

What's on a prophage: analysis of *Salmonella* spp. prophages identifies a diverse range of cargo with multiple virulence- and metabolism-associated functions

Caroline R. Yates,¹ Anthony Nguyen,² Jingqiu Liao,³ Rachel A. Cheng¹

AUTHOR AFFILIATIONS See affiliation list on p. 16.

ABSTRACT The gain of mobile elements, such as prophages, can introduce cargo to the recipient bacterium that could facilitate its persistence in or expansion to a new environment, such as a host. While previous studies have focused on identifying and characterizing the genetic diversity of prophages, analyses characterizing the cargo that prophages carry have not been extensively explored. We characterized prophage regions from 303 *Salmonella* spp. genomes (representing 254 unique serovars) to assess the distribution of prophages in diverse *Salmonella*. On average, prophages accounted for 3.7% (0.1%–8.8%) of the total genomic content of each isolate. Prophage regions annotated as Gifsy 1 and Salmon Fels 1 were the most commonly identified intact prophages, suggesting that they are common throughout the *Salmonella* genus. Among 21,687 total coding sequences (CDSs) from intact prophage regions in subsp. *enterica* genomes, 7.5% (median; range: 1.1%–47.6%) were categorized as having a function not related to prophage integration or phage structure, some of which could potentially provide a functional attribute to the host *Salmonella* cell. These predicted functions could be broadly categorized into CDSs involved in: (i) modification of cell surface structures (i.e., glycosyltransferases); (ii) modulation of host responses (e.g., SodC/SodA, SopE, ArtAB, and typhoid toxin); (iii) conferring resistance to heavy metals and antimicrobials; (iv) metabolism of carbohydrates, amino acids, and nucleotides; and (v) DNA replication, repair, and regulation. Overall, our systematic analysis of prophage cargo highlights a broader role for prophage cargo in influencing the metabolic, virulence, and resistance characteristics of *Salmonella*.

IMPORTANCE Lysogenic bacteriophages (phages) can integrate their genome into a bacterial host's genome, potentially introducing genetic elements that can affect the fitness of the host bacterium. The functions of prophage-encoded genes are important to understand as these genes could be mobilized and transferred to a new host. Using a large genomic dataset representing >300 isolates from all known subspecies and species of *Salmonella*, our study contributes important new findings on the distribution of prophages and the types of cargo that diverse *Salmonella* prophages carry. We identified a number of coding sequences (CDSs) annotated as having cell surface-modifying attributes, suggesting that prophages may have played an important role in shaping *Salmonella*'s diverse surface antigen repertoire. Furthermore, our characterization of prophages suggests that they play a broader role in facilitating the acquisition and transfer of CDSs associated with metabolism, DNA replication and repair, virulence factors, and to a lesser extent, antimicrobial resistance.

KEYWORDS *Salmonella*, prophage, bacteriophage, antimicrobial resistance genes

Editor Drusilla L. Burns, U.S. Food and Drug Administration, Silver Spring, Maryland, USA

Address correspondence to Rachel A. Cheng, racheng@vt.edu.

The authors declare no conflict of interest.

See the funding table on p. 16.

Received 13 January 2024

Accepted 22 April 2024

Published 22 May 2024

Copyright © 2024 Yates et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Salmonella is a diverse foodborne pathogen that accounts for over 90 million illnesses globally each year (1, 2). Within the genus *Salmonella*, there are two species, *Salmonella bongori* and *Salmonella enterica*; *S. enterica* can be divided further into six recognized subspecies *enterica* (I), *salamae* (II), *arizonae* (IIIa), *diarizonae* (IIIb), *houtenae* (IV), and *indica* (VI), and an additional proposed subspecies *londinensis* (VII) (3). *Salmonella enterica* subsp. *enterica* can be further categorized into four major phylogenetic clades (A–D); this classification has been useful for understanding the distribution of horizontally acquired elements among diverse *Salmonella* serovars (4). *Salmonella*'s lipopolysaccharide (LPS) and flagellar structures constitute the O and H antigens, respectively, that are the basis for the Kauffman–White–Le Minor serotype classification scheme that remains useful for surveillance. There are over 2,600 recognized serovars, with 1,586 belonging to *S. enterica* subsp. *enterica* (5). Importantly, some serovars may evolve to express the same serotype despite evolving from different most recent common ancestors, leading to so-called polyphyletic serovars (4).

S. enterica is associated with a wide variety of hosts and environments, which reflects the genetic diversity of this pathogen (6). Genes that have important relevance to human disease include those whose gene products facilitate (i) enhanced survival under physiological stress such as the environment encountered in the host gastrointestinal tract, (ii) utilization of host-generated metabolites, and (iii) resistance to antimicrobial treatments, all of which are known to vary by serovar (7, 8). Multiple studies have demonstrated that virulence potential can vary by serovar and strain, highlighting the limitations of current serotype-dependent surveillance frameworks, as well as the importance of understanding the genetic diversity of *Salmonella* (9, 10).

Bacteriophages (viruses that infect bacterial hosts) are the most abundant organisms in nature, having an estimated population of 10^{31} members (11). Temperate bacteriophages are capable of undergoing both lytic and lysogenic life cycles. During the lysogenic phase, bacteriophages are able to integrate their genome into the host bacterium's genome, forming a prophage (11). When the bacterial host experiences physiological stress (e.g., DNA damage, excessively high temperatures, antimicrobial treatment), prophage induction can occur, causing the bacteriophage to enter the lytic cycle producing virions that ultimately lyse and kill the host cell (12). Depending on the prophage, these genomic regions may be maintained in the genome (i.e., remain "intact"), or they may be degraded (leading to cryptic prophages). In addition to encoding genes necessary for their own replication and assembly, prophages are also important facilitators of horizontal gene transfer, often carrying CDSs that have been acquired from prior hosts as their "cargo" (13).

S. enterica is known to encode a variety of prophages, including Gifsy 1 and Salmon Fels 2 (4, 14). Some studies have demonstrated a role for prophages in facilitating acquisition of virulence factors, such as SodC, which has been shown previously to enhance *Salmonella* survival during macrophage infection (15, 16). Although previous work by Brussow et al. (17) has indicated that prophages likely provide some benefit to their bacterial host, the specific types of cargo that prophages carry, and the potential functional benefit of these cargo, remain poorly understood. Therefore, we aimed to (i) systematically characterize the diversity and distribution of prophages across seven *S. enterica* subspecies and *S. bongori* to contextualize the distribution of prophages in genetically diverse *Salmonella* spp. isolates, and (ii) assess the predicted functions of CDSs within intact prophage regions to understand what potential benefits these CDSs confer to lysogenized *Salmonella*.

RESULTS

S. enterica subsp. *enterica* genomes encode diverse prophages, which account for an average of 4% of the total genomic content

The genomic content of many bacterial pathogens has been shown to be impacted by horizontal gene transfer driven in part by lysogenic phages (13, 18–20). To determine how prophages may have shaped the resulting genomes of the foodborne pathogen

S. enterica, we first characterized the prophage content of 242 *S. enterica* subspecies *enterica* genomes representing 217 serovars that are associated with human and animal clinical infections in the United States (21). On average, 3.9% of the genome was annotated as prophage regions (range: 0.4%–8.8%), which was consistent across the four major subsp. *enterica* phylogenetic clades, A–D and for the two isolates representing serovars Poano and Lattenkamp ($P = 0.9$, Kruskal–Wallis; Fig. 1A), which could not be assigned to a major phylogenetic clade. All 242 genomes were annotated as having at least one prophage region, with a total of 1,918 prophage regions identified.

Annotation with Phaster also provided predictions about the completeness of the prophage region, including whether a prophage is predicted to be intact, incomplete, or questionable (Fig. 1B) (22). Not surprisingly, the majority of the prophage regions were annotated as incomplete (987 regions; 51.5%), followed by regions that were predicted to be intact (569 regions, 29.7%) or questionable (362 regions; 18.9%). We chose to focus our analysis on the 569 regions that were considered to be “intact” as we hypothesized that these may be the most likely regions that could be excised and transferred to a new host. Among these 242 genomes, 206 (85.1%) contained at least one intact prophage region (average: two intact prophages; range: 0–8). Additionally, these 206 genomes represented all major phylogenetic clades previously characterized in *Salmonella* (4), suggesting that our data set includes diverse representation of potentially mobilizable prophage regions throughout *S. enterica* subsp. *enterica*. Among these intact prophages, 83 unique prophages (based on the phage name from blast identification included in Phaster output) were identified, with 22 of those prophages being detected in more than five genomes (representing 74.6% of all intact prophages), defined here as common prophages (Fig. 2). Gifsy 1 and Salmon Fels 1 (both predicted to be intact in 67 genomes) represented the most common intact prophages in *S. enterica* subsp. *enterica*. Mapping the presence of these prophages onto the *S. enterica* phylogeny revealed a sporadic distribution by phylogenetic clade (Fig. 2; Data Set S1), suggesting a complex pattern of evolutionary events accompanying the acquisition, maintenance, and loss of these prophages.

Over half of the most common intact prophages identified in *S. enterica* subsp. *enterica* are also found in at least one other non-*enterica* subspecies

To better understand the distribution of these prophages across the *Salmonella* genus, we next examined their presence in an additional 61 genomes representing *S. bongori* (31 genomes) and representative genomes for the remaining *S. enterica* subspecies (*salamae*, *arizonae*, *diarizonae*, *houtenae*, *indica*, and *londinensis* [subsp. VII], five genomes per subspecies). Similar to what we observed for the subsp. *enterica* genomes, within these non-subsp. *enterica* genomes, an average of 3.1% of genomic content was annotated as prophage (range: 0.1–8.1%, Fig S1A). Among 380 total prophage regions identified (average: 6, range: 1–13 prophage regions per genome), 110 (28.9%; Fig S1B) were predicted to be intact with phage Salmon 118970 sal3 (detected in 14 genomes) being the most common of the 31 unique prophages identified in the non-subsp. *enterica* genomes. Overall, 15 of the 22 common intact prophages in subsp. *enterica* were also found to be intact in non-subsp. *enterica* genomes represented here, demonstrating that many of these prophages are broadly present across the *Salmonella* genus.

No significant associations were found between common prophages and O antigen groups

The O antigen (terminal portion of LPS) has been demonstrated previously to be a receptor for multiple bacteriophages that infect *Salmonella* (23). Given the apparent sporadic distribution of the intact prophages (Fig. 2) and the previous observation that O groups evolved many times throughout the *Salmonella*'s evolutionary history (24), we next determined if there were any associations between O groups and the presence of specific prophages. Subspecies *enterica* genomes harboring intact prophages represented 29 unique O groups (average: 7; range: 1–36 genomes per O group; Fig. 3).

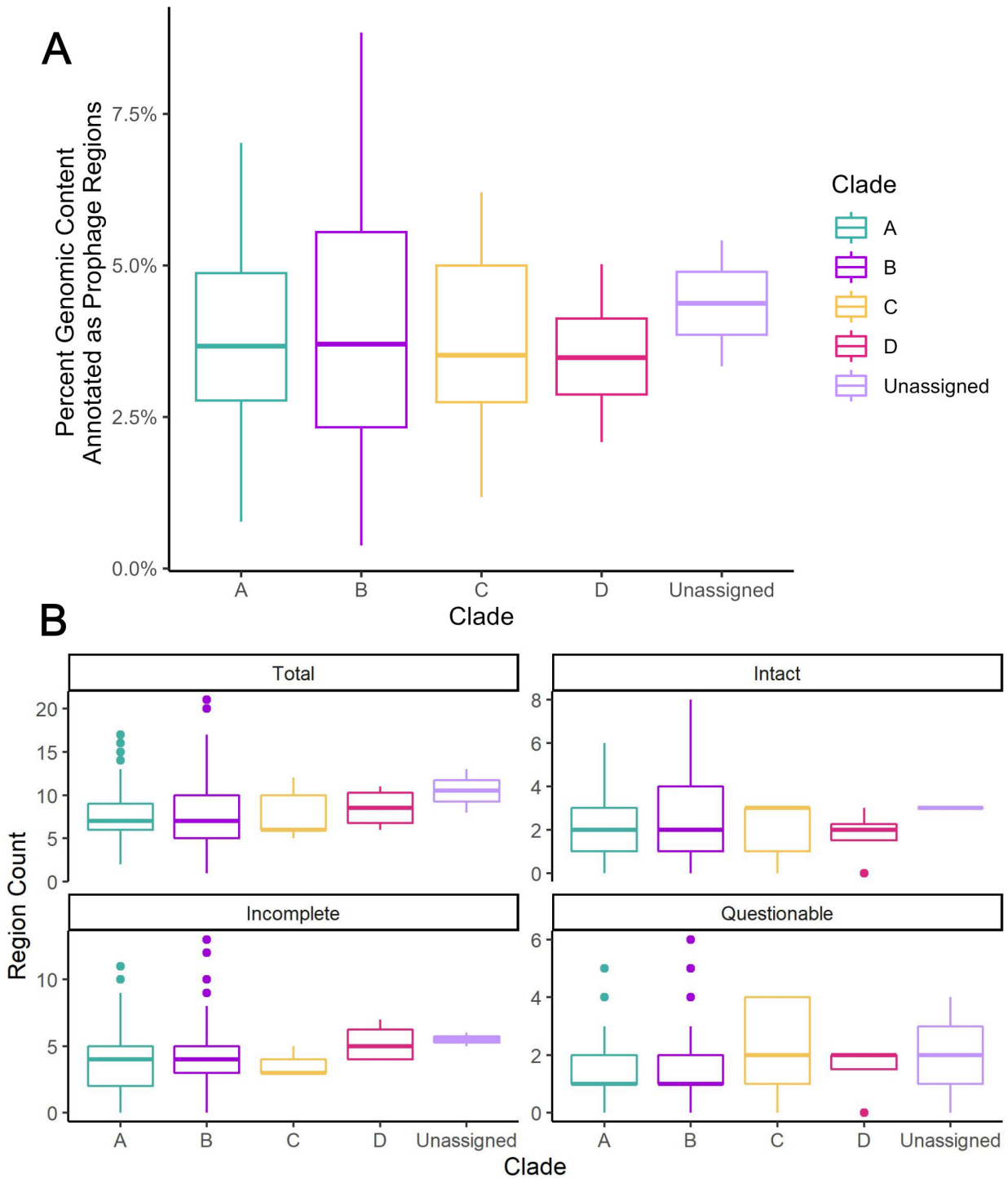


FIG 1 Distribution of prophage regions among 242 *S. enterica* subsp. *enterica* genomes representing 217 unique serovars and 25 poly or paraphyletic serovars. Prophage regions within the 242 *S. enterica* genomes (clade A [$n = 152$ genomes], clade B [$n = 79$ genomes], clade C [$n = 5$ genomes]), clade D [$n = 4$ genomes], and clade Unassigned [$n = 2$ genomes]) were annotated as intact, incomplete, and questionable by Phaster. For two genomes representing serovars Lattenkamp and Poano, the topology of the core SNP phylogeny indicated that these two genomes do not belong to one of the major clades (A, B, C, D), so these genomes were categorized as “clade unassigned.” (A) Percent of the genomic content annotated as prophage regions by phylogenetic clade for 242 *Salmonella enterica* genomes. The percents were calculated using the total length (in base pairs) annotated as prophage by Phaster and the total base pairs in the genome. (B) Box plots displaying counts of the number of total, intact, incomplete, and questionable prophage regions among all *S. enterica* subsp. *enterica* genomes.

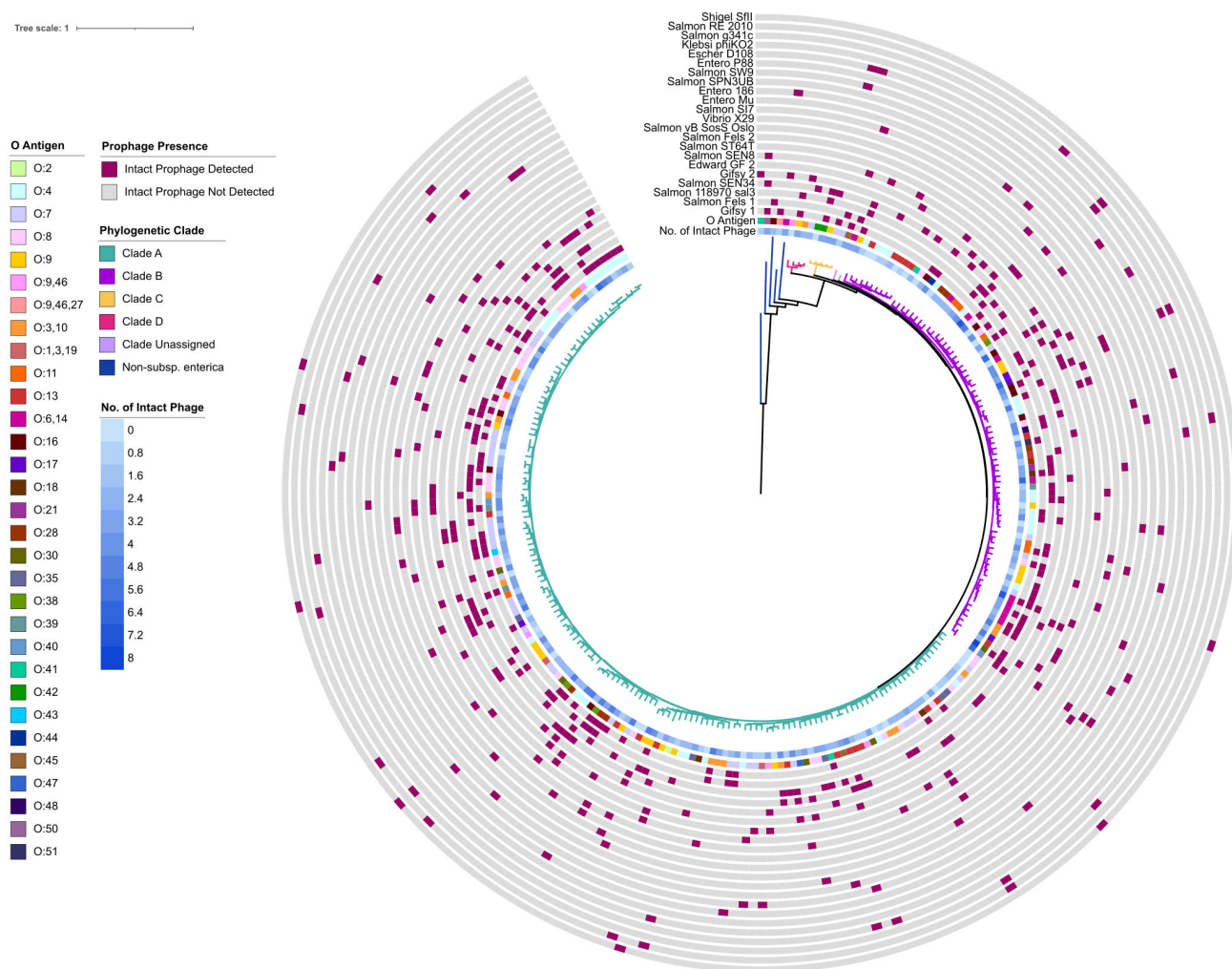


FIG 2 Distribution of intact prophage regions identified from *Salmonella* spp. genomes. A maximum likelihood phylogeny was inferred from core SNPs identified with kSNP3 for genomic assemblies representing 242 *S. enterica* subsp. *enterica* genomes and five additional assemblies representing subspecies *salamae* (II), *arizonae* (IIIa), *diarizonae* (IIIb), *houtenae* (IV), and *indica* (VI). Tree branches are colored to reflect the genome’s clade, which was assigned based on the tree topology and bootstrap support. The number of intact prophage regions identified with Phaster are indicated as the innermost line, with darker squares indicating higher numbers of intact prophage. The isolate O Group is indicated with a colored strip. For prophages with counts above five (i.e., present in more than five subsp. *enterica* genomes), presence (pink box) or absence (gray box) is shown.

Comparison of the presence/absence pattern of prophages by O group suggested that the distribution of most of these common prophages did not have a strong correlation with specific O groups (Fig. 3). Indeed, the proportion of genomes in each O group having a specific intact prophage did not vary significantly (Fisher’s exact tests; for *P*-values see Data Set S2), although for some O groups, the proportion of genomes having a specific intact prophage was notably higher (e.g., 67% of genomes in Group O:7 had intact Salmon vB Sos Oslo, and 57% of genomes in Group O:4 had intact Entero 186), suggesting that there may be a relationship between the presence of these prophages and isolates belonging to the O:4 and O:7 serogroups. Overall, this suggests that among the common intact prophages identified in subsp. *enterica* genomes, most are not associated with a specific O group.

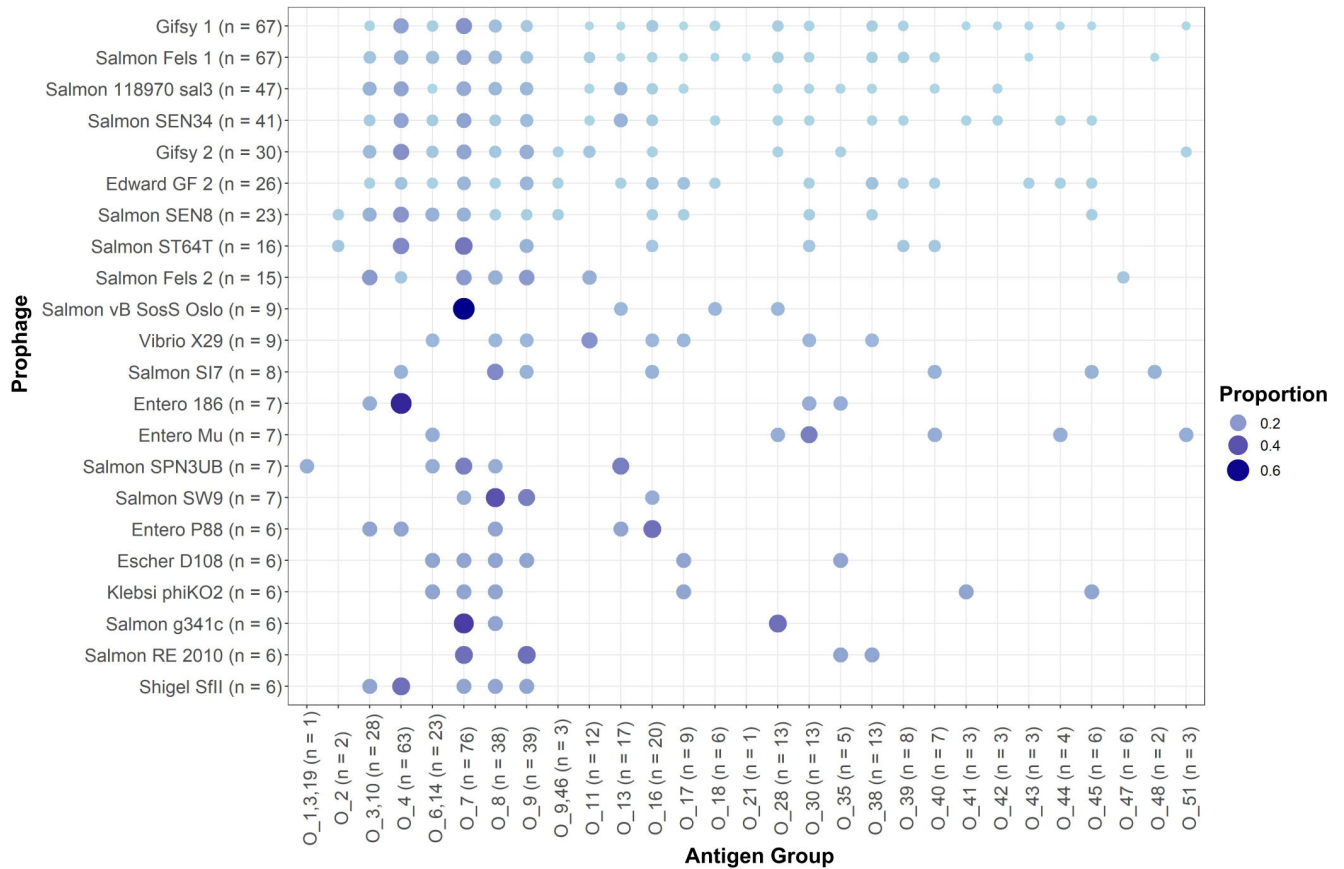


FIG 3 Distribution of the intact prophage regions across *Salmonella* serogroups. The bubble plot shows the number of prophages (y-axis; for prophages present in more than five *Salmonella* genomes) with *S. enterica* subsp. *enterica* genomes categorized by the O antigen group of the isolate (x-axis; number reflects the number of isolates in the serogroup included in the data set), represented as colored circles. The diameter and shading of the circle reflect the proportion of isolates in a given serogroup that were annotated as having an intact prophage region for the given prophage (y-axis), with larger, darker circles representing serogroups that had a high proportion of isolates with a given prophage.

Intact prophage regions harbor coding sequences with a diverse range of predicted functions

During prophage excision, it is possible for bacterial host DNA to be excised and packaged into phage capsids, enabling potential horizontal transfer of cargo to a new host upon infection. To systematically identify and categorize CDSs that could provide some accessory function to the host *Salmonella* (see Materials and Methods for details; see Data Set S3 for exact terms), we first categorized intact prophage regions into clusters of orthologous groups (COGs) using EggNOG-mapper v2 (25). EggNOG annotated a total of 21,687 CDSs among the 569 intact prophages from 206 subsp. *enterica* genomes. After removing CDSs annotated as “Unknown Function” (COG S; $n = 12,111$) and “COG unassigned” (COG “-”; $n = 4,015$), the curated data set contained 5,561 remaining CDS annotations. Overall, these 5,561 CDSs were assigned to 38 unique COG categories, including those that represent a combination of single category annotations (i.e., COG EG encompasses functions related to Amino Acid Metabolism [E] and Carbohydrate Metabolism [G]); Fig. 4A). COG categories Mobilome (COG X); Replication, Recombination, and Repair (COG L); and Transcription (COG K) had the highest proportion of CDSs assigned, which was consistent for all clades, except Clade D (Fig. S2). Carbohydrate Transport and Metabolism (G) and Cell Wall/Membrane/Envelope Biogenesis (M) were also among the COG categories with the highest number of CDSs assigned suggesting that prophage cargo may play an important role in carbohydrate metabolism and cell wall modifications.

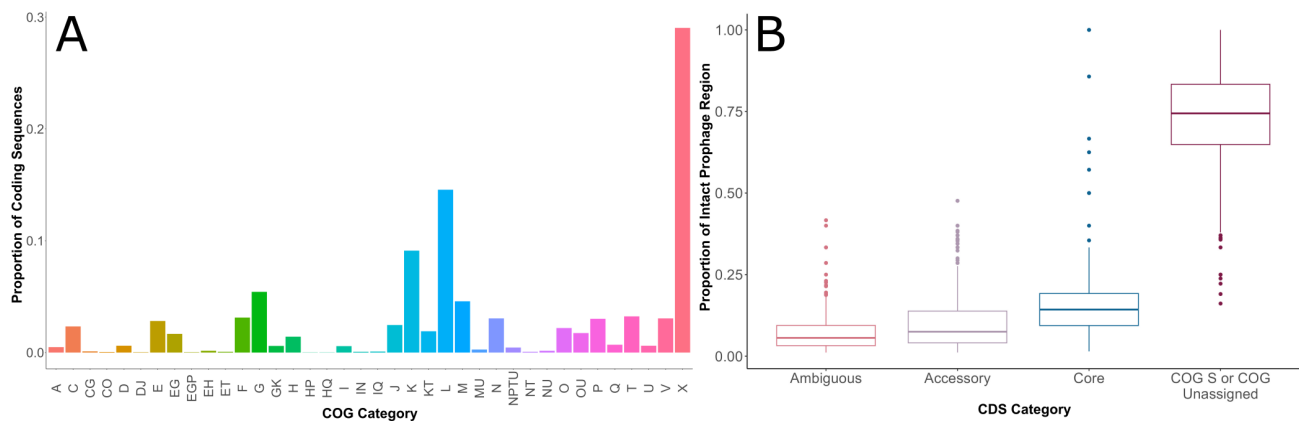


FIG 4 Classifications of CDSs from 569 intact prophage regions within 206 *S. enterica* subsp. *enterica* genomes. (A) Proportion of CDS assigned to each COG category for all phylogenetic clades for the 5,561 CDSs in the curated data set; the total number of genomes included in this analysis (206 genomes) represents the number of genomes from which intact prophage regions were identified with Phaster. COG annotations were determined using EggNOG. Data for COG categories “S” (Function Unknown) and “-” (No COG Category Assigned) are not shown as these COGs were not included in the curated CDS data set; raw data for this analysis is included in Data Set S2. COG categories are as follows: RNA Processing (A), Energy Production (C), Energy Production/Carbohydrate Metabolism (CG), Energy Production/Posttranslational Modification (CO), Cell Division (D), Cell Division/Translation (DJ), Amino Acid Metabolism (E), Amino Acid Metabolism/Carbohydrate Metabolism (EG), Amino Acid Metabolism/Carbohydrate Metabolism/Inorganic Ion Metabolism (EGP), Amino Acid Metabolism/Coenzyme Metabolism (EH), Amino Acid Metabolism/Signal Transduction (ET), Nucleotide Metabolism (F), Carbohydrate Metabolism (G), Carbohydrate Metabolism/Transcription (GK), Coenzyme Metabolism (H), Coenzyme Metabolism/Inorganic Ion Metabolism (HP), Coenzyme Metabolism/Secondary Metabolite Metabolism (HQ), Lipid Metabolism (I), Lipid Metabolism/Cell Motility (IN), Lipid Metabolism/Secondary Metabolite Metabolism (IQ), Translation (J), Transcription (K), Transcription/Signal Transduction (KT), Replication (L), Cell Wall/Membrane Biogenesis (M), Cell Wall/Membrane Biogenesis/Intracellular Trafficking (MU), Cell Motility (N), Cell Motility/Inorg Ion Met/Sigal Transduction/Intracell Trafficking (NPTU), Cell Motility/Signal Transduction (NT), Cell Motility/Intracellular Trafficking (NU), Posttranslational Modification (O), Posttranslational Modification/Intracellular Trafficking (OU), Inorganic Ion Metabolism (P), Secondary Metabolite Metabolism (Q), Signal Transduction (T), Intracellular Trafficking (U), Defense Mechanisms (V), Mobilome (X). (B) CDSs were categorized as ambiguous, accessory, core, or COG S or COG Unassigned based on their predicted function by EggNOG. All 21,687 CDSs from 569 intact prophages were analyzed to determine the proportion of each CDS category by region. Classifications were based on EggNOG annotations and manual curation as defined in the methods.

While the majority of CDSs within a given intact prophage could not be assigned a definitive function, 7.5% of CDSs on intact prophage regions (median; range: 1.1%–47.7%) were assigned a function that was considered to be “accessory” to the prophage (see Materials and Methods for a discussion of how these were assigned), which we hypothesized included some CDSs that could potentially provide a functional benefit to the host bacterial cell (Fig. 4B). These 1,661 accessory CDSs (representing 305 unique EggNOG definition terms) could be broadly categorized into (i) metabolic; (ii) cell surface structures; (iii) DNA replication, repair, and regulation; and (iv) virulence-associated functions. While CDSs were classified as accessory if their annotated function was not explicitly related to phage structural proteins or integration, some of these CDSs, especially those involved in DNA replication, repair, and regulation, could be part of the prophage replication module. Based on the distribution of these accessory CDSs, 73 were common (i.e., found in at least five different prophages or genomes; Data Set S3), which we then characterized further.

Two-thirds of accessory CDSs on intact prophage regions are associated with replication, transcription, or metabolism

Not surprisingly, 26 of the 73 most common accessory CDSs had an annotated function related to DNA replication and repair (e.g., DNA polymerase θ subunit, endonucleases), transcription (e.g., sigma factors, transcriptional regulators, methyltransferases), and translation (e.g., tRNA synthetase); six of these CDSs consistently mapped to the ends of prophage regions (Fig. 5; see Materials and Methods for details about how this was determined). Metabolism-associated CDSs were the second most populous category, accounting for 22 of the 73 most commonly occurring accessory CDSs on intact

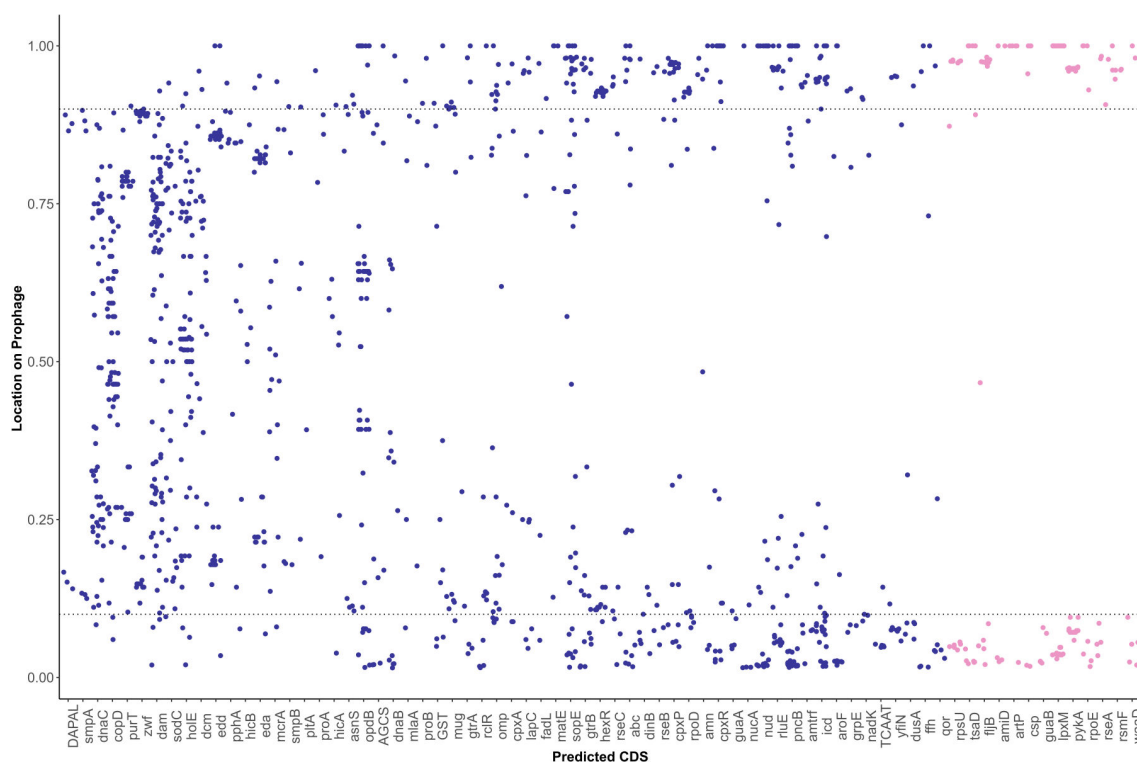


FIG 5 Position of accessory coding sequences on intact prophage regions. The position of the most common (presence in five or more intact prophage regions) accessory CDSs on intact prophages was calculated as a proportion where 0.0 indicates the first coding sequence annotated from the prophage, and 1.0 indicates the last coding sequence from that prophage (see Materials and Methods). Pink circles represent predicted CDS types where over 90% of the annotations were found in either the first or last 10% of the prophage; blue circles represent predicted CDS types with under 90% annotations found at either end of the prophage.

prophages (Data Set S3); only two of these CDSs were located on the ends of prophage regions. These CDSs were associated with metabolism of a wide variety of compounds including carbohydrates (glucose-6-phosphate), amino acids, nucleotides, cofactors, and metal ions (Data Set S3). CDSs for carbohydrate metabolism included isocitrate dehydrogenase (*icd*) and genes associated with glucose-6-phosphate oxidation, while those for nucleotide metabolism encoded enzymes for purine biogenesis (*purT*, *proAB*, and *guaA*). CDSs related to siderophore synthesis and export (*iroC*, *fepBCDG*, and *entABCEHS*), zinc transport (*znuABC*), and copper resistance protein CopD were also identified (Fig. 5; Data Set S3). These data suggest that the majority of accessory CDSs on prophages facilitate DNA replication and repair and metabolism of a variety of compounds.

Intact prophages carry CDSs associated with modification of LPS and cell surface proteins

Prophages have been previously shown to carry genes encoding proteins that modify the host cell to prevent infection with other bacteriophages that could use the same surface receptor (26, 27). Fifteen unique CDSs associated with functions related to modification of cellular surface structures were identified in the most commonly occurring 73 accessory CDSs, with five consistently occurring at the ends of the region (Fig. 5; Data Set S3). *fljB* encoding the phase 2 flagellin was almost exclusively found at the end of prophage regions identified as Salmon SEN 8 and Salmon Fels 2, suggesting either horizontal transfer of this locus or integration of these prophages adjacent to *fljB* in multiple serovars. CDSs for outer membrane porins (detected in 25 prophage regions) were also prominent among cell structure-associated CDSs. The GTR glucosyltransferase

operon *gtrABC* was found intact in five prophage regions (Salmon ST160, Salmon ST64T, and Entero UAB Phi20), or as an incomplete operon in an additional five intact prophage regions (Data Set S3) (28). Also identified were CDSs associated with maintenance of the bacterial outer membrane, including those involved in outer membrane lipid asymmetry and the synthesis of lipid A (*waaD* and *lpxM*), but these were also almost exclusively found at the ends of prophage regions (Fig. 5). These results highlight that prophages carry a variety of CDSs associated with cell surface modification, with some having a predicted function related to modification of LPS (O antigen), potentially altering the serotype of the lysogenized cell.

***sopE* and *sodC* are the predominant virulence-associated CDSs carried on intact prophages**

Our analysis also identified CDSs that were found previously to play a role in virulence, namely, *sopE* and *sodC*. *sopE*, encoding a guanine nucleotide exchange factor, was identified on 44 different intact regions representing nine different prophages, while *sodC* was found on three different intact prophages from 29 subsp. *enterica* genomes, most commonly on regions annotated as Gifsy 2 ($n = 22$; 75.9%) but also on regions annotated as Gifsy 1 and Salmon Fels 1 (Data Set S3). *sodA*, encoding a superoxide dismutase with a function related to that of SodC (29), was detected on two different prophages (Yersin L 413C and Entero fIAA91 ss), neither of which was annotated as having *sodC*. The ADP-ribosylating toxin subunit encoded by *artA* or *pltA* was identified in five genomes on three different prophages (Salmon 118970 sal 3, Vibrio X29, and Gifsy 1; Data Set S3). Overall, these results suggest that genes encoding superoxide dismutases (SodC and, to a lesser extent, SodA) and SopE are the predominant virulence factors carried by intact prophages in *Salmonella*, although ArtAB and typhoid toxins can be horizontally acquired due to lysogenization with prophages.

While some intact prophages carry antimicrobial resistance genes, fewer than 3% do

Previous studies reported the presence of antimicrobial resistance genes (ARGs) on prophages, although the role of prophages in facilitating their transmission is contested (30, 31). EggNOG classification suggested the presence of three types of ARGs found on intact prophage regions: *acrB*, *mdtK*, and *bacA*, which were found within COG V (Defense Mechanisms). To more definitively characterize these and other ARGs, we executed a local blastn search to identify and extract ARGs on all prophage regions using reference sequences from the Comprehensive Antibiotic Resistance Database (CARD) (32). A total of 18 ARGs, including three that were predicted to be hypothetically disrupted coding sequences (HDCs), were identified on intact prophage regions in subsp. *enterica* genomes with 25 additional ARGs detected on incomplete or questionable prophage regions (Fig. 6A). The 15 intact ARGs on intact prophage regions encompassed three unique ARGs and one antimicrobial resistance mechanism type (Fig. 6B). Unique ARGs that were found more than once were not associated with a specific prophage, with the exception of TEM-212 where both observations were found on the same prophage within the same genome (Table 1), suggesting that, in most cases, different prophages transfer different ARGs. The distribution of the ARGs was independent of the phylogenetic clade of the genome (Fig. 6C), and none were identified in genomes from Clade D ($n = 4$) or Clade Unassigned ($n = 2$), though this may be a result of the smaller number of genomes included to represent these phylogenetic groups. Therefore, while prophages can carry ARGs, intact ARGs on intact prophage make up a small percentage (0.9%, $n = 15$) of accessory cargo, suggesting that prophage-mediated acquisition of ARGs may not be the primary mode of recent acquisition for *Salmonella*.

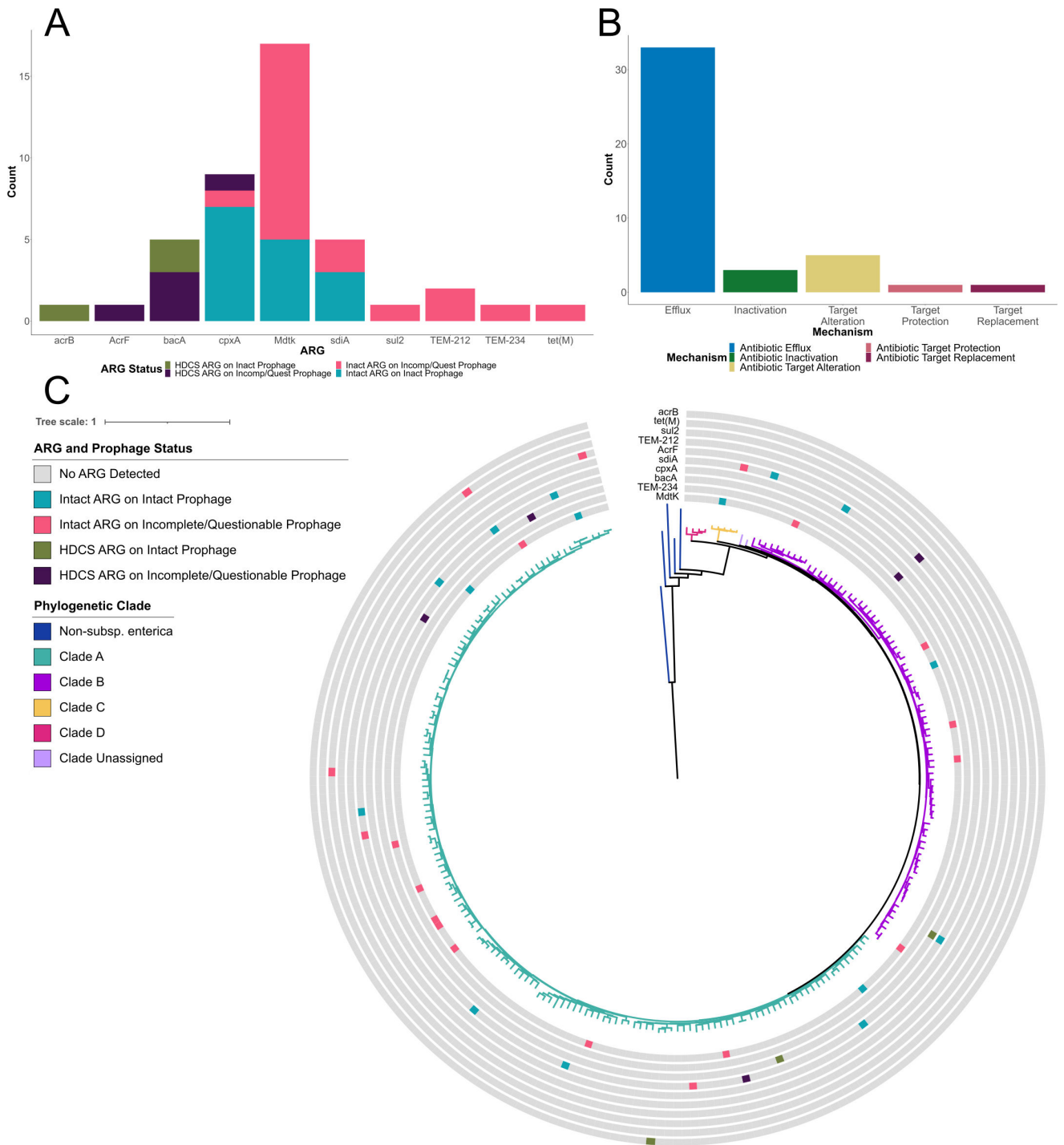


FIG 6 Analysis of ARGs found on intact, incomplete, and questionable prophage regions. The CARD database was used to identify and characterize ARGs ($n = 43$) on all prophage regions for the 242 *S. enterica* genomes. (A) Stacked bar plot shows the counts for the 10 unique AMR types identified. Colors indicate the status of the ARG (intact vs hypothetically disrupted coding sequence) and prophage (intact, incomplete, or questionable). (B) Bar plot shows all 43 ARGs (including those found on non-intact prophage regions) categorized by antimicrobial mechanism type. (C) Presence/absence distribution of ARG (colored [presence] or gray [absence] boxes) mapped onto the phylogenetic tree described in Fig. 2. ARGs are colored according to their status (intact or HDCS) and status of the prophage region they were found on (intact or incomplete/questionable).

TABLE 1 ARGs detected on intact, incomplete, and questionable prophage regions

ARG mechanism	CARD short name ^a	Count ^b	Proportion intact ARGs ^c	Associated prophages
Antibiotic efflux	MdtK	17	1	Salmon 118970 sal3, Cronob ENT39118, Gifsy 1, Clostr PhiS63, Edward GF 2, Erwini EtG, Lactob phiAQ113, Salmon vB SosS Oslo, Shigel Sf6
	cpxA	9	0.88	Yersin L 413C, Entero fiAA91 ss, Entero Wphi, Escher P2
	sdiA	5	1	Salmon SEN34, Entero mEp390, Entero SfV, Salmon Fels 2
	AcrF	1	0	Entero mEpX1
	acrB	1	0	Pseudo PaMx74
Antibiotic inactivation	TEM-234	1	1	Salmon SJ46
	TEM-212	2	1	Salmon SJ46
Antibiotic target alteration	bacA	5	0	Cronob ESSI 2, Entero 186, Escher 500465 1
Antibiotic target protection	tet(M)	1	1	Escher 500465 2
Antibiotic target replacement	sul2	1	1	Escher SH2026Stx1

^aComprehensive Antibiotic Resistance Database (32).

^bTotal count of intact and HDCS ARGs identified by CARD on intact and non-intact prophage regions. Categories of ARGs include intact ARG on intact prophage ($n = 15$), HDCS ARG on intact prophage ($n = 3$), intact ARG on non-intact prophage ($n = 20$), and HDCS ARG on non-intact prophage ($n = 5$).

^cARGs were classified as (i) intact if they did not have a premature stop codon and had a coverage >0.8 or (ii) hypothetically disrupted coding sequences if they had a premature stop codon or a coverage of <0.8 . The number of intact ARGs detected for each antibiotic resistance ontology term was divided by the total number of ARGs detected for that term.

DISCUSSION

Mobile genetic elements, such as prophages and plasmids, are key mediators of horizontal gene transfer in bacteria (18, 33). Recent characterizations of prophage regions in diverse serovars have offered insight into the evolutionary history of *Salmonella* and have identified the potential role of prophage cargo in mediating *Salmonella*'s interactions with its many hosts (13, 14, 34). The work presented here demonstrates that prophage regions across *Salmonella* spp. are considerably diverse and facilitate the transfer of CDSs associated with a variety of functions. While we identified a number of CDSs representing traditional virulence factors (SodC, SodA, and SopE), our analyses also identified a number of CDSs with a predicted function related to cell surface structure modification, heavy metal resistance, and metabolism of a variety of compounds, which may play an important role in facilitating *Salmonella*'s host–pathogen interactions.

Many intact prophages are broadly distributed throughout the *Salmonella* genus

Previous comparative genomic studies identified differences in prophage content as one of the key differences between clades within a serovar (14, 35, 36). Nearly one-third of the prophages found within our *S. enterica* subsp. *enterica* genomes were also found among genomes from *S. bongori* and *S. enterica* subsp. *salamae*, *arizonae*, *diarizonae*, *houtena*, *indica*, and *londinensis*. While serovars of *S. enterica* subsp. *enterica* cause roughly 99% of human salmonellosis (37), analyses focused on *S. bongori* and other *S. enterica* non-*enterica* subsp. will be beneficial for understanding the evolution of *Salmonella*'s prophage repertoire and frequency of mobilization of different prophages, and, more importantly, the cargo that they carry.

Prophages carry cargo with a diverse range of functions

Prophages are known to facilitate acquisition of toxins in many pathogenic bacteria, including the Shiga toxin encoded by Shiga toxin-producing *Escherichia coli* (STEC), botulinum neurotoxins BoNTX in *Clostridium botulinum*, and the cholera toxin in *Vibrio cholerae* (38–40). Previous comparative genomic studies of prophages in *Salmonella* established that prophages can facilitate the transfer of virulence factors, with the most well-studied examples being *sodC*, encoding a superoxide dismutase, and *sopE*, encoding a guanine nucleotide exchange factor associated with membrane ruffling

in infected eukaryotic cells (16, 41). In our data set, we identified *sopE* and *sodC* on intact prophage regions in 39 and 29 subsp. *enterica* genomes, respectively. This further demonstrates that *sopE* and *sodC* can be carried on multiple prophages and are found broadly across subsp. *enterica* lineages. We also identified ArtAB toxin and one instance of typhoid toxin being found on different intact prophages in five serovars from subsp. *enterica* clades A and B, confirming that non-DT104 lineages can harbor prophages encoding *artAB* as well (42–44). A recent analysis of ArtB toxin subtypes suggested that *artB* is encoded on a variety of prophages (45). Together, these findings continue to challenge our understanding of the potential role of phage-encoded toxins in *Salmonella*'s pathogenesis and virulence potential.

In addition to these traditional virulence factors, we also identified a number of CDSs that may facilitate metabolic adaptations that could provide some fitness benefit allowing *Salmonella* to propagate in a new host or environment. Previous studies demonstrated that CDSs within prophage regions can affect host metabolic activity, such as the *dgo* operon in facilitating D-galactonate metabolism and *pckA* involved in gluconeogenesis (13). The broad range of functional cargo identified here suggests that prophages in *Salmonella* may confer a wide variety of adaptations that might allow lysogenized strains to persist or survive in different environmental conditions.

Prophages likely played an important role in the evolution of *Salmonella* surface antigens

Past studies have shown that prophages played an important role in facilitating horizontal acquisition of surface-modifying enzymes, such as glycosyltransferases (13, 46). Prophages are known to facilitate horizontal acquisition of CDSs that modify the O antigen, thereby preventing superinfection (infection with other phages that might use the same receptor) (47, 48). In our data set, we found that multiple unique intact prophages carry CDSs that are annotated as modifying the O antigen (e.g., *gtrABC*). Variation in the O antigen due to prophage-mediated acquisition of surface-modifying enzymes (e.g., glycosyltransferases, acetyltransferases) has an impact on the resulting O antigen structure. Indeed, lysogenization with P22 results in glucosylation of the galactose subunit of the basal O antigen, resulting in recognition with O1 antiserum (representing the “1” O antigen in the White–Kauffman–Le Minor serotyping scheme) (5, 28). The *gtrABC* operon is known to glycosylate different moieties on sugars present at the terminal end of the LPS (the “O antigen”). Analyses of the O antigen diversity suggest that *gtrABC* carried by prophages may account for the structural modifications of the O antigen in antigen group O:6,14 (48). Our results support that in addition to P22, other prophages identified by Phaster as Enterovirus UAB Phi20, Salmon ST160, and Salmon ST64T encode *gtrABC* (49). The gain and loss of prophage cargo, or more specifically the cell surface-modifying cargo that they carry, could be one possible explanation for the emergence of so-called polyphyletic serovars, which has important implications for surveillance efforts that rely on serological typing (4, 50). Although our analyses did not identify any significant associations between the presence of intact prophages and specific O antigen groups, some prophages (such as Salmon vB Sosl Oslo) do carry cargo that is involved in the synthesis of Kdo₂-lipid A (51). Additional studies using larger numbers of genomes of representative isolates expressing the same serotype would be helpful in further identifying prophages whose cargo result in conversion to a different serotype.

Surprisingly, Phaster identified 20 CDSs for the flagellin FljB subunit (initially identified as FliC by EggNOG) on five different intact prophages; all but one of these CDSs mapped to the ends of the prophage region. While these results may suggest that prophages can carry the phase 2 flagella locus, it is also probable that the attachment site of these prophages is adjacent to the *fljAB-hin* locus. Indeed, Soyer et al. hypothesized that imprecise excision of a prophage adjacent to *fljAB* and *hin* led to the loss of the phase 2 antigen in several lineages of monophasic Typhimurium (also known as I 4,[5],12:i:-) (52). Transfer of *fliC* from an *S. Montevideo* strain into the ancestor of *S. Mbandaka*

is hypothesized to have given rise to serovar Lubbock, although the authors did not specify whether this horizontal event was mediated by prophages (53). Together, our observation of the *fljAB-hin* locus being present on a prophage or directly adjacent to a prophage, and other reports of prophage excision potentially leading to the loss of this locus in monophasic serovars, suggests that prophages may also influence *Salmonella* flagellar antigens, although experimental data are needed to definitively assert this.

Our analyses suggest that prophages and their cargo contributed importantly to the diversification of *Salmonella* spp. Experimental analyses examining whether the common intact prophages identified here can indeed be excised will be helpful for understanding the transmission of the CDSs identified in this study and their potential role in facilitating *Salmonella*'s interactions with its hosts and environments. Evolutionary analyses estimating the gain and loss events that led to the distribution of prophages throughout *Salmonella* will enable a more complete understanding of the role that these phages played and continue to play in shaping *Salmonella*'s genetic diversity. Finally, analyses characterizing the conservation and loss of these prophages among genomes within a serotype or standardizing prophage nomenclature are needed, especially in light of emerging evidence of "hypervirulent" subtypes within a serotype and the potential use of virulence markers in serovar-independent risk assessments (9). Overall, through characterization of the diversity and distribution of prophages, the entangled role of prophages in driving the interactions and evolution of *Salmonella* is beginning to be unpacked.

MATERIALS AND METHODS

Genome selection criteria

Genomic assemblies for a previously reported set of 217 *S. enterica* subsp. *enterica* serovars, represented by 242 genomes, were analyzed in this study (Data Set S1) (21). Serovar selection criteria included (i) human case prevalence in the United States from 2006 to 2016, (ii) nonhuman clinical and animal non clinical case prevalence in 2012, and (iii) novel or rare clades of *S. enterica* subsp. *enterica* as outlined in a prior study (21). Genomes within these serovars were then selected based on prevalence of whole-genome sequence (WGS) data availability using the National Center for Biotechnology Information (NCBI) Pathogen Detection Database to identify the most prevalent single-nucleotide polymorphism (SNP) cluster (<https://www.ncbi.nlm.nih.gov/pathogens/>). For serovars that were previously described as paraphyletic or polyphyletic clades (4), a single genome was chosen per clade. Within each SNP cluster, genomes were selected based on their having (i) the highest N_{50} value among isolates in the SNP cluster, (ii) their being sequenced on an Illumina platform (required for the various pipelines used), and (iii) their having a genome size of 4.5–5.7 million base pairs (expected genome size for *Salmonella* spp. isolates). Additional genomes representing *S. enterica* subspecies *salamae*, *arizonae*, *diarizonae*, *houtenae*, and *indica* (one genome per subspecies) were included for inferring phylogenetic structure. Assemblies for selected genomes were downloaded from the NCBI Pathogen Detection Database; all genomes were assembled with SKESA version 2.2 (54) and annotated with PGAP version 0.2.0 (55). Serotype information was confirmed with the *Salmonella in silico* typing resource (SISTR) version 1.0.2 (56).

Identification of core genome SNPs, inference of core genome phylogeny, and clade classification

Clade designations for *S. enterica* subsp. *enterica* genomes were assigned based on phylogenies inferred from core SNPs. kSNP version 3.1 with a kmer size of 19 was used to identify and align 13,926 core SNPs in all genomes (57). IQ-TREE v.2.0.7 was used to infer phylogenies using the model GTR + G + ASC (general time reversible with gamma distribution and ascertainment bias correction) and 1,000 ultra-fast (UF)

bootstraps repetitions (58, 59). Clades were assigned letters A–D as defined previously by Worley et al. (4). For two serovars (Lattenkamp and Poano), the phylogeny inferred from core SNPs could not be used to definitively assign these genomes to a clade, and therefore, they are presented as “clade unassigned” in the data set as they potentially belong to a novel phylogenetic clade (4).

Identification and characterization of prophage regions

Phaster (<https://phaster.ca/>) was used with default settings to identify, extract, and annotate prophage regions from all *Salmonella* spp. genomes (22, 60). Phaster uses three levels (“intact,” “incomplete,” and “questionable”) to rate the completeness of the prophage region that it identifies (22, 60). The output summary annotated prophage regions as intact, incomplete, or questionable and included the common phage name, region length, total protein number and phage hit protein number. The common intact prophages (defined here as presence in >5 subsp. *enterica* genomes) were identified using the phage name assigned by Phaster, which is based on blast scores obtained by comparing the prophage region to a database of bacteriophage genomes. Prophages were counted by their presence in the genome, not by number of regions, due to there being 19 genomes that each contained two regions annotated as the same prophage (i.e., two regions annotated as Gifsy 1 in the same assembly).

Characterization of prophages from non-subspecies *enterica* genomes

Assemblies for 61 genomes representing *S. bongori* and *S. enterica* subsp. II, IIIa, IIIb, IV, VI, and VII were included (for accession numbers see Data Set S1). All available *S. bongori* assemblies available on NCBI ($n = 32$; date accessed, 15 July 2021) were downloaded. One assembly was excluded due to poor sequence quality (GCA_003452365.1, $n = 446$ contigs), and therefore, a total of 31 assemblies representing 10 unique serotypes were included in the final data set. Representative assemblies for the remaining *S. enterica* subspecies were selected based on the number of isolates in each NCBI Pathogen Detection Database SNP cluster (date accessed, 15 July 2021). Assemblies with the highest N_{50} values from each of the top five most populous SNP clusters representing unique serotypes were selected for each subspecies (i.e., II, IIIa, IIIb, IV, and VI); for subspecies VII, due to the small number of assemblies available, selecting unique serotypes was not possible, and therefore, four of the five selected assemblies have an identical serotype. Prophage regions were characterized using Phaster following the same methods for the *S. enterica* subsp. *enterica* genomes, and the presence of the common prophages identified among intact *S. enterica* subsp. *enterica* genomes were compared in the non-subsp. *enterica* genomes.

Characterization of coding sequences encoded on intact *S. enterica* subsp. *enterica* prophage regions

CDSs for intact prophage regions from *S. enterica* subsp. *enterica* were identified and annotated with Prokka (61). To assign a putative function, CDSs were then annotated using EggNOG-mapper v2 (<http://eggnog-mapper.embl.de/>) with percent identity, query coverage, and subject coverage set to 60% (25, 62); a threshold of 60% was determined based on results from a preliminary analysis using intact prophage regions from five randomly selected genomes in the data set (Data Set S3). EggNOG output for the 206 subsp. *enterica* genomes with intact prophage regions detected by Phaster was manually curated using the following steps (note that all raw data are available in Data Set S3). First, CDSs annotated as COGs “-” (representing CDSs with “no COG category assigned”) or “S” (representing CDSs with “Function Unknown”) were removed ($n = 16,126$ CDSs), producing the curated data set of 5,561 CDSs. Second, the unique “description” terms for each CDS within each COG were identified from EggNOG output within the curated data set. The “description” annotation was chosen for the primary analysis from the annotations provided due to its completeness across the curated data set. Third, individual CDSs were classified as “ambiguous,” “core,” or “accessory”; core CDSs were those with

(i) COG definitions for COG X (mobilome), (ii) terms related to the prophage life cycle and maintenance consistent with those considered phage by Phaster (e.g., capsid, head, integrase, plate, tail, fiber, coat, terminase, portal, protease, and lysin), or (iii) were annotated as phage related by any EggNOG annotation (see Data Set S3 for full list of EggNOG definitions that we considered “core” CDSs) (22). The remaining CDSs were defined as “ambiguous” or “accessory” depending on their EggNOG annotations. Ambiguous CDSs contained EggNOG annotations where the functional role of the CDS was unclear due to the following: (i) more than two of the 13 annotation categories being incomplete, (ii) functional annotations being inconsistent, or (iii) CDSs being described as putative, hypothetical, locus tag, or domain of unknown function. All other CDSs were considered accessory CDSs (see Data Set S3 for a complete list of all CDSs and their assigned category). For *gtrC* (accession number NC_004348.1) and *sopE* (accession number AF153829.1), a blastn search was executed to confirm the identity and presence of these CDSs since they are known to be found on intact prophages but were not annotated as “GtrC or SopE” by EggNOG.

Identification of COG X CDSs

The NCBI Database of COGs classifies prophage and transposon terms into COG X (Mobilome). A manual analysis was performed to determine the presence of any COG Definitions within the CDS data that were annotated as a different COG category by EggNOG. CDSs were considered COG X if the “eggNOG_Ogs” contained a COG X Definition as defined by NCBI (<https://www.ncbi.nlm.nih.gov/research/cog/cogcategory/X/>; date accessed, 10 September 2023). Of the 12,687 total CDSs annotated by EggNOG, 4,385 were defined as COG X. During the analysis of genetic content within prophage regions, any COG X terms were considered “core CDSs” (Data Set S3).

Identification of putatively prophage adjacent CDSs

Phaster additionally searches all prophage regions for a CDS with an annotated function “integrase” for *attB* sites, which are short nucleotide repeats that the software predicts to be the attachment site where the phage genome integrated into the bacterial genome. Due to the methods implemented by Phaster to identify *attB* sites, we also reasoned that for some prophage genomes, it is possible that CDSs adjacent to the prophage regions could potentially be misclassified as being located on a prophage. The locations of CDSs within the curated data set belonging to the common 73 accessory CDSs were mapped using data from the EggNOG annotations of intact prophage regions identified by Phaster (22, 25). For each intact region, the total number of CDSs per region was identified, and each CDS was assigned a location in the region based on the query number from EggNOG output. For an individual CDS, the location (based on the query number) was divided by the total number of CDSs in the region. For each predicted CDS type, the number of CDSs with location values ≤ 0.1 or ≥ 0.9 (arbitrarily selected to represent the first 10% or last 10% of the prophage region) were divided by the total number of counts for the predicted CDS to get an overall location value. Predicted CDSs with an overall location ≥ 0.9 were considered to be “putatively prophage adjacent CDSs” as over 90% of the observations for that predicted CDS were in the first 10% or last 10% of the region (i.e., at the “ends” of region).

Analysis of ARGs in prophage regions

All prophage regions identified in *S. enterica* subsp. *enterica* genomes were compared to CARD databases for the detection of ARGs using blastn (32). Genes were classified into putative functional (no premature stop codon and has a coverage >0.8) (intact), non-functional (premature stop codon or $0.3 < \text{coverage} < 0.8$) (hypothetically disrupted coding sequences), and absence (no hits or coverage <0.3) (63). Hits were removed if

they (i) were annotated as “absent,” or (ii) the gene coverage was not complete. In the case of duplicates, the hit with the best bitscore was selected.

Data analysis and visualization

Data analysis and statistics were conducted using R Statistical Software version 4.1.2 (64). The tidyverse package version 2.0.0 and readxl package version 1.4.2 were used for data analysis (65, 66). The RVAideMemoire package version 0.9.83 was used for Fisher’s exact statistical tests (67). The ggplot2 package version 3.4.4 (68) and the Interactive Tree Of Life (iTOL) (69) were used for data visualization.

AUTHOR AFFILIATIONS

¹Department of Food Science and Technology, Virginia Tech, Blacksburg, Virginia, USA

²Computational Modeling and Data Analytics Program, Virginia Tech, Blacksburg, Virginia, USA

³Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, Virginia, USA

AUTHOR ORCID*s*

Rachel A. Cheng  <http://orcid.org/0000-0002-5932-7011>

FUNDING

Funder	Grant(s)	Author(s)
U.S. Department of Agriculture (USDA)	2020-67034-31905	Rachel A. Cheng

DATA AVAILABILITY

Sequence data are publicly available in the NCBI database; see Data Set S1 for accession numbers. Data for this article can be accessed in the supplemental material (Data Sets S1 to S3).

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Data Set S1 (mSphere00031-24-s0001.xlsx). Isolate and prophage metadata.

Data Set S2 (mSphere00031-24-s0002.xlsx). Table of *P*-values for analysis of presence of specific prophages within O antigen groups.

Data Set S3 (mSphere00031-24-s0003.xlsx). Tables from analysis and classification of CDSs within intact regions.

Figure S1

(mSphere00031-24-s0004.docx). Summary of prophage regions among 303 *Salmonella* spp. genomes

Figure S2 (mSphere00031-24-s0005.docx). Proportion of COG categories for 5,561 CDSs by phylogenetic clade.

REFERENCES

- Majowicz SE, Musto J, Scallan E, Angulo FJ, Kirk M, O’Brien SJ, Jones TF, Fazil A, Hoekstra RM, ftiCoEDBoI S. 2010. The global burden of nontyphoidal *Salmonella* gastroenteritis. *Clin Infect Dis* 50:882–889. <https://doi.org/10.1086/650733>
- Stanaway JD, Parisi A, Sarkar K, Blacker BF, Reiner RC, Hay SI, Nixon MR, Dolecek C, James SL, Mokdad AH. 2019. The global burden of nontyphoidal salmonella invasive disease: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Infect Dis* 19:1312–1324. [https://doi.org/10.1016/S1473-3099\(19\)30418-9](https://doi.org/10.1016/S1473-3099(19)30418-9)
- Pearce ME, Langridge GC, Lauer AC, Grant K, Maiden MCJ, Chattaway MA. 2021. An evaluation of the species and subspecies of the genus *Salmonella* with whole genome sequence data: proposal of type strains and epithets for novel *S. enterica* subspecies VII, VIII, IX, X and XI. *Genomics* 113:3152–3162. <https://doi.org/10.1016/j.ygeno.2021.07.003>

4. Worley J, Meng J, Allard MW, Brown EW, Timme RE. 2018. *Salmonella enterica* phylogeny based on whole-genome sequencing reveals two new clades and novel patterns of horizontally acquired genetic elements. *mBio* 9:e02303-18. <https://doi.org/10.1128/mBio.02303-18>
5. Issenhuth-Jeanjean S, Roggentin P, Mikoleit M, Guibourdenche M, de Pinna E, Nair S, Fields PI, Weill F-X. 2014. Supplement 2008-2010 (no. 48) to the White-Kauffmann-Le minor scheme. *Res Microbiol* 165:526–530. <https://doi.org/10.1016/j.resmic.2014.07.004>
6. Virginia Department of Health. 2016. Descriptive epidemiology of reportable diseases and selected maps, salmonellosis. Virginia Department of Health
7. Harvey RR, Friedman CR, Crim SM, Judd M, Barrett KA, Tolar B, Folster JP, Griffin PM, Brown AC. 2017. Epidemiology of *Salmonella enterica* serotype Dublin infections among humans, united states, 1968-2013. *Emerg Infect Dis* 23:1493–1501. <https://doi.org/10.3201/eid2309.170136>
8. Khajanchi BK, Foley SL. 2022. Antimicrobial resistance and increased virulence of *Salmonella*. *Microorganisms* 10:1829. <https://doi.org/10.3390/microorganisms10091829>
9. Chen R, Cheng RA, Wiedmann M, Orsi RH. 2022. Development of a genomics-based approach to identify putative hypervirulent nontyphoidal *Salmonella* isolates: *Salmonella enterica* serovar Saintpaul as a model. *mSphere* 7:e0073021. <https://doi.org/10.1128/msphere.00730-21>
10. Shah DH, Casavant C, Hawley Q, Addwebi T, Call DR, Guard J. 2012. *Salmonella enteritidis* strains from poultry exhibit differential responses to acid stress, oxidative stress, and survival in the egg albumen. *Foodborne Pathog Dis* 9:258–264. <https://doi.org/10.1089/fpd.2011.1009>
11. Stone E, Campbell K, Grant I, McAuliffe O. 2019. Understanding and exploiting phage–host interactions. *Viruses* 11:567. <https://doi.org/10.3390/v11060567>
12. Howard-Varona C, Hargreaves KR, Abedon ST, Sullivan MB. 2017. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *ISME J* 11:1511–1520. <https://doi.org/10.1038/ismej.2017.16>
13. Wahl A, Battesti A, Ansaldi M. 2019. Prophages in *Salmonella enterica*: a driving force in reshaping the genome and physiology of their bacterial host? *Mol Microbiol* 111:303–316. <https://doi.org/10.1111/mmi.14167>
14. Mottawea W, Duceppe M-O, Dupras AA, Usongo V, Jeukens J, Freschi L, Emond-Rheault J-G, Hamel J, Kukavica-Ibrulj I, Boyle B, Gill A, Burnett E, Franz E, Arya G, Weadge JT, Gruenheid S, Wiedmann M, Huang H, Daigle F, Moineau S, Bekal S, Levesque RC, Goodridge LD, Ogunremi D. 2018. *Salmonella enterica* prophage sequence profiles reflect genome diversity and can be used for high discrimination subtyping. *Front Microbiol* 9:836. <https://doi.org/10.3389/fmicb.2018.00836>
15. Sly LM, Guiney DG, Reiner NE. 2002. *Salmonella enterica* serovar Typhimurium periplasmic superoxide dismutases SodCI and SodCII are required for protection against the phagocyte oxidative burst. *Infect Immun* 70:5312–5315. <https://doi.org/10.1128/IAI.70.9.5312-5315.2002>
16. Figueroa-Bossi N, Bossi L. 1999. Inducible prophages contribute to *Salmonella* virulence in mice. *Mol Microbiol* 33:167–176. <https://doi.org/10.1046/j.1365-2958.1999.01461.x>
17. Canchaya C, Proux C, Fournous G, Bruttin A, Brüssow H. 2003. Prophage genomics. *Microbiol Mol Biol Rev* 67:238–276. <https://doi.org/10.1128/MMBR.67.2.238-276.2003>
18. Zhang Y, Miao J, Zhang N, Wang X, Li Z, Richard OA, Li B. 2023. The analysis of the function, diversity, and evolution of the *Bacillus* phage genome. *BMC Microbiol*. 23:1–15. <https://doi.org/10.1186/s12866-023-02907-9>
19. McCarthy AJ, Witney AA, Lindsay JA. 2012. *Staphylococcus aureus* temperate bacteriophage: carriage and horizontal gene transfer is lineage associated. *Front Cell Infect Microbiol* 2:6. <https://doi.org/10.3389/fcimb.2012.00006>
20. Frazão N, Sousa A, Lässig M, Gordo I. 2019. Horizontal gene transfer overrides mutation in *Escherichia coli* colonizing the mammalian gut. *Proc Natl Acad Sci U S A* 116:17906–17915. <https://doi.org/10.1073/pnas.1906958116>
21. Cheng RA, Orsi RH, Wiedmann M. 2022. The number and type of chaperone-usher fimbriae reflect phylogenetic clade rather than host range in *Salmonella*. *mSystems* 7:e00115–22. <https://doi.org/10.1128/mSystems.00115-22>
22. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 44:W16–W21. <https://doi.org/10.1093/nar/gkw387>
23. Salgado CJ, Zayas M, Villafane R. 2004. Homology between two different *Salmonella* phages: *Salmonella enterica* serovar Typhimurium phage P22 and *Salmonella enterica* serovar Anatum b var. 15 + phage^{ε34}. *Virus Genes* 29:87–98. <https://doi.org/10.1023/B:VIRU.0000032792.86188.fb>
24. Cheng RA, Orsi RH, Wiedmann M. 2021. Phylogeographic clustering suggests that distinct clades of *Salmonella enterica* serovar Mississippi are endemic in Australia. *mSphere* 6:e00485–21. <https://doi.org/10.1128/mSphere.00485-21>
25. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-Mapper V2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol* 38:5825–5829. <https://doi.org/10.1093/molbev/msab293>
26. Labrie SJ, Samson JE, Moineau S. 2010. Bacteriophage resistance mechanisms. *Nat Rev Microbiol* 8:317–327. <https://doi.org/10.1038/nrmicro2315>
27. Bondy-Denomy J, Qian J, Westra ER, Buckling A, Guttman DS, Davidson AR, Maxwell KL. 2016. Prophages mediate defense against phage infection through diverse mechanisms. *ISME J* 10:2854–2866. <https://doi.org/10.1038/ismej.2016.79>
28. Davies MR, Broadbent SE, Harris SR, Thomson NR, van der Woude MW. 2013. Horizontally acquired glycosyltransferase operons drive salmonellae lipopolysaccharide diversity. *PLoS Genet* 9:e1003568. <https://doi.org/10.1371/journal.pgen.1003568>
29. Tsolis RM, Bäumlér AJ, Heffron F. 1995. Role of *Salmonella typhimurium* Mn-superoxide dismutase (SodA) in protection against early killing by J774 macrophages. *Infect Immun* 63:1739–1744. <https://doi.org/10.1128/iai.63.5.1739-1744.1995>
30. Chen Q, Dharmaraj T, Cai PC, Burgener EB, Haddock NL, Spakowitz AJ, Bollyky PL. 2022. Bacteriophage and bacterial susceptibility, resistance, and tolerance to antibiotics. *Pharmaceutics* 14:1425. <https://doi.org/10.3390/pharmaceutics14071425>
31. Enault F, Briet A, Bouteille L, Roux S, Sullivan MB, Petit M-A. 2017. Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. *ISME J* 11:237–247. <https://doi.org/10.1038/ismej.2016.90>
32. Alcock BP, Huynh W, Chaili R, Smith KW, Raphenya AR, Wlodarski MA, Edalatmand A, Petkau A, Syed SA, Tsang KK, et al. 2023. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the comprehensive antibiotic resistance database. *Nucleic Acids Res* 51:D690–D699. <https://doi.org/10.1093/nar/gkac920>
33. Bobay L-M, Rocha EPC, Touchon M. 2013. The adaptation of temperate bacteriophages to their host genomes. *Mol Biol Evol* 30:737–751. <https://doi.org/10.1093/molbev/mss279>
34. Gao R, Naushad S, Moineau S, Levesque R, Goodridge L, Ogunremi D. 2020. Comparative genomic analysis of 142 bacteriophages infecting *Salmonella enterica* subsp. *enterica*. *BMC Genomics* 21:374. <https://doi.org/10.1186/s12864-020-6765-z>
35. Yu J, Xu X, Wang Y, Zhai X, Pan Z, Jiao X, Zhang Y. 2022. Prophage-mediated genome differentiation of the *Salmonella* Derby ST71 population. *Microb Genom* 8:000817. <https://doi.org/10.1099/mgen.0.000817>
36. Drahovská H, Mikasová E, Szemes T, Ficek A, Sásik M, Majtán V, Turna J. 2007. Variability in occurrence of multiple prophage genes in *Salmonella* Typhimurium strains isolated in Slovak Republic. *FEMS Microbiol Lett* 270:237–244. <https://doi.org/10.1111/j.1574-6968.2007.00674.x>
37. Brenner FW, Villar RG, Angulo FJ, Tauxe R, Swaminathan B. 2000. *Salmonella* nomenclature. *J Clin Microbiol* 38:2465–2467. <https://doi.org/10.1128/JCM.38.7.2465-2467.2000>
38. Krüger A, Lucchesi PMA. 2015. Shiga toxins and stx phages: highly diverse entities. *Microbiology (Reading)* 161:451–462. <https://doi.org/10.1099/mic.0.000003>
39. Sakaguchi Y, Hayashi T, Kurokawa K, Nakayama K, Oshima K, Fujinaga Y, Ohnishi M, Ohtsubo E, Hattori M, Oguma K. 2005. The genome sequence of *Clostridium botulinum* type C neurotoxin-converting phage and the molecular mechanisms of unstable lysogeny. *Proc Natl Acad Sci U S A* 102:17472–17477. <https://doi.org/10.1073/pnas.0505503102>
40. Waldor MK, Mekalanos JJ. 1996. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* 272:1910–1914. <https://doi.org/10.1126/science.272.5270.1910>

41. Hardt WD, Chen LM, Schuebel KE, Bustelo XR, Galán JE. 1998. *S. typhimurium* encodes an activator of Rho GTPases that induces membrane ruffling and nuclear responses in host cells. *Cell* 93:815–826. [https://doi.org/10.1016/s0092-8674\(00\)81442-7](https://doi.org/10.1016/s0092-8674(00)81442-7)
42. Uchida I, Ishihara R, Tanaka K, Hata E, Makino SI, Kanno T, Hatama S, Kishima M, Akiba M, Watanabe A, Kubota T. 2009. *Salmonella enterica* serotype Typhimurium DT104 ArtA-dependent modification of pertussis toxin-sensitive G proteins in the presence of [32P]NAD. *Microbiology (Reading)* 155:3710–3718. <https://doi.org/10.1099/mic.0.028399-0>
43. Saitoh M, Tanaka K, Nishimori K, Makino SI, Kanno T, Ishihara R, Hatama S, Kitano R, Kishima M, Sameshima T, Akiba M, Nakazawa M, Yokomizo Y, Uchida I. 2005. The artAB genes encode a putative ADP-ribosyltransferase toxin homologue associated with *Salmonella enterica* serovar Typhimurium DT104. *Microbiology (Reading)* 151:3089–3096. <https://doi.org/10.1099/mic.0.27933-0>
44. Carroll LM, Piacenza N, Cheng RA, Wiedmann M, Guldemann C. 2023. A multidrug-resistant *Salmonella enterica* Typhimurium DT104 complex lineage circulating among humans and cattle in the USA lost the ability to produce pertussis-like toxin ArtAB. *Microb Genom* 9:mgen001050. <https://doi.org/10.1099/mgen.0.001050>
45. Ojajakor A, Gibbs RN, Chen Z, Gao X, Fowler CC. 2022. The evolutionary diversification of the *Salmonella artAB* toxin locus. *Front Microbiol* 13:1016438. <https://doi.org/10.3389/fmicb.2022.1016438>
46. Trofeit L, Sattler E, Künz J, Hilbert F. 2023. *Salmonella* prophages, their propagation, host specificity and antimicrobial resistance gene transduction. *Antibiotics (Basel)* 12:595. <https://doi.org/10.3390/antibiotics12030595>
47. Rakhuba DV, Kolomiets EI, Dey ES, Novik GI. 2010. Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell. *Pol J Microbiol* 59:145–155.
48. Liu B, Knirel YA, Feng L, Perepelov AV, Senchenkova SN, Reeves PR, Wang L. 2014. Structural diversity in *Salmonella* O antigens and its genetic basis. *FEMS Microbiol Rev* 38:56–89. <https://doi.org/10.1111/1574-6976.12034>
49. Kintz E, Davies MR, Hammarlöf DL, Canals R, Hinton JCD, van der Woude MW. 2015. A BTP1 prophage gene present in invasive non-typhoidal *Salmonella* determines composition and length of the O-antigen of the lipopolysaccharide. *Mol Microbiol* 96:263–275. <https://doi.org/10.1111/mmi.12933>
50. Cheng RA, Eade CR, Wiedmann M. 2019. Embracing diversity: differences in virulence mechanisms, disease severity, and host adaptations contribute to the success of nontyphoidal *Salmonella* as a foodborne pathogen. *Front Microbiol* 10:1368. <https://doi.org/10.3389/fmicb.2019.01368>
51. Raetz CRH, Reynolds CM, Trent MS, Bishop RE. 2007. Lipid A modification systems in Gram-negative bacteria. *Annu Rev Biochem* 76:295–329. <https://doi.org/10.1146/annurev.biochem.76.010307.145803>
52. Soyer Y, Moreno Switt A, Davis MA, Maurer J, McDonough PL, Schoonmaker-Bopp DJ, Dumas NB, Root T, Warnick LD, Gröhn YT, Wiedmann M. 2009. *Salmonella enterica* serotype 4,5,12:i:-, an emerging *Salmonella* serotype that represents multiple distinct clones. *J Clin Microbiol* 47:3546–3556. <https://doi.org/10.1128/JCM.00546-09>
53. Bugarel M, Cook PW, den Bakker HC, Harhay D, Nightingale KK, Loneragan GH. 2019. Complete genome sequences of four *Salmonella enterica* strains (including those of serotypes Montevideo Mbandaka, and Lubbock) isolated from peripheral lymph nodes of healthy cattle. *Microbiol Resour Announc* 8:e01450-18. <https://doi.org/10.1128/MRA.01450-18>
54. Souvorov A, Agarwala R, Lipman DJ. 2018. SKESA: strategic k-mer extension for scrupulous assemblies. *Genome Biol.* 19:153. <https://doi.org/10.1186/s13059-018-1540-z>
55. Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, Badretdin A, Coulouris G, Chitsaz F, Derbyshire MK, Durkin AS, Gonzales NR, Gwartz M, Lanczycki CJ, Song JS, Thanki N, Wang J, Yamashita RA, Yang M, Zheng C, Marchler-Bauer A, Thibaud-Nissen F. 2021. RefSeq: expanding the prokaryotic genome annotation pipeline reach with protein family model curation. *Nucleic Acids Research* 49:D1020–D1028. <https://doi.org/10.1093/nar/gkaa1105>
56. Yoshida CE, Kruczkiewicz P, Laing CR, Lingohr EJ, Gannon VPJ, Nash JHE, Taboada EN. 2016. The *Salmonella* in silico typing resource (SISTR): an open web-accessible tool for rapidly typing and subtyping draft salmonella genome assemblies. *PLoS One* 11:e0147101. <https://doi.org/10.1371/journal.pone.0147101>
57. Gardner SN, Slezak T, Hall BG. 2015. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics* 31:2877–2878. <https://doi.org/10.1093/bioinformatics/btv271>
58. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 37:1530–1534. <https://doi.org/10.1093/molbev/msaa015>
59. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol.* 35:518–522. <https://doi.org/10.1093/molbev/msx281>
60. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. *Nucleic Acids Res* 39:W347–52. <https://doi.org/10.1093/nar/gkr485>
61. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
62. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 47:D309–D314. <https://doi.org/10.1093/nar/gky1085>
63. Liao J, Guo X, Li S, Anupaju SMB, Cheng RA, Weller DL, Sullivan G, Zhang H, Deng X, Wiedmann M. 2023. Comparative genomics unveils extensive genomic variation between populations of *Listeria* species in natural and food-associated environments. *ISME Commun* 3:85. <https://doi.org/10.1038/s43705-023-00293-x>
64. Team RC. 2021. R: a language and environment for statistical computing, R foundation for statistical computing, Vienna, Austria. <https://www.R-project.org/>.
65. Wickham H, Bryan J. 2023. readxl: read excel files, V R package version 1.4.2. <https://CRAN.R-project.org/package=readxl>.
66. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemond G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K, Ooms J, Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. 2019. Welcome to the tidyverse. *JOSS* 4:1686. <https://doi.org/10.21105/joss.01686>
67. HERVE M. 2023. RVAideMemoire: testing and plotting procedures for biostatistics, vR package version 0.9-83. <https://CRAN.R-project.org/package=RVAideMemoire>.
68. Wickham H. 2016. ggplot2: elegant graphics for data analysis, vVersion 3.4.4. Springer-Verlag New York,.
69. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) V5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49:W293–W296. <https://doi.org/10.1093/nar/gkab301>