 0.0148-0.0610 (HUM_TYPH_NY_04_S5_0370), 0.0174-0.0622 (BOV_TYPH_NY_99_A4_0023) and 0.0163-0.0620 (BOV_TYPH_NY_99_S3_0910; Supplementary Figure S12).

Phenotypic assays. All strain stocks (Supplementary Table S9) were maintained in CRYOBANK® tubes (Mast Ltd., Reinfeld, Germany) at -80°C. Strains were streaked out from stocks on tryptic soy agar (TSA; Merck KGaA, Darmstadt, Germany) and incubated overnight at 37°C. Single colonies from those plates were inoculated in 5 mL of tryptic soy broth (TSB; Merck KGaA, Darmstadt, Germany) and incubated for 16 - 18 h at 37°C with shaking at 200 rpm. The resulting overnight cultures were diluted 1/100 into 5 mL of fresh, pre-warmed TSB, followed by incubation at 37°C with shaking at 200 rpm to allow cultures to reach mid log phase (defined as OD_{600} of 0.4; 1-2 x10^8 CFU/mL). These cultures were used as input into the three

different phenotypic assays (exposure to ruminal fluid, acid stress, and bile stress; discussed in detail below). Bacterial enumeration before and after stress exposure was performed by direct colony counts of tilt plates according to Kühbacher et al. [99].

To evaluate exposure to ruminal fluid (RF), approximately 2 L of RF was acquired from a Jersey cow with a ruminal fistula on each experimental day prior to the experiments (same collection time was used for each experiment). The RF was immediately filtered through a cellulose filter (Labsolute® Type 80, Th. Geyer GmbH& Co. KG., Renningen, Germany) to remove any large debris, and the pH was measured, ranging from 7.20 to 7.62. Mid-log phase cultures were prepared and inoculated into the RF at two different concentrations. One hundred µl of culture suspensions were inoculated into 5mL of the RF at final concentrations of 10^8 (high) and 10^5 (low) CFU/mL and incubated for 1 h at 37°C without shaking with enumeration by direct colony counting on XLT-4 agar (Oxoid Ltd., Basingstoke, UK) prior and after RF exposure (Supplementary Table S10). The absence of *Salmonella* in the RF at the start of the experiments was confirmed by plating on XLT-4 agar.

Acid stress resistance of the different strains at pH 3.5 with and without prior adaption was tested using an adopted protocol from Horlbog et al. [100]. The pH of the TSB was adjusted with hydrochloric acid solution (1M and 6 M HCL; Merck KGaA, Darmstadt, Germany) immediately prior to the experiment. 1 mL aliquots of mid log phase cultures were transferred to reaction tubes and centrifuged at 14,000 x g for 10 min. For the non-adapted acid stress experiments, the pellets were resuspended in 1mL TSB pH 3.5 and incubated for 1 h at 37°C without shaking. For acid adaption, 1 mL of the same cultures were pelleted, resuspended in 1 mL TSB adjusted to pH 5.5 and incubated for 1h at 37°C (without shaking). Afterwards, the cultures were centrifuged again, resuspended in 1mL TSB pH 3.5, and incubated for 1 h at 37°C

without shaking. Bacteria enumeration was performed before and after the one-hour incubation at pH 3.5 (Supplementary Table S11).

Susceptibility to bile salts (cholic acid and deoxycholic acid in a mixture of 1:1, Bile Salts No.3, Thermo Fisher Scientific Inc., Waltham, USA) was tested in two different concentrations: 14.5 mmol/L corresponding to 0.6% [101] and 26.0 mmol/L corresponding to 1.1% [102] were chosen to represent reasonable physiological states in the duodenum. Bile salts were added, and the pH of the TSB was adjusted to 5.5 (TSB-bile) immediately prior to the experiment. Mid log phase cultures were centrifuged, resuspended in TSB-bile, incubated for 1h 37°C without shaking, and enumerated by direct colony counting prior and after bile exposure (Supplementary Table S12).

For each stress assay, base-ten logarithmic fold change (FC) values were calculated as follows: FC = log CFU/g at the start of the experiments – log CFU/g after the stress assay. Analysis of Variance (ANOVA) for the interpretation of the phenotypic assays were conducted using the “aov” function in R’s “stats” package, with the FC values for the respective assay treated as a response. Figures were designed using the ggplot2 package.

Data availability. Strain metadata, genome quality metrics, and Enterobase accession numbers for all publicly available genomes queried in this study are available in Supplementary Table S1. Strain metadata, genome quality metrics, Food Microbe Tracker IDs, and NCBI BioSample accession numbers for the 13 NYS DT104 strains queried in this study (including those queried via phenotypic assays) are available in Supplementary Table S2. LSD2 results (for the U.S. and combined DT104 data sets) and BEAST2 results (for the U.S. DT104 data set) are available as Supplementary Data.

RESULTS

Human- and bovine-associated DT104 from the U.S. harbor *artAB* on prophage Gifsy-1. Within the set of 230 human- and bovine-associated DT104 genomes derived from strains isolated in the U.S. (62 and 168 genomes from human and bovine sources, respectively; Figure 1A) [24], *artAB* was present in over 75% of genomes (177 of 230, 77.0%; Figure 2, Table 1, and Supplementary Figures S2-S4). *artAB* presence and absence was strongly associated with the presence and absence of anti-inflammatory effector *gogB* (two-sided Fisher's Exact Test raw $P < 2.2\times10^{-16}$, infinite odds ratio [OR]), as co-occurrence was observed in all 177 *artAB*-harboring genomes (100.0%; Figure 2, Table 1, and Supplementary Figures S2-S4). Additionally, *artAB* and *gogB* presence was strongly associated with the presence of prophage Gifsy-1 (NCBI Nucleotide Accession NC_010392.1; two-sided Fisher's Exact Test raw $P < 2.2\times10^{-16}$, infinite OR; Figure 2, Table 1, and Supplementary Figures S2-S4). Subsequent investigation confirmed that, for all 177 *artAB*-harboring U.S. DT104 genomes, *artAB* was located within the Gifsy-1 prophage region (classified as "intact" via PHASTER; Table 2 and Supplementary Table S5).

gogB was largely harbored within regions annotated as Gifsy-1 (126 of 180 *gogB*-harboring genomes; 70.0%), although only 51 of these Gifsy-1 regions were annotated as intact prophage (via PHASTER, 28.3% of *gogB*-harboring genomes; Table 2 and Supplementary Table S5). Occasionally, *gogB* was detected elsewhere in the genome: three genomes harbored *gogB* within regions annotated as prophage Gifsy-2 (3 of 180 *gogB*-harboring genomes, 1.7%; Table 2 and Supplementary Table S5), while *gogB* was detected outside of annotated prophage regions within the remaining 51 *gogB*-harboring genomes (28.3% of *gogB*-harboring genomes; Table 2 and Supplementary Table S5).

Only three genomes (bovine-associated BOV_TYPH_Washington_2007_SRR1519881, BOV_TYPH_Minnesota_2010_SRR1089590, and BOV_TYPH_Minnesota_2008_SRR1177378) possessed an intact Gifsy-1 prophage but did not possess *artAB* (1.7% of genomes in which an intact Gifsy-1 was detected), although all three genomes possessed *gogB* (*gogB* was detected within an incomplete Gifsy-1 prophage region in the two genomes from Minnesota, while the genome from Washington did not harbor *gogB* within an annotated prophage region; Figure 2, Tables 1 and 2, Supplementary Figures S2-S4, and Supplementary Table S5). Of the 168 bovine-associated DT104 genomes from the U.S., 150 (89.3%) possessed *artAB*, *gogB*, and Gifsy-1, while 153 (91.1%) possessed *gogB* and Gifsy-1 (Figure 2, Table 1, and Supplementary Figures S2-S4). Interestingly, of 62 human-associated genomes, only 27 (43.5%) possessed *artAB*, *gogB*, and Gifsy-1 (Figure 2, Table 1, and Supplementary Figures S2-S4), indicating that Gifsy-1/*artAB*/*gogB* may share a negative association with human-associated DT104 from the U.S. (two-sided Fisher's Exact Test raw $P < 4.1\times10^{-12}$, OR = 10.6; Table 1); however, no genes within the U.S. DT104 pan-genome shared a significant association with bovine or human host when accounting for population structure (treeWAS FDR-corrected $P > 0.10$).

Overall, 90 genes within the U.S. DT104 pan-genome were associated with Gifsy-1 presence or absence (two-sided Fisher's exact test FDR-corrected *P*-value < 0.05; Supplementary Figure S11 and Supplementary Table S7). The presence and absence of 30 genes shared a perfect association with Gifsy-1 presence and absence (i.e., these genes were absent from all U.S. DT104 genomes that did not possess Gifsy-1 and were present in all U.S. DT104 genomes that did possess Gifsy-1; FDR-corrected $P < 0.05$ and OR of infinity); in addition to *gogB*, these genes included numerous phage-associated proteins (Supplementary Table S7).

A MDR DT104 lineage circulating among cattle and humans across the U.S. lost prophage Gifsy-1 in close temporal proximity to a period of rapid DT104 population growth in the 1980s. To gain insight into the evolutionary relationships of *artAB*-negative U.S. DT104 strains, a time-scaled phylogeny was constructed using human- and bovine-associated U.S. DT104 genomes (Figure 2 and Supplementary Figures S2-S4). The common ancestor shared by all MDR U.S. bovine- and human-associated DT104 isolates included in this study was predicted to have existed circa 1975 (estimated node age 1974.9, node height 95% highest posterior density [HPD] interval [1958.1, 1986.4]; Figure 2 and Supplementary Figures S2-S4), which is consistent with observations in previous studies [13, 103], in which DT104 was predicted to have acquired its MDR phenotype in the 1970s. The mean evolutionary rate estimated for the U.S. DT104 queried here was 1.75×10^{-7} substitutions/site/year (95% HPD interval [1.38×10^{-7}, 2.11×10^{-7}]), which is similar to evolutionary rates estimated in previous studies of DT104 isolates from other world regions [13, 36] (Supplementary Figure S1, Supplementary Table S6, and Supplementary Data).

Notably, over 75% of all U.S. DT104 *artAB*-negative genomes (42 of 53 *artAB*-negative genomes, 79.2%) were members of a single, well-supported clade (posterior probability = 1.0, referred to hereafter as the "U.S. *artAB*-negative major clade"; Figure 2 and Supplementary Figures S2-S4). In addition to lacking *artAB*, all members of the U.S. *artAB*-negative major clade lacked Gifsy-1 and 50 additional genes, which were present in over half of U.S. DT104 genomes not included in the U.S. *artAB*-negative major clade, including *gogB*, a chitinase, and many phage-associated proteins (Figure 2, Supplementary Figures S2-S4, and Supplementary Table S8). Strains within the U.S. *artAB*-negative major clade were reported to have been isolated between 1997 and 2018 (the most recent year included in this study) from at least 11

different states across the U.S. (for two isolates, the U.S. state in which the strain was isolated was unknown; Figures 1B and 2 and Supplementary Figures S2-S4). Most strains within the U.S. *artAB*-negative major clade were isolated from human clinical cases (*n* = 30 of 42 U.S. *artAB*-negative major clade strains, 71.4%), and nearly half of all U.S. DT104 strains isolated from human sources were members of this clade (*n* = 30 of 62 U.S. DT104 strains from human sources, 48.4%); bovine strains within the U.S. *artAB*-negative major clade were isolated from bovine clinical cases or beef products (*n* = 12 of 42 strains, 28.6%; Figure 2, Supplementary Figures S2-S4, and Supplementary Table S1).

Based on results of ancestral state reconstruction using *artAB* presence/absence, the loss of Gifsy-1, *artAB*, *gogB*, and other Gifsy-1-associated genes among members of the U.S. *artAB*-negative major clade was estimated to have occurred between 1985 and 1987 (estimated node ages 1985.0 and 1987.2, node height 95% HPD intervals [1979.0, 1990.2] and [1981.7, 1992.1], respectively; Figure 2 and Supplementary Figures S6-S8). Interestingly, this predicted loss event occurred in close temporal proximity to a rapid increase in the effective population size of U.S. DT104, which occurred in the mid-to-late 1980s (Figure 3 and Supplementary Figure S5). Following this predicted rapid increase in the 1980s, the U.S. DT104 effective population size was predicted to have increased again in the mid-to-late 1990s, peaking circa 2000 (Figure 3 and Supplementary Figure S5).

Loss of *artAB* and *gogB* within the global DT104 population occurs sporadically. The absence of Gifsy-1, *artAB*, and/or *gogB* among DT104 strains was not strictly a U.S. phenomenon: Gifsy-1, *artAB*, and *gogB* were not detected in 19 and three genomes out of (i) 290 DT104 strains collected from numerous sources around the world [13] and (ii) 243 DT104 strains isolated from cattle and humans in Scotland [36], respectively (representing 6.6% and

1.2% of strains in their respective study that were included in our analysis here; Figure 4 and Supplementary Figures S9 and S10). However, the Gifsy-1/*artAB*/*gogB* loss event associated with the U.S. *artAB*-negative major clade represented the single largest Gifsy-1/*artAB*/*gogB* loss event (n = 42; Figure 4 and Supplementary Figures S9 and S10).

Among all 752 DT104 genomes queried here, the presence and absence of *artAB* and *gogB* was correlated with that of Gifsy-1 (two-sided Fisher's Exact Test raw $P < 2.2\times10^{-16}$ for each, ORs of 2069.8 and infinity, respectively), as well as each other (two-sided Fisher's Exact Test raw $P < 2.2\times10^{-16}$, infinite OR; Figure 4, Table 1, and Supplementary Figure S9). However, unlike the 177 *artAB*-harboring U.S. genomes queried here, *artAB* were not always detected within prophage regions annotated as Gifsy-1 in the other genomes (Table 2 and Supplementary Table S5).

***In vitro* response to human- and bovine-associated gastrointestinal stress factors is not correlated with the presence of *artAB, gogB,* and Gifsy-1 in U.S. DT104.** The (i) loss of Gifsy-1/*artAB*/*gogB* associated with the U.S. *artAB*-negative major clade in close temporal proximity to a predicted rapid increase in the DT104 effective population size, plus (ii) the over-representation of human strains in this clade led us to hypothesize that ArtAB and/or GogB production (or some other genomic element harbored on Gifsy-1) may influence the dynamics of DT104 in the digestive tracks of human and animal hosts. Thus, we used phenotypic assays that simulate human and/or bovine digestion-associated stress conditions to compare the phenotypes of Gifsy-1/*artAB*/*gogB*-negative members of the U.S. *artAB*-negative major clade to those of the most closely related Gifsy-1/*artAB*/*gogB*-positive U.S. DT104 strains available (Supplementary Table S9).

As the first three compartments of the bovine digestive tract differ massively from that in the human gut, the phenotype of Gifsy-1/*artAB*/*gogB*-positive and -negative strains was investigated in fresh bovine ruminal fluid (RF) obtained from a donor cow (Supplementary Table S10). DT104 concentrations were reduced by 3.4 log CFU (SD=0.2) when inoculated into RF at a final concentration of 10^5 CFU/mL, whereas DT104 numbers were reduced by 1.3 log CFU (SD = 0.2) when inoculated at a final concentration of 10^8 CFU/mL. While the inoculation density did significantly affect survival (ANOVA raw $P < 0.001$), the phenotype in RF was not associated with the presence or absence of Gifsy-1/*artAB*/*gogB* (ANOVA raw $P > 0.05$; Figure 5).

The presence of Gifsy-1/*artAB*/*gogB* also did not significantly influence acid stress survival at pH 3.5 (ANOVA raw $P > 0.05$; Figure 5 and Supplementary Table S11). While prior adaptation at an intermediate pH 5.5 significantly increased survival at pH 3.5 as expected (ANOVA raw $P < 0.01$), there was no significant difference in acid adaptation between Gifsy-1/*artAB*/*gogB*-positive and -negative strains (ANOVA raw $P > 0.05$; Figure 5). Both groups showed a concentration-dependent reduction in growth/survival at the two tested bile concentrations of 0.6% and 1.1% (ANOVA raw $P = 0.01$ for the difference in fold change at the two concentrations), but there was no Gifsy-1/*artAB*/*gogB*-dependent phenotype in the response of DT104 strains to bile stress (ANOVA raw $P > 0.05$; Figure 5 and Supplementary Table S12).

DISCUSSION

A DT104 lineage distributed across multiple U.S. states lost its ability to produce toxin ArtAB in close temporal proximity to the global DT104 epidemic. Bacterial ADP-ribosylating toxins play important roles in the virulence of numerous pathogens [19, 104]. While

the illness caused by *Salmonella enterica* is not considered to be a toxin-mediated disease in the classical sense (e.g., as is the case for *Clostridium botulinum* or *Vibrio cholerae*) [19], some *Salmonella* lineages are capable of producing ADP-ribosylating toxins, allowing them to alter host immune responses and promote pathogenesis [17, 19, 105-107]. ArtAB is one such toxin with a variable presence among *Salmonella* lineages; genes encoding ArtAB have been detected in at least 45 different serotypes and are correlated with the presence of typhoid toxin genes, although in DT104 this is not the case [108]. Additionally, in the majority of these serotypes, *artA* is predicted to be a pseudogene and the selective advantage of maintaining *artB* appears to be related to its use as an alternative binding subunit for the typhoid toxin [19, 109].

A previous study of ArtAB-producing DT104 strains [18] found that ArtAB production among DT104 appears to be the norm rather than the exception, as 237 of 243 strains (97.5%) in the study were *artAB*-positive [18]. We observed similar findings here, as *artAB* was detected in 678 of 752 DT104 genomes (90.2%). However, we additionally showed that *artAB* loss events appear sporadically throughout the DT104 phylogeny (Figures 2 and 4 and Supplementary Figures S9 and S10). Among U.S. DT104, these loss events usually coincided with Gifsy-1 excision, although not exclusively (i.e., three strains did not possess *artAB*, but possessed Gifsy-1; Figure 2 and Supplementary Figures S2-S4).

Most notably, we observed a MDR DT104 clade circulating among cattle and humans across 11 U.S. states, which lost Gifsy-1, along with the ability to produce ArtAB and GogB, (i.e., the U.S. *artAB*-negative major clade; Figure 2). The U.S. *artAB*-negative major clade was predicted to have lost Gifsy-1/*artAB*/*gogB* circa 1985-1987, which is in close temporal proximity to the predicted rapid increase in the U.S. DT104 effective population size, which occurred in the mid-to-late 1980s (Figure 3). Our results are consistent with a previous study of DT104 from

multiple world regions, which also identified periods of dramatic population growth in the 1980s and 1990s [13]. This rapid increase in population size is notable, as it coincides with the global MDR DT104 epidemic, which occurred among humans and animals throughout the 1990s [13, 14, 36]. However, it is essential to note that any potential association between the virulence and/or fitness of MDR DT104 and Gifsy-1/*artAB*/*gogB* loss among DT104 is merely speculative at this point (discussed in detail below); while previous studies of DT104 have shown that prophage excision and *artAB* loss occur in response to DNA damage and other stressors [17, 19], future studies are needed to better understand the roles that Gifsy-1, *artAB*, and *gogB* play in DT104 evolution.

Members of the U.S. *artAB*-negative major clade do not have a phenotypic advantage relative to other U.S. DT104 when exposed to ruminal fluid-, acid-, and bile-associated stressors *in vitro*. *Salmonella enterica* encounters numerous stressors within the gastrointestinal tracts of humans and animals, including (but not limited to) low pH, low oxygen, exposure to bile, and the host immune system [110-112]. Furthermore, the gastrointestinal environment that *Salmonella enterica* encounters can differ between hosts; for example, the first three compartments of the bovine digestive tract differ massively from those of the human gut, as they essentially serve as massive microbial fermentation chambers [113]. Here, we evaluated DT104 survival when exposed to three stressors encountered in the human and/or bovine gastrointestinal tracts: (i) ruminal fluid (RF; bovine rumen), (ii) low pH (bovine abomasum and human stomach), and (iii) exposure to bile (bovine and human duodenum); we discuss each step in detail below.

In the bovine digestion process, the RF, including the complex community of ruminal microbiota [114], presents an early line of defense against potential pathogens, like *Salmonella* spp. In RF, the kill rate of DT104 strains was dependent on the inoculation density. The high

inoculation rate (10^8 CFU/mL) was chosen to test the ability of the ruminal microbiota to efficiently kill or impede *Salmonella*. The lower inoculation rate of 10^5 CFU/mL was chosen for its dynamic range to measure either growth or decrease of *Salmonella* concentration. An interaction of the complex ruminal microbiota with the inoculated *Salmonella* is conceivable in two ways: either the microbiota exhibit strategies to produce antimicrobial compounds against *Salmonella* species [115, 116], or through competition for nutrients, e.g., iron [117]. The fact that the ruminal microbiota was less effective at killing DT104 at the high inoculation rate suggests that their defense mechanisms against DT104 are limited and/or the system started to be overrun by the high numbers of DT104.

Gastric acids in the stomach (or abomasum) are the next line of host defense, which *Salmonella* must overcome during gastrointestinal passage [118]. Like most enteric pathogens, *Salmonella enterica* is well adapted to the acid conditions of the stomach [119]. Our experiments confirmed that acid adaptation with HCl at pH 5.5 lead to much higher survival rates at pH 3.5. Well-known mechanisms such as decreased membrane conductivity for H^+, increased proton extrusion or changes in the cell envelope composition [120-122] could be responsible for this.

Upon leaving the stomach, enteric pathogens are confronted with bile. Bile salts show antimicrobial activity by dissolving membrane lipids, dissociating integral membrane proteins [123], and lead to general cell damage by misfolding and denaturation of proteins [124, 125] and DNA damage [126, 127]. *Salmonella enterica* is able to survive duodenal bile salt concentrations through DNA repair mechanisms [127], multiple changes in gene expression [128], and increased production of anti-oxidative enzymes [129]. Here, selected DT104 strains were able to survive at both tested bile salt concentrations (14.5 mmol/L and 26.0 mmol/L); however, no

significant differences were observed between strains that harbored Gifsy-1/*artAB*/*gogB* and those that did not (Figure 5).

In summary, all tested strains were able to adapt to and survive in the *in vitro* gastrointestinal conditions tested here, while the presence or absence of Gifsy-1/*artAB*/*gogB* showed no fitness effects (Figure 5). However, it is important to note that we have only tested the most obvious stress conditions associated with the gastrointestinal tract; thus, it is conceivable that Gifsy-1/*artAB*/*gogB* loss among members of the U.S. *artAB*-negative major clade confer an advantage in conditions associated with the infection chain that were not tested here, including interactions with different host cell types.

Future research is needed to understand the roles that Gifsy-1, ArtAB, and GogB play in DT104 virulence. The results presented here indicate that prophage-mediated ArtAB production among DT104 can undergo temporal changes. Most notably, we identified the U.S. *artAB*-negative major clade, the common ancestor of which lost the ability to produce ArtAB in close temporal proximity to the global MDR DT104 epidemic. However, the ecological and/or evolutionary significance of this loss-of-function event remains unclear. Although phenotypic assessments have demonstrated a role for DT104-encoded ArtAB in both cell culture and a mouse model [18], the true benefit of this toxin in the context of human and bovine salmonellosis has not been investigated. It has been previously shown that reactive oxygen species (ROS) induce production of ArtAB [23], which may suggest that *artAB* is expressed in response to immune cell derived ROS. Furthermore, as treatment with ArtA increases intracellular levels of cAMP in macrophage-like 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.

The *in vitro* stress assays performed in this study aimed to mimic the stressors that DT104 encounters in the gastrointestinal tracts of humans and ruminants. Given the over-representation of human-associated Gifsy-1/*artAB*/*gogB*-negative strains observed here, one may be tempted to speculate that Gifsy-1, *artAB*, and/or *gogB* absence may confer members of the U.S. *artAB*-negative major clade with a competitive advantage in the human host gastrointestinal tract; however, no Gifsy-1/*artAB*/*gogB*-dependent phenotype of DT104 was observed under the tested conditions (Figure 5). Despite this, it may be possible that Gifsy-1/*artAB*/*gogB* absence may confer some advantage(s) to U.S. *artAB*-negative major clade strains in environmental conditions, which were not tested in this study, including those outside of the host (e.g., high osmotic pressure and competitive microbiota in manure or wastewater, food safety measures like disinfectants, antimicrobials and food processing) [111]. However, at the present, this is merely speculation; future studies are needed to evaluate whether Gifsy-1/*artAB*/*gogB* loss among members of the U.S. *artAB*-negative major clade is merely coincidental or indicative of some evolutionarily advantageous phenotype.

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).

Acknowledgments. Figure 1 was created with BioRender.com.

REFERENCES

1. **Bobay LM, Touchon M, Rocha EP**. Pervasive domestication of defective prophages by bacteria. *Proc Natl Acad Sci U S A* 2014;111(33):12127-12132.
2. **Fortier LC, Sekulovic O**. Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence* 2013;4(5):354-365.
3. **Pleska M, Lang M, Refardt D, Levin BR, Guet CC**. Phage-host population dynamics promotes prophage acquisition in bacteria with innate immunity. *Nat Ecol Evol* 2018;2(2):359-366.
4. **Ramisetty BCM, Sudhakari PA**. Bacterial 'Grounded' Prophages: Hotspots for Genetic Renovation and Innovation. *Front Genet* 2019;10:65.
5. **Owen SV, Canals R, Wenner N, Hammarlöf DL, Kröger C et al.** A window into lysogeny: Revealing temperate phage biology with transcriptomics. *bioRxiv* 2019:787010.
6. **Haaber J, Leisner JJ, Cohn MT, Catalan-Moreno A, Nielsen JB et al.** Bacterial viruses enable their host to acquire antibiotic resistance genes from neighbouring cells. *Nat Commun* 2016;7:13333.
7. **Wang X, Kim Y, Ma Q, Hong SH, Pokusaeva K et al.** Cryptic prophages help bacteria cope with adverse environments. *Nat Commun* 2010;1:147.
8. **Liu B, Knirel YA, Feng L, Perepelov AV, Senchenkova SyN et al.** Structural diversity in *Salmonella* O antigens and its genetic basis. *FEMS Microbiology Reviews* 2014;38(1):56-89.
9. **Czajkowski R**. May the Phage be With You? Prophage-Like Elements in the Genomes of Soft Rot Pectobacteriaceae: *Pectobacterium* spp. and *Dickeya* spp. *Frontiers in Microbiology*, Original Research 2019;10(138).
10. **Hiley L, Fang NX, Micalizzi GR, Bates J**. Distribution of Gifsy-3 and of variants of ST64B and Gifsy-1 prophages amongst *Salmonella enterica* Serovar Typhimurium isolates: evidence that combinations of prophages promote clonality. *PLoS One* 2014;9(1):e86203.
11. **World Health Organization**. 2018. *Salmonella* (non-typhoidal). https://www.who.int/news-room/fact-sheets/detail/salmonella-(non-typhoidal) [accessed 5 May 2021 2021].
12. **Wang X, Biswas S, Paudyal N, Pan H, Li X et al.** Antibiotic Resistance in *Salmonella* Typhimurium Isolates Recovered From the Food Chain Through National Antimicrobial Resistance Monitoring System Between 1996 and 2016. *Front Microbiol* 2019;10:985.
13. **Leekitcharoenphon P, Hendriksen RS, Le Hello S, Weill FX, Baggesen DL et al.** Global Genomic Epidemiology of *Salmonella enterica* Serovar Typhimurium DT104. *Appl Environ Microbiol* 2016;82(8):2516-2526.
14. **Threlfall EJ**. Epidemic *salmonella* typhimurium DT 104--a truly international multiresistant clone. *J Antimicrob Chemother* 2000;46(1):7-10.
15. **Helms M, Ethelberg S, Molbak K, Group DTS**. International *Salmonella* Typhimurium DT104 infections, 1992-2001. *Emerg Infect Dis* 2005;11(6):859-867.

727 16. **Allen CA, Fedorka-Cray PJ, Vazquez-Torres A, Suyemoto M, Altier C et al.** *In Vitro* and *In*
728 *Vivo* Assessment of *Salmonella enterica* Serovar Typhimurium DT104 Virulence. *Infection and*
729 *Immunity* 2001;69(7):4673-4677.
730 17. **Saitoh M, Tanaka K, Nishimori K, Makino SI, Kanno T et al.** The *artAB* genes encode a
731 putative ADP-ribosyltransferase toxin homologue associated with *Salmonella enterica* serovar
732 Typhimurium DT104. *Microbiology* 2005;151(Pt 9):3089-3096.
733 18. **Tamamura Y, Tanaka K, Uchida I**. Characterization of pertussis-like toxin from
734 *Salmonella* spp. that catalyzes ADP-ribosylation of G proteins. *Sci Rep* 2017;7(1):2653.
735 19. **Cheng RA, Wiedmann M**. The ADP-Ribosylating Toxins of *Salmonella*. *Toxins (Basel)*
736 2019;11(7).
737 20. **Weiss AA, Hewlett EL**. Virulence factors of *Bordetella pertussis*. *Annu Rev Microbiol*
738 1986;40:661-686.
739 21. **Hewlett EL, Sauer KT, Myers GA, Cowell JL, Guerrant RL**. Induction of a novel
740 morphological response in Chinese hamster ovary cells by pertussis toxin. *Infection and*
741 *immunity* 1983;40(3):1198-1203.
742 22. **Carbonetti NH**. Pertussis toxin and adenylate cyclase toxin: key virulence factors of
743 *Bordetella pertussis* and cell biology tools. *Future microbiology* 2010;5(3):455-469.
744 23. **Uchida I, Ishihara R, Tanaka K, Hata E, Makino S et al.** *Salmonella enterica* serotype
745 Typhimurium DT104 ArtA-dependent modification of pertussis toxin-sensitive G proteins in the
746 presence of [32P]NAD. *Microbiology* 2009;155(Pt 11):3710-3718.
747 24. **Carroll LM, Huisman JS, Wiedmann M**. Twentieth-century emergence of antimicrobial
748 resistant human- and bovine-associated *Salmonella enterica* serotype Typhimurium lineages in
749 New York State. *Sci Rep* 2020;10(1):14428.
750 25. **Zhou Z, Alikhan N-F, Mohamed K, Fan Y, Achtman M**. The user's guide to comparative
751 genomics with EnteroBase, including case studies on transmissions of micro-clades of
752 *Salmonella*, the phylogeny of ancient and modern *Yersinia pestis* genomes, and the core
753 genomic diversity of all *Escherichia*. *bioRxiv* 2019:613554.
754 26. **Alikhan NF, Zhou Z, Sergeant MJ, Achtman M**. A genomic overview of the population
755 structure of *Salmonella*. *PLoS Genet* 2018;14(4):e1007261.
756 27. **Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database C**.
757 The sequence read archive. *Nucleic Acids Res* 2011;39(Database issue):D19-21.
758 28. **Kodama Y, Shumway M, Leinonen R, International Nucleotide Sequence Database C**.
759 The Sequence Read Archive: explosive growth of sequencing data. *Nucleic Acids Res*
760 2012;40(Database issue):D54-56.
761 29. **Yoshida CE, Kruczkiewicz P, Laing CR, Lingohr EJ, Gannon VP et al.** The *Salmonella In*
762 *Silico* Typing Resource (SISTR): An Open Web-Accessible Tool for Rapidly Typing and Subtyping
763 Draft *Salmonella* Genome Assemblies. *PLoS One* 2016;11(1):e0147101.
764 30. **Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J**. RhierBAPS: An R
765 implementation of the population clustering algorithm hierBAPS. *Wellcome Open Res*
766 2018;3:93.
767 31. **Bolger AM, Lohse M, Usadel B**. Trimmomatic: a flexible trimmer for Illumina sequence
768 data. *Bioinformatics* 2014;30(15):2114-2120.

32. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M et al.** SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19(5):455-477.
33. **Andrews S**. FastQC: a quality control tool for high throughput sequence data version 0.11.5. 2019. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
34. **Gurevich A, Saveliev V, Vyahhi N, Tesler G**. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;29(8):1072-1075.
35. **Ewels P, Magnusson M, Lundin S, Kaller M**. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 2016;32(19):3047-3048.
36. **Mather AE, Reid SW, Maskell DJ, Parkhill J, Fookes MC et al.** Distinguishable epidemics of multidrug-resistant *Salmonella* Typhimurium DT104 in different hosts. *Science* 2013;341(6153):1514-1517.
37. **Arndt D, Marcu A, Liang Y, Wishart DS**. PHAST, PHASTER and PHASTEST: Tools for finding prophage in bacterial genomes. *Brief Bioinform* 2019;20(4):1560-1567.
38. **Arndt D, Grant JR, Marcu A, Sajed T, Pon A et al.** PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 2016;44(W1):W16-21.
39. **Seemann T**. ABRicate: Mass screening of contigs for antimicrobial resistance or virulence genes. 2018. https://github.com/tseemann/abricate.
40. **Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ et al.** Validating the AMRFinder Tool and Resistance Gene Database by Using Antimicrobial Resistance Genotype-Phenotype Correlations in a Collection of Isolates. *Antimicrob Agents Chemother* 2019;63(11).
41. **Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O et al.** *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 2014;58(7):3895-3903.
42. **Chen L, Yang J, Yu J, Yao Z, Sun L et al.** VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res* 2005;33(Database issue):D325-328.
43. **Owen SV, Wenner N, Canals R, Makumi A, Hammarlof DL et al.** Characterization of the Prophage Repertoire of African *Salmonella* Typhimurium ST313 Reveals High Levels of Spontaneous Induction of Novel Phage BTP1. *Front Microbiol* 2017;8:235.
44. **Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J et al.** BLAST+: architecture and applications. *BMC Bioinformatics* 2009;10:421.
45. **Seemann T**. Snippy: Rapid haploid variant calling and core genome alignment version 4.3.6. 2019. https://github.com/tseemann/snippy.
46. **Li H**. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 2013:1303.3997.
47. **Li H, Durbin R**. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25(14):1754-1760.
48. **Li H**. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;34(18):3094-3100.
49. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J et al.** The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25(16):2078-2079.
50. **Quinlan AR, Hall IM**. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;26(6):841-842.

812 51. **Quinlan AR**. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc*
813 *Bioinformatics* 2014;47:11 12 11-34.
814 52. **Li H**. A statistical framework for SNP calling, mutation discovery, association mapping
815 and population genetical parameter estimation from sequencing data. *Bioinformatics*
816 2011;27(21):2987-2993.
817 53. **Garrison E, Marth G**. Haplotype-based variant detection from short-read sequencing.
818 *arXiv* 2012:1207.3907.
819 54. **Cleary JG, Braithwaite R, Gaastra K, Hilbush BS, Inglis S et al.** Comparing Variant Call
820 Files for Performance Benchmarking of Next-Generation Sequencing Variant Calling Pipelines.
821 *bioRxiv* 2015:023754.
822 55. **Tan A, Abecasis GR, Kang HM**. Unified representation of genetic variants. *Bioinformatics*
823 2015;31(13):2202-2204.
824 56. **Cingolani P, Platts A, Wang le L, Coon M, Nguyen T et al.** A program for annotating and
825 predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of
826 *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 2012;6(2):80-92.
827 57. **Seemann T**. samclip: Filter SAM file for soft and hard clipped alignments version 0.2.
828 2019. https://github.com/tseemann/samclip.
829 58. **Li H**. Seqtk: a fast and lightweight tool for processing sequences in the FASTA or FASTQ
830 format version 1.2-r102-dirty 2019. https://github.com/lh3/seqtk.
831 59. **Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T et al.** SNP-sites: rapid efficient
832 extraction of SNPs from multi-FASTA alignments. *Microb Genom* 2016;2(4):e000056.
833 60. **Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA et al.** Rapid phylogenetic
834 analysis of large samples of recombinant bacterial whole genome sequences using Gubbins.
835 *Nucleic Acids Res* 2015;43(3):e15.
836 61. **Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ**. IQ-TREE: a fast and effective
837 stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*
838 2015;32(1):268-274.
839 62. **Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS**. ModelFinder: fast
840 model selection for accurate phylogenetic estimates. *Nat Methods* 2017;14(6):587-589.
841 63. **Kimura M**. Estimation of evolutionary distances between homologous nucleotide
842 sequences. *Proc Natl Acad Sci U S A* 1981;78(1):454-458.
843 64. **Minh BQ, Nguyen MA, von Haeseler A**. Ultrafast approximation for phylogenetic
844 bootstrap. *Mol Biol Evol* 2013;30(5):1188-1195.
845 65. **Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS**. UFBoot2: Improving the
846 Ultrafast Bootstrap Approximation. *Mol Biol Evol* 2018;35(2):518-522.
847 66. **Rambaut A, Lam TT, Max Carvalho L, Pybus OG**. Exploring the temporal structure of
848 heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* 2016;2(1):vew007.
849 67. **To T-H, Jung M, Lycett S, Gascuel O**. Fast Dating Using Least-Squares Criteria and
850 Algorithms. *Systematic Biology* 2015;65(1):82-97.
851 68. **Rambaut A**. FigTree: a graphical viewer of phylogenetic trees version 1.4.4. 2016.
852 http://tree.bio.ed.ac.uk/software/figtree/.
853 69. **Treangen TJ, Ondov BD, Koren S, Phillippy AM**. The Harvest suite for rapid core-genome
854 alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol*
855 2014;15(11):524.

70. **Bruen TC, Philippe H, Bryant D**. A simple and robust statistical test for detecting the presence of recombination. *Genetics* 2006;172(4):2665-2681.
71. **Letunic I, Bork P**. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 2021;49(W1):W293-W296.
72. **Bouckaert R, Heled J, Kuhnert D, Vaughan T, Wu CH et al.** BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 2014;10(4):e1003537.
73. **Bouckaert R, Vaughan TG, Barido-Sottani J, Duchene S, Fourment M et al.** BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 2019;15(4):e1006650.
74. **Bouckaert R**. 2014. Correcting for constant sites in BEAST2. https://groups.google.com/forum/#!topic/beast-users/QfBHMOqImFE [accessed July 26, 2022].
75. **Bouckaert RR, Drummond AJ**. bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evol Biol* 2017;17(1):42.
76. **Drummond AJ, Ho SY, Phillips MJ, Rambaut A**. Relaxed phylogenetics and dating with confidence. *PLoS Biol* 2006;4(5):e88.
77. **Drummond AJ, Rambaut A, Shapiro B, Pybus OG**. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 2005;22(5):1185-1192.
78. **R Core Team**. R: A Language and Environment for Statistical Computing version 4.1.2. R Foundation for Statistical Computing, Vienna, Austria; 2021. https://www.R-project.org/.
79. **Wickham H**. *ggplot2: Elegant Graphics for Data Analysis*: Springer-Verlag New York; 2016.
80. **Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y**. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* 2017;8(1):28-36.
81. **Yu G, Lam TT, Zhu H, Guan Y**. Two Methods for Mapping and Visualizing Associated Data on Phylogeny Using Ggtree. *Mol Biol Evol* 2018;35(12):3041-3043.
82. **R Hackathon**. phylobase: Base Package for Phylogenetic Structures and Comparative Data version 0.8.10. 2021. https://CRAN.R-project.org/package=phylobase.
83. **Yu G**. treeio: Base Classes and Functions for Phylogenetic Tree Input and Output version 1.18.1. 2021. https://guangchuangyu.github.io/software/treeio.
84. **Revell LJ**. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 2012;3(2):217-223.
85. **Seemann T**. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30(14):2068-2069.
86. **Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G et al.** Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol* 2020;21(1):180.
87. **Katoh K, Standley DM**. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30(4):772-780.
88. **Katoh K, Misawa K, Kuma K, Miyata T**. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002;30(14):3059-3066.
89. **Collins RE, Higgs PG**. Testing the infinitely many genes model for the evolution of the bacterial core genome and pangenome. *Mol Biol Evol* 2012;29(11):3413-3425.
90. **Baumdicker F, Hess WR, Pfaffelhuber P**. The infinitely many genes model for the distributed genome of bacteria. *Genome Biol Evol* 2012;4(4):443-456.

900 91. **Zamani-Dahaj SA, Okasha M, Kosakowski J, Higgs PG**. Estimating the Frequency of
901 Horizontal Gene Transfer Using Phylogenetic Models of Gene Gain and Loss. *Mol Biol Evol*
902 2016;33(7):1843-1857.
903 92. **Huerta-Cepas J, Szklarczyk D, Heller D, Hernandez-Plaza A, Forslund SK et al.** eggNOG
904 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on
905 5090 organisms and 2502 viruses. *Nucleic Acids Res* 2019;47(D1):D309-D314.
906 93. **Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J**. eggNOG-
907 mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the
908 Metagenomic Scale. *Mol Biol Evol* 2021;38(12):5825-5829.
909 94. **Benjamini Y, Hochberg Y**. Controlling the False Discovery Rate: A Practical and Powerful
910 Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)*
911 1995;57(1):289-300.
912 95. **Collins C, Didelot X**. A phylogenetic method to perform genome-wide association
913 studies in microbes that accounts for population structure and recombination. *PLoS Comput*
914 *Biol* 2018;14(2):e1005958.
915 96. **Vangay P, Fugett EB, Sun Q, Wiedmann M**. Food microbe tracker: a web-based tool for
916 storage and comparison of food-associated microbes. *J Food Prot* 2013;76(2):283-294.
917 97. **Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S et al.** Roary: rapid large-scale
918 prokaryote pan genome analysis. *Bioinformatics* 2015;31(22):3691-3693.
919 98. **Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P et al.** vegan: Community
920 Ecology Package. R package version 2.5-7. https://CRAN.R-project.org/package=vegan. 2021.
921 99. **Kuhbacher A, Cossart P, Pizarro-Cerda J**. Internalization assays for *Listeria*
922 *monocytogenes*. *Methods Mol Biol* 2014;1157:167-178.
923 100. **Horlbog JA, Kent D, Stephan R, Guldimann C**. Surviving host - and food relevant
924 stresses: phenotype of *L. monocytogenes* strains isolated from food and clinical sources. *Sci Rep*
925 2018;8(1):12931.
926 101. **Fausa O**. Duodenal bile acids after a test meal. *Scand J Gastroenterol* 1974;9(6):567-570.
927 102. **Guariglia-Oropeza V, Orsi RH, Guldimann C, Wiedmann M, Boor KJ**. The *Listeria*
928 *monocytogenes* Bile Stimulon under Acidic Conditions Is Characterized by Strain-Specific
929 Patterns and the Upregulation of Motility, Cell Wall Modification Functions, and the PrfA
930 Regulon. *Front Microbiol* 2018;9:120.
931 103. **Carroll LM, Pierneef R, Mathole M, Matle I**. Genomic Characterization of Endemic and
932 Ecdemic Non-typhoidal *Salmonella enterica* Lineages Circulating Among Animals and Animal
933 Products in South Africa. *Front Microbiol* 2021;12:748611.
934 104. **Simon NC, Aktories K, Barbieri JT**. Novel bacterial ADP-ribosylating toxins: structure and
935 function. *Nat Rev Microbiol* 2014;12(9):599-611.
936 105. **Guiney DG, Fierer J**. The Role of the *spv* Genes in *Salmonella* Pathogenesis. *Front*
937 *Microbiol* 2011;2:129.
938 106. **Spano S, Ugalde JE, Galan JE**. Delivery of a *Salmonella* Typhi exotoxin from a host
939 intracellular compartment. *Cell Host Microbe* 2008;3(1):30-38.
940 107. **Pollard DJ, Young JC, Covarelli V, Herrera-Leon S, Connor TR et al.** The Type III
941 Secretion System Effector SeoC of *Salmonella enterica* subsp. *salamae* and *S. enterica* subsp.
942 *arizonae* ADP-Ribosylates Src and Inhibits Opsonophagocytosis. *Infect Immun*
943 2016;84(12):3618-3628.

944 108. **Gaballa A, Cheng RA, Harrand AS, Cohn AR, Wiedmann M**. The Majority of Typhoid
945 Toxin-Positive *Salmonella* Serovars Encode ArtB, an Alternate Binding Subunit. *mSphere*
946 2021;6(1).
947 109. **Fowler CC, Stack G, Jiao X, Lara-Tejero M, Galan JE**. Alternate subunit assembly
948 diversifies the function of a bacterial toxin. *Nat Commun* 2019;10(1):3684.
949 110. **Flint A, Butcher J, Stintzi A**. Stress Responses, Adaptation, and Virulence of Bacterial
950 Pathogens During Host Gastrointestinal Colonization. *Microbiol Spectr* 2016;4(2).
951 111. **Burgess CM, Gianotti A, Gruzdev N, Holah J, Knochel S et al.** The response of foodborne
952 pathogens to osmotic and desiccation stresses in the food chain. *Int J Food Microbiol*
953 2016;221:37-53.
954 112. **Horn N, Bhunia AK**. Food-Associated Stress Primes Foodborne Pathogens for the
955 Gastrointestinal Phase of Infection. *Front Microbiol* 2018;9:1962.
956 113. **Hofmann RR**. Evolutionary steps of ecophysiological adaptation and diversification of
957 ruminants: a comparative view of their digestive system. *Oecologia* 1989;78(4):443-457.
958 114. **McCann JC, Wickersham TA, Loor JJ**. High-throughput Methods Redefine the Rumen
959 Microbiome and Its Relationship with Nutrition and Metabolism. *Bioinform Biol Insights*
960 2014;8:109-125.
961 115. **Kalmokoff ML, Cyr TD, Hefford MA, Whitford MF, Teather RM**. Butyrivibriocin AR10, a
962 new cyclic bacteriocin produced by the ruminal anaerobe *Butyrivibrio fibrisolvens* AR10:
963 characterization of the gene and peptide. *Can J Microbiol* 2003;49(12):763-773.
964 116. **Oyama LB, Girdwood SE, Cookson AR, Fernandez-Fuentes N, Prive F et al.** The rumen
965 microbiome: an underexplored resource for novel antimicrobial discovery. *NPJ Biofilms*
966 *Microbiomes* 2017;3:33.
967 117. **Deriu E, Liu JZ, Pezeshki M, Edwards RA, Ochoa RJ et al.** Probiotic bacteria reduce
968 *salmonella* typhimurium intestinal colonization by competing for iron. *Cell Host Microbe*
969 2013;14(1):26-37.
970 118. **Singh A, Barnard TG**. Adaptations in the physiological heterogeneity and viability of
971 *Shigella dysenteriae*, *Shigella flexneri* and *Salmonella* typhimurium, after exposure to simulated
972 gastric acid fluid. *Microb Pathog* 2017;113:378-384.
973 119. **Smith JL**. The role of gastric acid in preventing foodborne disease and how bacteria
974 overcome acid conditions. *J Food Prot* 2003;66(7):1292-1303.
975 120. **Foster JW, Hall HK**. Adaptive acidification tolerance response of *Salmonella*
976 typhimurium. *J Bacteriol* 1990;172(2):771-778.
977 121. **Koutsoumanis KP, Sofos JN**. Comparative acid stress response of *Listeria*
978 *monocytogenes*, *Escherichia coli* O157:H7 and *Salmonella* Typhimurium after habituation at
979 different pH conditions. *Lett Appl Microbiol* 2004;38(4):321-326.
980 122. **Torres MA, Terraf MCL, Minahk CJ, Delgado MA**. Stability of the *Salmonella*
981 Typhimurium rcsC11 mutant under different stress conditions. *Microbiology (Reading)*
982 2020;166(2):157-168.
983 123. **Hofmann AF, Hagey LR**. Bile acids: chemistry, pathochemistry, biology, pathobiology,
984 and therapeutics. *Cell Mol Life Sci* 2008;65(16):2461-2483.
985 124. **Begley M, Gahan CG, Hill C**. The interaction between bacteria and bile. *FEMS Microbiol*
986 *Rev* 2005;29(4):625-651.

125. **Merritt ME, Donaldson JR**. Effect of bile salts on the DNA and membrane integrity of enteric bacteria. *J Med Microbiol* 2009;58(Pt 12):1533-1541.
126. **Prieto AI, Ramos-Morales F, Casadesus J**. Bile-induced DNA damage in *Salmonella enterica*. *Genetics* 2004;168(4):1787-1794.
127. **Prieto AI, Ramos-Morales F, Casadesus J**. Repair of DNA damage induced by bile salts in *Salmonella enterica*. *Genetics* 2006;174(2):575-584.
128. **Hernandez SB, Cota I, Ducret A, Aussel L, Casadesus J**. Adaptation and preadaptation of *Salmonella enterica* to Bile. *PLoS Genet* 2012;8(1):e1002459.
129. **Walawalkar YD, Vaidya Y, Nayak V**. Response of *Salmonella* Typhi to bile-generated oxidative stress: implication of quorum sensing and persister cell populations. *Pathogens and Disease* 2016;74(8):ftw090.
130. **Di Lorenzo P**. usmap: US Maps Including Alaska and Hawaii version R package version 0.6.0. 2022. https://CRAN.R-project.org/package=usmap.

TABLES

Table 1. Presence and absence of *artAB*, *gogB*, and Gifsy-1 among the U.S., Scottish, and global DT104 data sets, plus all data sets combined.

Data Set(s)	Host(s)	Total # of Genomes	*artAB* Present (%)[a]	*gogB* Present (%)[a]	Gifsy-1 Present (%)[b]
U.S. [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%)
Scottish [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%)
Global [e]					
	All	290	271 (93.4%)	271 (93.4%)	265 (91.4%)
Combined					
	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 219 U.S. human- and bovine-associated DT104 genomes acquired as described previously [24] and characterized here, plus 11 bovine- and human-associated U.S. DT104 genomes from the global DT104 data set.
[d] Refers to a set of 243 Scottish human- and bovine-associated DT104 genomes sequenced and characterized previously [36].
[e] Refers to a set of 290 DT104 genomes collected from various sources around the world, which were sequenced and characterized previously [13].

Table 2. Location of *artAB* and *gogB* in DT104 genomes within the U.S., Scottish, and global DT104 data sets, plus all data sets combined.

		# 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]
Genes[a]	**Data Set (# of Genomes with Gene/ Total # of Genomes; %)**	**Intact (%)[c]**	**Incomplete (%)[c]**	**Intact (%)[c]**	**Incomplete (%)[c]**	**Intact (%)[c]**	**Incomplete (%)[c]**	
artAB								
	U.S.[d] (177/230; 77.0%)	177 (100.0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	Scottish[e] (240/243; 98.8%)	21 (8.8)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	219 (91.3)
	Global[f] (271/290; 93.4%)	263 (97.0)	0 (0)	0 (0)	0 (0)	2 (0.7)	0 (0)	6 (2.2)
	Combined (678/752; 90.2%)	451 (66.5)	0 (0)	0 (0)	0 (0)	2 (0.3)	0 (0)	225 (33.2)
gogB								
	U.S.[d] (180/230; 78.3%)	51 (28.3)	75 (41.7)	1 (0.6)	2 (1.1)	0 (0)	0 (0)	51 (28.3)
	Scottish[e] (240/243; 98.8%)	50 (20.8)	85 (35.4)	0 (0)	0 (0)	0 (0)	0 (0)	105 (43.8)
	Global[f] (271/290; 93.4%)	11 (4.1)	67 (24.7)	2 (0.7)	1 (0.4)	0 (0)	0 (0)	190 (70.1)
	Combined (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 "Data Set" column used as a denominator.
[d] Refers to the set of 219 U.S. human- and bovine-associated DT104 genomes acquired as described previously [24] and characterized here, plus 11 bovine- and human-associated U.S. DT104 genomes from the global DT104 data set.
[e] Refers to a set of 243 Scottish human- and bovine-associated DT104 genomes sequenced and characterized previously [36].
[f] Refers to a set of 290 DT104 genomes collected from various sources around the world, which were sequenced and characterized previously [13].

FIGURES

Figure 1. Geographic and source origins (i.e., human or bovine) of (A) all 230 human- and bovine-associated United States (U.S.) DT104 genomes queried in this study, 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 [130].

Figure 2. Bayesian time-scaled phylogeny constructed using 146 human- and bovine-associated DT104 genomes collected in the United States (U.S.). 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 S2-S4.

Figure 3. Coalescent Bayesian Skyline plot constructed using 146 U.S. bovine- and human-associated *S.* Typhimurium DT104 genomes. 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).

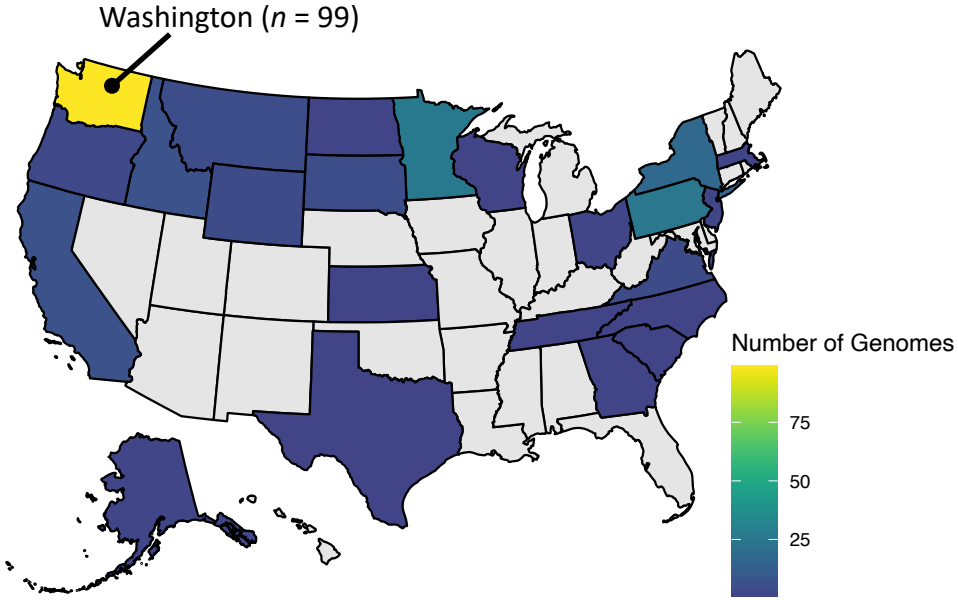
Figure 4. Time-scaled maximum likelihood (ML) phylogeny constructed using the combined 752-genome DT104 data set. Tip label colors denote the study with which each genome is affiliated. The heatmap encompassing the phylogeny denotes the presence and absence of selected virulence factors and intact prophage. 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 S9.

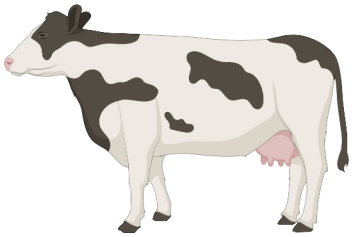
Figure 5. Response of DT104 isolates (*n* = 6) to environmental stress factors correlated with the presence of Gifsy-1/*artAB*/*gogB*. Base-ten logarithmic fold change (FC) was calculated as follows: FC = log CFU/g at the start of the experiments – log CFU/g after the stress assay. (A) (A) Log FC of 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.

U.S. DT104 genomes

(full data set; *n* = 230)





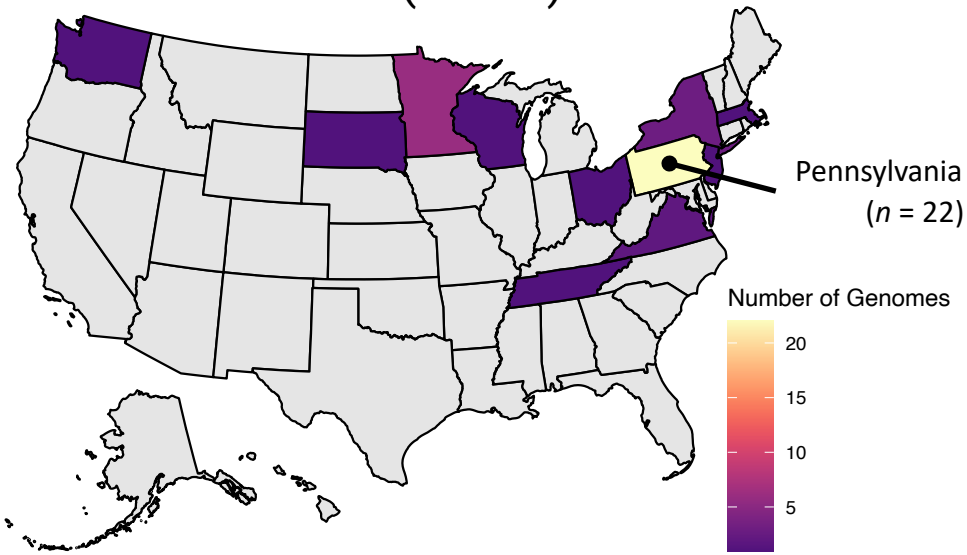
n = 62

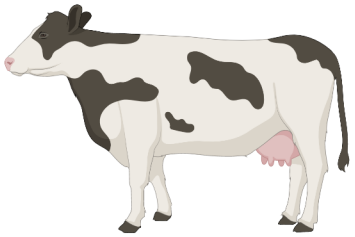
n = 168

B.

U.S. *artAB*-negative major clade

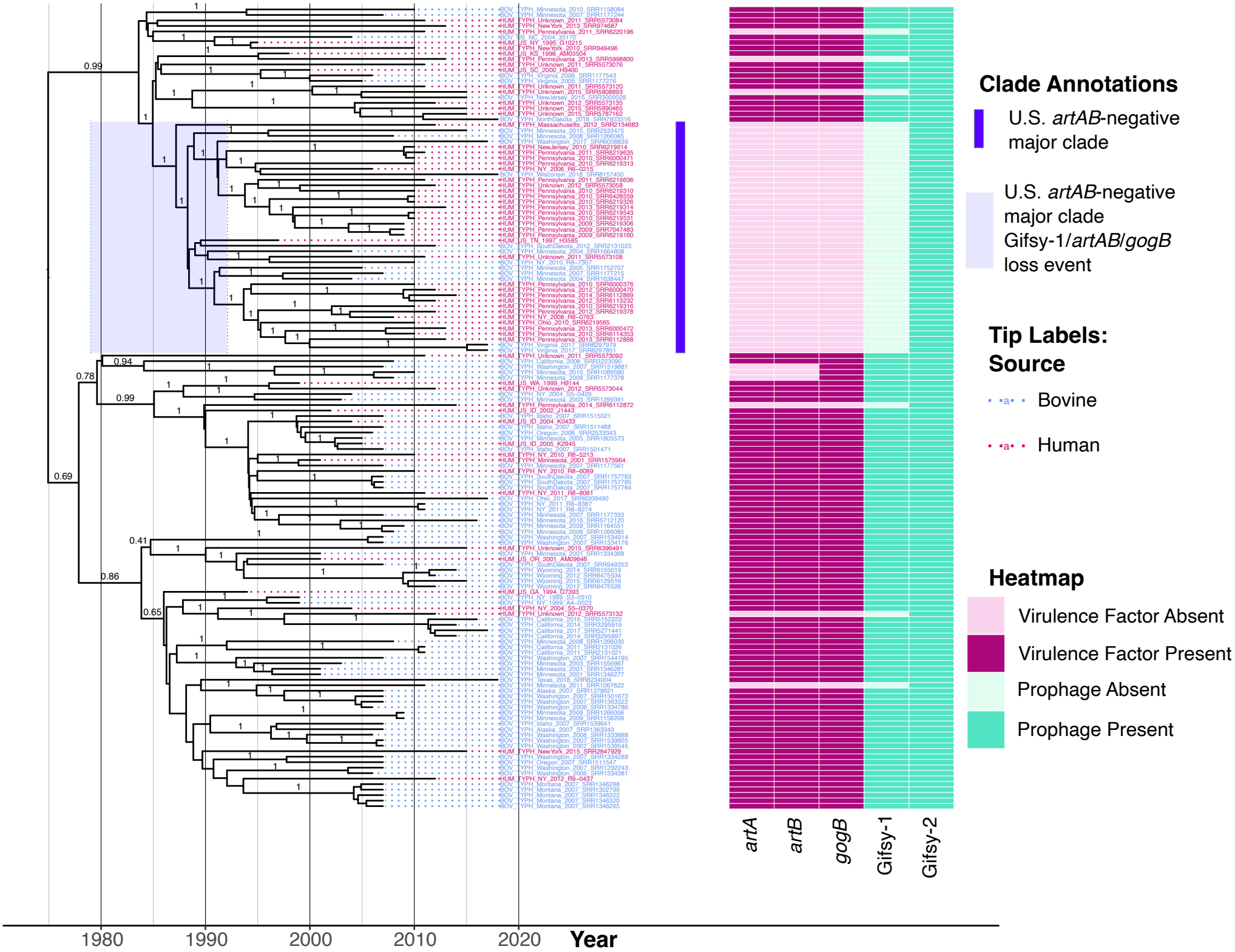
(*n* = 42)

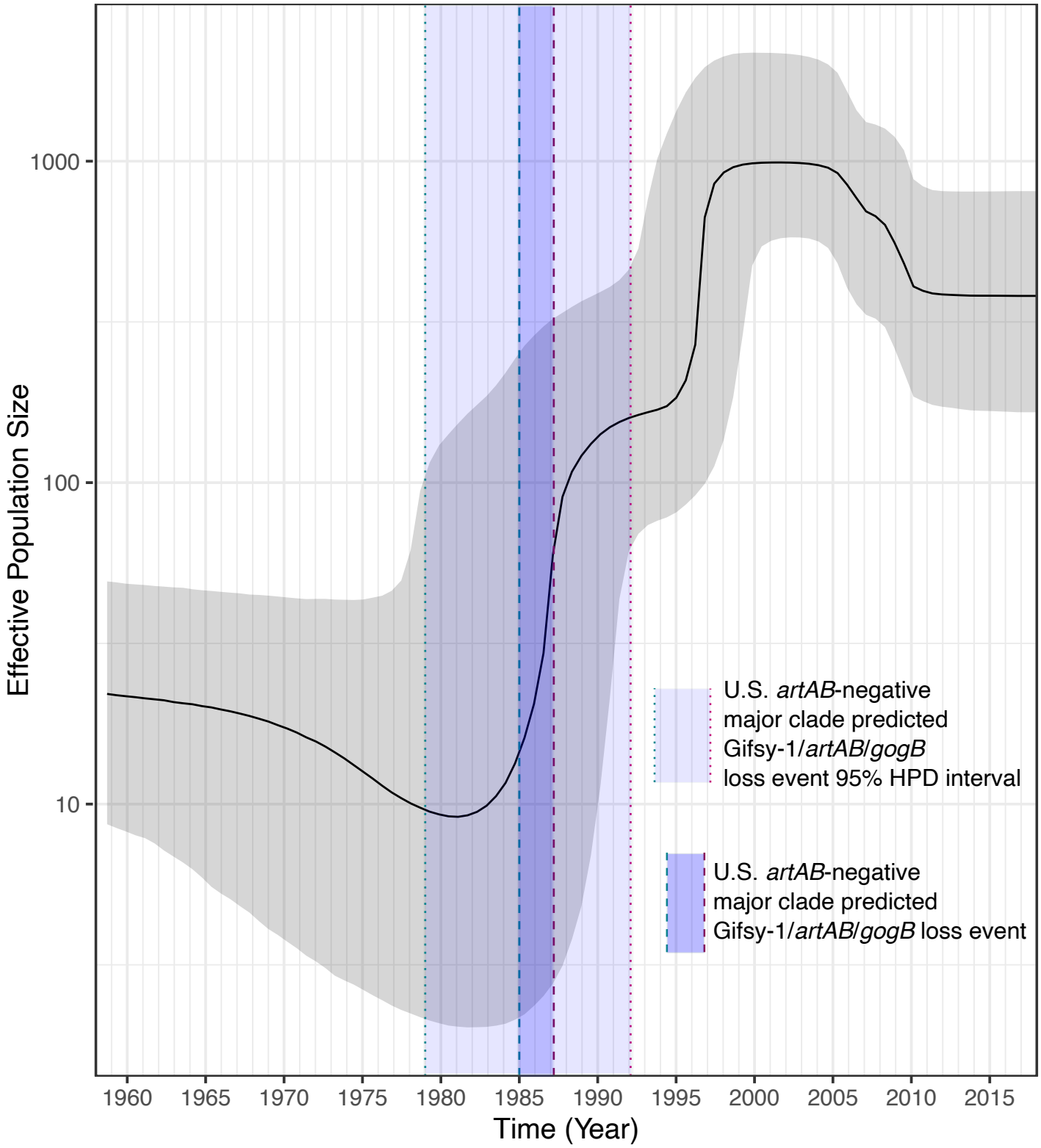


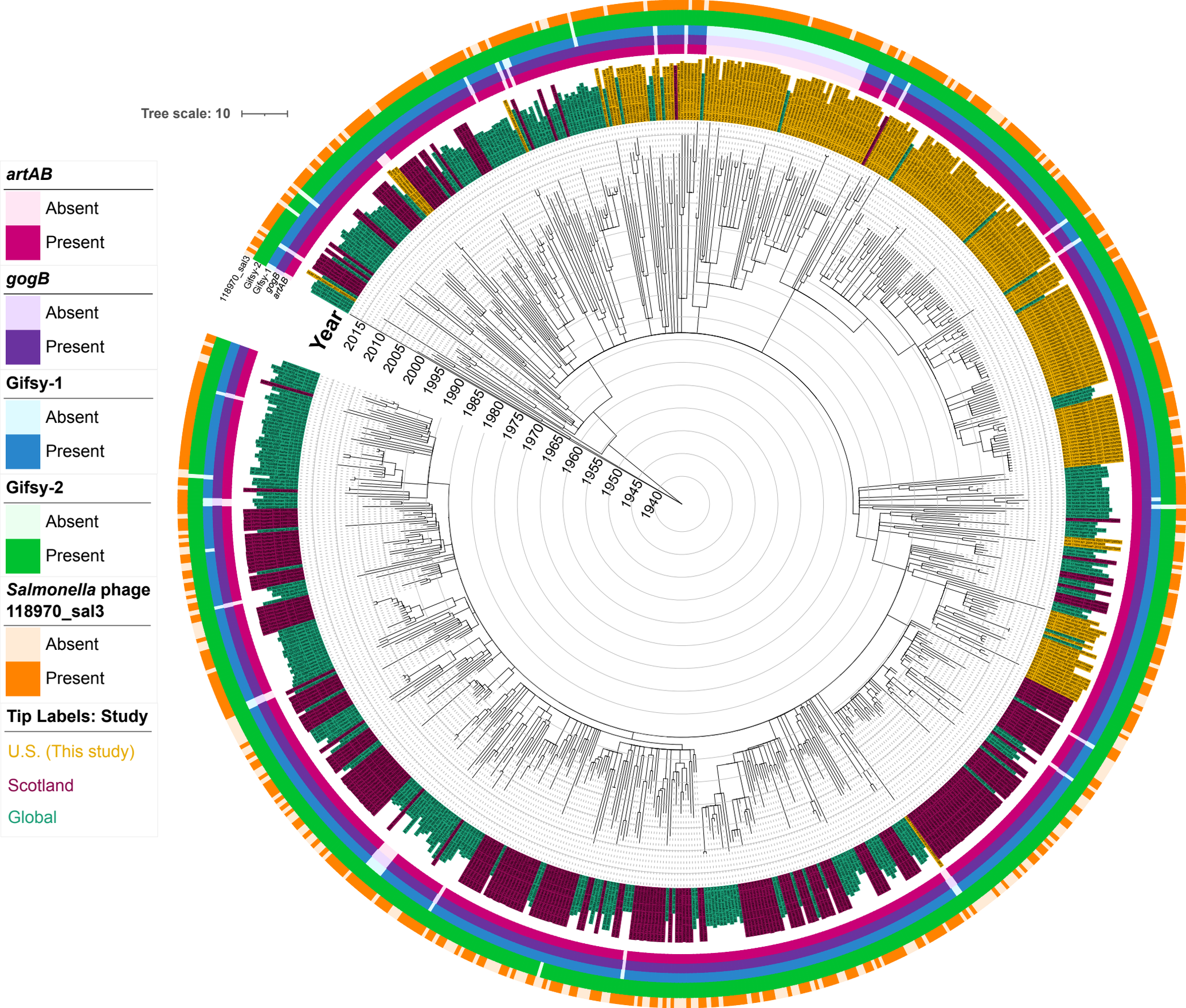


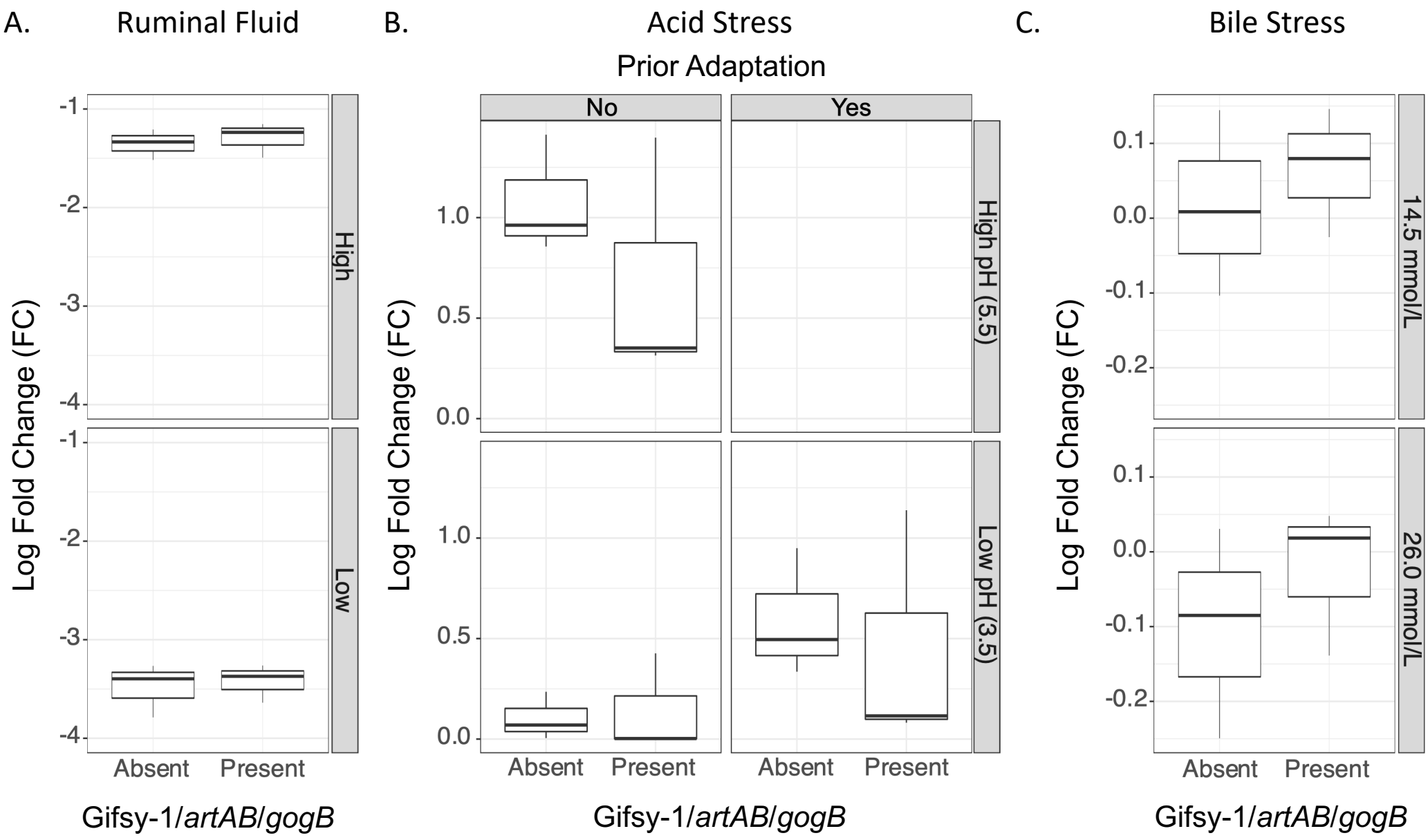
n = 30

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


Tree scale: 10
artAB
Absent
Present
gogB
Absent
Present
Gifsy-1
Absent
Present
Gifsy-2
Absent
Present
Salmonella phage 118970_sal3
Absent
Present
Tip Labels: Study
U.S. (This study)
Scotland
Global
118970_sal3
Gifsy-2
Gifsy-1
gogB
artAB
Year
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