Infection, Genetics and Evolution 90 (2021) 104771

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Genomic signatures of adaptation to natural settings in non-typhoidal *Salmonella enterica* Serovars Saintpaul, Thompson and Weltevreden

J.R. Aguirre-Sanchez [a], J.R. Ibarra-Rodriguez [a], I.F. Vega-Lopez [b], J. Martínez-Urtaza [c], C. Chaidez-Quiroz [a,*]

[a] Centro de Investigación en Alimentación y Desarrollo, Culiacán, Mexico
[b] Parque de Innovación Tecnológica de la Universidad Autónoma de Sinaloa, Culiacán, Mexico
[c] Departement of Genetics and Microbiology, Universitat Autònoma de Barcelona, Barcelona, Spain

ARTICLE INFO

Keywords:
Salmonella
Comparative genomics
Sequencing
Environmental

ABSTRACT

Salmonella enterica is a pathogenic bacterium responsible for intestinal illness and systemic diseases such as typhoid and paratyphoid fevers. Among clinical manifestation classification, non-typhoidal *Salmonella* is mainly known as foodborne pathogen associated with the consumption of fecal contaminated food and water. Even though *Salmonella* hosts include humans and warm-blooded animals, it has been found in non-host environments as river water where the bacteria use different strategies to fitness the environment persisting and establishment. Now with the availability of WGS and bioinformatics tools, we can explore bacterial genomes with higher resolution to increase our understanding of specific genetic signatures among environmental and clinical isolates, being the goal of this work. Pangenome construction allowed the detection of specific environmental and clinical gene clusters related to metabolism and secretion systems as the main signature respectively. Specifically, D-galactonate degradation pathway was observed mainly in environmental genomes while T3SS and flagellum genes were detected for all clinical but not for all environmental isolates. Gene duplication and pseudogenes accumulation were detected as the main adaptation strategy for environmental isolates; thus, isolation source may play an important role in genome plasticity, conferring a selective advantage to survive and persist for environmental *Salmonella* isolates. Intact prophage sequences with cargo genes were observable for both isolation sources playing an important role in virulence contribution.

1. Introduction

Salmonella is a worldwide prevalent pathogenic bacterium with disease symptoms ranging from enterocolitis to typhoid fever (Stevens et al., 2009). It has been estimated that *Salmonella* causes around 1.35 million infections resulting in 420 deaths in the United States each year (CDC, 2018) and it's typically acquired through the consumption of contaminated food or water with fecal material (Wray and Wray, 2000).

Salmonella life cycle comprises the colonization of animals and humans, although it is also ubiquitous in natural settings where it can reside until reaching a new host. The journey of *Salmonella* begins when contaminated food is ingested and reaches a low-pH stomach receptacle. Those able to survive the stomach acidic conditions make their way to the small intestine specifically in the ileum where *Salmonella* starts colonizing enterocytes, M cells, and dendritic cells in the intestinal epithelium by gaining nutrients and space (Garai et al., 2012; Velge et al., 2012). After colonization takes place, *Salmonella* reproduces by binary fission to a large scale and the resulted clonal complex is released into the environment through fecal material to start the cycle again (Liu et al., 2018; Ranjbar et al., 2018).

Once excreted, *Salmonella* finds itself battling for survival facing limited nutrient availability, osmotic stress, variations in temperature, and pH (Winfield and Groisman, 2003). However, *Salmonella* is capable of sensing and responding to these stressful conditions. Otherwise, the bacterium could face potentially life-threatening conditions like those reported by Winfield and Groisman (2003) (Kenyon et al., 2002). These adverse conditions not only have a negative impact on *Salmonella* growth and survival but also influence virulence and resistance to several antimicrobials agents (Beceiro et al., 2013). Once the risk of perishing is overcome, the aquatic environment becomes the best niche for *Salmonella* to reside and reproduce (Raudales et al., 2014; Steele and Odumeru, 2004). Consequently, water becomes a major source of food

* Corresponding author at: CIAD, Culiacán, Carretera El Dorado Km 5.5, Campo el Diez, 80110 Culiacán, Sinaloa, Mexico.
E-mail address: chaqui@ciad.mx (C. Chaidez-Quiroz).

https://doi.org/10.1016/j.meegid.2021.104771
Received 27 August 2020; Received in revised form 4 February 2021; Accepted 11 February 2021
Available online 13 February 2021

contamination.

Several authors have proposed that *Salmonella* overcomes such stressful conditions by different survival strategies; for example, starvation-stress response (SSR) modulated by the scarcity of nutrients, carbon, and energy source shortage (C-starvation) (Spector, 1998; Spector and Cubitt, 1992). SSR has demonstrated that it accomplishes two main functions: combats the long-term effects of C-starvation and provides general resistance to other environmental stresses such as extreme temperature and pH (Kenyon et al., 2002). Two other conditions that exert protection to *Salmonella* in the environment are the formation of biofilm and the viable but nonculturable state. These conditions promote many cellular changes, including cell leakage, depletion of energy pools, altered expression of genes, and DNA replication (Morishige et al., 2017; Winfield and Groisman, 2003) increasing the survival rate in the environment.

The horizontal gene transfer (HGT) mediated by phages also allows *Salmonella* to diversify and survive in a variety of ecological niches by the incorporation of foreign genetic material into a recipient cell. This is known to be responsible for the acquisition of environmental, antimicrobial, and virulence genes playing an important role in the evolution of many bacterial species (Porwollik and McClelland, 2003).

Once *Salmonella* is settled down in the environment it can reach back human hosts causing food-borne diseases due to the consumption of fresh produce and animal products that come in contact with contaminated water (Greene et al., 2008; Hanning et al., 2009). Nevertheless, genetic differences between isolates from environmental and clinical settings have been poorly explored due to the limitations of the available techniques so far (Achtman et al., 2012; Page et al., 2017). The advent of whole-genome sequencing (WGS) technologies as well as the increasing number of bacterial genomes deposited in public repositories has enabled a deeper investigation to identify differential genomic signatures between strains isolated from distinct ecological niches and explore how genome plasticity and HGT are contributing to the survival of *Salmonella* outside the host (Balloux et al., 2018; Kwong et al., 2015).

Under this context, the present study aims to identify and characterize differences at a genomic level among *Salmonella enterica* isolated from environmental and clinical settings using different bioinformatics approaches and comparative genomics to gain a better understanding about the strategies of *Salmonella* to survive under the adverse conditions outside the host as a critical step in its life cycle and an intermediary stage to reach the human host and cause disease.

2. Materials and methods

2.1. Bacterial isolates and sequencing

A total of 34 *Salmonella enterica* strains isolated from Culiacan river and asymptomatic domestic animals belonging to serovars Saintpaul, Thompson, and Weltevreden (20, 6, and 8, respectively) were used. These strains belonged to the collection of the National Laboratory for Food Safety and Research (LANIIA) and were isolated in the course of previous studies (Jiménez et al., 2011, 2014). Additionally, 3 clinical genomes for each of the serovars Saintpaul (CFSAN004173, CFSAN004174 and CFSAN004175), Thompson (RM1984, PNUSAS019450 and PNUSAS019186) and Weltevreden (2511STDY5462388, 2511STDY5462413 and 25STDY5712385) were downloaded from National Center for Biotechnology Information (NCBI) (Yao et al., 2017). DNA extraction for strains from LANIIA's private collection was performed using the kit DNA easy Blood & Tissue Culture according to the manufacturer's instructions. After extraction, libraries were prepared with the Nextera XT DNA sample kit, and genomes were sequenced using Illumina Miseq platform. Three *S. enterica* belonging to serovars Saintpaul (CFSAN047351), Thompson (CFSAN047352) and Weltevreden (CFSAN047349) from the environmental source was also sequenced using nanopore technologies and assembled *de novo* by default settings in Canu program (Koren et al., 2017). A hybrid assembly using both Nanopore and Miseq data was performed by comparing SPAdes (Bankevich et al., 2012) and Canu assemblies using Mauve to generate complete genomes for these isolates.

2.2. Assembling and annotation

Reads were quality trimmed using Trimommatic v0.32 (Bolger et al., 2014). The first 20 pb were removed from each read and a 4-base wide sliding window was used to cut when the average Phred quality score per base was below 15. Reads less than 50 bp were removed from the dataset. Draft genomes were assembled *de novo* for each strain using A5-miseq v20160825 (Coil et al., 2015). Moreover, the sequence of three *S. enterica* belonging to serovars Saintpaul (CFSAN047351), Thompson (CFSAN047352) and Weltevreden (CFSAN047349) from the environmental source were closed by a combination of long-read (using MinION platform), and short-read (Illumina Miseq) sequencing as described in the previous work (Gonzalez-Escalona et al., 2019).

The generated assemblies were annotated using Prokka v1.11 (Seemann, 2014) and the ST groups identification was performed using Ariba v2.14.4 (Hunt et al., 2017) based on the PubMLST database scheme for *Salmonella* spp. (Jolley et al., 2018). To explore *Salmonella* metabolic pathways in clinical and environmental isolates the amino acid sequence in FASTA format for each genome was annotated with BlastKOALA (Kanehisa et al., 2016) then KEGG Mapper (Kanehisa and Sato, 2020) was used to reconstruct the metabolic pathways based on the previous genome annotation.

2.3. Pan-genome construction

We used the anvio's pan-genome workflow (Eren et al., 2015) to construct the pan-genome for each serovar based on the identification of gene clusters among the genomes. This workflow consisted of three main steps: the first one included the generation on anvio's genomes storage using the program "Anvi-gen-genomes-storage" to store DNA and amino acid sequences from external genomes, as well the annotation of genes. The second step consisted of the generation of an Anvi'o pan database using the program "anvi-pan-genome" to run the pangenomic analysis and finally the distribution visualization of gene clusters across genomes. To compute the *Salmonella enterica* serovars Saintpaul, Thompson, and Weltevreden pangenome we first generated an anvi'o genomes storage database from the FASTA file of 23, 9 and 11 genomes for each serovar respectively using the environmental and clinical genomes. We ran the pangenome analysis with the genomes storage database for each serovar using the flag –use-ncbi-blast for amino acid sequence similarity search. Additionally, the parameters –minbit 0.5 and –mcl-inflation 10 were used to define the sensitivity of the algorithm during the identification of gene clusters for the comparison of very closely related genomes. Finally, the anvi'o pangenome that stored all results was visualized. Within the context of anvi'o analysis, proteins clusters refer to the group of translated DNA sequences from predicted open read frames that share a close degree of homology among all genomes, while gene clusters represent a set of genes found within the genomes that encode similar proteins. Gene calls were clustered into bins based on their frequency across genomes: the first one corresponds to core bin (genes presented in 95 to 100% of the genomes), shell bin (genes between 15 and 95%) and finally cloud bin (presented in 0 to 15%) (Page et al., 2015).

2.4. SNP calling and phylogenetic tree

A core genome alignment based on SNPs was constructed for each of the three serovars using Parsnp with a random clinical genome as a reference. The resulting alignments were explored visually with Gingr to identify the core percentage shared among all strains. A multi-FASTA file was generated for each alignment with HarvestTools (Treangen

et al., 2014) using as input the gingr file. A maximum-likelihood tree was built from the core genome SNP alignment using RAxML (Stamatakis, 2014) with the general time-reversible GTR model and gamma distribution with 100 bootstrap replicates to assess statistical support. Each resulting tree was midpoint rooting, visualized, and edited on iTOL v4 (Letunic and Bork, 2007). Additionally, a SNP difference matrix among each serovar was performed using CSI Phylogenetic v1.4. (Kaas et al., 2014) selecting a minimum SNP quality of 30 with a minimum depth of 30× for each SNP position and ignoring heterozygous SNPs.

2.5. Antimicrobial, virulence gene and plasmid identification

AMRFinderPlus v3.1.1 (Feldgarden et al., 2019) command-line version was used for the detection of acquired or intrinsic antimicrobial resistance gene products using the NCBI database as part of its Pathogen Detection Pipeline. For this purpose, we ran the recommended command line using as input the FASTA assembly, the annotation file (GFF), and the protein sequence. Also, we added the plus option to detect virulence factors and stress-response genes. To detect point mutations in antimicrobial resistance genes (AMR) we specified *Salmonella* spp. as taxon. Additionally, ResFinder v3.2 (Zankari et al., 2012) was used to corroborate and analyze the acquired AMR and chromosomal mutations. For this purpose, we selected genus *Salmonella* with the option to show only know mutations above 90% of identity with a minimum alignment of 60%. This was tested in all closed environmental and clinical isolates. Nevertheless, 2 draft clinical genomes were used for serovar Thompson due to the lack of closed genomes.

To accurately detect virulence genes in the genomes we used the VFanalyzer to construct orthologous groups and then compare it to reference genomes from The Virulence Factors Database (Liu et al., 2019). We used option 3 to upload private data selecting *Salmonella* as a genus. PlasmidFinder v2.1 (Carattoli et al., 2014) was used for the rapid identification of plasmids in *Enterobacteriaceae* species that are of interest for epidemiological investigations considering plasmid associated with the spread of antimicrobial resistance. For the analysis, we select a 95% identity and 60% as coverage as a minimum.

2.6. Phage identification and CRIPSR region prediction

To explore HGT mediated by phages in the closed environmental and clinical isolates PHASTER web service (Arndt et al., 2016) was used. Intact phage sequences (with score > 90) for all genomes were selected to analyze the contribution to virulence or antimicrobial resistance using the VFDB database (Liu et al., 2019) and ResFinder v3.2 (Zankari et al., 2012). CRISPR regions were predicted using CRISPRCasFinder (Couvin et al., 2018) with 100 bp of flanking regions and performing CAS gene detection.

2.7. Secretion system search

Secretion system associated proteins were detected by MacSyFinder (Abby et al., 2014) using the TXSScan (Abby et al., 2016) model proteins profile selecting ordered replicon as type of dataset and all protein secretion systems to perform the detection (Falgellum, pT4SSi, pT4SSt, T1SS, T2SS, T3SS, T4P, T4SS_typeB, T4SS_typeC, T4SS_typeF, T4SS_typeG, T4SS_typeI, T4SS_typeT, T5aSS, T5bSS, T5cSS, T6SSiii, T6SSii, T6SSi, T9SS and Tad). Additionally, maximal e-value was adjusted to 1×10^{-4}.

3. Results

3.1. Pan-genome analysis

Pangenomic analysis of the 23, 9, and 11 isolated genomes from both environmental and clinical sources belonging to serovars Saintpaul, Thompson, and Weltevreden resulted in 5135, 4428 and 5499 protein clusters, respectively (Table 1). For serovar Saintpaul, we detected 4275 and 860 protein clusters for bins core and cloud respectively. *S.* Thompson protein clusters were distributed in core, shell, and cloud bins with 4200, 71, and 157 clusters. Finally, Weltevreden holds 4273, 295, and 917 protein clusters associated with bins core, shell, and cloud (Table 1). A similar percentage of annotated genes with clusters of orthologous groups (COGs) functions was found for serovars Saintpaul and Thompson (80.7 and 83.6%) which most of that annotations are harbored in the core bin. On the other hand, 71.5% percent of annotated clusters were detected for Weltevreden which are also included in the core bin principally.

A total of 101,793, 39,120, and 53,223 genes for serovars Saintpaul, Thompson, and Weltevreden were identified by the gene call annotation in the contig databases for each genome (Table 1) respectively. More than 92% of the genes detected are included in the core.

Specific genes were identified in *S.* Saintpaul genome of isolates from environmental sources (Fig. 1). These genes were related to iron and magnesium acquisition systems ABC type, which are important for enteropathogenic bacteria by the catalysis of oxide-reduction reactions and the iron uptake for enzyme cofactors (Drago-Serrano, 2009). Moreover, additional copies of metabolic genes implied in the Etner-Doudoroff pathway as triosephosphate isomerase (tpiA) and phosphoglycerate mutase (gpmA) (Flamholz et al., 2013) were detected. Concerning to clinical genomes, specific segments shared for all genomes were identified. These segments included genes related to phages, integrases, transposases, pilus, and type II secretory pathway fundamentally.

Genomes of *S.* Thompson isolates were more uniform with a reduced number of accessory genes (Fig. 2). Nevertheless, a specific region was detected in genomes from environmental sources which was not present in the two clinical genomes. This region was represented by the shell bin and included genes like YdaS, a member of Cro superfamily coding for a DNA binding protein to the xenobiotic response which is responsible for signal transduction of exogenous environment pollutants (Kuramoto et al., 2003), as well as genes related to the mobilome mediated by prophages (structural proteins for tail and capsid) and transposons. Also, genes associated with carbohydrate transport and metabolism (designed as G category for COGs) required for transcriptional repression during the growth on glycerol (Sherwood et al., 2009) were found. Additionally, genomes from environmental isolates also showed the presence of DNA replication proteins that protect cells against DNA damage resulted from alkylating and supercoiling activity were identify (Huang and Huang, 2018). The clpP gene was identified, which plays a central role in Salmonella to grow under various stressful conditions such as low pH, elevated temperature, and high salt concentrations (Thomsen et al., 2002). Just for one clinical genome a specific segment in the cloud bin was detected with genes associated with type II and IV secretory pathway, extracellular structures for pilus assembly, and phage related genes.

Finally, serovar Weltevreden pangenome showed in shell bin a specific gene content for only the environmental genomes. This unique block consisted of genes that codify transposases, recombinases for cutting and rejoining of DNA, SOS-response transcriptional repressors, and phage related genes. Moreover, a collinear block is representative for all environmental and one clinical genome, the genes in the segment are related with mobilome.

Concerning to the cloud bin, it comprises many specific gene clusters for clinical and environmental strains. The specific bits for clinical isolates included genes for T3SS as well as genes related to phage structure and integrases. The specific singleton in the clinical genome 2511STDY5712388 holds the gene RadD involved in radiation resistance and double-strand break repair (Chen et al., 2015), as well as the cytolethal distending toxin subunit B been a virulence factor in promoting the infection persistence (Pons et al., 2019). In this component of the genome of environmental isolates were identife some specific regions which genes categorized as hypothetical proteins due to the short

Table 1
The number of gene clusters and genes detected across *S. enterica* serovars Saintpaul, Thompson, and Weltevreden from environmental and clinical isolates.

Serovar pangenome	Number of protein clusters				Number of genes calls			
	Bins			Total	Bins			Total
	Core	*Shell*	*Cloud*		*Core*	*Shell*	*Cloud*	
S. enterica Saintpaul	4275	0	860	5135	99,985	0	1808	101,793
S. enterica Thompson	4200	71	157	4428	38,421	523	176	39,120
S. enterica Weltevreden	4273	295	917	5499	48,915	2871	1283	53,069

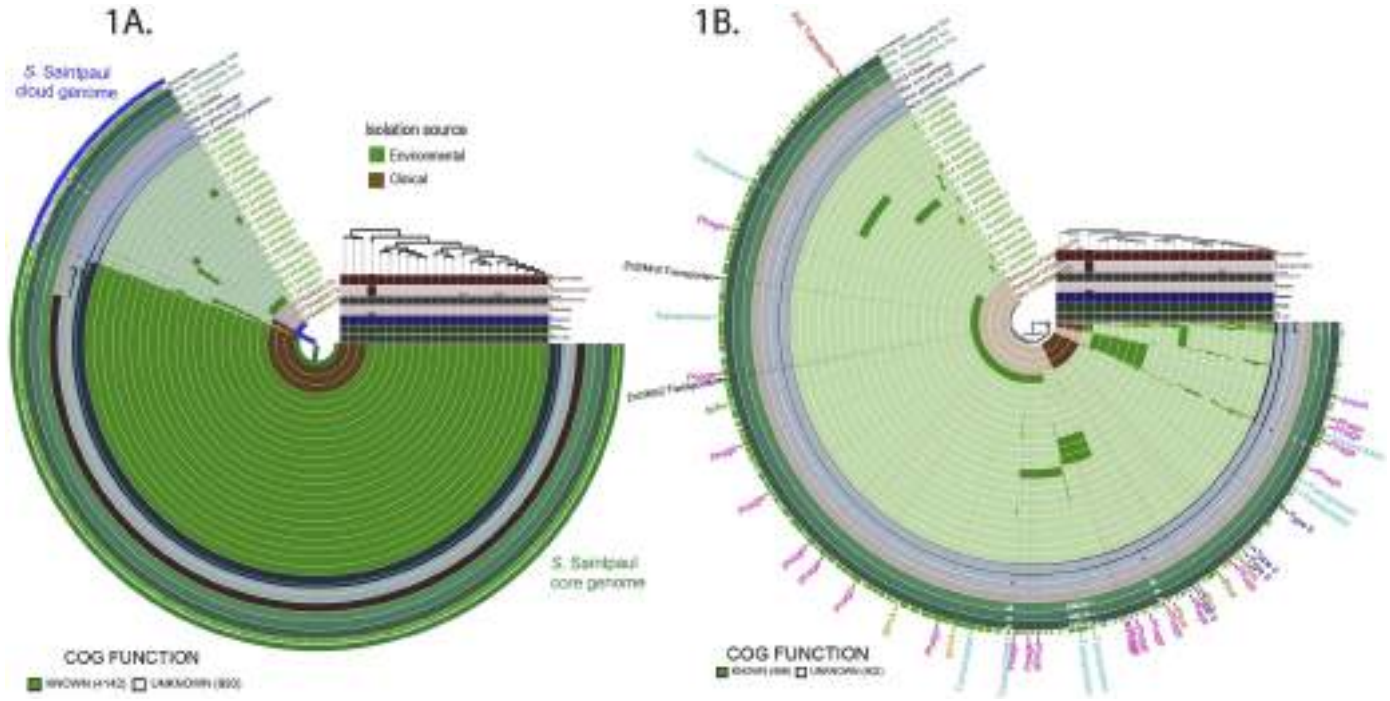

Fig. 1. (A) Pangenome of *S.* Saintpaul genomes isolated from environmental and clinical sources. Rings in green represent environmental genomes and those in brown the clinical. The outer colored ring indicates the classification of gene clusters. The number of the known and unknown gene function are indicated below. (B) Cloud split pangenome. Labels in the outer ring represent the annotated genes included in cloud pangenome. The outer colored ring indicates the classification of gene clusters. The number of the known and unknown gene functions are indicated below as COG function.

fragment of the ORF detected (20 bp). The only annotated genes corresponded to copies of genes lexA, which main function are the transcriptional repressor for the regulation of DNA repair in many bacteria (Smollett et al., 2012). Besides, the activator or repressor for mercury resistance operon merR was identified.

Interestingly, all environmental genomes displayed multiple copies of genes that play key roles in reesitance to environmental stress in *Salmonella enterica* serovars. For example, genes soxS and OxyR, which have been reported in mechanism associated with oxidative, and bile resistance or genes SlyA and RcsC involved in antimicrobial peptide resistance (Spector and Kenyon, 2012). Additionally, copies of genes related to biofilm formation as fimA were identified. As part of core genome for all the serovars for both environmental and clinical sources, genes associated with C-starvation (rpoE, rpoS, cprX and cAMP-CRP) and biofilm formation (csgDEFG-scgBAC and bcsABZC) VBNC (rpoS) were detected but in single copies.

Until 230 metabolic pathways associated with carbohydrate, lipid, energy, nucleotide, amino acids, cofactors and vitamins, glycan, terpenoids and polyketides metabolism were annotated and identified according to KEGG Mapper for all genomes (Fig. 3). Module D-galactonate degradation belonging to galactose metabolism was differentially detected in all environmental strains and some few clinical isolates (33%).

3.2. Antimicrobial resistance gene and plasmid replicons identification

All the clinical and the environmental isolates were analyzed *in silico* to determine antimicrobial resistance genes and point mutations inducing resistance. A total of 6 genes (with identity >95%) for antimicrobial resistance, efflux pump, and stress genes were detected by Resfinder and ARMFinder (Table 2). Genes GolS, GolT, mdsA, mdsB, and aac(6′)-Iaa (gold stress response, efflux pump complex, and aminoglycoside resistance) were found in all isolates. arsD gene was the only one distinctively detected in environmental Thompson and Weltevreden genomes. Its role is the response to metal arsenic stress.

Chromosomal point mutations were explored for genes pmrB, pmrA, gyrB, 16S_rrsD, parC, gyrA, and parE. Only one-point mutation was detected in parC gene inducing a codon change from threonine to serine conferring resistance to nalidixic acid and ciprofloxacin. This mutation was identified in serovars Thompson and Weltevreden for both sources. No point mutations for genes listed previously were determined for serovar Saintpaul in either source.

An *in silico* typing and identification of plasmid replicons from a collection of 5599 fully sequenced plasmids associated with the members of the family Enterobacteriaceae was performed, including those ones for *Salmonella* Typhimurium. Plasmid replicons Incl1 and IncFII(S) were respectively identified in all the three clinical genomes of serovars Saintpaul and Weltevreden. Only one clinical genome of serovar

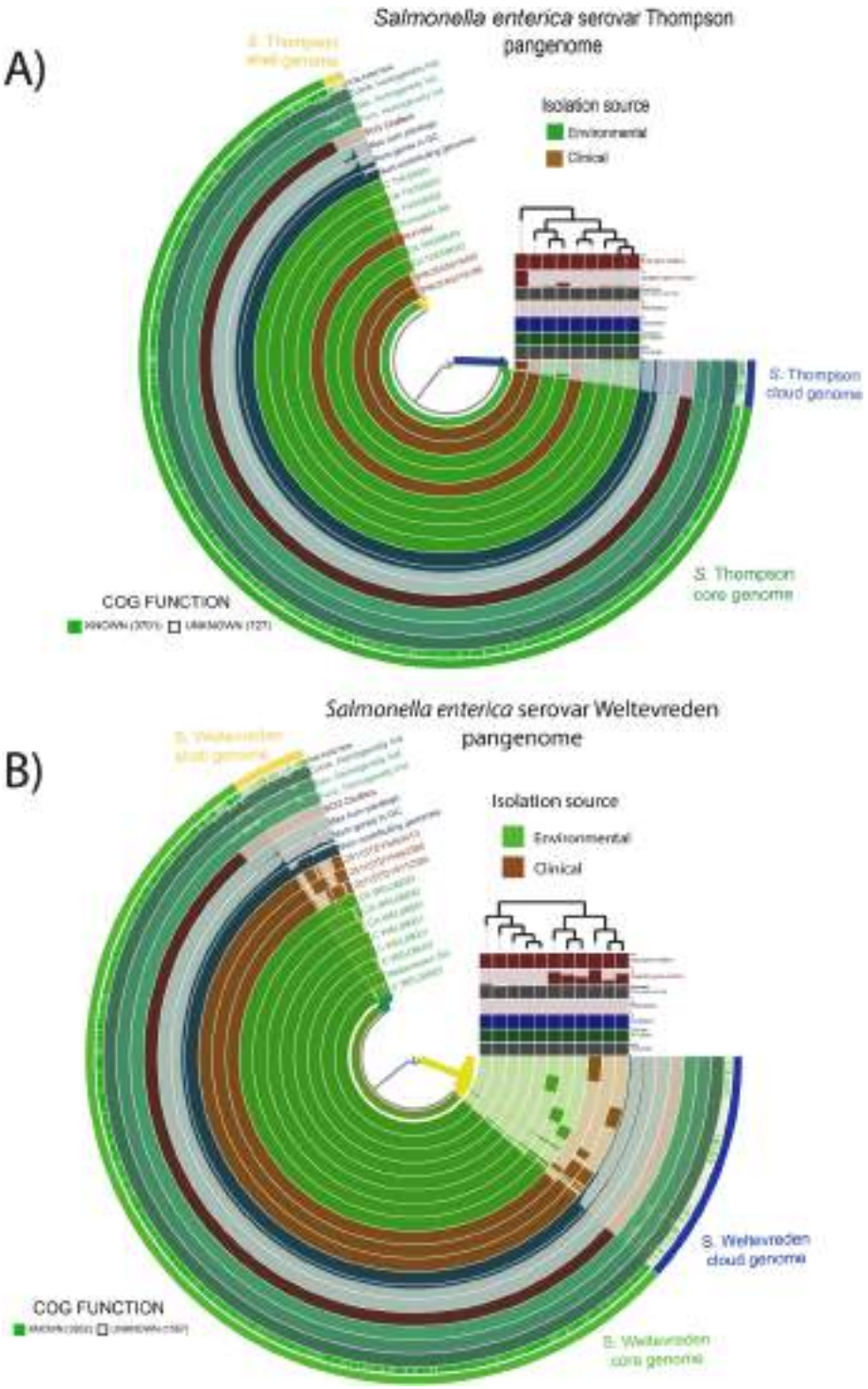

Fig. 2. (A) Pangenome of *S.* Thompson genomes isolated from clinical and environmental sources. (B) Pangenome of *S.* Weltevreden genomes isolated from clinical and environmental sources. Rings in green represent environmental genomes and those in brown the clinical.

Thompson showed the presence of replicon IncF(p96A). For environmental isolates, the replicon IncFII(S) was only detected for serovar Weltevreden. All matched replicons had an identity percentage higher than 97%.

3.3. Identification of virulence genes

Virulence genes related to fimbrial adherence determinants, macrophage inducible genes, magnesium uptake, non-fimbrial adherence determinants, secretion system (SPI-1 and SPI-2), serum resistance and stress adaptation were analyzed for all genomes (Fig. 4). A similar virulence profile was determined for genomes belonging to the same serovar. Interestingly, the SodCl stress adaptation gene was not found in clinical genomes but was present in all the environmental isolates. Additionally, genes cdtB and pltA (codifies for typhoid toxin) were detected in clinical genomes (2 in S. Weltevreden and 1 in S. Thompson) and in one environmental genomes of serovar Weltevreden.

Differences in the virulence profiles among serovars were identified.

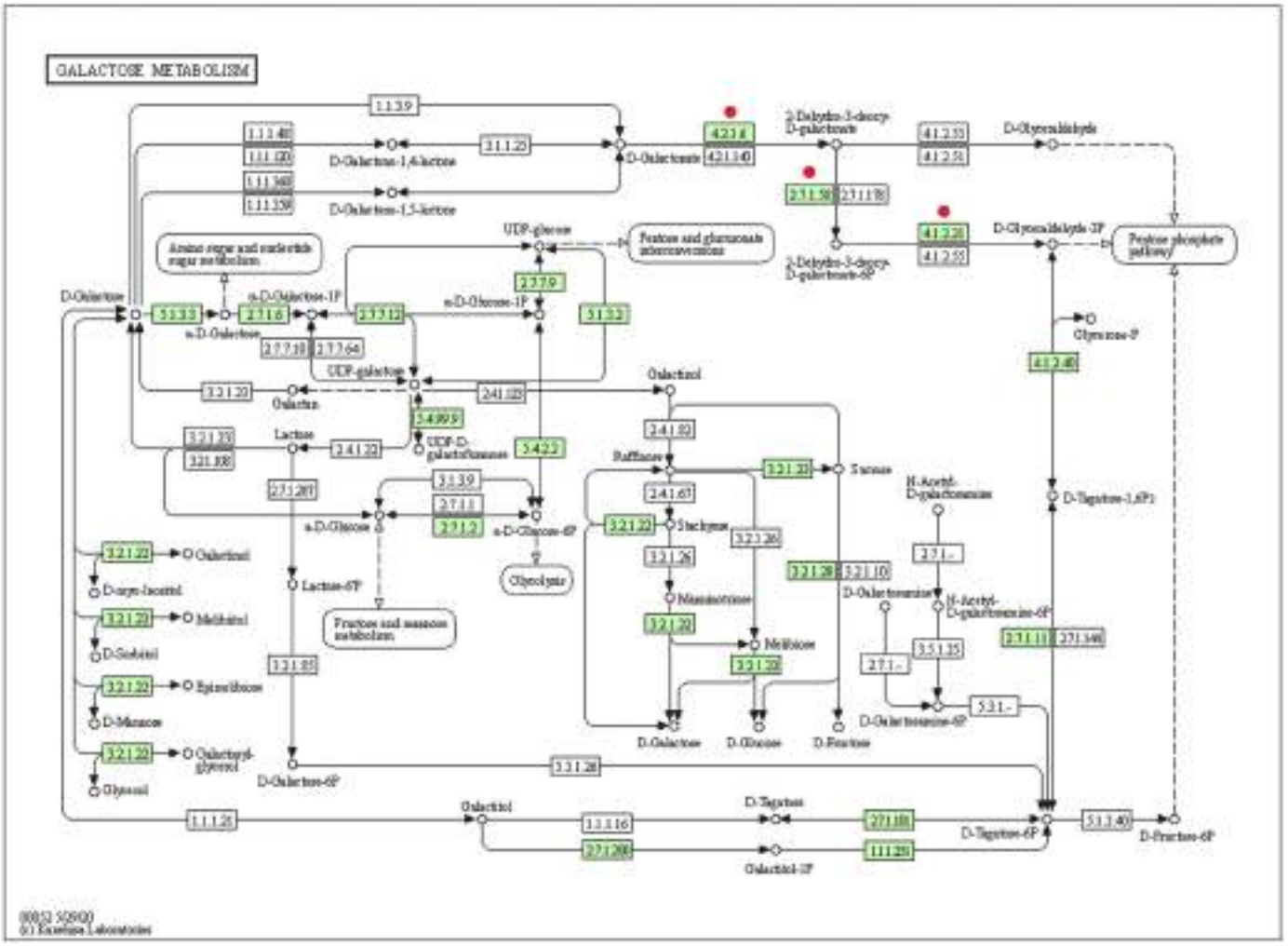

Fig. 3. Galactose metabolism pathway. Green colored labels represent participation in galactose metabolism. Blue circles above labels indicate enzymes codified by operon dgoinvolved in D-galactonate degradation in environmental genomes.

Table 2
Antibiotic resistance genes detected for closed clinical and environmental isolates for serovars Saintpaul, Thompson and Weltevreden. Symbol "+" indicates presence and "−"means absence.

Genomes			Resistance and stress response genes					
			GolS	GolT	mdsA	mdsB	arsD	aac(6′)-Iaa
S. Saintpaul	Environmental	CFSAN047351	+	+	+	+	−	+
	Clinical	CFSAN004173	+	+	+	+	−	+
		CFSAN004174	+	+	+	+	−	+
		CFSAN004175	+	+	+	+	−	+
S. Thompson	Environmental	CFSAN047352	+	+	+	+	+	+
	Clinical	PNUSAS019186	+	+	+	+	−	+
		PNUSAS019450	+	+	+	+	−	+
		RM1984	+	+	+	+	−	+
S. Weltevreden	Environmental	CFSAN047349	+	+	+	+	+	+
	Clinical	2511STDY5462413	+	+	+	+	−	+
		2511STDY5462413	+	+	+	+	−	+
		2511STDY5712385	+	+	+	+	−	+

For example, *S.* Weltevreden lacked the long polar fimbria (lpfABCDE); the operon peg (pegABCD) was detected exclusively for serovar Thompson; the fimbrial adherence operon stk (stk ABCDEFG) was present in Thompson and Saintpaul; operon stc (ABCD) was found in serovars Saintpaul and Weltevreden. Nevertheless, all the essential genes related to the SPI-1, SPI-2, and translocated effectors for both pathogenic islands were found conserved among all three serovars.

The serum resistance gene (rck) was identified in only one clinical Thompson genome. All of the genomes had macrophage inducible genes (mig-14 and mig-5), magnesium uptake (mgtB and mgtC), nonfimbrial adherence determinants (misL, ratB, shdA, and sinH) and regulation genes (phoP and phoQ). No capsular Vi antigen genes were detected for any genome.

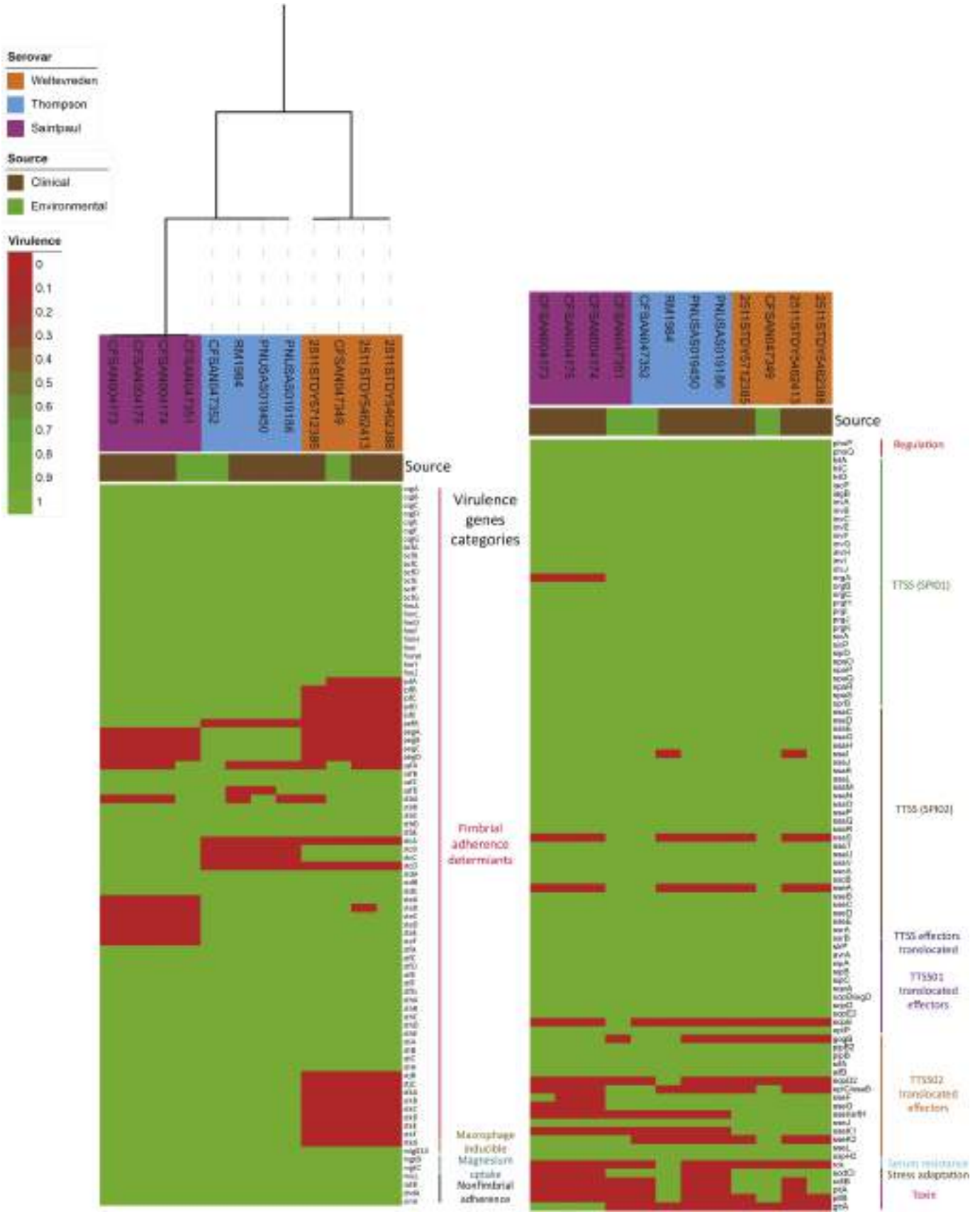

Fig. 4. Phylogenetic tree with the presence of virulence genes. Colored labels indicated *Salmonella* serovars. The identity scale is represented in red (absence) to green (presence) range. Virulence genes and virulence category labels are indicated in next to the heatmap.

3.4. Phage identification and CRIPSR region prediction

A total of 37 intact lysogenic phages were identified by PHASTER (Table 3). The most common families included relatives of Gifsy-1 (NC_010392) and Gifsy-2 (NC_010393) phages. Gifsy-1 like phage was found in all clinical and environmental *Salmonella* serovars and Gifsy-2 like phage were present in all the Weltevreden genomes and one environmental Saintpaul genome. Additionally, some particular phage sequences were found exclusively in a specific serovar. For example, the Entero_PsP3 like phage (NC_0053408) and Salmon_Fels_1 (NC_010391) were identified in all the Weltevreden genomes. Additionally, the Salmon_ST64T (NC_004143) was found in the Saintpaul genomes. To gain a better understanding of the role of phages in virulence and environmental adaptation, we searched for carried non-phage cargo genes (Table 4). We detected 7 cargo genes related to *Salmonella* virulence, these include sopE, sseL/srfH, ctdB, pltA, shdA, zur/yjbk and gogB. Interestingly, the cytholethal distending toxin subunit B was detected for almost all genomes belonging to serovar Weltevreden. *S. Typhi* toxin (pertussis-like toxin) pltA was detected for one Weltevreden genome belonging to both strain sources. Genes spoE and gogB consider as TTSS-1 and TTSS-2 translocated effector were detected in one environmental Saintpaul genome, and in one each clinical and environmental Thompson genomes. Clinical Weltevreden genomes hold the biggest virulence cargo genes with additionally sdhA, a nonfimbrial adherence determinant and ssel/srfh a TTSS-2 translocated effector.

Type I-E class 1 CRISPR-Cas system which uses a multisubunit crRNA-effector for immunity and characterized by the presence of Cas3, Cas10, and CSF1 was detected for all *Salmonella* genomes (Table 4). Two CRISPER arrays with different numbers of spacers were detected across all genomes. Spacer sequences were also analyzed to detect target prophages with Blastp using UniProt phage sequences (https://www.uniprot.org/). Prophage targets were observed in 38% (9/24) of the total genomes with the most common prophage targets being tail and holins proteins.

3.5. Detection of secretion systems in Salmonella enterica genomes

An exploration of 22 models of the protein secretion system was performed for all clinical and environmental genomes. We detected 7 models with the presence of mandatory genes (essential in a system to be functional) related to flagellum, T1SS, T3SS, T4SS_typeF, T5SS, T5Bss, and T5cSS (Fig. 5). Differences among environmental and clinical genomes were observed. For example, all clinical isolates held the mandatory genes related with the flagellum (fliF, fliL, fliN/M, flip, fliQ, fliR, flhB and flhA genes) and T3SS (sctC, sctJ, sctN, sctQ, sctR, sctS, sctT, sctU and sctV genes) which are a crucial requirement for *Salmonella enterica* and other Enterobactericeae members to be able to infect a host. Nevertheless, the T4SS_typeF had a 10% higher frequency in genomes from environmental isolated than from the clinical ones. This secretion system encodes proteins involve in signaling and host cell interaction and is essentially characterized by various stages of pseudogenization revealed by gene duplication and deletions. System T5SS subtypes T5bSS (two-partner passenger-translocators) and T5cSS (trimeric autotransporters) were identified for one Weltevreden clinical isolate. These subtypes are characterized by the use of the Sec translocase to pass through the inner membrane to the periplasmatic space. Finally, mandatory genes abc and omf belonging to T1SS were detected for all both environmental and clinical genomes.

4. Discussion

Surface water is an inhospitable environment for *Salmonella*, therefore the ability to survive, resist, and adapt are conditioning factors that can determine the continuity of its life cycle (Liu et al., 2018). During its passage through the environment, *Salmonella* manages to be part of the microbial community of diverse environments, among which stand out the soil and the water where it has been reported that it manages to stay for long periods (Gorski et al., 2011; Martinez-Urtaza et al., 2004). However, the behavior of *Salmonella* in adverse aquatic environments has been poorly studied (Levantesi et al., 2012). The metabolic and genetic changes that adverse aquatic environments exert on the *Salmonella* colonization and adaptation processes are yet to be discovered. There have been efforts to address this question. Medrano-Félix et al. (2018) reported the metabolic profiles and diversity of carbohydrates used by *Salmonella* isolates under river water conditions (Medrano-Félix et al., 2018) Also, Estrada-Acosta et al. (2018) determined how aquatic environments affect Salmonella's capacity to infect hosts (Estrada-Acosta et al., 2018). High-resolution genomic signatures could be another piece in the puzzle to understand Salmonella's ability to colonize natural environments as river water and at what cost.

The construction of pangenome enabled the identification of serovars-specific genetic signature in isolates from both environmental and clinical sources. Also, accessory genome detection and quantification increased the understanding of the isolation sources, virulence and

Table 3
Intact lysogenic phages and cargo genes detected in clinical and environmental genomes.

Salmonella serovar	Sample	Source	No of phage	Best match (kb)	Virulence genes on phage
Saintpaul	CFSAN047351	Environmental	2	Entero_P88_NC_026014(21), Phage_Gifsy_2_NC_010393(47)	sopE
	CFSAN004173	Clinical	2	Phage_Gifsy_1_NC_010392(39), Salmon_ST64T_NC_004348(24)	–
	CFSAN004174	Clinical	2	Phage_Gifsy_1_NC_010392(39), Salmon_ST64T_NC_004348(24)	–
	CFSAN004175	Clinical	2	Phage_Gifsy_1_NC_010392(39), Salmon_ST64T_NC_004348(24)	–
Weltevreden	CFSAN047349	Environmental	6	Entero_PsP3_NC_005340(35), Gifsy_1_NC_01039(26), Haemop_HP1_NC_001697(18), Haemop_HP1_NC_001697(19), Phage_Gifsy_2_NC_010393(46), Salmon_Fels_1_NC_010391(27),	sseI/srfH, cdtB, pltA
	2511STDY5462388	Clinical	8	Entero_PsP3_NC_0053408(36), Phage_Gifsy_1_NC_010392(25), Phage_Gifsy_2_NC_010393(47), Salmon_118970_sal3_NC_031940(27), Salmon_epsilon15_NC_004775(36), Salmon_Fels_1_NC_010391(20), Salmon_g341c_NC_013059(48), Salmon_SEN34_NC_028699(33)	sseI/srfH, cdtB
	2511STDY5662413	Clinical	8	Entero_PsP3_NC_005340(36), Escher_phiV10_NC_007804(36), Gifsy_1_NC_01039(25), Phage_Gifsy_2_NC_010393(46), Salmon_118970_sal3_NC_031940(61), Salmon_Fels_1_NC_010391(20), Salmon_SEN34_NC_028699(33), Shigel_SfIV_NC_022749(1)	shdA, sseI/srfH, zur/yjbK
	2511STDY5712385	Clinical	4	Entero_PsP3_NC_005340(35), Phage_Gifsy_1_NC_010392(23), Phage_Gifsy_2_NC_010393(36), Salmon_Fels_1_NC_010391(29)	sseI/srfH, cdtB, pltA
Thompson	CFSAN047352	Environmental	1	Phage_Gifsy_1_NC_010392(39)	gogB
	PNUSAS19186	Clinical	1	Salmon_Fels_2_NC_010463(28)	–
	RM1984	Clinical	1	Phage_Gifsy_1_NC_010392(39)	gogB
	PNUSAS19450	Clinical	0	–	–

Table 4
Prophage target identified in spacer sequences in CRIPSR arrays in *Salmonella* serovars.

S. enterica serovar	Sample	Source	CRISPR number	Cas cluster	Spacer		
					no	target	gene
Saintpaul	CFSAN047351	Environmental	2	CAS-TypeIE	22	prophage	Phage tail protein
					13	–	–
	CFSAN004173	Clinical	2	CAS-TypeIE	20	prophage	Phage tail protein
					15	–	–
	CFSAN004174	Clinical	2	CAS-TypeIE	20	prophage	Phage tail protein
					15	–	–
	CFSAN004175	Clinical	2	CAS-TypeIE	20	prophage	Phage tail protein
					15	–	–
Weltevreden	CFSAN047349	Environmental	2	CAS-TypeIE	34	–	–
					44	–	–
	2511STDY5462388	Clinical	2	CAS-TypeIE	40	prophage	phage holin
					25	prophage	Phage tail protein
	2511STDY5662413	Clinical	2	CAS-TypeIE	32	prophage	phage holin
					29	prophage	Phage tail protein
	2511STDY5712385	Clinical	2	CAS-TypeIE	38	–	–
					25	prophage	Phage tail protein
Thompson	CFSAN047352	Environmental	2	CAS-TypeIE	14	–	–
					18	–	–
	PNUSAS19186	Clinical	2	CAS-TypeIE	17	–	–
					28	–	–
	PNUSAS19450	Clinical	2	CAS-TypeIE	19	–	–
					23	–	–
	RM1984	Clinical	2	CAS-TypeIE	26	–	–
					18	–	–

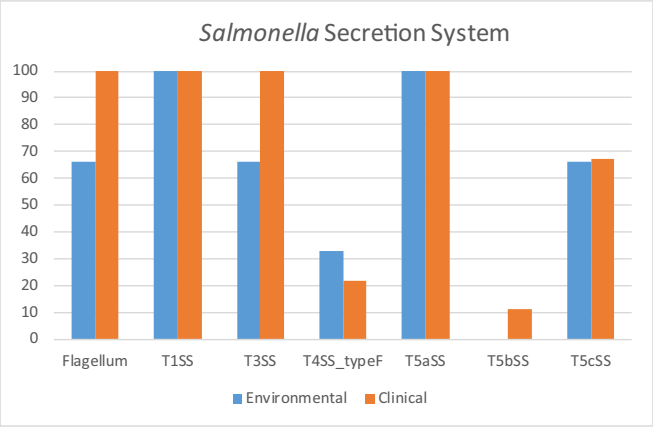

Fig. 5. *Salmonella enterica* serovars Saintpaul, Thompson, and Weltevreden secretion system detection. Environmental genomes are denoted in blue and clinical in orange. Each bar is labeled with the secretion systems detected.

resistance potential difference (Guimarães et al., 2015). To this end, we determined that serovar Saintpaul holds the biggest number of cloud-related genes with detectable environmental and clinical gene clusters, followed by *S.* Weltevreden and serovar Thompson. Serovar Weltevreden showed the biggest genetic variation in the pangenome among clinical and environmental isolates, this could be to the submissive genome structure characterized by the ability to routinely acquired and lose genetic material as previously reported for this serovar (Makendi et al., 2016). However, a similar number of core genes clustered among and within all serovars were detected. Therefore, a basic set of genes present in all Salmonella could be a testimony to the conservative nature of evolution (Lapierre and Gogarten, 2009). Genes related to the basic biology of *Salmonella* such transport and metabolism of carbohydrates, cell wall and envelope biogenesis, DNA process (replication, transcription and translation), intracellular trafficking, secretion and vesicular transport were found as the main component of core genome (Anjum et al., 2005).

Environmental stressful conditions may trigger, in shell and cloud genomes, a high level of gene duplication, which might relate to the evolutionary process in many prokaryotic organisms. Therefore, the faster the bacteria settle in the environment, the lesser need for duplication of genes is generated (Bratlie et al., 2010). Furthermore, a small number of protein families (~ 5021 groups) detected in the serovars under study, reflects the flexibility of those genes in their ability to adapt new function indicating the that the preferred mode of adaptation consists in the exploration of new strategies by gene duplication or mutation(Lapierre and Gogarten, 2009). Singletons of pseudogenes were detected for all three serovars, supporting the notion that the evolution of *S. enterica* serovars towards becoming host-adapted and host-restricted is characterized by the accumulation of pseudogenes (Waldner et al., 2012). The superoxide dismutase gene is a genome-constituent of *Salmonella*. In aquatic environments, the gene expression results in the prevention of the oxidation process set by phagocytic superoxide (Tidhar et al., 2015). Moreover, environmental isolates must survive predation from bacteriophage and predatory bacteria, under this context we detected biofilm formation genes which have been reported as a defense mechanism by a membrane rearrangement to avoid being ingested (Erken et al., 2013). Also, T3SS has been described as an *E. coli* and *V. parahaemolyticus* strategy to survive inside trophozoites of *A. castellanii* and *C. roenbergensis* which are predominant protozoa in natural environments such river and ocean water (Matz et al., 2011; Siddiqui et al., 2011), these strategies may additionally help *Salmonella* to predation evasion and survival in non-host environments.

D-galactose operon, present in all environmental isolate genomes, plays a crucial role in the degradation process of the D-galactonate. Humans are not the principal source of D-galactose contrary only a mere fraction is produced. Most production comes from nature, mainly prokaryotic cells (Singh et al., 2019). The metabolic flexibility by *Salmonella* during D- galactonate degradation (Selzer, 2016) makes the bacteria survive, establish, and persist in river water with high possibility to reach back human hosts by means of consumption of contaminated food or water.

Nevertheless, type II and IV secretory pathways genes, toxins, pilus assembly, and mandatory genes detected mainly in clinical isolates, contribute to bacterial colonization by the invasion of phagocytic and non-phagocytic cells (Ibarra and Steele-Mortimer, 2009). T3SS is an essential secretion system for actin rearrangements required for

Salmonella internalization, pathogenicity, and survival inside the host. Moreover, this system has been found as crucial for other pathogenic bacteria as *Yersinia* and *Shigella* (Green and Mecsas, 2016). Additionally, Salmonella easies the cellular internalization with T1SS, fimbriae adherence, and, flagella. For example, lpf operon detected for serovars Saintpaul and Thompson plays an important role in *Salmonella* adhesion to Peyer's patches, the portal of entry for all non-typhoidal *Salmonella* infections (Wagner and Hensel, 2011). The presence/absence of virulence operons found in this work suggest that fimbriae represent a source of diversity among *Salmonella* serovars as previously reported for several authors (Dufresne et al., 2018). Even though T5SS is not described as an essential system for Salmonella, it may contribute to bacterial virulence. The secretion of substrates may serve as toxins and receptor binding proteins (Wells and Henderson, 2013).

Bacteria infected with lysogenic phages produce a more diversified progeny including increases in virulence factors (Gupta et al., 2019). Moreover, Gifsy-1 and Gifsy-2 like phage are the most common phages found in *Salmonella* genomes and had been reported in other *Salmonella* serovars as Typhimurium playing an important role in the contribution of virulence (Ho and Slauch, 2001). SsrB and gogB were identified as important cargo genes of Gifsy-1 and Gifsy-2 which have a principal role in activating global regulon of acquired genes (Worley et al., 2000) and increasing *Salmonella* survival after phagocytosis (Kutter and Sulakvelidze, 2004), respectively. Our study supports the notion that phages have an important role in the pathogenicity in *Salmonella* serovars.

CRISPR-Cas is an immunity strategy to avoid phage attack based in a set of Cas proteins with functional domains related to nucleases, helicases, and polynucleotide-binding proteins (Medina-Aparicio et al., 2018), due to the high death rate (4–50% killed bacteria) caused by phage lysis. Here we report that *Salmonella* genomes have the CRISPR array 1 (CRISPR1) associated with the Type I-E set of cas genes, which had been exclusively reported for *Salmonella* (Shariat et al., 2015) and other Enterobactericeae members as *E coli*. Spacer contained in the CRISPR arrays provide a memory of the foreign elements that have invaded *Salmonella*. The 38% of spacers found in this study belong to phages, representing higher values than reported for other studies with 10% associated values. Non identified spacer sequences are believed to correspond to plasmids and their own genome sequence as shown in other studies (Medina-Aparicio et al., 2018). Differences in spacer numbers were detected among isolation sources, for example, clinical isolates belonging to serovar Thompson displayed higher numbers associated with spacers. This could be explained under a microbial adaptation context considering that environmental isolates rely on HGT mechanism to fit the non-host environment and the acquisition of CRIPR elements had been reported as a barrier for HGT, the reason why bacteria may eliminate or reduce CRISPR elements (Westra et al., 2016). Nevertheless, further studies may be conducted.

By the employment of AMRfinder all studied serovars concurred with the same pattern of antibiotic resistance and stress response. Interestingly, the metallochaperone arsD which is a constituent gene in ars operon providing arsenic resistance in natural environments and industrial contaminated sites (Jackson and Dugas, 2003) was the only difference detected in the environmental Thompson and Weltevreden isolates respecting the clinical ones. This adaptive signature gives Salmonella ways to survive and establish themselves in non-host environments such as high arsenic river water (Joerger et al., 2010). The parC mutation observed in both isolates (environmental and clinical) analyzed in the present study is the same that *Salmonella* Typhimurium and *E. coli* poses. This mutation provides resistance to fluoroquinolones due to amino acid change in the ParC subunit of topoisomerase IV, which is the second target of quinolones, and a growing worldwide problem (Baucheron et al., 2004; Dalhoff, 2012).

5. Conclusion

In the spite of the limited number of analyzed genomes in this study, the application of WGS and bioinformatics approaches allowed a comprehensive analysis and full genomic comparison between *Salmonella* isolates from environmental and clinical sources. Genes associated with stress resistance and metabolic pathways were identified as environmental isolates signature that may represent an adaptation and survival advantage in non-host environments as river water, suggesting that the isolation source plays an important role in genome plasticity. In addition to HGT we identified gene duplication and pseudogenes accumulation as environmental survival strategies characterized by genome rearrangements. According to the accessory genome of clinical isolates, secretion system, flagellum, and virulence genes seem to be preserved among all genomes as essential elements to invade, internalize, and cause pathogenesis. The most common phages sequences as Gifsy_1and Gifsy_2 reported for *Salmonella* were detected, these held virulence cargo genes contributing of bacterial virulence, supporting the idea that horizontal gene transfer enhanced bacterial fitness.

Declarations of interest

None

Credit author statement

J. R. Aguirre-Sanchez: Formal analysis, Writing – original draft and Investigation. **J. Martínez-Urtaza**: Writing – review & editing. **J. R. Ibarra-Rodriguez**: Investigation. **I. F. Vega-Lopez**: Investigation. **C. Chaidez-Quiroz**: Writing – review & editing, Supervision.

Acknowledgements

We thank you the National Laboratory for Food Safety and Research (LANIIA) at Centro de Investigación y Desarrollo A. C. (CIAD) at Culiacan for sharing *Salmonella enterica* serovars Saintpaul, Thompson, and Weltvreden strains.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2021.104771.

References

Abby, S.S., Néron, B., Ménager, H., Touchon, M., Rocha, E.P.C., 2014. MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-Cas systems. PLoS One 9.

Abby, S.S., Cury, J., Guglielmini, J., Néron, B., Touchon, M., Rocha, E.P.C., 2016. Identification of protein secretion systems in bacterial genomes. Sci. Rep. 6, 1–14.

Achtman, M., Wain, J., Weill, F.-X., Nair, S., Zhou, Z., Sangal, V., Krauland, M.G., Hale, J. L., Harbottle, H., Uesbeck, A., 2012. Multilocus sequence typing as a replacement for serotyping in Salmonella enterica. PLoS Pathog. 8, e1002776.

Anjum, M.F., Marooney, C., Fookes, M., Baker, S., Dougan, G., Ivens, A., Woodward, M. J., 2005. Identification of core and variable components of the *Salmonella enterica* subspecies I genome by microarray. Infect. Immun. 73 https://doi.org/10.1128/IAI.73.12.7894-7905.2005, 7894 LP – 7905.

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., Wishart, D.S., 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 44, W16–W21.

Balloux, F., Brynildsrud, O.B., Van Dorp, L., Shaw, L.P., Chen, H., Harris, K.A., Wang, H., Eldholm, V., 2018. From theory to practice: translating whole-genome sequencing (WGS) into the clinic. Trends Microbiol. 26, 1035–1048.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V. M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477.

Baucheron, S., Chaslus-Dancla, E., Cloeckaert, A., 2004. Role of TolC and parC mutation in high-level fluoroquinolone resistance in Salmonella enterica serotype Typhimurium DT204. J. Antimicrob. Chemother. 53, 657–659.

Beceiro, A., Tomás, M., Bou, G., 2013. Antimicrobial resistance and virulence: a successful or deleterious association in the bacterial world? Clin. Microbiol. Rev. 26, 185–230.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Bratlie, M.S., Johansen, J., Sherman, B.T., Lempicki, R.A., Drabløs, F., 2010. Gene duplications in prokaryotes can be associated with environmental adaptation. BMC Genomics 11, 588.
Carattoli, A., Zankari, E., Garcìa-Fernandez, A., Larsen, M.V., Lund, O., Villa, L., Aarestrup, F.M., Hasman, H., 2014. PlasmidFinder and pMLST: in silico detection and typing of plasmids. Antimicrob. Agents Chemother. 58 (7), 3895–3903.
CDC (Center for Disease Control and Prevention), 2018. *Salmonella* Homepage. https://www.cdc.gov/salmonella/index.html.
Chen, S.H., Byrne, R.T., Wood, E.A., Cox, M.M., 2015. E scherichia coli rad D (yejH) gene: a novel function involved in radiation resistance and double-strand break repair. Mol. Microbiol. 95, 754–768.
Coil, D., Jospin, G., Darling, A.E., 2015. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. Bioinformatics 31, 587–589.
Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., Rocha, E.P.C., Vergnaud, G., Gautheret, D., Pourcel, C., 2018. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res. 46, W246–W251.
Dalhoff, A., 2012. Global fluoroquinolone resistance epidemiology and implictions for clinical use. Interdiscip. Perspect. Infect. Dis. 2012.
Drago-Serrano, M.E., 2009. Sistemas de adquisición de hierro en Salmonella enterica. Rev. Biomed. 20, 41–54.
Dufresne, K., Saulnier-Bellemare, J., Daigle, F., 2018. Functional analysis of the chaperone-usher fimbrial gene clusters of Salmonella enterica serovar Typhi. Front. Cell. Infect. Microbiol. 8, 26.
Eren, A.M., Esen, Ö.C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L., Delmont, T. O., 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ 3, e1319.
Erken, M., Lutz, C., McDougald, D., 2013. The rise of pathogens: predation as a factor driving the evolution of human pathogens in the environment. Microb. Ecol. 65, 860–868.
Estrada-Acosta, M.D., Ramirez, K., Medrano-Félix, J.A., Castro-Del Campo, N., López-Moreno, H.S., Jimenez Edeza, M., Martínez-Urtaza, J., Chaidez, C., 2018. Effect of river water exposition on adhesion and invasion abilities of Salmonella Oranienburg and Saintpaul. Int. J. Environ. Health Res. 28, 43–54.
Feldgarden, M., Brover, V., Haft, D.H., Prasad, A.B., Slotta, D.J., Tolstoy, I., Tyson, G.H., Zhao, S., Hsu, C.-H., McDermott, P.F., 2019. Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. Antimicrob. Agents Chemother. 63 e00483–19.
Flamholz, A., Noor, E., Bar-Even, A., Liebermeister, W., Milo, R., 2013. Glycolytic strategy as a tradeoff between energy yield and protein cost. Proc. Natl. Acad. Sci. 110, 10039–10044.
Garai, P., Gnanadhas, D.P., Chakravortty, D., 2012. Salmonella enterica serovars Typhimurium and Typhi as model organisms: revealing paradigm of host-pathogen interactions. Virulence 3, 377–388.
Gonzalez-Escalona, N., Aguirre-Sanchez, J.R., Ibarra-Rodriguez, J.R., Chaidez-Quiroz, C., Martinez-Urtaza, J., 2019. Closed genome sequences of three Salmonella enterica strains belonging to serovars Saintpaul, Weltevreden, and Thompson, isolated from Mexico. Microbiol. Resour. Announc. 8 e00656–19.
Gorski, L., Parker, C.T., Liang, A., Cooley, M.B., Jay-Russell, M.T., Gordus, A.G., Atwill, E.R., Mandrell, R.E., 2011. Prevalence, distribution, and diversity of Salmonella enterica in a major produce region of California. Appl. Environ. Microbiol. 77, 2734–2748.
Green, E.R., Mecsas, J., 2016. Bacterial secretion systems: an overview. Virulence Mech. Bact. Pathog. 213–239.
Greene, S.K., Daly, E.R., Talbot, E.A., Demma, L.J., Holzbauer, S., Patel, N.J., Hill, T.A., Walderhaug, M.O., Hoekstra, R.M., Lynch, M.F., 2008. Recurrent multistate outbreak of Salmonella Newport associated with tomatoes from contaminated fields, 2005. Epidemiol. Infect. 136, 157–165.
Guimarães, L.C., Florczak-Wyspianska, J., de Jesus, L.B., Viana, M.V.C., Silva, A., Ramos, R.T.J., de Soares, S.C., de Soares, S.C., 2015. Inside the Pan-genome - methods and software overview. Curr. Genomics 16, 245–252. https://doi.org/10.2174/1389202916666150423002311.
Gupta, S.K., Sharma, P., McMillan, E.A., Jackson, C.R., Hiott, L.M., Woodley, T., Humayoun, S.B., Barrett, J.B., Frye, J.G., McClelland, M., 2019. Genomic comparison of diverse Salmonella serovars isolated from swine. PLoS One 14.
Hanning, I.B., Nutt, J.D., Ricke, S.C., 2009. Salmonellosis outbreaks in the United States due to fresh produce: sources and potential intervention measures. Foodborne Pathog. Dis. 6, 635–648.
Ho, T.D., Slauch, J.M., 2001. OmpC is the receptor for Gifsy-1 and Gifsy-2 bacteriophages of Salmonella. J. Bacteriol. 183, 1495–1498. https://doi.org/10.1128/JB.183.4.1495-1498.2001.
Huang, Y.-H., Huang, C.-Y., 2018. SAAV2152 is a single-stranded DNA binding protein: the third SSB in Staphylococcus aureus. Oncotarget 9, 20239–20254. https://doi.org/10.18632/oncotarget.24427.
Hunt, M., Mather, A.E., Sánchez-Busó, L., Page, A.J., Parkhill, J., Keane, J.A., Harris, S. R., 2017. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. Microb. Genom. 3.
Ibarra, J.A., Steele-Mortimer, O., 2009. Salmonella–the ultimate insider. Salmonella virulence factors that modulate intracellular survival. Cell. Microbiol. 11, 1579–1586.
Jackson, C.R., Dugas, S.L., 2003. Phylogenetic analysis of bacterial and archaeal arsC gene sequences suggests an ancient, common origin for arsenate reductase. BMC Evol. Biol. 3, 18.
Jiménez, M., Martínez-Urtaza, J., Chaidez, C., 2011. Geographical and temporal dissemination of salmonellae isolated from domestic animal hosts in the Culiacan Valley, Mexico. Microb. Ecol. 61, 811–820.
Jiménez, M., Martinez-Urtaza, J., Rodriguez-Alvarez, M.X., Leon-Felix, J., Chaidez, C., 2014. Prevalence and genetic diversity of Salmonella spp. in a river in a tropical environment in Mexico. J. Water Health 12, 874–884.
Joerger, R.D., Hanning, I.B., Ricke, S.C., 2010. Presence of arsenic resistance in Salmonella enterica serovar Kentucky and other serovars isolated from poultry. Avian Dis. 54, 1178–1182.
Jolley, K.A., Bray, J.E., Maiden, M.C.J., 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST. org website and their applications. Wellcome Open Res. 3.
Kaas, R.S., Leekitcharoenphon, P., Aarestrup, F.M., Lund, O., 2014. Solving the problem of comparing whole bacterial genomes across different sequencing platforms. PLoS One 9.
Kanehisa, M., Sato, Y., 2020. KEGG mapper for inferring cellular functions from protein sequences. Protein Sci. 29, 28–35.
Kanehisa, M., Sato, Y., Morishima, K., 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J. Mol. Biol. 428, 726–731.
Kenyon, W.J., Sayers, D.G., Humphreys, S., Roberts, M., Spector, M.P., 2002. The starvation-stress response of Salmonella enterica serovar Typhimurium requires σE-, but not CpxR-regulated extracytoplasmic functions. Microbiology 148, 113–122.
Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., Phillippy, A.M., 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27, 722–736.
Kuramoto, N., Baba, K., Gion, K., Sugiyama, C., Taniura, H., Yoneda, Y., 2003. Xenobiotic response element binding enriched in both nuclear and microsomal fractions of rat cerebellum. J. Neurochem. 85, 264–273. https://doi.org/10.1046/j.1471-4159.2003.01679.x.
Kutter, E., Sulakvelidze, A., 2004. Bacteriophages: Biology and Applications. CRC Press.
Kwong, J.C., McCallum, N., Sintchenko, V., Howden, B.P., 2015. Whole genome sequencing in clinical and public health microbiology. Pathology 47, 199–210.
Lapierre, P., Gogarten, J.P., 2009. Estimating the size of the bacterial pan-genome. Trends Genet. 25, 107–110.
Letunic, I., Bork, P., 2007. Interactive tree of life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23, 127–128.
Levantesi, C., Bonadonna, L., Briancesco, R., Grohmann, E., Toze, S., Tandoi, V., 2012. Salmonella in surface and drinking water: occurrence and water-mediated transmission. Food Res. Int. 45, 587–602.
Liu, B., Zheng, D., Jin, Q., Chen, L., Yang, J., 2019. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. Nucleic Acids Res. 47, D687–D692.
Liu, H., Whitehouse, C.A., Li, B., 2018. Presence and persistence of Salmonella in water: the impact on microbial quality of water and food safety. Front. Public Health 6, 159.
Makendi, C., Page, A.J., Wren, B.W., Le Thi Phuong, T., Clare, S., Hale, C., Goulding, D., Klemm, E.J., Pickard, D., Okoro, C., 2016. A phylogenetic and phenotypic analysis of Salmonella enterica serovar Weltevreden, an emerging agent of diarrheal disease in tropical regions. PLoS Negl. Trop. Dis. 10, e0004446.
Martinez-Urtaza, J., Liebana, E., Garcia-Migura, L., Perez-Piñeiro, P., Saco, M., 2004. Characterization of Salmonella enterica serovar Typhimurium from marine environments in coastal waters of Galicia (Spain). Appl. Environ. Microbiol. 70, 4030–4034.
Matz, C., Nouri, B., McCarter, L., Martinez-Urtaza, J., 2011. Acquired type III secretion system determines environmental fitness of epidemic Vibrio parahaemolyticus in the interaction with bacterivorous protists. PLoS One 6, e20275.
Medina-Aparicio, L., Davila, S., Rebollar-Flores, J.E., Calva, E., Hernández-Lucas, I., 2018. The CRISPR-Cas system in Enterobacteriaceae. Pathog. Dis. 76 fty002.
Medrano-Félix, J.A., Castro-del Campo, N., de Peraza Garay, F.J., Martínez-Rodríguez, C. I., Chaidez, C., 2018. Carbon source utilization-based metabolic activity of Salmonella Oranienburg and Salmonella Saintpaul in river water. Water Environ. J. 32, 118–124.
Morishige, Y., Koike, A., Tamura-Ueyama, A., Amano, F., 2017. Induction of viable but nonculturable Salmonella in exponentially grown cells by exposure to a low-humidity environment and their resuscitation by catalase. J. Food Prot. 80, 288–294.
Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., Fookes, M., Falush, D., Keane, J.A., Parkhill, J., 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31, 3691–3693.
Page, A.J., Alikhan, N.-F., Carleton, H.A., Seemann, T., Keane, J.A., Katz, L.S., 2017. Comparison of classical multi-locus sequence typing software for next-generation sequencing data. Microb. Genom. 3.
Pons, B.J., Vignard, J., Mirey, G., 2019. Cytolethal distending toxin subunit b: a review of structure–function relationship. Toxins (Basel) 11, 595.
Porwollik, S., McClelland, M., 2003. Lateral gene transfer in Salmonella. Microbes Infect. 5, 977–989.
Ranjbar, R., Rahmati, H., Shokoohizadeh, L., 2018. Detection of Common Clones of *Salmonella enterica* Serotype Infantis from Human Sources in Tehran Hospitals Gastroenterology and Hepatology from Bed to Bench.
Raudales, R.E., Parke, J.L., Guy, C.L., Fisher, P.R., 2014. Control of waterborne microbes in irrigation: a review. Agric. Water Manag. 143, 9–28.
Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069.
Selzer, P.M., 2016. Host-Pathogen Interaction: Microbial Metabolism, Pathogenicity and Antiinfectives. John Wiley & Sons.

Shariat, N., Timme, R.E., Pettengill, J.B., Barrangou, R., Dudley, E.G., 2015. Characterization and evolution of Salmonella CRISPR-Cas systems. Microbiology 161, 374–386.
Sherwood, K.E., Cano, D.J., Maupin-Furlow, J.A., 2009. Glycerol-mediated repression of glucose metabolism and glycerol kinase as the sole route of glycerol catabolism in the haloarchaeon Haloferax volcanii. J. Bacteriol. 191, 4307–4315.
Siddiqui, R., Malik, H., Sagheer, M., Jung, S.-Y., Khan, N.A., 2011. The type III secretion system is involved in Escherichia coli K1 interactions with Acanthamoeba. Exp. Parasitol. 128, 409–413.
Singh, B., Arya, G., Kundu, N., Sangwan, A., Nongthombam, S., Chaba, R., 2019. Molecular and functional insights into the regulation of D-galactonate metabolism by the transcriptional regulator DgoR in Escherichia coli. J. Bacteriol. 201 e00281–18.
Smollett, K.L., Smith, K.M., Kahramanoglou, C., Arnvig, K.B., Buxton, R.S., Davis, E.O., 2012. Global analysis of the regulon of the transcriptional repressor LexA, a key component of SOS response in mycobacterium tuberculosis. J. Biol. Chem. 287, 22004–22014.
Spector, M.P., 1998. The starvation-stress response (SSR) of Salmonella. In: Advances in Microbial Physiology. Elsevier, pp. 233–279.
Spector, M.P., Cubitt, C.L., 1992. Starvation-inducible loci of Salmonella typhimurium: regulation and roles in starvation-survival. Mol. Microbiol. 6, 1467–1476.
Spector, M.P., Kenyon, W.J., 2012. Resistance and survival strategies of Salmonella enterica to environmental stresses. Food Res. Int. 45, 455–481.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.
Steele, M., Odumeru, J., 2004. Irrigation water as source of foodborne pathogens on fruit and vegetables. J. Food Prot. 67, 2839–2849.
Stevens, M.P., Humphrey, T.J., Maskell, D.J., 2009. Molecular insights into farm animal and zoonotic salmonella infections. Philos. Trans. R. Soc. B Biol. Sci. https://doi.org/10.1098/rstb.2009.0094.
Thomsen, L.E., Olsen, J.E., Foster, J.W., Ingmer, H., 2002. ClpP is involved in the stress response and degradation of misfolded proteins in Salmonella enterica serovar Typhimurium. Microbiology 148, 2727–2733.
Tidhar, A., Rushing, M.D., Kim, B., Slauch, J.M., 2015. Periplasmic superoxide dismutase SodCI of S almonella binds peptidoglycan to remain tethered within the periplasm. Mol. Microbiol. 97, 832–843.
Treangen, T.J., Ondov, B.D., Koren, S., Phillippy, A.M., 2014. The harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol. 15, 524.
Velge, P., Wiedemann, A., Rosselin, M., Abed, N., Boumart, Z., Chaussé, A.M., Grépinet, O., Namdari, F., Roche, S.M., Rossignol, A., Virlogeux-Payant, I., 2012. Multiplicity of Salmonella entry mechanisms, a new paradigm for Salmonella pathogenesis. Microbiologyopen 1, 243–258. https://doi.org/10.1002/mbo3.28.
Wagner, C., Hensel, M., 2011. Adhesive mechanisms of Salmonella enterica. In: Bacterial Adhesion. Springer, pp. 17–34.
Waldner, L., MacKenzie, K., Köster, W., White, A., 2012. From exit to entry: long-term survival and transmission of Salmonella. Pathogens 1, 128–155. https://doi.org/10.3390/pathogens1020128.
Wells, T.J., Henderson, I.R., 2013. Type 1 and 5 secretion systems and associated toxins. In: Escherichia Coli. Elsevier, pp. 499–532.
Westra, E.R., Dowling, A.J., Broniewski, J.M., van Houte, S., 2016. Evolution and ecology of CRISPR. Annu. Rev. Ecol. Evol. Syst. 47, 307–331.
Winfield, M.D., Groisman, E.A., 2003. Role of nonhost environments in the lifestyles of Salmonella and Escherichia coli. Appl. Environ. Microbiol. 69, 3687–3694.
Worley, M.J., Ching, K.H.L., Heffron, F., 2000. Salmonella SsrB activates a global regulon of horizontally acquired genes. Mol. Microbiol. 36, 749–761.
Wray, C., Wray, A., 2000. Salmonella in Domestic Animals (Cabi).
Yao, K., Muruvanda, T., Allard, M.W., Hoffmann, M., 2017. Complete genome sequences of three Salmonella enterica subsp. enterica serovar Saintpaul isolates associated with a 2013 multistate outbreak in the United States. Genome Announc. 5.
Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F.M., Larsen, M.V., 2012. Identification of acquired antimicrobial resistance genes. J. Antimicrob. Chemother. 67, 2640–2644.